3 Maximum-Likelihood Methods in Quantum Mechanics

Zdeněk Hradil[1], Jaroslav Řeháček[1], Jaromír Fiurášek[1,2], and Miroslav Ježek[1]

[1] Department of Optics, Palacky University, Olomouc, Czech Republic, hradil@aix.upol.cz
[2] Ecole Polytechnique, CP 165, Université Libre de Bruxelles, 1050 Bruxelles, Belgium

Yet your mistrust cannot make me a traitor:
Tell me whereon the likelihood depends.

W. Shakespeare: As you like it

Abstract. Maximum Likelihood estimation is a versatile tool covering wide range of applications, but its benefits are apparent particularly in the quantum domain. For a given set of measurements, the most likely state is estimated. Though this problem is nonlinear, it can be effectively solved by an iterative algorithm exploiting the convexity of the likelihood functional and the manifold of density matrices. This formulation fully replaces the inverse Radon transformation routinely used for tomographic reconstructions. Moreover, it provides the most efficient estimation strategy saturating the Cramer-Rao lower bound asymptotically. In this sense it exploits the acquired data set in the optimal way and minimizes the artifacts associated with the reconstruction procedure. The idea of maximum likelihood reconstruction is further extended to the estimation of quantum processes, measurements, and discrimination between quantum states. This technique is well suited for future applications in quantum information science due to its ability to quantify very subtle and fragile quantum effects.

3.1 Introduction

Our world is controlled by chance and everybody must cope with it. Events may but need not happen. Success and surviving depends on the ability of any living creature to set the risk associated with a particular alternative and to chose the optimal one. Chance has probably played a very important role in the formation of life on the Earth as Darwin noticed centuries ago. But chance manifests itself even in less sophisticated aspects of everyday life. Insurance companies, share markets, lotteries, bookmakers, and betting agencies count on chance. Chance has been commercialized and converted to money.

Its is intriguing to note that the chance in classical physics is caused by the enormous complexity of the world, due to which deterministic solutions are intractable in reality. In this sense, the origin of chance may be explained

Z. Hradil, J. Řeháček, J. Fiurášek, M. Ježek, Maximum-Likelihood Methods in Quantum Mechanics, Lect. Notes Phys. **649**, 59–112 (2004)
http://www.springerlink.com/

and quantified. The notion of chance acquires dramatically different meaning in the quantum domain, since quantum theory is intrinsically a probabilistic theory. Chance expressed by probabilities is here an indispensable ingredient of any effect and deserves special attention. All information about the quantum world comes through the probabilities. According to the Copenhagen interpretation of quantum theory, probabilities are coded in a quantum state describing the available knowledge about the system. Quantum theory handles observable events on the most fundamental level currently available, predicting the statistics of quantum phenomena.

The statistical nature of the quantum world is revealed in the experimental reality. When performing repeatedly an experiment on the ensemble of identically prepared systems, the experimenter cannot deterministically control their result due to the unavoidable fluctuations. But what can be observed is the statistics of certain events, expressed either in the form of probabilities or in the form of average values and their moments. Quantum theory answers the question: What statistics can be expected provided that the quantum state is known? This demonstrates the central role played by the quantum state. Its knowledge makes it possible to predict statistical results of any measurement performed on the same system. The determination of the quantum state itself represents an inverse problem, and as it is well known, inverse problems are more involved than direct ones. The inverse problem of determining the quantum state is called quantum state reconstruction or quantum tomography.

Though the history of this problem can be traced back to the early days of quantum mechanics, namely to the Pauli problem [1], experimental utilization had to wait until quantum optics has opened a new era. The theoretical predictions of Vogel and Risken [2] were closely followed by the experimental realization of the suggested algorithm by Smithey et. al. [3]. Since that time many improvements and new techniques have been proposed, an overview can be found in [4]. Reconstruction of quantum states is now considered a standard technique used in various branches of contemporary physics [5–7]. Methods of this kind treat the problem of quantum state reconstruction as a special case of Radon transformation used in medicine X-ray tomography. Chapter by Raymer and Beck of this volume is devoted to various aspects of this wide spread treatment, which will be called *standard* in the following.

The purpose of this chapter is to contribute to the mosaic of quantum state reconstruction techniques presented in the book, and to present an alternative to standard schemes. The idea behind this is surprisingly simple. In accordance with the elucidation of the role of chance in our world, it could be paraphrased, loosely speaking, as the following strategy: Always bet on the most likely interpretation. This method, is known in mathematical statistics as the maximum likelihood (ML) estimation. It was proposed in the 1920s by R. Fisher.

The principle of maximum likelihood is not a rule that requires justification – it does not need to be proved. Nowadays it is widely used in many

applications in optics, noise analysis, communications, geophysics, nuclear physics etc. Among the most curious and spectacular applications that we can find, is the discovery of tunnels in the demilitarized zone between North and South Korea [8]. What makes this technique so attractive and powerful is its efficiency. It plays the prominent role among the other estimation techniques since it may reach the ultimate limit of accuracy predicted by Cramer-Rao bound. In this chapter we will stress out another remarkable feature, namely that ML estimation is predetermined for solving inverse quantum problems. It naturally reproduces the structure of quantum theory such as the closure or uncertainty relations. This fundamental reason justifies its usage, though it leads to nonlinear algorithms and its numerical implementation is demanding on computing resources.

In Sect. 3.2 various reconstruction techniques will be compared. Particularly, the essence of the standard quantum state reconstruction will be briefly reviewed pointing out weak points of this technique. This will serve as motivation for the ML estimation, which will be further advocated as a versatile and natural tool for solving quantum problems. For the sake of completeness, ML quantum-state tomography will be briefly compared with alternative statistical approaches to this problem, such as Bayesian and Maximum Entropy (MaxEnt) estimations.

In Sect. 3.3 basics of ML estimation of the quantum state will be presented using the Jensen inequality. Since this will serve as a paradigm for further applications, an elementary exposition on the introductory level will be given. As a result, the extremal equation will be derived and its interpretation will be given and compared to standard reconstructions.

In Sect. 3.3.4 the ML technique will be treated as a problem of a statistical distance. The extremal equation will be re-derived using variational approach, which is more suitable for further generalizations. The numerical aspects of the extremal solution will be detailed on the example of a reconstruction of two entangled qubits.

Section 3.4 will link the problem of the state reconstruction with the operational information. As will be shown, the information content acquired in the course of measurement may be quantified by means of Fisher information matrix. Loosely speaking, it grasps the width of the likelihood functional.

In Sect. 3.5 we will apply the technique of ML estimation to the reconstruction of quantum processes. Section 3.6 will address the estimation of quantum measurements, and finally Sect. 3.7 will consider the problem of quantum discrimination motivated by the ML estimation technique.

3.2 Overview of Quantum-State Reconstruction Techniques

To address the problem of quantum state reconstruction, let us consider a generic quantum measurement. The formulation will be developed for the

case of finite dimensional quantum systems. The reader can think of a spin 1/2 system for the sake of simplicity. Applications to infinite dimensional Hilbert spaces will be mentioned separately.

Assume that we are given a finite number N of identical copies of the system, each in the same but unknown quantum state described by the density operator ρ. Given those systems our task is to identify the unknown *true* state ρ as accurately as possible from the results of measurements performed on them. For simplicity we will assume all measurements in the sense of von Neumann. The generalization to a positive valued operator measure (POVM) is straightforward.

Let us consider, for concreteness, that N particles prepared in the same state have been observed in M different output channels of the apparatus. For spin one-half particles those channels could be for instance the $M = 6$ output channels of the Stern-Gerlach apparatus oriented along x, y, and z directions, respectively.

Provided that each particular output

$$|y_j\rangle\langle y_j|, \qquad j = 1, \ldots, M \tag{3.1}$$

has been registered n_j times, $\sum_j n_j = N$, the relative frequencies are given as $f_j = n_j/N$. Using these ingredients, the true state ρ is to be inferred. For the sake of simplicity, the measurement performed will be assumed as complete, i.e.

$$H \equiv \sum_j |y_j\rangle\langle y_j| = 1. \tag{3.2}$$

This condition is satisfied for Stern-Gerlach measurements mentioned above. Later we will release this condition and consider incomplete measurements too. The probabilities of occurrences of various outcomes are generated by the true quantum state ρ according to the well-known quantum rule,

$$p_j = \langle y_j|\rho|y_j\rangle. \tag{3.3}$$

Let us now briefly discuss various approaches to the quantum state reconstruction.

3.2.1 Standard Reconstruction

If the probabilities p_j of getting a sufficient number of different outcomes $|y_j\rangle$ were known, it would be possible to determine the true state ρ directly by inverting the linear relation (3.3). This is the philosophy behind the "standard" quantum tomographic techniques. For example, in the rather trivial case of a spin one half particle, the probabilities of getting three linearly independent projectors determine the unknown state uniquely. Here, however, a serious problem arises. Since only a finite number of systems can be investigated,

there is no way to find out those probabilities. The only data one has at his or her disposal are the relative frequencies f_j, which sample the *principally* unknowable probabilities p_j. It is obvious that for a small number of runs, the true probabilities p_j and the corresponding detected frequencies f_j may differ substantially. As a result of this, the modified realistic problem,

$$f_j = \langle y_j|\rho|y_j\rangle, \tag{3.4}$$

has generally no solution on the space of semi-positive definite hermitian operators describing physical states. This linear equation for the unknown density matrix may be solved for example by means of pattern functions, see e.g. [4, 9], what could be considered as a typical example of the standard approach suffering from the above mentioned drawbacks. In the case of homodyne detection the discrete index j should be replaced by the pair of continuous variables x and ϕ denoting the detected value and phase of a quadrature operator. The quantum state of a single-mode quantum field represented by the Wigner function $W(\alpha)$ can then be formally written as the following integral transformation

$$W(\alpha) = \int d\phi dx K(x,\phi) f(x,\phi), \tag{3.5}$$

where $K(x,\phi)$ denotes a kernel of the integral transformation. Notice, that such a standard solution is linear with respect to data. No matter how ingeniously the inversions in the standard reconstruction schemes are done, there always remains some problems caused by the application of this treatment in quantum theory. All measurements in practice are inevitably limited as far as the amount and accuracy of data is concerned. The continuous parameter can be scanned only in a limited number of positions, and such a measurement done with a finite ensemble will always be affected by fluctuations. Moreover, the kernels involved in the standard reconstruction of the type (3.5) are often singular. When the standard approach is applied to real data, serious problems with positivity of reconstructed density matrix may appear. Though these simple techniques give us a rough picture of the unknown state, they are not able to provide the full quantum description.

3.2.2 ML Estimation

Having measurements done and their results registered, the experimenter's knowledge about the measured system is increased. Since quantum theory is probabilistic, it makes little sense to pose the question: "What quantum state is determined by that measurement?" A more appropriate question is: "What quantum states seem to be most likely for that measurement?"

Quantum theory predicts the probabilities of individual detections, see (3.3). From them one can construct the total joint probability of registering data $\{n_j\}$. We assume that the input system (particle) is always detected in

one of M output channels, and this is repeated N times. Subsequently, the overall statistics of the experiment is multinomial,

$$\mathcal{L}(\rho) = \frac{N!}{\prod_i n_i!} \prod_j \langle y_j|\rho|y_j\rangle^{n_j}, \tag{3.6}$$

where $n_j = Nf_j$ denotes the rate of registering a particular outcome j. In the following we will omit the multinomial factor from expression (3.6) as it has no influence on the results. Physically, the quantum state reconstruction corresponds to a synthesis of various measurements done under different experimental conditions, performed on the ensemble of identically prepared systems. For example, the measurement might be a subsequent recording of an unknown spin of the neutron (polarization of the photon) using different settings of the Stern Gerlach apparatus, or the recording of the quadrature operator of light in rotated frames in quantum homodyne tomography. The likelihood functional $\mathcal{L}(\rho)$ quantifies the degree of belief in the hypothesis that for a particular data set $\{n_j\}$ the system was prepared in the quantum state ρ. The ML estimation simply selects the state for which the likelihood attains its maximum value on the manifold of density matrices.

The mathematical formulation of ML estimation in quantum theory will be worked out below showing that it provides a feasible statistical inversion. This is guaranteed by the convexity of both the set of density matrices and likelihood functional. As will be demonstrated, provided that the standard method yields a physically sound (semipositive) solution, the same density matrix is also a ML solution. On the contrary to standard approaches, ML principle is a highly nonlinear fitting procedure. Let us clarify some physical reasons for why it is so.

For simplicity, let us consider the textbook example of a spin 1/2 system and simple Stern-Gerlach (SG) detections. Let us assume that several projections have been registered. How to use all of them without discarding any information? Some observations may appear incompatible with each other due to the fluctuations and noises involved. That is why the overdetermined system of (3.4) is inapplicable in the presence of noise. Moreover, different SG projections cannot be treated on equal footing. They are observing various "faces" of the spin system. Such measurements, even when done with equal numbers of particles, possess different errors. It is not difficult to see why. The measured frequencies fluctuate around the true probabilities according to the binomial statistics, and significantly, their root-mean square errors depend on the overlap between the projections and the true (but unknown!) spin state. Various SG measurements are also incompatible in the sense of quantum theory, because projections to different directions do not commute. Such data cannot be obtained in the same measurement, but may be collected subsequently in the course of repeated detections, where, in general, different observations will be affected by different errors. But these errors in turn depend on the unknown state. The estimation procedure must therefore

predict the unknown state and consider the data fluctuations simultaneously. This indicates the nonlinearity of the optimal estimation algorithm. As will be demonstrated later, ML estimation does this job.

3.2.3 Other Statistical Approaches

In this short subsection let us mention two other statistically motivated approaches to the quantum state reconstruction problem. The heart of the Bayesian statistical inference is Bayes's rule [10]

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)}, \tag{3.7}$$

where $p(D|H)$ is the probability of data D given hypothesis H, $p(H)$ is the prior probability that hypothesis H is true, $P(H|D)$ is the posterior probability that hypothesis H is true given data D, and the normalization $p(D) = \sum_H p(D|H)p(H)$ is the marginal probability of data D. Bayes's rule shows how the acquired data D updates our knowledge, $p(H) \to p(H|D)$. This formulation is classical in the sense that it involves classical probabilities, but it can also be extended to to quantum domain, see [11] and the chapter by Fuchs and Schack in this volume. On the conceptual level this can be done easily associating hypothesis H with a quantum state, $H \to \rho$, and data D with a particular output of the measurement. Invoking the linearity of the quantum theory, the quantum state associated with the posterior distribution $p(\rho)$ on the space of density matrices reads

$$\rho = \int d\rho p(\rho)\rho. \tag{3.8}$$

The linkage of the Bayesian reasoning to ML seems to be clear – it represents just a rough approximation replacing a probability distribution by a single peak localized at the most likely state. This is true, but the problem of quantum Bayesian inference is more complicated than it looks. Namely, it is not clear what should be chosen as the proper measure for the integration over the manifold of mixed states. Beside this, when resorting to Bayesian estimation the inferred state is not going to be the maximally likely one in general. This is a bit counterintuitive, since why it should be better to throw away the chance and diminish the probability of success resorting to a less probable interpretation?

Another statistically motivated approach is the maximum entropy inference devised by Jaynes [12] and since then applied to many physical problems [13]. The chapter by Bužek also in this volume is a detailed exposition of this approach. Its formulation hinges on the entropy as a measure of information. Loosely speaking, the MaxEnt principle looks for maximally unbiased solutions fulfilling the given set of constraints. In tomographic applications, the constraints depend on the registered data. In this sense the

MaxEnt reconstruction provides the most pessimistic guess compatible with the observations made. Its mathematical formulation is conceptually simple. Entropy of a quantum state is is defined as follows,

$$S = -k\mathrm{T}r[\rho \log \rho], \tag{3.9}$$

k being an arbitrary constant (Boltzmann constant k_B in statistical physics, here we will set it to $k = 1$). As constraints we usually fix the expectation values of some set of observables or projectors,

$$a_i = \mathrm{T}r[\rho A_i], \tag{3.10}$$

a_i are real numbers inferred from the measurement, and A_i are the corresponding observables (q-numbers). The normalization, $\mathrm{Tr}\rho = 1$, can be taken as an additional constraint with $a_0 = 1$ and $A_0 = \hat{1}$. MaxEnt solution can always be written in the form,

$$\rho = \exp[\sum_i \lambda_i A_i], \tag{3.11}$$

where the Lagrange multipliers λ_i can be determined from the constraints (3.10) by solving a set of coupled nonlinear equations. This can be difficult, particularly when the operators A_i do not commute.

To compare the MaxEnt and ML estimations let us have a closer look on the constraints corresponding to the sampled probabilities ($A_i = |i\rangle\langle i|$). It is not difficult to see that such constraints are equivalent to problem (3.4) of the standard reconstruction scheme. Consider the case of a p-dimensional Hilbert space, where the unknown quantum state is uniquely specified by $p^2 - 1$ parameters. This is also the maximum number of constraints which may be satisfied simultaneously. But even in the case of under-determined problems, some constraints may still appear as inconsistent due to the fluctuations present in the observed data. This is a serious obstacle for the straightforward application of the MaxEnt principle to the sampled probabilities. To get a physically correct reconstruction, the constraints have to be released.

3.3 ML Quantum-State Estimation

3.3.1 Extremal Equation

The mathematical formulation of ML estimation will be developed here. The likelihood functional corresponding to detected data (3.6) should be maximized on the manifold of density matrices. This will guarantee getting a correct and physically sound interpretation of obtained results. No additional assumptions are needed. Here we will formulate the ML extremal equation for a generic case of projective measurements $|y_i\rangle\langle y_i|$; generalization to measurements described by elements of a probability valued operator measure is straightforward as will be shown later.

ML estimation consist in maximizing the likelihood functional,

$$\mathcal{L}(\rho) = \prod_i \langle y_i|\rho|y_i\rangle^{n_i}. \tag{3.12}$$

Here, and in the following, we omit the unimportant multinomial multiplicative factor. In order to maximize (3.12), the Jensen inequality between the geometric and arithmetic averages will be adopted,

$$\prod\nolimits_i \left[\frac{x_i}{a_i}\right]^{f_i} \leq \sum\nolimits_i f_i \frac{x_i}{a_i}, \tag{3.13}$$

where $x_i \geq 0$ and $a_i > 0$ are auxiliary positive nonzero numbers and $f_i = n_i/N$. The equality sign is reached if and only if all the numbers x_i/a_i are equal. We will use boldface to denote vectors, i.e. $\mathbf{a}$ and $\mathbf{x}$. Setting $x_i = \langle y_i|\rho|y_i\rangle$, the dimension of these vectors is given by the number of independent projections, parameters a_i being a subject of further considerations. This inequality may be easily adopted for the maximization of likelihood since,

$$\big(\mathcal{L}(\rho)\big)^{1/N} = \prod\nolimits_i \Big(\langle y_i|\rho|y_i\rangle\Big)^{f_i} \leq \prod\nolimits_j a_j^{f_j} \mathrm{Tr}\{\rho R(\mathbf{y},\mathbf{a})\} \tag{3.14}$$

$$\leq \lambda(\mathbf{y},\mathbf{a}) \prod\nolimits_i a_i^{f_i}. \tag{3.15}$$

Operator R is a positive semi-definite operator defined by its expansion into the measured, generally non-orthogonal projectors as:

$$R(\mathbf{y},\mathbf{a}) = \sum\nolimits_i \frac{f_i}{a_i} |y_i\rangle\langle y_i|. \tag{3.16}$$

Relation (3.14) follows simply from the definition of the likelihood (3.12) and from the Jensen inequality (3.13) for any allowed vector $\mathbf{a}$. The last inequality (3.15) represents the maximization of the previous expression over all density matrices. Parameter $\lambda(\mathbf{y},\mathbf{a})$ representing the largest eigenvalue of the operator R defines the upper bound of the likelihood. Let us clarify under what conditions the equality sign will be achieved in the chain of inequalities (3.14) and (3.15). Fixing the parameters a_i, the most general extremal state ρ_e should have its support in the subspace corresponding to the maximal eigenvalue λ,

$$R(\mathbf{y},\mathbf{a})\rho_e = \lambda(\mathbf{y},\mathbf{a})\rho_e. \tag{3.17}$$

This guarantees the equality sign in the inequality (3.15). Setting further, $a_i = \langle y_i|\rho_e|y_i\rangle$, the equality sign in (3.14) will be achieved too. Put together, the extremal equation for the ML density matrix reads,

$$R\rho = \rho, \tag{3.18}$$

where the operator

$$R = \sum\nolimits_i \frac{f_i}{\langle y_i|\rho|y_i\rangle} |y_i\rangle\langle y_i| \tag{3.19}$$

is state dependent, Lagrange multiplier $\lambda = 1$ by normalization, and the subscript "e" was omitted for brevity. Notice also, that there are many equivalent forms of this extremal equation. Particularly, the extremal state is seen to commute with the operator R, $[R, \rho] = 0$, since ρ is hermitian. An alternative form of extremal equation reads,

$$\sum\nolimits_i \frac{f_i}{\langle y_i|\rho|y_i\rangle} \Pi_i = 1_\rho, \tag{3.20}$$

where 1_ρ denotes the identity operator defined on the support of the (unknown) extremal density matrix, and $\Pi_i = |y_i\rangle\langle y_i|$ are POVM elements corresponding to measurement.

3.3.2 Example of Simple Measurements

Let us show some special solutions linking the ML estimation to standard reconstruction schemes. The latter, based on linear inversion, is obtained as the solution of equations $f_i = \langle y_i|\rho|y_i\rangle$. If a semi-positive definite solution exists, such a density matrix will also maximize the ML functional. This follows from the Gibbs inequality, [14]

$$\sum_i f_i \ln \langle y_i|\rho|y_i\rangle \leq \sum_i f_i \ln f_i. \tag{3.21}$$

Consider the textbook example [15] of a measurement represented by a single nondegenerate hermitian operator A, $A|a\rangle = a|a\rangle$. Its discrete orthonormal eigenvectors, $\langle a|a'\rangle = \delta_{aa'}$ provide the closure relation $\sum_a |a\rangle\langle a| = \hat{1}$. The probability of measuring a particular value of A on a system that has been prepared in a quantum state ρ is $p_a = \langle a|\rho|a\rangle$. When this measurement is repeated on N identical copies of the system, each outcome a occuring n_a times, the relative frequencies $f_a = \frac{n_a}{N}$ will sample the true probabilities p_a reproducing them only in the limit $N \to \infty$. Experimenter's knowledge can then be expressed in the form of a diagonal density matrix

$$\rho_{est} = \sum_a f_a |a\rangle\langle a|. \tag{3.22}$$

This state is apparently semi-positive definite, and as a consequence of inequality (3.21) it also maximizes the likelihood functional. In formula (3.22) we have just stated the result of the experiment $\{n_a\}$ in one special way; in fact, no tomography is involved there. Similar knowledge is obtained by compatible observations, i.e. by measurements described by commuting operators with a common diaginalizing basis $\{|a\rangle\}$.

In order to find also non-diagonal elements of the unknown density matrix, some other measurements non-compatible with A should be adopted. This can be done by projecting the unknown state onto a set of non-commuting eigenvectors of operators A_j, $j = 1, 2, \ldots$. Such a measurement is going to give us more information than just the diagonal elements of the density matrix in some a priori *given* basis. Maximum-likelihood synthesis and interpretation of such general observations will be given in the following section.

3.3.3 Interpretation of the ML Solution

An intriguing property of ML extremal equation (3.20) is that it resembles the closure relation $\sum_j |y_j\rangle\langle y_j| = \hat{1}$. Indeed, in terms of re-normalized projectors $\Pi_i^{'} = (f_i/p_i)\Pi_i$, the extremal equation reads

$$\sum_i \Pi_i^{'} = 1_\rho. \tag{3.23}$$

Moreover, the expectation values of such re-normalized projectors taken with the extremal state reproduce identically the detected frequencies

$$\mathrm{T}r(\rho\Pi_i^{'}) \equiv f_i, \quad \forall i. \tag{3.24}$$

Maximum likelihood estimation thus can be interpreted as a renormalization of observations done, which apart from being able to fit the registered data exactly provides a resolution of unity and thus represents a genuine quantum measurement.

This can now be contrasted to the standard formulation of the estimation problem which is based on the solution of a linear set of equations,

$$\mathrm{Tr}(\rho\Pi_i) = f_i, \quad \forall i, \tag{3.25}$$

while the closure relation,

$$\sum_i \Pi_i \equiv 1, \tag{3.26}$$

is fulfilled as identity. Notice that roles of the closure relation and probability rule in ML approach are *reversed* with respect to that in the standard formulation.

Let us stress once more that the standard reconstruction has often no solution on the set of density matrices, whereas the ML formulation has always a solution. If the standard reconstruction yields a density matrix, then this state is also maximum-likely quantum state, but not vice versa.

Another interesting feature of the ML reconstruction technique is that the method itself, through the closure relation (3.23), defines "the field of view" of the tomographic scheme as the subspace, where the reconstruction is done.

No such interpretation of the standard solution is possible. We consider this mutual relationship between the standard and ML reconstructions as a strong argument in favor of the latter method.

In view of the above analysis one may wonder why standard methods are so popular. The answer is that in many applications, the ensembles of measured systems are so huge, that fluctuations, which scale as $1/N$, can entirely be neglected. Standard methods are then preferred due to their simplicity and much lower computing costs.

As will be shown later in Sect. 3.4, the ignoring of the positivity constraint always lowers the accuracy of the estimation procedure. However, this undesirable effect may be suppressed by enlarging the ensemble of probed particles, at least in applications where the positivity is not crucial. That is why the X-ray tomography in medicine is such a wonderful tool, though the standard back projection based on the deterministic inverse Radon transformation is used on fluctuating data. However, in applications with much smaller count rates, such as in neutron radiography discussed by Badurek *et al.* in this volume, the ill posed nature of inverse problems fully reveals itself by the appearance of severe artifacts. The same may and usually does happen in quantum tomography.

3.3.4 Maximum Likelihood as a Statistical Distance

Probabilistic interpretation of quantum theory suggests that a statistical treatment of the observed data [17–19] is more natural and appropriate than the deterministic one. Maximum likelihood estimation [17], which is the main subject of this chapter looks for the most likely quantum state. The likelihood (3.12) quantifies our degree of belief in the given hypothesis ρ. However, one can also think of the likelihood as a sort of statistical distance $D[\mathbf{f}, \mathbf{p}(\rho)]$ between the theory p_j and data f_j. Via its minimization we find the density matrix ρ_e that generates through (3.3) probabilities p_j lying as "close" to the observed frequencies f_j as possible. This statistical distance is known in the theory of statistics as the relative entropy or Kullback-Leibler divergence [21]:

$$D[\mathbf{f}, \mathbf{p}] = -\sum_j f_j \ln p_j. \tag{3.27}$$

Indeed, adopting metric (3.27) is equivalent to finding the maximum of the likelihood functional (3.12).

At first sight it might seem that there is no reason to *prefer* one particular metric to another one – different metrics leading to different results. This ambiguity can be resolved by considering the formal description of the reconstruction process [20]. If the whole measurement and subsequent reconstruction is looked at as a single generalized measurement, see Sect. 3.3.3, then the *relation* between the actually performed measurement and resulting probability operator measure becomes particularly simple and easy to interpret if Kulback-Leibler distance (3.27) is adopted.

ML methods are well-known in the field of inverse problems and they have found many applications in reconstructions and estimations so far [22–25]. Unfortunately, except in most simple cases, the maximization of the likelihood functional is a challenging problem on its own.

In the following we will show two different but related ways to maximize the likelihood (3.12). The first one is based on splitting the complex problem of estimating the density operator into two simpler tasks: finding the optimal eigenvalues and eigenvectors. The second approach relies on a direct application of the calculus of variations [17, 26].

Both these routes will provide alternative derivations of the extremal equation (3.18) of Sect. 3.3.1, and give us some hints how to solve it iteratively.

3.3.5 Maximization of the Likelihood

In the classical signal processing an important role is played by linear and positive (LinPos) problems [27, 28]. Since these are closely related to the problem of quantum state reconstruction it is worthwhile to recall how the positive and linear problems can be dealt with using the ML approach.

Let us consider that the probabilities p_j of getting outcomes y_j are given by the following linear and positive relation

$$p_j = \sum_i r_i h_{ij}, \quad \mathbf{p}, \mathbf{r}, \mathbf{h} > \mathbf{0}. \tag{3.28}$$

Here $\mathbf{r}$ is the vector describing the "state" of the system. For example, the reconstruction of a one-dimensional object from the noiseless detection of its blurred image could be accomplished by inverting the relation (3.28), where $\mathbf{r}$ and $\mathbf{p}$ would be the normalized intensities of the object and image, and $\mathbf{h}$ would describe the blurring mechanism. Again here the presence of noise ($f_j \neq p_j$) tends to spoil the positivity of the reconstructed intensity $\mathbf{r}$.

The solution to LinPos problems in the sense of ML can be found using the expectation-maximization (EM) algorithm [27, 28],

$$r_i^{(n)} = r_i^{(n-1)} \sum_j \frac{h_{ij} f_j}{p_j(\mathbf{r}^{(n-1)})}, \tag{3.29}$$

which if initialized with a positive vector $\mathbf{r}^0$ ($r_i^0 > 0 \, \forall i$) is guaranteed to converge to the global minimum of the Kullback-Leibler divergence $D[\mathbf{f}, \mathbf{p}]$ for any input data $\mathbf{f}$.

This algorithm is convenient from the point of view of the numerical analysis. It is certainly much more convenient than the direct multidimensional maximization of the corresponding ML functional $\ln \mathcal{L} = \sum_j f_j \ln p_j$ [29]. This brings us back to the problem of quantum state reconstruction. It would be nice to have a similar iterative algorithm for dealing with the problem

(3.4), or equivalently for maximizing the ML functional (3.12). On the one hand it is clear that the problem of quantum state reconstruction is not a LinPos problem, since the quantum rule (3.3) cannot be rewritten to the form of (3.28) with a known positive kernel **h**. As a consequence of this, the EM algorithm cannot be straightforwardly applied here. On the other hand the reconstruction of the elements of the density matrix becomes a LinPos problem if the eigenbasis diagonalizing the density matrix is known. In this case the unknown density matrix can be parametrized as follows

$$\rho = \sum_k r_k |\phi_k\rangle\langle\phi_k|, \qquad \rho|\phi_k\rangle = r_k|\phi_k\rangle, \tag{3.30}$$

where r_i are eigenvalues of ρ, the only parameters which remain to be determined from the performed measurement. Using (3.30), relation (3.3) can easily be rewritten to the form of LinPos problem (3.28).

Unlike the standard quantum measurement of a single observable, a series of sequential measurements of many non-commuting operators determines not only the diagonal elements of ρ but also the diagonalizing basis itself. This hints on splitting the quantum state reconstruction into two subsequent steps: the reconstruction of the eigenvectors of ρ in a fixed basis, which represents the classical part of the problem, followed by the "rotation" of the basis $\{|\phi_i\rangle\}$ in the "right" direction using the unitary transformation

$$|\phi'_k\rangle\langle\phi'_k| = U|\phi_k\rangle\langle\phi_k|U^\dagger. \tag{3.31}$$

Its infinitesimal form reads

$$U \equiv e^{i\epsilon G} \approx 1 + i\epsilon G. \tag{3.32}$$

Here G is a hermitian generator of the unitary transformation and ϵ is a positive real number small enough to make the second equality in (3.32) approximately satisfied.

Consider now the total change of the log-likelihood caused by the change of diagonal elements of density matrix and rotation of basis. Keeping the normalization condition $\mathrm{Tr}\rho = 1$, the first order contribution to the variation reads

$$\delta \ln \mathcal{L} = \sum_k \delta r_k (\langle\phi_k|R|\phi_k\rangle - 1) + i\epsilon \,\mathrm{Tr}\left\{G[\rho, R]\right\}. \tag{3.33}$$

Operator R was defined earlier in (3.19). We remind the reader that it is a semi-positive definite operator depending on ρ.

Inspection of (3.33) reveals a simple two-step strategy to make the likelihood of the new state ρ' as high as possible [within the limits of the validity of the linearization (3.32), of course.] In the first step, the first term on the right hand side of (3.33) is maximized by estimating the eigenvalues of the density matrix keeping its eigenvectors $|\phi_k\rangle$ constant. The iterative EM algorithm

(3.29) described above can straightforwardly be applied to this LinPos problem. As the second step, the likelihood can further be increased by making the second term on the right hand side of (3.33) positive. This is accomplished by a suitable choice of the generator G of the unitary transformation (3.32). Recalling the natural norm induced by the scalar product defined on the space of operators, $(A, B) = \mathrm{Tr}\{A^\dagger B\}$, the generator G may be chosen as

$$G = i[\rho, R]. \tag{3.34}$$

This choice guarantees the non-negativity of its contribution to the likelihood and is optimal in the sense of the above introduced scalar product. Notice that this derivation holds only if the second order contribution in ϵ to (3.32) and (3.33) is negligible. From this, an upper bound on the value of ϵ can be derived. Parameter ϵ can then be adaptively changed in each U step in order to minimize the computing time.

Now we have at our disposal all ingredients comprising the EMU quantum state reconstruction algorithm [30]. Starting from some strictly positive density matrix ρ^0, this initial guess is improved, first by finding new eigenvalues using the EM iterative algorithm (3.29), and then again by finding new eigenvectors by a suitable unitary (U) transformation of the old ones according to (3.31-3.32) and (3.34). Repetition of these two steps, each monotonically increasing the likelihood of the current estimate, resembles climbing a hill. Convexity of the likelihood functional (3.12), $\mathcal{L}(\alpha\rho_1 + (1-\alpha)\rho_2) \geq \alpha\mathcal{L}(\rho_1) + (1-\alpha)\mathcal{L}(\rho_2)$, $\alpha \in \langle 0, 1\rangle$, then guarantees that the global maximum is always attained eventually.

The EMU algorithm naturally leads to the previously introduced extremal equation for the density matrix [17, 26]. The stationary point of EMU algorithm is characterized by the vanishing variation of the log likelihood (3.33). Since the variations $\delta r_k, \epsilon$ are arbitrary parameters, this is equivalent to the Lagrange-Euler equation for density matrix,

$$R\rho_e = \rho_e, \tag{3.35}$$

that was already derived in Sect. 3.3.1 by other means.

For the sake of completeness, let us present one more derivation based on the calculus of variations [26]: For generality let us assume that a generalized measurement described by POVM elements $\{Pi_j\}$, $\sum_j \Pi_j = 1$ was performed repeatedly N times. The theoretical probability of observing outcome j is $p_j = \mathrm{Tr}\rho\Pi_j$. Let us now maximize the logarithm of the likelihood functional on the space of positive operators subject to normalization $\mathrm{Tr}\rho = 1$. The latter constraint will be dealt with using an undetermined Lagrange multiplier λ while the positivity is ensured by decomposing ρ as follows, $\rho = A^\dagger A$. A necessary condition for ρ to be the maximum likely state is now that the functional,

$$F = \sum_j f_j \ln\left(\mathrm{Tr}\{A^\dagger A\Pi_j\}\right) - \lambda\mathrm{Tr}\{A^\dagger A\} \tag{3.36}$$

is stationary in the independent variable A. Varying A to $A + \delta A$, where δA is a small variation, the value of F will change as follows:

$$\delta F = \sum_j \frac{f_j}{p_j} \mathrm{Tr}\{\Pi_j A^\dagger \delta A\} - \lambda \mathrm{Tr}\{A^\dagger \delta A\}. \tag{3.37}$$

Since ρ is an extremal point, δF must vanish for all δA, which means that

$$\sum_j \frac{f_j}{p_j} \Pi_j A^\dagger = \lambda A^\dagger. \tag{3.38}$$

On the left hand side we recognize the operator R introduced before in (3.19) with projectors $|y_j\rangle\langle y_j|$ replaced by POVM elements Π_j. Multiplying this equation by A from the right side we get

$$R\rho = \lambda\rho. \tag{3.39}$$

Lagrange multiplier λ is determined simply by taking trace and using the normalization condition, which yields $\lambda = 1$. This completes the alternative derivation of the extremal equation (3.35) for the maximum likelihood quantum-state estimate.

Except in the most simple cases it is impossible to solve (3.35) by analytical means. This is, of course, a consequence of its strong nonlinearity in ρ. To find the solution numerically one can use the above described EMU algorithm which combines the classical EM algorithm with a rotation of the eigenbasis of ρ in each step. Another possibility is to apply iterations directly to the extremal equation (3.35). Since for a general ρ, the left hand side of (3.35) is not even a hermitian operator it is first necessary to rewrite (3.35) to an explicitly positive semidefinite form. Notice that if (3.35) holds so will its adjoint $\rho = \rho R$. Using this in (3.35) we get

$$R\rho R = \rho, \tag{3.40}$$

where on both sides are now positive semidefinite operators. Starting from some unbiased density matrix such as the maximally mixed state, (3.40) can be used to find the first approximation, and this procedure can be repeated until the stationary point of transformation (3.40) is attained. Unfortunately, unlike for classical EM algorithm the convergence of this quantum algorithm is not guaranteed in general. However it is not difficult to prove its convergence for sufficiently small steps when the transformation R is "diluted" as follows $R \to (I + \alpha R/2)/(1 + \alpha/2)$, where I is the identity operator and α is a small positive number. Considering now the ith step of the algorithm (3.40), the current density matrix ρ^i is transformed as follows:

$$\rho^{i+1} = (1 - \alpha)\rho^i + \alpha \frac{\rho^i R^i + R^i \rho^i}{2}. \tag{3.41}$$

Terms quadratic in α have been neglected. Denoting

$$\rho' = \frac{\rho^i R^i + R^i \rho^i}{2}, \tag{3.42}$$

it is seen that $(i+1)$th iteration ρ^{i+1} is a convex combination of ρ^i and ρ'. Loosely speaking it is obtained by moving a "distance" α from the old density matrix along the line connecting it with a "density matrix" ρ'. Now we will show that $\mathcal{L}(\rho^{i+1}) \geq \mathcal{L}(\rho^i)$, which is the same as $\ln \mathcal{L}(\rho^{i+1}) \geq \ln \mathcal{L}(\rho^i)$. Evaluating log-likelihood at ρ^{i+1}, expanding it in α and neglecting higher order terms we get

$$\ln \mathcal{L}(\rho^{i+1}) = \ln \mathcal{L}(\rho^i) + \alpha\big[\mathrm{Tr}\{R^i \rho'\} - 1\big]. \tag{3.43}$$

It remains to show that the last term on the right hand side is positive:

$$\mathrm{Tr}\{R^i \rho'\} = \mathrm{Tr}\{R^i \rho^i R^i\} = \mathrm{Tr}\{R^i \rho^i R^i\}\mathrm{Tr}\{\rho^i\} \geq \mathrm{Tr}^2\{R^i \rho^i\} = 1. \tag{3.44}$$

In the second to last step the Cauchy-Schwarz inequality has been used, and the last step follows from the definition of operator R. So we proved that a sufficiently small step always increases the likelihood. Since this functional has no side minima (convexity), the algorithm converges monotonically to the global maximum.

All results discussed so far have to be modified in the case of incomplete detections. Provided that $H \neq 1$ in (3.2), the closure relation may be always recovered in the form

$$\sum_j H^{-1/2} |y_j\rangle\langle y_j| H^{-1/2} = 1. \tag{3.45}$$

This corresponds to the renormalization of the true probabilities $p_j = \langle y_j|\rho|y_j\rangle$ as follows: $p_j \to p_j / \sum_i p_i$. This formulation incorporates the case of incomplete detection. Notice, that the extremal equation (3.35) remains valid, provided we introduce renormalized operators

$$R \to R' = (H')^{-1/2} R (H')^{-1/2}, \quad \rho_e \to \rho'_e = (H')^{1/2} \rho_e (H')^{1/2},$$

where we defined $H' = H / \sum_j p_j$. All the conclusions derived for complete measurements may be extended to this case of incomplete measurement as well.

This formulation coincides with the ML estimation provided that the experiment is governed by Poissonian statistics. Assume that n_i samples the mean number of particles np_i, where p_i is as before the prediction of quantum theory for a detection in the i-th channel and n is an unknown mean number of particles. The relevant part of log-likelihood corresponding to the Poissonian statistics reads $\ln \mathcal{L} \propto \sum_i n_i \ln(np_i) - n \sum_i p_i$. The extremal equation for n can easily be formulated as the condition $n = \sum_i n_i / \sum_i p_i$. On inserting this estimate of the unknown mean number of Poissonian particles into the log-likelihood we reproduce the renormalized likelihood function.

3.3.6 Maximum Likelihood as a Statistical Distance Continued ...

Having shown several ways how to maximize likelihood we can now come back to the justification of the Kullback-Leibler divergence as a plausible statistical distance. Let us assume the existence of a quantum measure $D(\mathbf{d}, \mathbf{p})$, which parameterizes the distance between measured data and probabilities predicted by quantum theory. We will search for the state(s) located in the closest neighborhood of the data. Repeating steps leading to (3.38) and (3.39) we get a new extremal equation,

$$\sum_j \frac{\partial D}{\partial p_j} \Pi_j \rho = \lambda \rho. \tag{3.46}$$

Of course, different measures D yield different extremal equations whose solutions are also different in general. Lagrange multiplier λ is determined from the normalization condition $\mathrm{Tr}\rho = 1$ as before,

$$\lambda = \sum_i \frac{\partial D}{\partial p_j} p_j.$$

As any composed function $G[D(\mathbf{f}, \mathbf{p})]$ fulfills the same extremal equation (3.46) with the Lagrange multiplier rescaled as $\lambda \frac{dG}{dD}$, we will consider, without the loss of generality, the normalization condition $\lambda = 1$.

The extremal equation (3.46) has the form of a decomposition of the identity operator on the subspace, where the density matrix is defined by

$$\sum_i \frac{\partial D}{\partial p_j} \Pi_j = 1_\rho. \tag{3.47}$$

This resembles the definition of a POVM characterizing a new generalized measurement [31]. To link the above extremalization with quantum theory, let us postulate the natural condition for the quantum expectation values,

$$\mathrm{Tr}\left(\frac{\partial D}{\partial p_j} \Pi_j \rho \right) = f_j. \tag{3.48}$$

This assumption seems to be reasonable. The synthesis of sequential non–compatible observations may be regarded as a new measurement scheme, namely the measurement of the quantum state. The quantum measure D then fulfills the differential equation

$$\frac{\partial D}{\partial p_j} p_j = f_j. \tag{3.49}$$

and singles out the solution in the form

$$D(\mathbf{f}, \mathbf{p}) = \sum_j f_j \ln p_j. \tag{3.50}$$

This is nothing else than the log-likelihood or Kullback–Leibler relative information [21]. Formal requirements of quantum theory, namely the interpretation of the extremal equation as a POVM, result in the concept of maximum likelihood in mathematical statistics. In this sense maximum likelihood estimation may be considered as a new quantum measurement.

3.3.7 Example: Entangled Photons

Let us illustrate ML quantum-state estimation on a simple example of a two-photon entangled state generated by the spontaneous downconversion source of White et al. [32]; for more examples see Badurek *et al.* in this volume. In their experiment, White et al. measured the nominal Bell state $(|HH\rangle + |VV\rangle)/\sqrt{2}$ along sixteen distinct directions: $\{|y_j\rangle\} = \{|HH\rangle, |HV\rangle, |VH\rangle, |VV\rangle, |HD\rangle, |HL\rangle, |DH\rangle, |RH\rangle, |DD\rangle, |RD\rangle, |RL\rangle, |DR\rangle, |DV\rangle, |RV\rangle, |VD\rangle, |VL\rangle\}$. H, V, D, R, and L being horizontal, vertical, diagonal, right circular, and left circular polarization, respectively. Counted numbers of coincidences along these directions can be found in [32]. Let us use their experimental data to estimate the *true* state of entangled photons.

Due to various sources of errors, the true state is expected to differ from the nominal state. Notice that the chosen measurements are not complete, that is $\hat{H} = \sum_j |y_j\rangle\langle y_j| \neq \hat{1}$. This has been taken into account, see the discussion after (3.45).

Starting from the maximally mixed state $(|HH\rangle\langle HH| + |VV\rangle\langle VV| + |HV\rangle\langle HV| + |VH\rangle\langle VH|)/4$, new eigenvalues and eigenvectors of density matrix are found using (3.29) and (3.31). This has been repeated until a stationary point of the iteration process has been attained. The diagonal representation of the reconstructed density matrix reads

$$\rho_e^{\mathrm{ML}} = 0.962\,|\phi_1\rangle\langle\phi_1| + 0.038\,|\phi_2\rangle\langle\phi_2|. \tag{3.51}$$

The other two eigenvalues are zero. The eigenvectors $|\phi_1\rangle$ and $|\phi_2\rangle$ are given in Table 3.1.

The reconstructed density matrix (3.51) agrees well with the qualitative reasoning given in [32]. Namely, the reconstructed state is almost a pure state – a slightly rotated Bell state. The apparent incompatibility of the nominal state with the registered data was interpreted in [32] as the result of possible slight misalignments of the axes of analysis systems with respect to the axes

Table 3.1. Eigenvectors of the reconstructed density matrix.

	$\|\phi_1\rangle$	$\|\phi_2\rangle$
$\|VV\rangle$	$0.696 - 0.027i$	$0.630 + 0.071i$
$\|VH\rangle$	$-0.050 - 0.020i$	$-0.284 + 0.174i$
$\|HV\rangle$	$-0.040 + 0.015i$	$-0.150 - 0.247i$
$\|HH\rangle$	$0.712 - 0.062i$	$-0.634 - 0.035i$

of the downconversion source. This is, of course, reflected in the reconstructed state (3.51), which quantifies such misalignments and might serve for hunting down the errors, and calibrating the experimental setup. For such purposes the error analysis of the presented reconstruction technique becomes crucial. The results of numerical simulations suggest that the fidelity of the reconstructed state corresponding to tens of thousands of detections (like in [32]) is typically better than 0.99. More detailed discussion of estimation errors will be presented in the next section.

Notice also that the reconstructed density matrix (3.51) is semi-positive definite. This should be contrasted with the result of the standard reconstruction given in [32]. The deterministic inversion of (3.4) has the following diagonal representation: $r_1 = 1.022$, $r_2 = 0.068$, $r_3 = -0.065$, $r_4 = -0.024$. The corresponding eigenvectors need not be specified here. Apparently, the standard reconstruction is an unphysical, non-positive matrix. It is worth noticing that the negative eigenvalues are comparable in magnitude with the non-diagonal elements of ρ_e^{ML} in H-V basis, see Table 3.1. This is a nice example of a situation when standard methods fail even though rather high number of particles (tens of thousands) has been registered. ML reconstruction provides always physically sound results. Moreover, it represents genuine quantum measurement of entangled state.

It is also interesting to see how large an improvement upon the standard direct inversion can be expected if more computationally demanding ML estimation is used. Let us illustrate this again on the simple example of two entangled qubits.

To compare two different estimators some measure of the quality of reconstruction is needed. A natural requirement is that the reconstructed state should come as close to the true one as possible. The word "close" here is however somewhat ambiguous. A number of different measures of distance could be used each leading to slightly different results. Let us here mention Jozsa fidelity (generalized overlap) $F = \mathrm{Tr}^2\{\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}\}$ [33] , relative entropy $E_r = \mathrm{Tr}\{\rho_1 \log \rho_2\}$, and trace-class distance $D = \mathrm{Tr}\{|\rho_1 - \rho_2|\}/2$ to name a few. All these measures have good operational meaning. The fidelity is closely related to Wootters statistical distance [34] and also quantifies probability of success of the error-free state discrimination [35, 36]. E_r is linked to the probability of mistaking ρ_1 for ρ_2 in optimum discrimination and large N limit [37]. The last distance D quantifies the reliability of the predictions based on the reconstructed state. It measures how much on the average the probabilities of particular outcomes of measurements predicted by ρ_1 differ from the probabilities predicted by ρ_2. D has the advantage that unlike the fidelity and relative entropy it defines distance also between two non-positive operators. This is important because the direct inversion often yields unphysical results, as we have seen above, and we want to compare the results as they are without any tampering with them.

The results of two simulations are summarized in Fig. 3.1. In the upper panel, the errors of the standard and ML estimation are compared for

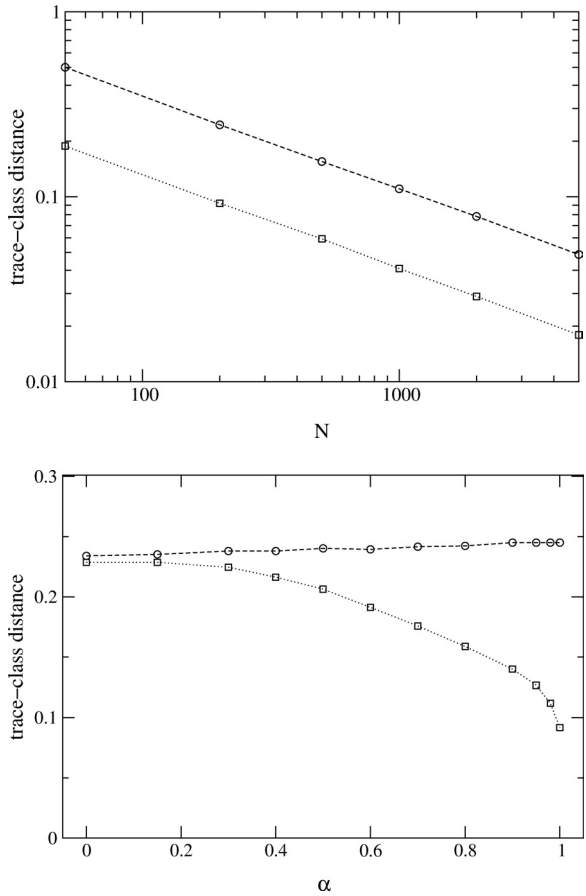

Fig. 3.1. Errors of the ML (squares) and standard (circles) reconstructions in dependence on the (simulated) experimental situation. Upper panel: the source generates Bell states $\bar{\rho} = (|HH\rangle + |VV\rangle)(\langle HH| + \langle VV|)/2$; its intensity N is the independent variable. Lower panel: the intensity is kept constant at $N = 200$ photon pairs; the purity and the degree of entanglement are varied, $\bar{\rho} = (1 + \alpha\sigma_1 \otimes \sigma_1 - \alpha\sigma_2 \otimes \sigma_2 + \alpha\sigma_3 \otimes \sigma_3)/4$.

a different number of detected photon pairs (measuring time). One of the four Bell states was chosen as the true state. Notice that the relative ML improvement upon the standard method remains constant over a wide range of intensities and numerically it is close to the Euler number $e \approx 2.71$. This is a significant improvement worth of technically more involved ML reconstruction procedure. Lower panel shows the dependence of estimation errors on the purity (or degree of entanglement) of the true state for a fixed intensity of the source of photon pairs. Apparently, ML estimation is significantly better for highly pure states while it is just a shade better for highly mixed states.

The improvement increases with moving the true state closer to the boundary of the convex space of physically allowed states. This can be understood as follows: Under the given conditions the standard and ML reconstructions would coincide provided the standard reconstruction would fall within the physically allowed region. This is simply because for the 16 measured observables there is a one to one correspondence between the registered data and the (not necessarily positive) operator representing the source. Only when the standard procedure fails to yield a physical state its result differs from the ML estimate. In that case $\mathcal{L}$ has maximum outside the closed set on which it should be maximized and we thus search for the maximum likely state on the boundary. Obviously the probability that this happens becomes larger for states lying closer to the boundary. Figure 3.1 demonstrates that the ML reconstruction scheme is particularly efficient in that case.

Here one remark seems to be in order. The chosen system of two qubits is a rather simple one of dimension 4. For more complicated systems the boundary of the set of density matrices becomes a *very* complicated set of states and, loosely speaking, almost any state is located near some border. It is reassuring that ML scheme is capable of handling those cases well.

3.4 Estimation Errors and Fisher Information

In the previous section we have discussed several unique features of ML inference. Among them, the high efficiency of ML estimators is perhaps the most important one. Error of ML estimation is tightly connected to an interesting measure of information named after R. A. Fisher [38]. Let us first illustrate this useful concept on the simple example of the estimation of a single parameter θ. Consider a result x of a measurement occurs with probability $p(x|\theta)$. Now suppose that y has been registered. From this result we are to estimate the true value of parameter θ. Any function $\theta(x)$ of data, called estimator, can be used for this purpose. If this experiment was repeated with the same setting of θ, some other value might have been observed which would in general lead to a different estimated value of θ. Of course different estimators can yield very different errors when applied to the same measured data. An important result by Cramér and Rao [39, 40] states that no matter in how clever way an estimator is designed, its error is always bounded from below by

$$(\Delta\theta)^2 \geq \frac{1}{F}, \tag{3.52}$$

where F is the Fisher information defined as

$$F = \left\langle \left(\frac{d}{d\theta} \ln p(x|\theta) \right)^2 \right\rangle_x , \tag{3.53}$$

and the angle brackets denote averaging over data. Inequality (3.52) is called the Cramér-Rao lower bound (CRLB) and its importance stems from the

fact that it provides the ultimate resolution of estimation. Another important theorem due to Fisher [38] says that ML estimators attain this bound asymptotically for large amount of registered data (detected particles). So if many observations are collected, ML estimation is the most efficient estimation strategy.

Let us now evaluate the overall performance of ML quantum-state tomography with the help of CRLB. First we will chose a convenient parameterization of the quantum system under observation. Any density matrix can be decomposed in an orthonormal basis $\{\Gamma_k\}$ of traceless hermitian operators defined on the Hilbert space of the system

$$\rho = 1/p + \sum_{k=1}^{p^2-1} a_k \Gamma_k, \tag{3.54}$$

where p is the dimension of the Hilbert space. For the system of n qubits, a natural basis is the one generated by Pauli spin matrices $\Gamma_k = \sigma_{k_1} \otimes \ldots \otimes \sigma_{k_n}/n$; $k = k_1 + 2^1 k_2 + \ldots + 2^{n-1} k_n$; $k_1, \ldots, k_n = 0, 1, 2, 3$, though other choices are possible [41]. Generators Γ_k have the property of $\mathrm{Tr}\{\Gamma_i \Gamma_j\} = \delta_{ij}$. Real parameters a_k are now the unknown parameters that are to be estimated from the measurement. As in the case of a single parameter, their errors are bounded by CRLB,

$$\mathrm{Var}(\mathrm{a_k}) \geq (\mathbf{F}^{-1})_{kk}, \tag{3.55}$$

where $\mathbf{F}$ is a multidimensional generalization of the Fisher information,

$$F_{kl} = \left\langle \frac{\partial \log \mathcal{L}}{\partial a_k} \frac{\partial \log \mathcal{L}}{\partial a_l} \right\rangle_{\mathbf{n}}. \tag{3.56}$$

$\mathcal{L}$ is the joint probability of observing data $\mathbf{n}$. At the same time this function is the likelihood of the given vector $\mathbf{a}$ parameterizing the quantum state.

Any measurement can be described by POVM elements $\{\Pi_j\}$, where index j labels independent output channels. Statistics of such a generalized measurement is multinomial, see (3.6). The relevant part of the total probability now reads,

$$\mathcal{L} = \prod_j \left(\mathrm{Tr}\rho\Pi_j\right)^{n_j}. \tag{3.57}$$

Let us evaluate the Fisher information matrix in this most general case. Using (3.54) and (3.57) in definition (3.56), we get

$$F_{kl} = \sum_i \sum_j \frac{\mathrm{Tr}\{\Gamma_k \Pi_i\}\mathrm{Tr}\{\Gamma_l \Pi_j\}}{p_i p_j} \langle n_i n_j \rangle. \tag{3.58}$$

Averaging with respect to multinomial distribution with a total of N repetitions of the measurement yields $\langle n_i n_j \rangle = N^2 p_i p_j - N p_i p_j + N p_i \delta_{ij}$, where

δ_{ij} is the Kronecker symbol. Only the last term gives a nonzero contribution since $d(\sum_j p_j)/da_k = 0$ due to the normalization to unity. The Fisher information matrix thus simplifies to

$$F_{kl} = N \sum_j \mathrm{Tr}\{\Gamma_k \Pi_j\}\mathrm{Tr}\{\Gamma_l \Pi_j\}/p_j. \tag{3.59}$$

Unfortunately the variances of a_k alone do not provide enough information needed for placing "error-bars" on the reconstructed state, nor do they suffice for determining how much the predictions based on it could be in error. This is due to the possible correlations of fluctuations of different a_k that might arise in the reconstruction procedure. To avoid such correlations let us decompose the state of the system in a new basis Γ' – such that the corresponding Fisher matrix becomes diagonal. This has a clear physical meaning: On registering counts $\mathbf{n}$, the probability $P(\mathbf{n}|\mathbf{a})$ characterizes the likelihood $\mathcal{L}$ of various states $\mathbf{a}$. In the asymptotical limit of a large amount of accumulated data the likelihood can be approximate by Gaussian distribution. In terms of the Fisher matrix it reads:

$$\log \mathcal{L} = \log P(\mathbf{n}|\mathbf{a}) \approx -\sum_{kl} (a_k - \tilde{a}_k)(a_l - \tilde{a}_l)\mathrm{F}_{kl}, \tag{3.60}$$

where $\tilde{\mathbf{a}}$ specifies the maximum-likelihood solution. Let us define the error volume as the set of density matrices which have likelihoods that do not drop below a certain threshold, $\mathcal{L} \geq$ const. According to (3.60) the error volume is an ellipsoid, whose axes lie in the directions of the diagonalizing basis Γ', their lengths being given by the eigenvalues of the inverse of Fischer information matrix. Obviously, all the states inside the ellipsoid are still likely solution of our inverse problem. The rotation of the error ellipsoid is described by unitary transformation of the generators,

$$\Gamma'_k = U^\dagger \Gamma_k U = \sum_i u_i^k \Gamma_k. \tag{3.61}$$

Orthogonal $(p^2 - 1)$ dimensional matrix of coefficients u_i^k will be denoted by $\mathbf{U}$ to distinguish it from the operator U. In the new basis the state of the system reads,

$$\rho - 1/p = \sum_k a'_k \Gamma'_k = \sum_k a'_k U^\dagger \Gamma_k U = \sum_k \sum_i a'_k u_i^k \Gamma_k. \tag{3.62}$$

Similarly, the measurement can be decomposed in the Γ basis,

$$\Pi_j = \sum_k q_k^j \Gamma_k. \tag{3.63}$$

The transformation properties of Fisher information matrix are simple. Using (3.61) and (3.63) in (3.59), we find that under the rotation of basis F

transforms as follows:

$$F'_{kl} = \sum_m \sum_n u^k_m \left(\sum_j \frac{q^j_m q^j_n}{p_j} \right) u^l_n. \tag{3.64}$$

The term in parenthesis can be recognized as the old Fisher information matrix, so finally we get

$$F' = \mathbf{U} F \mathbf{U}^T. \tag{3.65}$$

As seen, the transition from the old basis to the new one coinciding with the axes of the error ellipsoid (F' becomes diagonal) is provided by matrix $\mathbf{U}$ composed from the eigenvectors of the Fisher matrix. They depend on the measurement that has been done and on the true probabilities which can be (approximatively) calculated from the maximum-likely state. The fluctuations of a'_k, now being independent, then can be propagated to all the quantities of interest using the standard methods of error analysis.

From this point of view, the synthesis of all the quantum observations is equivalent to the registration of the orthogonal observables Γ' defining those axes. In this new representation the Fisher matrix attains the diagonal form, which means that the estimates of the transformed quantum-state "coordinates" $\bar{a}'_k = \mathrm{Tr}\{\bar{\rho}\Gamma'\}$ fluctuate independently. This hints at the possibility to form a single number quantifying the performance of the reconstruction scheme as a whole by adding together those independent errors:

$$\epsilon = \sum_k \langle (\Delta a'_k)^2 \rangle = \sum_k \langle (\Delta a_k)^2 \rangle \geq \mathrm{Tr}\{F^{-1}\}. \tag{3.66}$$

Notice that this number does not depend on the chosen operator basis. Error ϵ has an interesting geometrical interpretation: Hilbert-Schmidt distance

$$d(\rho, \bar{\rho}) = \mathrm{Tr}\{(\rho - \bar{\rho})^2\} \tag{3.67}$$

is a natural metric that can be defined on the space of hermitian operators. Let us evaluate the Hilbert-Schmidt distance between the true state $\bar{\rho}$ and its estimate ρ. Decomposing both states as $\bar{\rho} = \sum_k \bar{a}_k \Gamma_k$, and $\rho = \sum_k a_k \Gamma_k$, we get

$$d(\rho, \bar{\rho}) = \sum_k (a_k - \bar{a}_k)^2. \tag{3.68}$$

The mean distance (error) is then given by averaging d over many repetitions of the estimation procedure, each yielding slightly different estimates of the true state,

$$\langle d \rangle = \sum_k \langle (\Delta a_k)^2 \rangle. \tag{3.69}$$

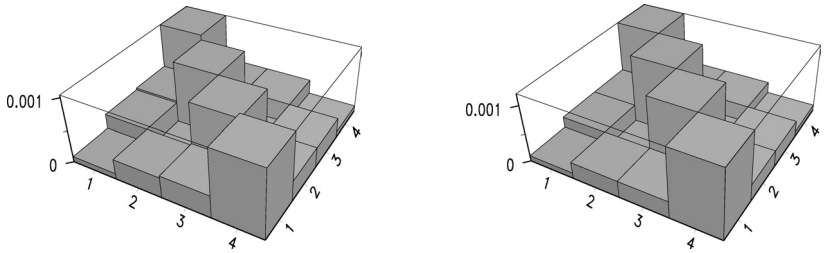

Fig. 3.2. Squared errors (variances) of the elements of the reconstructed density matrix. Left panel: simulation; right panel: CRLB. $\bar{\rho} = (1+0.8\sigma_1 \otimes \sigma_1 - 0.8\sigma_2 \otimes \sigma_2 + 0.8\sigma_3 \otimes \sigma_3)/4$; $N = 2000$ pairs. We note that the diagonal is situated horizontally.

But according to (3.66) this quantity is bounded from below by the trace of the inverse Fisher matrix,

$$\langle d \rangle \geq \mathrm{Tr}\{F^{-1}\}. \tag{3.70}$$

Thus the mean error of the optimal estimation from the measurement of the chosen set of observables is given by the trace of the inverse of the Fisher matrix. In particular, $\mathrm{Tr}\{F^{-1}\}$ quantifies the performance of ML estimation schemes per particle in the limit of a large amount of collected data $N \gg 1$. Its reciprocal then quantifies information about the unknown quantum state acquired in the measurement.

Also notice that since each new observed quantum system contributes equally to the Fisher information, see (3.59), it is possible to define total information about the quantum state gained per particle as follows:

$$I = \frac{1}{N\mathrm{Tr}\{F^{-1}\}} \tag{3.71}$$

This quantity measures the quality of a given reconstruction scheme. One can use it for comparing performances of different sets of tomographic measurements or for investigating their invariance properties.

As an example, let us compare errors calculated from the Fisher information with actual errors of the quantum estimation of a state of two entangled qubits. The chosen set of 16 projections is the same as that discussed in the previous section; for details see [32]. The results for a slightly mixed entangled true state and observation of 2000 pairs of photons are shown in Fig. 3.2. Right panel shows variances of the estimated elements of the density matrix in $|HH\rangle$, $|HV\rangle$, $|VH\rangle$ and $|VV\rangle$ basis predicted by CRLB. Left panel shows actual errors that were obtained by averaging over many repeated estimations. Notice that the reconstructed values of the diagonal elements are much more reliable than the rest. This is perhaps caused by the choice of the set of measured observables since the four basis vectors can be found among them. It is also clear that Fisher information in this case gives reliable error

estimates. In a completely analogous fashion one could use Fisher information to derive error estimates for more complicated inverse problems such as the estimation of quantum operations or measurement that will be described in the following.

3.5 Estimation of Quantum Processes

The quantum-state tomography discussed in the preceeding sections provides a precise quantitative method of characterization of sources of quantum states. However, the preparation of quantum states is often only the initial part of experiments. Typically, one would like to manipulate the quantum states by applying some unitary transformations, or transmit the quantum states via some channel to a distant location. The holy grail of the quantum information science is to construct a quantum computer which would provide an ultimate control over the evolution of quantum states. In practice, errors will unavoidably occur, stemming from decoherence, losses, etc. Thus, the quantum state ρ will typically become entangled with the environment and the evolution of ρ cannot be described as a unitary transformation of ρ. Therefore, a more general theory of transformations of quantum states must be invoked.

A very general framework is provided by the formalism of quantum operations (also called quantum processes or quantum channels) [42]. Assuming that the environment interacting with ρ is not initially entangled or correlated with the device that prepares the state ρ, the most general transformation of a quantum state ρ allowed by the laws of the quantum mechanics is the so-called linear completely positive (CP) map $\mathcal{E}$. The condition that the environment and the state preparation device are not correlated is satisfied in the vast majority of the experiments and the formalism of quantum operations is thus applicable in practically all cases.

Given current interest in the quantum-information processing it is thus of paramount importance to develop tools for characterization of quantum processes. The concept of quantum process tomography has been introduced independently by Nielsen and Chuang [43] and by Poyatos *et al.* [44]. The reconstruction of quantum process has many practical applications, ranging from the probing of quantum communication channels to the evaluation of the performance of quantum gates and eventually debugging of quantum computers [43–50]. Several experimental demonstrations of the quantum process tomography in NMR [51, 52] and quantum optics systems [53–55] have been reported recently.

Most proposed quantum-process reconstruction techniques are based on direct linear inversion of experimental data. These methods are conceptually simple but may yield unphysical results. On the other hand, all necessary properties of the deterministic quantum transformations, namely the complete positivity and trace preservation can be naturally incorporated within

the maximum-likelihood approach as the appropriate constraints [56, 57]. In what follows we shall show that the ML estimation of quantum process can be formally formulated in a very similar way as the ML estimation of quantum states.

The discussion is greatly facilitated by the Jamiolkowski isomorphism [58] that associates a positive-semidefinite operator E with a CP map $\mathcal{E}$. This formalism is briefly reviewed below. The main formal difference between estimation of processes and states lies in a higher number of constraints involved in the former case. The trace-preservation condition gives rise to altogether d^2 real constraints, where d is the dimension of the Hilbert space $\mathcal{H}$ of the input states. Rigorous ML estimation should take all these constraints into account properly. We shall demonstrate how this can be accomplished in practice and how the ML estimate of E can be efficiently numerically computed from the acquired experimental data. We will also touch upon several topics of current interest, such as the probing of quantum processes by entangled states and the reconstruction of trace-decreasing CP maps, which describe probabilistic (conditional) quantum operations.

3.5.1 Jamiolkowski Isomorphism

The deterministic quantum operation $\mathcal{E}$ must satisfy the following conditions (see [42] for an excellent detailed discussion):

(i) *Positivity.* If $\rho_{\rm in} \geq 0$ then $\rho_{\rm out} \equiv \mathcal{E}(\rho_{\rm in}) \geq 0$, the map transforms density operators onto density operators.

(ii) *Complete positivity.* The positivity is not a sufficient condition for $\mathcal{E}$ to be a physical map. The reason is that the input state $\rho_{\rm in}$ may be a part of a maximally entangled state $|\psi\rangle_{AB}$. Let $\mathcal{H}$ and $\mathcal{K}$ denote the Hilbert spaces of input and output states, respectively, and denote $d = \dim\mathcal{H}$. The map $\mathcal{E}$ must preserve positive semidefiniteness of ρ_{AB} when applied to one part of ρ_{AB}. We say that $\mathcal{E}$ is completely positive iff the extended map $\mathcal{M} = \mathcal{I}_{h,A} \otimes \mathcal{E}_B$ is positive for all h, where h is a dimension of the auxiliary Hilbert space $\mathcal{H}'_A$, $\mathcal{I}$ denotes an identity operation and $\mathcal{M}$ acts on operators on the Hilbert space $\mathcal{H}'_A \otimes \mathcal{H}_B$. In fact, it suffices to take $h = d$.

(iii) *Trace preservation.* The deterministic maps must preserve the total probability, hence $\mathrm{Tr}(\rho_{\rm out}) = \mathrm{Tr}(\rho_{\rm in})$ must hold for all $\rho_{\rm in}$.

There are several ways how the CP map can be described mathematically. Physically, every linear trace-preserving CP map can be realized as a unitary operation U_E on an extended Hilbert space $\mathcal{H} \otimes \mathcal{A}$ of the system and ancilla, where the ancilla is initially prepared in a blank pure state $|0\rangle_A$,

$$\rho_{\rm out} = \mathrm{Tr}_A[U_E\, \rho_{\rm in} \otimes |0\rangle_A\langle 0|\, U_E^\dagger]. \tag{3.72}$$

There is a big freedom in the choice of U_E so this description of $\mathcal{E}$ is not very suitable for our purposes.

Another widely used formalism is the Kraus decomposition which states that every CP map can be expressed in terms of a set of operators A_j as follows,

$$\rho_{\rm out} = \sum_j A_j \rho_{\rm in} A_j^\dagger.$$

The trace preservation condition amounts to the constraint $\sum_j A_j^\dagger A_j = I$. The Kraus decomposition is particularly useful when studying quantum noise and quantum error correction. However, the decomposition is not unique and infinitely many sets of operators $\{A_j\}$ can describe the same map $\mathcal{E}$.

The mathematical representation of CP maps that we shall employ relies on the isomorphism between linear CP maps $\mathcal{E}$ from operators on the Hilbert space $\mathcal{H}$ to operators on the Hilbert space $\mathcal{K}$ and positive semidefinite operators E on the Hilbert space $\mathcal{H} \otimes \mathcal{K}$. This representation is sometimes referred to as Jamiolkowski isomorphism. Its main advantage is that it eliminates all the free parameters and provides a compact description of the CP map $\mathcal{E}$.

Let us investigate what happens if the map $\mathcal{E}$ is applied to one part of the maximally entangled state on $\mathcal{H} \otimes \mathcal{H}$. Let us define

$$|\psi\rangle_{AB} = \sum_{j=1}^{d} |j\rangle_A |j\rangle_B \tag{3.73}$$

and consider the positive semidefinite operator E on $\mathcal{H} \otimes \mathcal{K}$,

$$E = \mathcal{I}_A \otimes \mathcal{E}_B(|\psi\rangle_{AB}\langle\psi|). \tag{3.74}$$

By definition, E is positive semidefinite, $E \geq 0$. One immediately finds that

$$E = \sum_{j,k} |j\rangle\langle k| \otimes \mathcal{E}(|j\rangle\langle k|). \tag{3.75}$$

It is an easy exercise, which we leave for the reader, to verify that the input-output transformation can be expressed in terms of the operator E as

$$\rho_{\rm out} = {\rm Tr}_{\mathcal{H}}[E\rho_{\rm in}^T \otimes I], \tag{3.76}$$

where T denotes transposition in the basis $|j\rangle$ and ${\rm Tr}_{\mathcal{H}}$ denotes the partial trace over the input Hilbert space. The deterministic quantum transformations preserve the trace of the transformed operators, ${\rm Tr}_{\mathcal{K}}[\rho_{\rm out}] = {\rm Tr}_{\mathcal{H}}[\rho_{\rm in}]$. Since this must hold for any $\rho_{\rm in}$ the operator E must satisfy the condition

$${\rm Tr}_{\mathcal{K}}[E] = I_{\mathcal{H}}, \tag{3.77}$$

where $I_{\mathcal{H}}$ is an identity operator on space $\mathcal{H}$. The condition (3.77) effectively represents $({\rm dim}\mathcal{H})^2$ real constraints as advertised earlier.

3.5.2 Reconstruction of Trace-Preserving CP Map

Having established the mathematical formalism, we can now proceed to the reconstruction of the CP map. The experimental setup that we have in mind is as follows. Some sources prepare various input states ρ_m that are used for the determination of the quantum process. Here we assume that we have full knowledge of ρ_m. Later on, we will remove this assumption and will analyze a more complex scenario when one simultaneously reconstructs the states and the process. The states ρ_m are sent through the quantum channel $\mathcal{E}$. Measurements described by POVMs Π_{ml} are carried out on each corresponding output state $\mathcal{E}(\rho_m)$. Let f_{ml} denote the relative frequency of detection of the POVM element Π_{ml}. These frequencies approximate the theoretical detection probabilities

$$p_{ml} = \mathrm{Tr}[\mathcal{E}(\rho_m)\Pi_{ml}] = \mathrm{Tr}\left[E\,\rho_m^{\mathrm{T}} \otimes \Pi_{ml}\right], \tag{3.78}$$

where we used (3.76). The quantum process E should be reconstructed from the knowledge of the probe states ρ_m and the measured frequencies f_{ml}.

With the help of the Jamiolkowski isomorphism we may formulate the exact maximum-likelihood principle for estimated CP map E in a very simple and transparent form. The estimated operator E should maximize the constrained log-likelihood functional [56, 57]

$$\mathcal{L}_{\mathrm{c}}[f_{ml}, p_{ml}(E)] = \sum_{m,l} f_{ml} \ln p_{ml} - \mathrm{Tr}[\Lambda E], \tag{3.79}$$

where $\Lambda = \lambda \otimes I_{\mathcal{K}}$ and λ is the (hermitian) matrix of Lagrange multipliers that account for the trace-preservation condition (3.77). The extremal equations for E can be obtained by varying functional (3.79) with respect to E, similarly to the case of quantum state estimation. This leads to the extremal equation

$$E = \Lambda^{-1} K E, \tag{3.80}$$

where the operator K reads

$$K = \sum_{m,l} \frac{f_{ml}}{p_{ml}} \rho_m^{\mathrm{T}} \otimes \Pi_{ml}. \tag{3.81}$$

Further we have from (3.80) and Hermicity that $E = EK\Lambda^{-1}$. When we insert this expression in the right-hand side of (3.80), we finally arrive at symmetrical expression suitable for iterations,

$$E = \Lambda^{-1} K E K \Lambda^{-1}. \tag{3.82}$$

The Lagrange multiplier λ must be determined from the constraint (3.77). On tracing (3.82) over space $\mathcal{K}$ we obtain quadratic equation for λ which

may be solved as

$$\lambda = \left(\mathrm{Tr}_{\mathcal{K}}[KEK]\right)^{1/2}, \tag{3.83}$$

where we take the positive root, $\lambda > 0$. The system of coupled (3.82) and (3.83) may be conveniently solved numerically by means of repeated iterations, starting from some unbiased CP map, for example $E^{(0)} = I_{\mathcal{H}\otimes\mathcal{K}}/(\dim\mathcal{K})$. It is important to note that (3.82) preserves the positive semidefiniteness of E and also the constraint $\mathrm{Tr}_{\mathcal{K}}[E] = I_{\mathcal{H}}$ is satisfied at each iteration step.

The feasibility of this reconstruction technique has been confirmed by extensive numerical simulations for various single- and two-qubit CP maps $\mathcal{E}$, various sets of probe density matrices ρ_m and POVMs Π_{ml}. As an illustrative example, we describe the reconstruction of a single-qubit trace preserving CP map. The operator E is characterized by 16 real parameters. The trace-preservation condition yields four constraints, which leaves 12 independent parameters. The maximum-likelihood estimation then amounts to finding the maximum of the log-likelihood functional in a 12-dimensional space, with highly non-trivial boundary defined by $E \geq 0$.

In the numerical simulations, we have considered six different probe states — the eigenstates of three Pauli matrices σ_x, σ_y, and σ_z. $3N$ copies of each input state are used. On each corresponding output state, a spin projection along axes x, y and z is measured N times. The detected frequencies were generated by means of Monte Carlo simulations. The maximum likelihood estimate of the true process has been obtained by iteratively solving the nonlinear extremal equations (3.82) and (3.83.

As the first example, consider the partially depolarizing channel (DC)

$$E_{DC} = \eta\mathcal{I} + (1-\eta)\mathcal{O}, \tag{3.84}$$

where $\mathcal{I}$ is the identity operation while $\mathcal{O}$ denotes the totally depolarizing channel that maps all density matrices onto maximally mixed state, $\mathcal{O}(\rho) = I/2$. The results of the simulated quantum process tomography are shown in Fig. 3.3. For observations on $N = 100$ copies of each probe state, the reconstruction works very well and the estimated process is typically very close to the true one, as confirmed by the small values of the estimation errors displayed in Fig. 3.3.

Our second example is the amplitude damping channel (ADC) that describes a decay process. This channel naturally arises when the qubit is represented by two levels of an atom, where $|1\rangle$ is the ground state while $|0\rangle$ is the excited state. Even in vacuum, the atom may emit a photon and decay to the ground state due to the coupling of the atom to the vacuum fluctuations of the electromagnetic field. If the spontaneous emission occurs with probability η^2, then in the basis $|00\rangle$, $|10\rangle$, $|01\rangle$, $|11\rangle$ the ADC is described by the operator

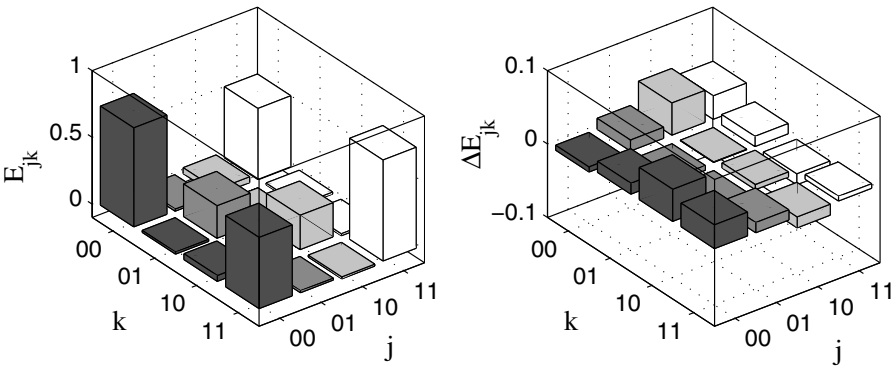

Fig. 3.3. Reconstruction of a depolarizing channel (3.84) with $\eta = 0.5$. $3N = 300$ copies of each of six input probe states have been used – see text for more details. The left panel shows the matrix elements of the reconstructed operator E and the right panel displays the difference between reconstructed and exact operators $\Delta E = E_{\rm est} - E_{\rm true}$. Only real parts of complex elements are shown.

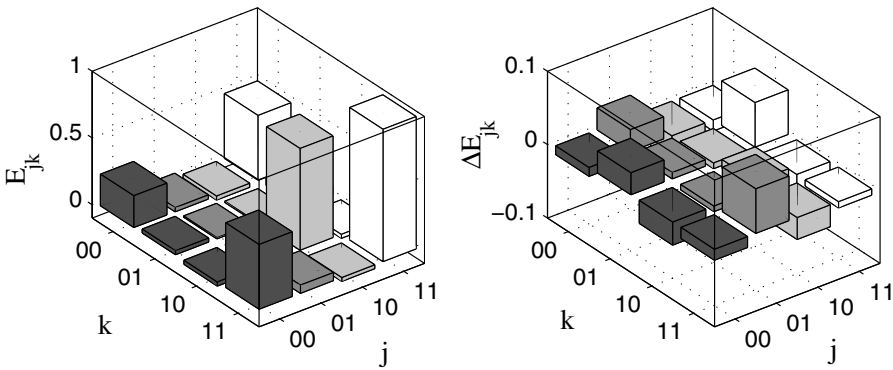

Fig. 3.4. Reconstruction of an amplitude damping channel (3.85) with $\eta = 0.5$. Similarly as in Fig. 3.3, only real parts of complex elements are shown.

$$E_{ADC} = \begin{pmatrix} \eta^2 & 0 & 0 & \eta \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1-\eta^2 & 0 \\ \eta & 0 & 0 & 1 \end{pmatrix}. \tag{3.85}$$

An example of the simulated tomography of the amplitude damping channel is given in Fig. 3.4. Similarly as in the case of the depolarizing channel, the maximum-likelihood estimate is in good agreement with the true process. An important point that should be emphasized is that in both examples, the estimated process is really a trace preserving completely positive map and all the constraints imposed by quantum mechanics are satisfied. In particular, $E \geq 0$

and $\mathrm{Tr}_{\mathcal{K}}[E] = I_{\mathcal{H}}$ holds. As discussed in previous sections, the maximum-likelihood estimation is asymptotically (for large number of probes N) the optimal estimation method because it saturates the Cramér-Rao bound. All these facts illustrate the advantages of this approach in comparison to simpler data processing techniques.

3.5.3 Entangled Probes

The above discussed reconstruction of a single-qubit CP map bears striking resemblance to the estimation of the entangled-two qubit state which was discussed in Sect. 3.3. This similarity is direct consequence of the Jamiolkowski isomorphism. Following the definition (3.74), the operator E representing the CP map can be in fact prepared physically in the laboratory if we first prepare a maximally entangled state on the Hilbert space $\mathcal{H} \otimes \mathcal{H}$ and then apply the CP map to one part of this entangled state. In this way, the quantum-process tomography can be transformed to the quantum-state tomography. More generally, this suggests that it may be useful to employ entangled quantum states as probes of an unknown quantum process [48, 49].

Let $\rho_{m,AB}$ denote the entangled state on the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ that serves as a probe of the CP map $\mathcal{E}$ that is applied to the subsystem A. A joint generalized measurement described by the POVMs $\mathit{\Pi}_{ml}$ if performed on the output Hilbert space $\mathcal{K} \otimes \mathcal{H}_B$. The log-likelihood functional still has the form (3.79), only the formula for the probability p_{ml} changes to

$$p_{ml} = \mathrm{Tr}_{\mathcal{H}_A \mathcal{H}_B \mathcal{K}}[(E \otimes I_{\mathcal{H}_B})(\rho_{m,AB}^{T_A} \otimes I_{\mathcal{K}})(I_{\mathcal{H}_A} \otimes \mathit{\Pi}_{ml})], \tag{3.86}$$

where T_A stands for the partial transposition in the subsystem A. Consequently, the operator K appearing in the extremal equations (3.82) and (3.83) must be calculated as follows,

$$K = \sum_{l,\,m} \frac{f_{ml}}{p_{ml}} \mathrm{Tr}_{\mathcal{H}_B}[(\rho_{m,AB}^{T_A} \otimes I_{\mathcal{K}})(I_{\mathcal{H}_A} \otimes \mathit{\Pi}_{ml})]. \tag{3.87}$$

Apart from these modifications of p_{ml} and K one can proceed as before and solve (3.82) and (3.83) by means of repeated iterations.

3.5.4 Probabilistic Operations

The transformations of the quantum states may be probabilistic. Such a conditional map $\mathcal{E}$ succeeds with probability

$$q_m \equiv \mathrm{Tr}[\mathcal{E}(\rho_m)] = \mathrm{Tr}[E\, \rho_m^T \otimes I] \tag{3.88}$$

and fails with probability $1 - q_m$. Since $q_m \leq 1$, the inequality $\mathrm{Tr}_{\mathcal{K}}[E] \leq I_{\mathcal{H}}$ must hold. Note that the re-normalized output state $\rho_{m,\mathrm{out}} = \mathcal{E}(\rho_m)/q_m$ is a

nonlinear function of ρ_m. Probabilistic transformations arise in many areas of quantum information processing and quantum state manipulation. We can mention the conditional generation of quantum states [59–61], entanglement distillation protocols [62,63], and the probabilistic scheme for quantum computing with linear optics and single photons proposed by Knill, Laflamme and Milburn [64]. Three alternative ways for manipulating CP decreasing maps will be presented.

The tomography of probabilistic CP maps can be in fact re-formulated as a tomography of a trace preserving CP map provided that we know in each case whether the probabilistic map succeeded or failed. The observation of a failure is a valid measurement outcome which can be associated with a POVM element $\Pi_\emptyset$. Without loss of generality, we may assume that the total output Hilbert space $\mathcal{K}_{\rm tot}$ is a direct sum of the Hilbert space $\mathcal{K}$ and a one-dimensional space $\mathcal{K}_{\rm fail}$ spanned by $|\emptyset\rangle$, $\mathcal{K}_{\rm tot} = \mathcal{K} \oplus \mathcal{K}_{\rm fail}$. Thus whenever the operation fails, we assume that the output state is $|\emptyset\rangle$. The state $|\emptyset\rangle$ serves as a sink for all unsuccessful trials and $\Pi_\emptyset = |\emptyset\rangle\langle\emptyset|$. The POVM $\{\Pi_{ml}\}_{l=1}^N$ that describes measurement of the output states in the Hilbert space $\mathcal{K}$ is completed by adding the element $\Pi_\emptyset$ so that the new POVM satisfies the closure relation on $\mathcal{K}_{\rm tot}$.

Instead of the trace-decreasing map $\mathcal{E}$ we shall reconstruct the extended trace-preserving map $\tilde{\mathcal{E}}$ that maps operators on $\mathcal{H}$ onto operators on $\mathcal{K}_{\rm tot}$. Let $f_{m\emptyset}$ denote the number of observed failures of the application of the map $\mathcal{E}$ to the probe state ρ_m. The constrained log-likelihood functional can be obtained as a simple extension of (3.79),

$$\mathcal{L}_{\rm c}[\tilde{E}] = \sum_{m,l} f_{ml} \ln p_{ml} + \sum_m f_{m\emptyset} \ln p_{m\emptyset} - {\rm Tr}[\Lambda \tilde{E}], \tag{3.89}$$

where $p_{ml} = {\rm Tr}[\tilde{E}\rho_m^T \otimes \Pi_{ml}]$ and $p_{m\emptyset} = {\rm Tr}[\tilde{E}\,\rho_m^T \otimes \Pi_\emptyset]$. The trace-preserving CP map $\tilde{\mathcal{E}}$ can be reconstructed with the use of the iterative algorithm described in the Sect. 3.5.2. From the estimated operator $\tilde{E}$ we can extract the sought after operator E by projecting $\tilde{E}$ onto the subspace $\mathcal{H} \otimes \mathcal{K}$.

In many experiments, however, we cannot determine how often the conditional map succeeded. This problem typically arises in the experiments with photon pairs generated by means of spontaneous parametric downconversion. This process is random and we do not know whether the photon pair was generated until we detect it. In certain experimental setups, this prevents us from measuring the number of unsuccessful events $f_{m\emptyset}$. Without knowing $f_{m\emptyset}$, we cannot use the trick with extension to the trace-preserving map and we must try to estimate directly the trace-decreasing map $\mathcal{E}$.

Consequently, the constraint (3.77) should be replaced with ${\rm Tr}_{\mathcal{K}} E \le I_{\mathcal{H}}$. If we do not know the frequency of failures, E can be determined only up to an overall normalization prefactor and the constraint ${\rm Tr}_{\mathcal{K}} E \le I_{\mathcal{H}}$ is thus irrelevant. In the log-likelihood (3.79) we must replace p_{ml} with the re-normalized

probabilities p_{ml}/q_m. We thus have to maximize

$$\mathcal{L} = \sum_{l,\,m} f_{ml}(\ln p_{ml} - \ln q_m), \tag{3.90}$$

under the constraint $E \geq 0$.

The extremal equation for E can be derived by introducing the decomposition $E = A^\dagger A$ and varying (3.90) with respect to A. We obtain $RE = SE$, where

$$S = \sum_{m,l} \frac{f_{ml}}{q_m} \rho_m^T \otimes I. \tag{3.91}$$

This operator replaces the Lagrange multiplier Λ. Since S is invertible, we can express $E = S^{-1}RE$ and symmetrize the extremal equation which yields

$$E = a^{-1} S^{-1} RERS^{-1}, \tag{3.92}$$

where $a = \mathrm{Tr}[S^{-1}RERS^{-1}]$ is a normalization factor chosen such that $\mathrm{Tr}[E] = 1$.

Alternatively, the ML estimate of E can be determined by the analogue of the EMU algorithm. This approach can be applied here because we are not bound by the constraints (3.77) and we can thus consider independently the variations of the eigenstates r_j^2 and eigenvalues $|e_j\rangle$ of $E = \sum_j r_j^2 |e_j\rangle\langle e_j|$. Assume small variations $r_j \to r_j + \delta r_j$ and $|e_j\rangle \to \exp(i\epsilon H)|e_j\rangle$, where H is a Hermitean operator. The corresponding variation of $\mathcal{L}$ is given by

$$\delta\mathcal{L} = 2\sum_j r_j \delta r_j \sum_{l,\,m} \langle e_j|X_{ml}|e_j\rangle + i\epsilon\,\mathrm{Tr}\left(H \sum_{m,l} [E, X_{ml}]\right) \tag{3.93}$$

where

$$X_{ml} = \frac{f_{ml}}{p_{ml}} \rho_m^T \otimes \Pi_{ml} - \frac{f_{ml}}{q_m} \rho_m^T \otimes I. \tag{3.94}$$

From (3.93) we can deduce δr_j and H that will increase the log-likelihood,

$$r_j \to r_j + \eta r_j \sum_{m,l} \langle e_j|X_{ml}|e_j\rangle, \qquad |e_j\rangle \to \exp(-\epsilon \sum_{m,l} [E, X_{ml}])|e_j\rangle. \tag{3.95}$$

At each iteration step, η and ϵ can be further optimized in order to achieve maximum increase of $\mathcal{L}$. Moreover, after each iteration, we can re-normalize $E \to E/\mathrm{Tr}[E]$. Note, that this does not change the value of $\mathcal{L}$.

3.5.5 Unknown Probes

Up to now quantum states and processes have been treated independently. Widely accepted strategy of how to approach a complex problem is to specify

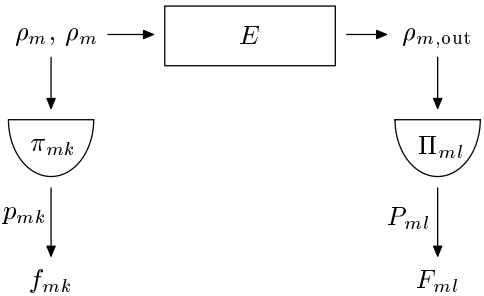

Fig. 3.5. Scheme of setup for the generalized measurement of quantum process using unknown quantum states as probes.

some partial subproblems, address them separately and merge the solutions. This technique usually gives a good answer in the technical sense. Though this is possible even in quantum theory, there are no fundamental reasons for such a factorization. To consider the full problem without splitting it into isolated subproblems is technically more advanced but could be advantageous. This strategy can be demonstrated on the synthesis of the problems treated separately in the previous paragraphs. Let us assume the estimation of the generic quantum process (CP map) E with the help of a set of probe states ρ_m, identity of which is also unknown [57]. What is only known to the experimenters are the output of certain measurements performed on the ensemble of probe states and on the ensemble of transformed probe states. All the physically relevant results can be derived exclusively from the acquired data, where input states and their transformation are inseparably involved.

In accordance with the theory presented above let us consider the set of probe states ρ_m on the space $\mathcal{H}$. By means of unknown quantum process E these states are transformed onto output states $\rho_{m,\mathrm{out}}$ in the space $\mathcal{K}$. The observation must be more complex now involving the detection on the ensemble of both the input and the output states. For this purpose the corresponding POVM elements will be denoted by π_{mk} and $\mathit{\Pi}_{ml}$. The diagram involving detected signals and measurements is shown in Fig. 3.5. Let f_{mk} denote the relative frequency of detection of the POVM element π_{mk} in the input space $\mathcal{H}$ and F_{ml} denote the relative frequency of detection of the POVM element $\mathit{\Pi}_{ml}$ in the output space $\mathcal{K}$. The frequencies f_{mk}, $\sum_k f_{mk} = 1$, and F_{ml}, $\sum_l F_{ml} = 1$, approximate the true probabilities p_{mk} and P_{ml} of individual outcomes, respectively,

$$p_{mk} = \mathrm{Tr}_{\mathcal{H}}[\rho_m \pi_{ml}]\,, \quad P_{ml} = \mathrm{Tr}_{\mathcal{K}}[\rho_{m,\mathrm{out}} \mathit{\Pi}_{ml}] = \mathrm{Tr}\big[E(\rho_m^{\mathrm{T}} \otimes \mathit{\Pi}_{ml})\big]\,, \quad (3.96)$$

where the relation (3.76) was used. The estimated process E and probe states ρ_m should maximize the constrained log-likelihood functional

$$\mathcal{L} = \sum_{m,k} f_{mk} \ln p_{mk} + \sum_{m,l} F_{ml} \ln P_{ml} - \sum_m \mu_m \mathrm{Tr}\,[\rho_m] - \mathrm{Tr}\,[\mathit{\Lambda} E]\,. \quad (3.97)$$

The additivity of log likelihood reflects the independence of observations performed on the input and output states. The Lagrange multipliers μ_m and $\Lambda = \lambda \otimes I_{\mathcal{K}}$ fix necessary constraints—the trace normalization of the states, $\mathrm{Tr}[\rho_m] = 1$, and the trace-preserving property (3.77) of the process E.

Maximization of this functional again leads to coupled extremal equations and iterative algorithm that preserves all necessary properties of estimated quantum states ρ_m and quantum process E. It can also be shown that combined ML strategies yield results superior to other methods [57].

3.6 Estimation of Quantum Measurements

Let us imagine that we possess an apparatus that performs some measurement on a certain quantum mechanical system. We are not sure which measurement the device carries out and we would like to calibrate it. This problem is in a sense complementary to the reconstruction procedures for quantum states discussed in the previous sections. Here, the role of the states and the measuring apparatus will be inverted and we shall use known states ρ_m to probe the measuring device.

3.6.1 Calibration of the Measuring Apparatus

Suppose that the apparatus can respond with L different measurement outcomes. As is well known from the theory of quantum measurement [31, 65], such a device is completely POVM whose L elements Π_l, $l = 1, \ldots, L$ govern the measurement statistics. Let us recall that the probability p_{ml} of the measurement readout Π_l when measuring the quantum state with density matrix ρ_m can be expressed as

$$p_{ml} = \mathrm{Tr}[\Pi_l \rho_m]. \tag{3.98}$$

The POVM elements are positive semi-definite hermitian operators, $\Pi_l \geq 0$, which decompose the identity operator,

$$\sum_{l=1}^{L} \Pi_l = I. \tag{3.99}$$

This ensures the probability normalization $\sum_{l=1}^{L} p_{ml} = 1$.

The general strategy to determine the POVM consists of performing a set of measurements on various known quantum states and then inferring Π_l from the collected experimental data [66]. The POVM can be easily estimated by direct linear inversion of (3.98). Let f_{ml} denote the total number of detections of Π_l for the measurements performed on the quantum state ρ_m. Assuming that the theoretical detection probability p_{ml} given by (3.98) can be replaced

with the corresponding relative frequency, we may write

$$\mathrm{Tr}\,[\Pi_l \rho_m] \equiv \sum_{i,j=1}^{d} \Pi_{l,ij}\rho_{m,ji} = \frac{f_{ml}}{\sum_{l'=1}^{L} f_{ml'}}, \tag{3.100}$$

where d is the dimension of the Hilbert space on which the operators Π_l act. Formula (3.100) establishes a system of linear equations for the matrix elements of the operators Π_l. If sufficient amount of data is available then (3.100) can be inverted (e.g. by the least squares method [67,68]) and we can determine Π_l. This approach is a direct analogue of linear reconstruction algorithms devised for quantum-state and quantum-process reconstructions. The linear inversion is simple and straightforward, but it does not guarantee the positivity of the reconstructed POVM elements. Consequently, the linear estimation may lead to unphysical POVM, predicting negative probabilities p_{ml} for certain input quantum states. To avoid such problems, one should resort to a more sophisticated reconstruction strategy. In what follows we show how to calibrate the measuring apparatus with the use of the maximum-likelihood (ML) estimation [69]. The reconstruction of the POVM is thus another example of the remarkable utility and versatility of the estimation methods based on the maximum-likelihood principle.

3.6.2 Maximum-Likelihood Estimation of the POVM

The estimated operators Π_l should maximize the log-likelihood functional [69]

$$\mathcal{L}[\{\Pi_l\}] = \sum_{l=1}^{L}\sum_{m=1}^{M} f_{ml} \ln p_{ml}, \tag{3.101}$$

where M is the number of different quantum states ρ_m used for the reconstruction and note that p_{ml} depends on Π_l through (3.98).

We proceed as before and derive the extremal equations for the most likely POVM. The constraint (3.99) has to be incorporated by introducing a hermitian operator λ whose matrix elements are the Lagrange multipliers. We thus have to find the maximum of the constrained log-likelihood functional

$$\mathcal{L}'[\{\Pi_l\}] = \mathcal{L}[\{\Pi_l\}] - \sum_{l=1}^{L} \mathrm{Tr}[\lambda \Pi_l]. \tag{3.102}$$

As usual, the extremal equations for Π_l can be derived by introducing the decomposition $\Pi_l = A_l^\dagger A_l$, varying $\mathcal{L}'$ with respect to A_l, and setting the variations equal to zero. After some algebraic manipulations, one obtains

$$\Pi_l = \lambda^{-1} R_l \Pi_l, \tag{3.103}$$

where

$$R_l = \sum_{m=1}^{M} \frac{f_{ml}}{p_{ml}} \rho_m.$$

The extremal equation can be symmetrized by substituting $\Pi_l = \Pi_l R_l \lambda^{-1}$ into the right-hand side of (3.103), and we have

$$\Pi_l = \lambda^{-1} R_l \Pi_l R_l \lambda^{-1}. \tag{3.104}$$

The Lagrange multiplier λ must be calculated self-consistently from the constraint (3.99). This yields

$$\lambda = \left(\sum_{l=1}^{L} R_l \Pi_l R_l \right)^{1/2}, \tag{3.105}$$

where the positive branch of the square root is taken. The extremal eqations (3.104) and (3.105) can be conveniently solved by means of repeated iterations. We emphasize that the conditions $\Pi_l \geq 0$ and $\sum_l \Pi_l = I$ are exactly fulfilled at each iteration step $\Pi_l \to \lambda^{-1} R_l \Pi_l R_l \lambda^{-1}$.

If there exists a POVM whose elements Π_l^e exactly solve linear equations (3.100) then the ML estimate coincides with Π_l^e. In this case it holds for all l, m that

$$p_{ml} = \frac{f_{ml}}{\sum_{l'=1}^{L} f_{ml'}}. \tag{3.106}$$

On inserting this expression into (3.103), we find after some algebra that the set of L equations (3.103) reduces to the formula for the operator of Lagrange multipliers

$$\lambda = \sum_{m=1}^{M} \sum_{l=1}^{L} f_{ml} \rho_m. \tag{3.107}$$

The principal advantage of the ML estimation lies in its ability to handle correctly any experimental data and to provide reliable estimates in cases when linear algorithms fail. As noted before, the linear inversions may provide unphysical estimates, namely operators Π_l, which 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 in the case of von Neumann measurement, when the operators Π_l are rank-one projectors and $d - 1$ eigenvalues of each Π_l are equal to zero. For sufficiently large number of measured data, the linear estimate of a matrix element of Π_l is random variable with Gaussian distribution centered at the true value. In the basis where the projector Π_l is diagonal, its $d - 1$ diagonal elements

fluctuate around zero. It follows that in most cases at least one diagonal element is negative and the linear inversion yields non-positive $\{\Pi_l\}$, 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 (3.100). The frequencies f_{ml} are fluctuating quantities with multinomial distribution characterized by probabilities p_{ml}. In the experiment we can, in principle, detect any f_{ml}. However, some sets of relative frequencies do not coincide with any theoretical probabilities (3.98) calculated for given quantum states ρ_m used for the calibration. In other words, sometimes there does not exist a POVM that would yield probabilities p_{ml} equal to detected relative frequencies and a direct linear inversion of (3.100) may then provide an unphysical result. The observation of several different quantum states by a single measuring apparatus is equivalent with the measurement of several noncommuting observables on many copies of a given quantum state. Thus the ML estimation of the quantum measurement can be interpreted as a synthesis of information from mutually incompatible observations [20, 26].

The determination of the quantum measurement can simplify considerably if we have some reliable *a-priori* information about the apparatus. In particular, the structure of the POVM may be fixed by the superselection rules. As an example let us consider a class of optical detectors that are sensitive only to the number of photons in a single mode of an electromagnetic field. The elements of the POVM describing a phase-insensitive detector are all diagonal in the Fock basis,

$$\Pi_l = \sum_n r_{ln} |n\rangle\langle n| \tag{3.108}$$

and the ML estimation reduces to the determination of the eigenvalues $r_{ln} \geq 0$, which is an instance of the LinPos problem. The extremal equation (3.103) simplifies to

$$r_{ln} = \frac{r_{ln}}{\lambda_n} \sum_{m=1}^{M} \frac{f_{ml}}{p_{ml}} \rho_{m,nn}, \qquad \lambda_n = \sum_{m=1}^{M} \sum_{l=1}^{L} \frac{f_{ml}}{p_{ml}} \rho_{m,nn} r_{ln}. \tag{3.109}$$

Here $p_{ml} = \sum_n \rho_{m,nn} r_{ln}$. We have thus recovered a generalized version of the expectation-maximization algorithm discussed in Sect. 3.3.

3.6.3 Stern-Gerlach Apparatus

Let us illustrate the reconstruction of the POVM by means of numerical simulations for 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.

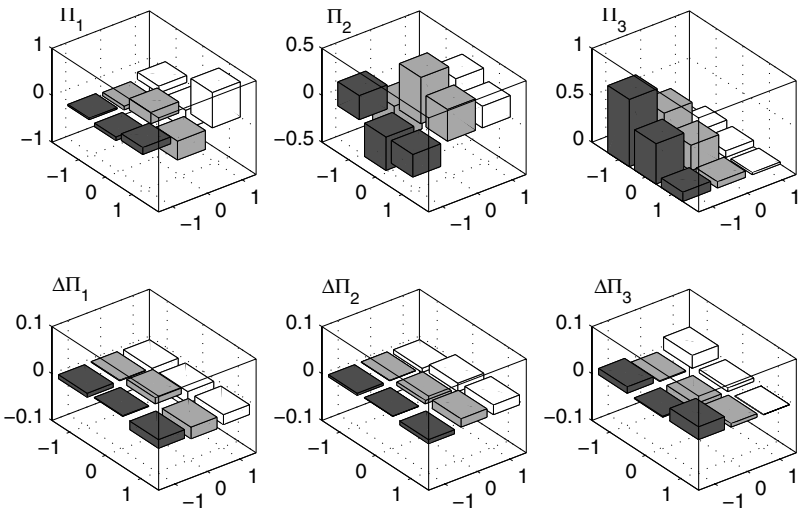

Fig. 3.6. Example of maximum-likelihood reconstruction of a POVM that characterizes a Stern Gerlach apparatus that measures spin along the $(1,0,1)/\sqrt{2}$ axis. Nine different pure states have been used for the reconstruction, and 100 copies of each state have been measured. The upper panels show the reconstructed POVM elements while the lower panels display the estimation errors $\Delta\Pi_j$. Only real parts of complex matrix elements are displayed.

Let S_x, S_y, and S_z denote the operators of spin projections onto axis x, y, and z, respectively. We choose the three eigenstates of S_z as the basis states, $S_z|s_z\rangle = s_z|s_z\rangle$, $s_z = -1, 0, 1$. In our numerical simulations, nine different pure quantum states are used for the calibration: three eigenstates of S_z and six superposition states

$$\frac{1}{\sqrt{2}}(|j_z\rangle + |k_z\rangle), \qquad \frac{1}{\sqrt{2}}(|j_z\rangle + i|k_z\rangle), \tag{3.110}$$

where $j_z, k_z = -1, 0, 1$ and $j_z < k_z$. The measurement on each state is performed N times. We consider a Stern-Gerlach apparatus that measures the projection of the spin component along direction $\boldsymbol{n}$. This von Neumann measurement has three outcomes and the POVM elements Π_l are projectors

$$\Pi_j = |j_{\boldsymbol{n}}\rangle\langle j_{\boldsymbol{n}}|, \qquad j_{\boldsymbol{n}} = -1, 0, 1, \tag{3.111}$$

where $S_{\boldsymbol{n}}|j_{\boldsymbol{n}}\rangle = j_{\boldsymbol{n}}|j_{\boldsymbol{n}}\rangle$ and $S_{\boldsymbol{n}} = n_x S_x + n_y S_y + n_z S_z$.

We have performed Monte Carlo simulations of the measurements and we have subsequently reconstructed the POVM from the simulated experimental data. The ML estimates $\Pi_{l,\mathrm{ML}}$ were obtained by an iterative solution of the extremal equations (3.103) and (3.105). A typical example of the ML reconstruction of the POVM is given in Fig. 3.6 for $\boldsymbol{n} = (1,0,1)/\sqrt{2}$. The linear estimates $\Pi_{l,\mathrm{lin}}$ were found by solving the system of (3.100) by means

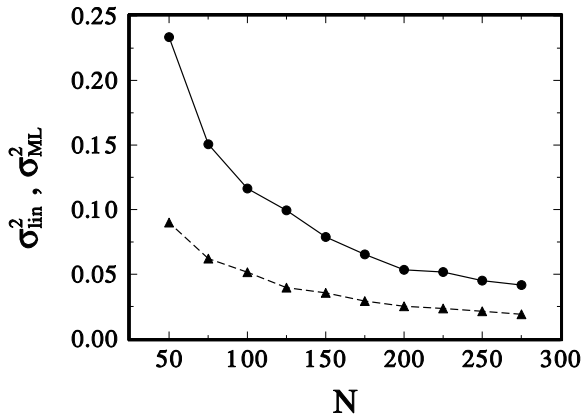

Fig. 3.7. Variances of linear (○) and ML (△) estimates versus the number of measurements N. The figure shows results for Stern-Gerlach apparatus measuring spin along the axis $\boldsymbol{n} = (1,1,1)/\sqrt{3}$.

of least squares inversion. In order to compare these two procedures, we define the variances of the estimates as

$$\sigma^2_{\mathrm{ML}} = \Big\langle \sum_l \mathrm{Tr}\left[\Delta\Pi^2_{l,\mathrm{ML}}\right] \Big\rangle_{\mathrm{ens}}, \qquad \sigma^2_{\mathrm{lin}} = \Big\langle \sum_l \mathrm{Tr}\left[\Delta\Pi^2_{l,\mathrm{lin}}\right] \Big\rangle_{\mathrm{ens}}, \tag{3.112}$$

where $\Delta\Pi_{l,\mathrm{ML}} = \Pi_{l,\mathrm{ML}} - \Pi_l$, $\Delta\Pi_{l,\mathrm{lin}} = \Pi_{l,\mathrm{lin}} - \Pi_l$, and $\langle\rangle_{\mathrm{ens}}$ denotes averaging over the ensemble of all possible experimental outcomes.

We have repeated the reconstruction of the POVM for 100 different simulated experimental data and the ensemble averages yielded σ^2_{ML} and σ^2_{lin}. The variances were determined for 10 different N and the results are shown in Fig. 3.7. We can see that the ML estimates exhibit significantly lower fluctuations than the linear ones. The ML estimation procedure guarantees that the reconstructed POVM elements are positive semidefinite operators. This restriction to physically allowed Π_l significantly improves the reconstruction accuracy. This is a considerable practical advantage of the ML estimation compared to linear inversions.

3.7 Discrimination Between Quantum States

Discrimination between a set of known quantum states ρ_j, $j = 1 \ldots N$ can be understood as a limiting case of quantum-state estimation with "multiple-delta-peaked" prior information. The problem becomes more interesting if we are allowed to choose the measurement at our will. The goal is then to optimize the discriminating apparatus with respect to some figure of merit.

It is no wonder that methods and tools similar to those developed in previous sections for the maximization of the likelihood can also be applied

to the quantum state discrimination problem as well. This similarity will be pursued in this section.

Like density matrices in quantum estimation, the generalized measurements we optimize here are subject to positivity constraints that make quantum discrimination a highly involved nonlinear problem.

Theoretical and experimental aspects of quantum discrimination are described in detail by Bergou *et al.* and Chefles in this volume. The reader is advised to consult those chapters for basic facts and definitions.

3.7.1 Minimum-Error Discrimination

The goal of minimum-error discrimination is to maximize the probability,

$$P_{\mathrm{S}} = \sum_{j=1}^{N} p_j \mathrm{Tr}[\rho_j \Pi_j], \tag{3.113}$$

of guessing the right state out of a set of N states, where the N-component POVM $\{\Pi_j\}$ describes one of Bob's general guessing strategies, and p_j is the prior probability of ρ_j being sent. Let us note in passing, that there are no fundamental reasons for taking error probability as the cost of decisions. Depending on applications, other measures of merit may turn out to be more appropriate, see e.g. [70].

In compact form the problem reads:

$$\begin{aligned} &\text{maximize } \mathrm{P_S} \text{ subject to constraints} \\ &\Pi_j \geq 0, \quad j = 1, \ldots, N, \\ &\textstyle\sum_j \Pi_j = 1. \end{aligned} \tag{3.114}$$

Unfortunately, attacking this problem by analytical means has a chance to succeed only in the simplest cases ($N = 2$) [31], or in cases with symmetric or linearly independent states [77–80]. In most situations one must resort to numerical methods. An iterative algorithm for finding the maximum of the average success rate over projection valued POVM was derived in [81]. Later, an iterative algorithm solving the problem (3.114) in its full generality was devised in [82].

We are going to seek the global maximum of the success functional P_{S} subject to the constraints given in (3.114). To take care of the first constraint we will decompose the POVM elements as follows $\Pi_j = A_j^\dagger A_j$, $j = 1, \ldots, N$. The other constraint (completeness) can be incorporated into our model using the method of uncertain Lagrange multipliers. Putting all these things together, the functional to be maximized becomes

$$\mathcal{L} = \sum_j p_j \mathrm{Tr}\{\rho_{\mathrm{j}} \mathrm{A}_{\mathrm{j}}^\dagger \mathrm{A}_{\mathrm{j}}\} - \mathrm{Tr}\{\lambda \sum_{\mathrm{j}} \mathrm{A}_{\mathrm{j}}^\dagger \mathrm{A}_{\mathrm{j}}\}, \tag{3.115}$$

where λ is a hermitian Lagrange operator. This expression is now varied with respect to N independent variables A_j to yield a necessary condition for the extremal point in the form of a set of N extremal equations for the unknown POVM elements,

$$p_j \rho_j \Pi_j = \lambda \Pi_j, \qquad j = 1, \ldots, N, \tag{3.116}$$

originally derived by Holevo in [83]. For our purposes it is advantageous to bring these equations to an explicitly positive semidefinite form,

$$\Pi_j = p_j^2 \lambda^{-1} \rho_j \Pi_j \rho_j \lambda^{-1}, \qquad j = 1, \ldots, N. \tag{3.117}$$

Lagrange operator λ is obtained by summing (3.117) over j,

$$\lambda = \left(\sum_j p_j^2 \rho_j \Pi_j \rho_j \right)^{1/2}. \tag{3.118}$$

The iterative algorithm comprised of the N+1equations (3.117) and (3.118) provides an elegant and efficient way to optimize Bob's discriminating measurement.

One usually starts from some "unbiased" trial POVM $\{\Pi_j^0\}$. After plugging it in (3.118) the first guess of the Lagrange operator λ is obtained. This operator is, in turn, used in (3.117) to get the first correction to the initial-guess strategy $\{\Pi_j^0\}$. The procedure gets repeated, until, eventually, a stationary point is attained. Both the positivity and completeness of the initial POVM are preserved in the course of iterating [82]. It is interesting to notice that when the initial POVM is chosen to be the maximally ignorant one, $\Pi_j^0 = 1/N$, $j = 1, \ldots, N$, the first correction is quite similar to the "pretty good" measurement [84]. For equally probable states, $p_j = 1/N, \forall j$, they are equivalent. Interestingly enough, the pretty good measurement is known to be optimal in certain cases [80].

Since (3.117) and (3.118) represent only a necessary condition for the extreme, one should always check the optimality of the stationary point by verifying the following set of conditions [31, 83]:

$$\lambda - p_j \rho_j \geq 0, \quad j = 1, \ldots, N. \tag{3.119}$$

It is worth mentioning that this condition can also be derived from the theory of the semidefinite programming (SDP) [85, 86]. SDP tools also provide an alternative means of solving the problem (3.114) numerically [87]. To see the link between the quantum discrimination and SDP theory let us remind that the dual problem of SDP is defined as follows:

$$\begin{aligned} &\text{maximize}: \quad -\mathrm{Tr} F_0 Z, \text{ subject to} \\ &Z \geq 0, \\ &\mathrm{Tr} F_i Z = c_i, \quad i = 1, \ldots, n, \end{aligned} \tag{3.120}$$

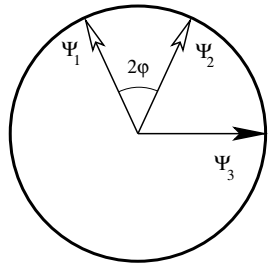

Fig. 3.8. A cut through the Bloch sphere showing the states to be discriminated.

where data are $n+1$ hermitian matrices F_i and a complex vector $c \in \mathbb{C}^n$, and Z is a hermitian variable. Our problem (3.114) reduces to this dual SDP problem upon the following substitutions:

$$F_0 = -\bigoplus_{j=1}^{n} p_j \rho_j, \quad Z = \bigoplus_{j=1}^{n} \Pi_j,$$
$$F_i = \bigoplus_{j=1}^{n} \Gamma_i, \quad c_i = \mathrm{Tr}\Gamma_i, \quad i = 1, \ldots, h^2. \tag{3.121}$$

Here operators $\{\Gamma_i, i = 1, \ldots, h^2\}$ comprise an orthonormal operator basis in the h^2-dimensional space of hermitian operators acting on the Hilbert space of our problem: $\mathrm{Tr}\Gamma_j\Gamma_k = \delta_{jk},\ j,k = 1, \ldots, h^2$. For simplicity, let us take Γ_1 proportional to the unity operator, then all c_i apart from c_1 vanish. In SDP the necessary condition (3.116) for the maximum of the functional (3.115) is called the complementary slackness condition. When inequalities in (3.119) hold it can be shown to be also sufficient.

The advantage of the SDP formulation of the quantum-state discrimination problem is that there are strong numerical tools designed for solving SDP problems, for their review see [87]. They make use of the duality of SDP problems. The optimal value is bracketed between the trial maximum of the dual problem and trial minimum of the primal problem. One then hunts the optimal value down by making this interval gradually smaller. SDP tools are much more complicated than the proposed algorithm (3.117)-(3.118) but they are guaranteed to converge to the real solution.

Let us illustrate the utility of our algorithm on a simple example of discriminating between three coplanar pure qubit states. The geometry of this problem is shown in Fig. 3.8. Ψ_1 and Ψ_2 are equal-prior states, $p_1 = p_2 = p/2$, symmetrically placed around the z axis; the third state lies in the direction of x or y. A similar configuration (with Ψ_3 lying along z) has been investigated in [79]. Exploiting the mirror symmetry of their problem the authors derived analytic expressions for POVMs minimizing the average error rate. For a given angle φ the optimum POVM turned out to have two or three nonzero elements depending on the amount of the prior information p.

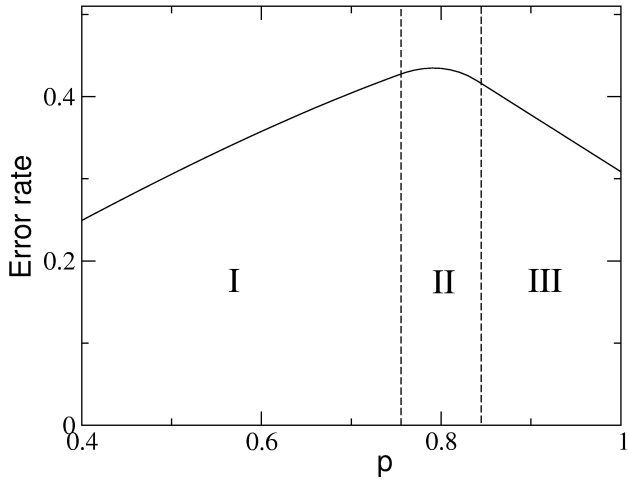

Fig. 3.9. Average error rate $(1 - P_S)$ in dependence on Bob's prior information p; $\varphi = \pi/16$. Regions I, II, and III are regions where the optimum discriminating device has two, three, and two output channels, respectively.

Our problem is a bit more complicated one due to the lack of mirror symmetry. Let us see whether the transition from the mirror-symmetric configuration to a non-symmetric one has some influence on the qualitative behavior of the optimal POVMs. Minimal error rates calculated using the proposed iterative procedure (3.117 and 3.118) for the fixed angle of $\varphi = \pi/16$ are summarized in Fig. 3.9. For large p (region III) the optimum strategy consist in the optimal discrimination between states Ψ_1 and Ψ_2. When ξ becomes smaller than a certain φ-dependent threshold (region II), state Ψ_3 can no longer be ignored and the optimum POVM has three nonzero elements. Simple calculation yields

$$p_{\mathrm{II,III}} = \frac{1}{1 + \sin\varphi\cos\varphi} \tag{3.122}$$

for the threshold value of the prior. However, when p becomes still smaller (region I), the optimum POVM will eventually become a two-element POVM again – the optimal strategy now being the optimal discrimination between states Ψ_1 and Ψ_3. This last regime is absent in the mirror-symmetric case. The transition between regions I and II is governed by a much more complicated expression than (3.122), and will not be given here.

The convergence properties of the algorithm are shown in Fig. 3.10 for three typical prior probabilities representing regions I, II, and III of Fig. 3.9. After a short transient period an exponentially fast convergence sets in. Sixteen-digit precision in the resulting error-rate is usually obtained after less than one hundred iterations. Let us close the example noting that already a few iterations are enough to determine the optimum discriminating

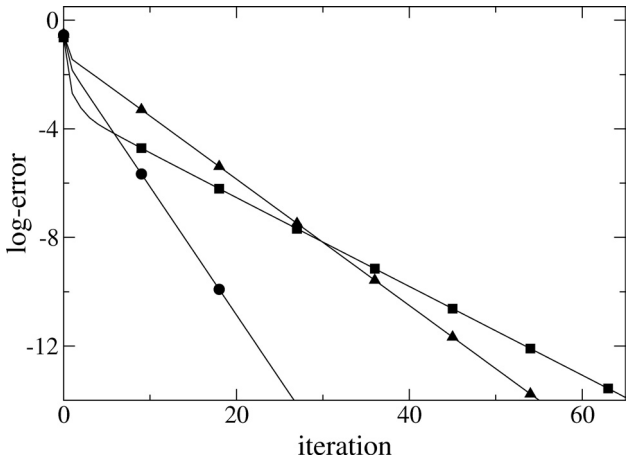

Fig. 3.10. Accuracy of the calculated error rate of the optimal POVM vs. the number of iterations. Convergence of the proposed algorithm is shown for three different priors: $p = 0.6$ (squares), $p = 0.8$ (triangles), and $p = 0.9$ (circles). Ordinate is labeled by the precision in decimal digits.

device to the precision the elements of the realistic experimental setup can be controlled within the laboratory.

3.7.2 Minimum-Error Discrimination Involving Inconclusive Results

The above discussed scenario can be considered as limiting cases of a more general scheme that involves certain fraction of inconclusive results P_{I} for which we maximize the success rate. The main feature of this scheme is that if we allow for inconclusive results then we can improve the relative (or re-normalized) success rate

$$P_{\mathrm{RS}} = \frac{P_{\mathrm{S}}}{1 - P_{\mathrm{I}}}. \tag{3.123}$$

In other words, with probability P_{I} Bob fails completely and he cannot say at all which state was sent to him. However, in the rest of the cases when he succeeds he can correctly guess the state with higher probability than if he would not allow for the inconclusive results. Thus, there exists a trade-off between probability P_{I} of inconclusive results and a re-normalized success rate. For pure linearly independent states this generalized scenario was discussed in papers [88, 89]. The analysis was extended to general mixed states in [90–92].

Let us assume that the quantum state sent to Bob is drawn from the set of N mixed states $\{\rho_j\}_{j=1}^{N}$ with the a-priori probabilities p_j. Bob's measurement

on the state may yield $N+1$ different results and is formally described by the POVM whose $N+1$ components satisfy

$$\Pi_j \geq 0, \quad j = 0, \ldots, N, \qquad \sum_{j=0}^{N} \Pi_j = 1. \tag{3.124}$$

The outcome Π_0 indicates failure and the probability of inconclusive results is thus given by

$$P_{\mathrm{I}} = \sum_{j=1}^{N} p_j \mathrm{Tr}[\rho_j \Pi_0]. \tag{3.125}$$

For a certain fixed value of P_{I} we want to maximize the relative success rate (3.123) that is equivalent to the maximization of the success rate (3.113). To account for the linear constraints (3.124) and (3.125) we introduce Lagrange multipliers λ and a where λ is hermitian operator and a is a real number. Taking everything together we should maximize the constrained success rate functional

$$\mathcal{L} = \sum_{j=1}^{N} p_j \mathrm{Tr}[\rho_j \Pi_j] - \sum_{j=0}^{N} \mathrm{Tr}[\lambda \Pi_j] + a \sum_{j=1}^{N} p_j \mathrm{Tr}[\rho_j \Pi_0]. \tag{3.126}$$

With the help of Cholesky decomposition and calculus of variation we arrive at the extremal equations for the optimal POVM,

$$(\lambda - p_j \rho_j)\Pi_j = 0, \qquad j = 1, \ldots, N, \tag{3.127}$$

$$(\lambda - a\sigma)\Pi_0 = 0, \tag{3.128}$$

where the operator σ introduced for the sake of notational simplicity reads

$$\sigma = \sum_{j=1}^{N} p_j \rho_j. \tag{3.129}$$

From the constraint $\mathrm{Tr}[\sigma \Pi_0] = P_{\mathrm{I}}$ we can express a in terms of λ,

$$a = P_{\mathrm{I}}^{-1} \mathrm{Tr}[\lambda \Pi_0]. \tag{3.130}$$

Furthermore, if we sum all (3.127) and also (3.128) and use the resolution of the identity (3.124), we obtain the following formula for λ,

$$\lambda = \sum_{j=1}^{N} p_j \rho_j \Pi_j + a\sigma\Pi_0. \tag{3.131}$$

The analytical solution of this extremal problem seems to be extremely complicated. Nevertheless, we can solve again the extremal equations numerically [90] as in the case of ambiguous discrimination. In principle, one could

iterate directly (3.127) and (3.128). However, the POVM elements Π_j should be positive semidefinite hermitian operators. All constraints can be exactly satisfied at each iteration step if the extremal equations are symmetrized. First we express $\Pi_j = p_j \lambda^{-1} \rho_j \Pi_j$ and combine it with its hermitian conjugate. We proceed similarly also for Π_0 and we get

$$\Pi_j = p_j^2 \lambda^{-1} \rho_j \Pi_j \rho_j \lambda^{-1}, \quad j = 1, \ldots, N, \tag{3.132}$$

$$\Pi_0 = a^2 \lambda^{-1} \sigma \Pi_0 \sigma \lambda^{-1}. \tag{3.133}$$

The Lagrange multipliers λ and a must be determined self-consistently so that all the constraints will hold. If we sum (3.132) and (3.133) and take into account that $\sum_{j=0}^{N} \Pi_j = 1$, we obtain

$$\lambda = \left[\sum_{j=1}^{N} p_j^2 \rho_j \Pi_j \rho_j + a^2 \sigma \Pi_0 \sigma \right]^{1/2}. \tag{3.134}$$

The fraction of inconclusive results calculated for the POVM after the iteration is given by

$$P_{\mathrm{I}} = a^2 \mathrm{Tr}[\sigma \lambda^{-1} \sigma \Pi_0 \sigma \lambda^{-1}]. \tag{3.135}$$

Since the Lagrange multiplier λ is expressed in terms of a, (3.135) forms a nonlinear equation for a single real parameter a (or, more precisely, a^2). This nonlinear equation can be very efficiently solved by Newton's method of halving the interval. At each iteration step for the POVM elements, we thus solve the system of coupled nonlinear equations (3.134) and (3.135) for the Lagrange multipliers. These self-consistent iterations typically exhibit an exponentially fast convergence [82, 90].

As the fraction of inconclusive results P_{I} is increased the success rate P_S decreases. However, the relative success rate P_{RS} grows until it achieves its maximum $P_{\mathrm{RS,max}}$. If $\{\rho_j\}_{j=1}^{N}$ are linearly independent pure states, then $P_{\mathrm{RS,max}} = 1$ because exact IDP scheme works and the unambiguous discrimination is possible. Generally, however the maximum is lower than unity. The analytical expression for this maximum can be found [90].

Let us illustrate the trade-off between probability of inconclusive results and relative success rate P_{RS} on explicit example. We consider the problem of optimal discrimination between two mixed qubit states ρ_1 and ρ_2. To simplify the discussion, we shall assume that the purities of these states as well as the a-priori probabilities are equal, $\mathcal{P}_1 = \mathcal{P}_2 = \mathcal{P}$, $p_1 = p_2 = 1/2$. The mixed states can be visualized as points inside the Poincaré sphere and the purity determines the distance of the point from the center of that sphere. Without loss of generality, we can assume that both states lie in the xz plane and are symmetrically located about the z axis,

$$\rho_{1,2} = \eta \psi_{1,2}(\theta) + \frac{1-\eta}{2} 1, \tag{3.136}$$

where $\psi_j = |\psi_j\rangle\langle\psi_j|$ denotes a density matrix of a pure state,

$$|\psi_{1,2}(\theta)\rangle = \cos\frac{\theta}{2}|0\rangle \pm \sin\frac{\theta}{2}|1\rangle, \tag{3.137}$$

and $\theta \in (0, \pi/2)$. The parameter η determines the purity of the mixed state (3.136), $\mathcal{P} = (1+\eta^2)/2$.

From the symmetry it follows that the elements Π_1 and Π_2 of the optimal POVM must be proportional to the projectors $\psi_1(\phi)$ and $\psi_2(\phi)$, where the angle $\phi \in (\pi/2, \pi)$ is related to the fraction of the inconclusive results. The third component Π_0 is proportional to the projector onto state $|0\rangle$. The normalization of the POVM elements can be determined from the constraint (3.124) and we find

$$\begin{aligned} \Pi_{1,2}(\phi) &= \frac{1}{2\sin^2(\phi/2)}\psi_{1,2}(\phi), \\ \Pi_0(\phi) &= \left(1 - \frac{1}{\tan^2(\phi/2)}\right)|0\rangle\langle 0|. \end{aligned} \tag{3.138}$$

The relative success rate for this POVM reads

$$P_{\mathrm{RS}} = \frac{1+\eta\cos(\phi-\theta)}{2(1+\eta\cos\theta\cos\phi)} \tag{3.139}$$

and the fraction of inconclusive results is given by

$$P_{\mathrm{I}} = \frac{1}{2}\left(1+\eta\cos\theta\right)\left(1 - \frac{1}{\tan^2(\phi/2)}\right). \tag{3.140}$$

The formulas (3.139) and (3.140) describe implicitly the dependence of the relative success rate P_{RS} on the fraction of the inconclusive results P_{I}. From (3.130) and (3.131) one can determine the Lagrange multipliers λ and a for the POVM (3.138) and check that the extremal equations (3.127), (3.128) are satisfied. The maximum $P_{\mathrm{RS,max}}$ is achieved if the angle ϕ is chosen as follows,

$$\cos\phi_{\max} = -\eta\cos\theta. \tag{3.141}$$

On inserting the optimal $\phi_{\max}$ back into (3.139) we get

$$P_{\mathrm{RS,max}} = \frac{1}{2}\left[1 + \frac{\eta\sin\theta}{\sqrt{1-\eta^2\cos^2\theta}}\right]. \tag{3.142}$$

The optimal POVM (3.138) can be also obtained numerically. We demonstrate the feasibility of iterative solution of the symmetrized extremal equations (3.132), (3.133), (3.134), and (3.135) for mixed quantum states (3.136) with the angle of separation $\theta = \pi/4$. The trade-off of the relative success

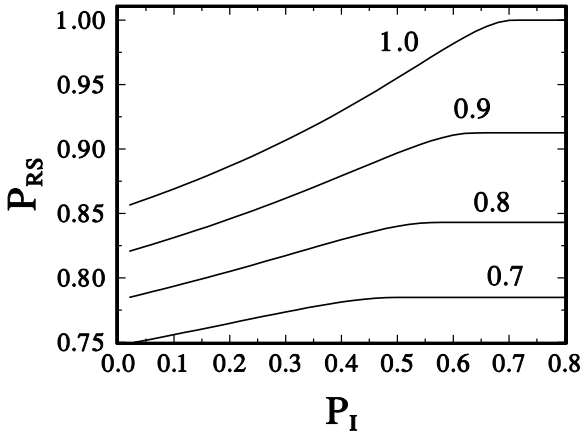

Fig. 3.11. Relative success rate P_{RS} versus the fraction of inconclusive results P_{I} for the optimal discrimination of two mixed states (3.136) with $\theta = \pi/4$ and four different parameters $\eta = 0.7$, $\eta = 0.8$, $\eta = 0.9$, and $\eta = 1.0$.

rate and the probability of inconclusive results is shown in Fig. 3.11 for various purities of the states being discriminated. For the given probability P_{I} of inconclusive results and the given purity of the states the extremal equations are solved self-consistently by means of repeated iterations. The success rate P_{S} is calculated from the obtained optimal POVM and re-normalized according to (3.123). The numerically obtained dependence of P_{RS} on P_{I} is in excellent agreement with the analytical dependence following from formulas (3.139) and (3.140). Typically, a sixteen digit precision is reached after several tens of iterations. The trade-off curves shown in Fig. 3.11 reveal the monotonous growth of P_{RS} until the maximal plateau (3.142) is reached.

Conclusions

We have presented a powerful reconstruction method that is capable of dealing with any experimental data. It stems from mathematical statistics but can be interpreted equally well in terms of quantum theory. Though the formulations of mathematical statistics and quantum theory are rather independent, they both overlap when describing measurement and information. Maximum likelihood plays a prominent role among other estimation techniques. It exploits the full information potential of registered data, preserves the structure of quantum theory such as completeness or uncertainty relations, and reaches the ultimate resolution asymptotically. This seems to be crucial for the future potential applications in quantum information science for quantifying all the subtle and fragile quantum effects.

Acknowledgments

This work was supported by Grant No. LN00A015 of the Czech Ministry of Educations and by Czech-Italian project No. 29, "Decoherence and Quantum Measurement. JF also acknowledges support from the EU under projects RESQ (IST-2001-35759) and CHIC (IST-2001-32150).

References

1. S. Weigert: Phys. Rev. **A 45**, 7688 (1992); *ibid.* **A 53**, 2078 (1996).
2. K. Vogel, H. Risken: Phys. Rev. A **40**, 2847 (1989).
3. D. T. Smithey, M. Beck, M. G. Raymer, A. Faridani: Phys. Rev. Lett. **70**, 1244 (1993).
4. U. Leonhardt: *Measuring of the Quantum State of Light* (Cambridge Press, Cambridge, 1997).
5. S. Wallentowitz, W. Vogel: Phys. Rev. Lett. **75**, 2932 (1996).
6. P. J. Bardoff, C. Leichtle, G. Schrade, W. P. Schleich: Phys. Rev. Lett. **77**, 2198 (1996).
7. D. Leibfried, D. M. Meekhof, B. E. King, C. Monroe, W. M. Itano, D.J. Wineland: Phys. Rev. Lett. **77**, 4281 (1996).
8. A. Witten: Inverse Problems **7**, L49 (1991).
9. D.-G. Welsch, W. Vogel, T. Opatrný: Homodyne Detection and Quantum State Reconstruction. In: *Progress in Optics*, vol 39, ed by E. Wolf (North Holland, Amsterdam, 1999).
10. J. M. Bernardo, A. F. M. Smith: *Bayesian Theory* (Wiley, Chichester, 1994); S. Sykora: J. Stat. Phys. **11**, 17 (1974).
11. R. Schack, T. A. Brun, C. M. Caves: Phys. Rev. A **64** 014305 (2001).
12. E. T. Jaynes: Information Theory and Statistical Mechanics. In: *1962 Brandeis Summer Lectures*, vol 3, ed by K. W. Ford (Benjamin, New York, 1963) p 181.
13. V. Buzek, R. Derka: Quantum observations. In *Coherence and Statistics of Photons and Atoms*, ed by J. Perina (Wiley, New York, 2001) pp 198 - 261.
14. C. M. Caves, P. D. Drummond: Rev. Mod. Phys. **66**, 481 (1994).
15. J. J. Sakurai: *Modern Quantum Mechanics* (Addison-Wesley, Reading, 1994).
16. V. Vedral, M. B. Plenio, M. A. Rippin, P. L. Knight: Phys. Rev. Lett. **78**, 2275 (1997).
17. Z. Hradil: Phys. Rev. A **55**, 1561(R) (1997).
18. K. R. W. Jones: Ann. Phys. **207**, 140 (1991).
19. V. Bužek, R. Derka, G. Adam, P. L. Knight: Ann. Phys. **266**, 454 (1998).
20. Z. Hradil, J. Summhammer: J. Phys. A: Math. Gen. **33**, 7607 (2000).
21. S. Kullback, R. A. Leibler: Ann. of Math. Stat. **22**, 79 (1951).
22. J. H. Shapiro, S. R. Shepard, N. C. Wong: Phys. Rev. Lett. **62**, 2377 (1989).
23. A. S. Lane, S. L. Braunstein, C. M. Caves: Phys. Rev. **A 47**, 1667 (1993).
24. J. Řeháček, Z. Hradil, M. Zawisky, S. Pascazio, H. Rauch, J. Peřina: Phys. Rev. A **60**, 473 (1999).
25. G. M. D' Ariano, M. G. A. Paris, M. F. Sacchi: Phys. Rev. A **62**, 023815 (2000).
26. Z. Hradil, J. Summhammer, H. Rauch: Phys. Lett. A **261**, 20 (1999).
27. A. P. Dempster, N. M. Laird, D. B. Rubin: J. R. Statist. Soc. B **39**, 1 (1977).

28. Y. Vardi and D. Lee: J. R. Statist. Soc. B **55**, 569 (1993).
29. K. Banaszek, G. M. D' Ariano, M. G. A. Paris, M. F. Sacchi: Phys. Rev. A **61**, 010304(R) (1999).
30. J. Řeháček, Z. Hradil, and M. Ježek: Phys. Rev. A **63**, 040303(R) (2001).
31. C. W. Helstrom: *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
32. A.G.White, D.F.V. James, P.H. Eberhard, and P.G. Kwiat: Phys. Rev. Lett. **83**, 3103 (1999).
33. R. Jozsa: J. Mod. Optics **41**, 2315 (1993).
34. W.K. Wootters: Phys. Rev. D **23**, 357 (1981).
35. D. Dieks: Phys. Lett. A **126**, 303 (1998).
36. A. Peres: Phys. Lett. A **128**, 19 (1998).
37. V. Vedral, M.B.Plenio, K. Jakobs, and P.L. Knight: Phys. Rev. A **56**, 4452 (1997).
38. R.A. Fisher: Proc. Camb. Phi. Soc. **22**, 700 (1925).
39. C.R. Rao: Bull. Calcutta Math. Soc. **37**, 81 (1945).
40. H. Cramér: *Mathematical methods of statistics* (Princeton University Press, Princeton, 1946).
41. D.F.V. James, P.G. Kwiat, W.J. Munro, and A.G. White: Phys. Rev. A **64**, 052312 (2001); e-print arXiv:quant-ph/0103121.
42. M.A. Nielsen and I.L. Chuang: *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
43. I. L. Chuang and M. A. Nielsen: J. Mod. Opt. **44**, 2455 (1997).
44. J. F. Poyatos, J. I. Cirac, and P. Zoller: Phys. Rev. Lett. **78**, 390 (1997).
45. G. M. D'Ariano and L. Maccone: Phys. Rev. Lett. **80**, 5465 (1998).
46. A. Luis and L. L. Sánchez-Soto: Phys. Lett. A **261**, 12 (1999).
47. R. Gutzeit, S. Wallentowitz, and W. Vogel: Phys. Rev. A **61**, 062105 (2000).
48. G.M. D'Ariano and P. Lo Presti: Phys. Rev. Lett. **86**, 4195 (2001).
49. D.G. Fischer, H. Mack, M.A. Cirone, and M. Freyberger: Phys. Rev. A **64**, 022309 (2001).
50. M. F. Sacchi: Phys. Rev. A **63**, 054104 (2001).
51. A. M. Childs, I. L. Chuang, and D. W. Leung: Phys. Rev. A **64**, 012314 (2001).
52. N. Boulant, T. F. Havel, M. A. Pravia, and D. G. Cory: Phys. Rev. A **67**, 042322 (2003).
53. F. De Martini, A. Mazzei, M. Ricci, and G.M. D'Ariano: Phys. Rev. A **67**, 062307 (2003).
54. J.B. Altepeter, D. Branning, E. Jeffrey, T.C. Wei, P.G. Kwiat, R.T. Thew, J.L. O'Brien, M.A. Nielsen, and A.G. White: Phys. Rev. Lett. **90**, 193601 (2003).
55. M.W. Mitchell, C.W. Ellenor, S. Schneider, and A.M. Steinberg: Phys. Rev. Lett. **91**, 120402 (2003).
56. J. Fiurášek and Z. Hradil: Phys. Rev. A **63**, 020101(R) (2001).
57. M. Ježek, J. Fiurášek, and Z. Hradil: Phys. Rev. A 68, 012305 (2003).
58. A. Jamiolkowski: Rep. Math. Phys. **3**, 275 (1972).
59. M. Dakna, J. Clausen, L. Knöll, and D.-G. Welsch: Phys. Rev. A **59**, 1658 (1999);
60. H. Lee, P. Kok, N.J. Cerf, and J.P. Dowling: Phys. Rev. A **65**, 030101 (2002).
61. J. Fiurášek: Phys. Rev. A **65**, 053818 (2002).
62. C.H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J.A. Smolin, and W.K. Wootters: Phys. Rev. Lett. **78**, 2031 (1997).

63. D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera: Phys. Rev. Lett. **77**, 2818 (1996).
64. E. Knill, R. Laflamme, and G.J. Milburn: Nature (London) **409**, 46 (2001).
65. A.S. Holevo: *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
66. A. Luis and L.L. Sanchez-Soto: Phys. Rev. Lett. **83**, 3573 (1999).
67. T. Opatrný, D.-G. Welsch, and W. Vogel: Phys. Rev. A **56**, 1788 (1997).
68. S.M. Tan: J. Mod. Opt. **44**, 2233 (1997).
69. J. Fiurášek: Phys. Rev. A **64**, 024102 (2001).
70. M. Ježek: Phys. Lett. A **299**, 441 (2002),
71. Z. Hradil and J. Summhammer: Phys. Rev. A **62**, 014101 (2000).
72. W.K. Wootters and W.H. Zurek: Nature (London) **299**, 802 (1982).
73. D. Dieks: Phys. Lett. A **92**, 271 (1982).
74. C.H. Bennett and G. Brassard: In: *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing* (IEEE, New York, Bangalore, India, 1984) p 175.
75. C.H. Bennett, G. Brassard, and N.D. Mermin: Phys. Rev. Lett. **68**, 557 (1992).
76. C. H. Bennett: Phys. Rev. Lett. **68**, 3121 (1992).
77. M. Sasaki, K. Kato, M. Izutsu, and O. Hirota: Phys. Rev. A **58**, 146 (1998).
78. S.M. Barnett: Phys. Rev. A **64**, 030303(R) (2001).
79. E. Andersson, S.M. Barnett, C.R. Gilson, and K. Hunter: Phys. Rev. A **6**5, 052308 (2002).
80. M. Ban, K. Kurokawa, R. Momose, and O. Hirota: Int. J. Theor. Phys. **36**, 1269 (1997).
81. C.W. Helstrom: IEEE Trans. **I**T-28, 359 (1982).
82. M. Ježek, J. Řeháček, and J. Fiurášek: Phys. Rev. A **6**5, 060301(R) (2002).
83. A.S. Holevo: J. Multivar. Anal. **3**, 337 (1973).
84. P. Hausladen and W. K. Wootters: J. Mod. Opt. **41**, 2385 (1994).
85. H.P. Yuen, R.S. Kennedy, and M. Lax: IEEE Trans. **I**T-21, 125 (1975).
86. Y.C. Eldar, A. Megretski, and G.C. Verghese: IEEE Trans. Inform. Theory **49**, 1007 (2003).
87. L. Vandenberghe and S. Boyd: Semidefinite programming, SIAM Review **38**, 49 (1996).
88. A. Chefles and S. M. Barnett: J. Mod. Opt. **45**, 1295 (1998).
89. C. W. Zhang, C. F. Li, and G. C. Guo: Phys. Lett. A **261**, 25 (1999).
90. J. Fiurášek and M. Ježek: Phys. Rev. A **67**, 012321 (2003).
91. Y. C. Eldar: IEEE Trans. Inform. Theory **49**, 446 (2003).
92. Y. C. Eldar: Phys. Rev. A **67**, 042309 (2003).