PHYSICAL REVIEW D VOLUME 26, NUMBER 1 1 JULY 1982

Evaluation of an incompletely measured spin density matrix

S. Dagan and Y. Dothan
Department of Physics and Astronomy, Tel Aviv University, Ramat Aviv, Israel
(Received 9 March 1981; revised manuscript received 24 July 1981)

A method for the determination of the density matrix of a resonant state using incomplete measurements pertaining to its decay is developed. In order to define the bounds on the density-matrix elements, the positivity and rank conditions are used. The method proceeds in two steps. First one obtains a particular density matrix which satisfies the positivity and rank conditions and reproduces the experimental data. This matrix is calculated by maximizing the information entropy. Then using this matrix as a starting point and the tolerance of the data presented by the experimental errors, the actual bounds on the density-matrix elements are found by means of a parametrization which assures both positivity and the rank condition. The method was successfully tested on data for f production in the reaction $\pi^- + p \to f + n$ at 12 GeV/c and the subsequent decay $f \to \pi^+ + \pi^-$.

I. INTRODUCTION

The density-matrix formalism is widely used in data analysis for extracting spin and helicity information. In order to determine the density matrix, one has to measure the polarization and the angular distribution of all the particles involved in the reaction. Unfortunately experimental technique today does not allow a full measurement of the polarization, leaving a model-independent evaluation of the density matrix as an underconstrained problem. Furthermore, most of the available experimental data consists of angular-distribution information, which is related only to the real part of the density matrix and not to its imaginary part.

Although the problem is generally underconstrained there are two additional conditions, which the density matrix is bound to satisfy: positivity and rank. The positivity condition arises from the fact that the density matrix is a positive-semidefinite matrix. However positivity is a nonlinear constraint and its implementation is complicated, so that in many cases positivity is totally or partially ignored. The rank condition is imposed by the physics of the reaction involved, namely, the rank of the density matrix is limited to be at most equal to a given bound. It is obvious that this condition is meaningful only if the dimension of the density matrix is higher than the upper bound of the rank and therefore can be useful for high-spin states. Strictly speaking, one should have sharply defined values for the mass and momentum of the resonances in order that the rank condition be valid. Therefore the application of the rank condition to experimental data should be done with care.

In a recent analysis[1] of ρ^0 and f production, using only angular-distribution data, it was shown that the application of positivity yields reasonably bounded values of the real part of the density-matrix elements. The method used there consists of a thorough exploration in a multidimensional space, defined by the density-matrix elements, limited by positivity and consistency with the experimental data. The efficiency of this method depends largely on the matrix which is used as a starting point. If this matrix is inconsistent with the experimental data, as expressed by a high value of χ^2, a considerable amount of the effort is spent on a search for a matrix with low χ^2. In addition, the method does not include a provision for rank limitation, which could be useful, especially for density matrices of high-spin states. The present study presents a solution of these two difficulties in the method.

Generally speaking, calculations involving large matrices are complicated and highly time consuming, even with the powerful modern computers. The above-mentioned analysis[1] deals with a system with mixed spins $J=0$, 1, and 2, which results in a 9×9 matrix. By including spin $J=3$, the matrix becomes 16×16, which already makes the evaluation of the density-matrix elements a difficult, if not practically unfeasible, task. We are proposing a method which not only meaningfully simplifies the procedure, but also reduces the computation time, especially for high-spin states.

In the present study we use the information-entropy functional and a suitable parametrization of the density matrix. The use of this functional for strictly positive-definite matrices in the case of incomplete measurements was studied by Wychmann.[2] We generalize the method for positive-semidefinite matrices, allowing in that way a rank limitation. We show that the maximization of the information-entropy functional (expressed in terms of the density matrix) constrained by the experimental data yields a unique solution of the density matrix. This solution is by definition positive semidefinite and reproduces the experimental data. The maximization can be done with

or without imposing a constraint on the rank of the matrix. The matrix obtained that way can be used as a starting point for the exploration of the range of the density matrix. For this purpose we use a parametrization of the matrix[3] which assures both the positivity condition and the rank condition (if applied).

The information-entropy functional in terms of the density matrix and the constraints imposed by the data are described in Sec. II. Section III deals with the actual technique of maximization of the entropy and the reconstruction of the density matrix, which corresponds to the solution. In order to clarify the problem a simple example with its solution is given in Sec. IV. Section V describes the parametrization of the density matrix, which ensures the positivity and rank conditions. Finally the results of the application of the method to actual data is discussed in Sec. VI. Technical details and mathematical proofs are given in Appendices A through D.

II. THE INFORMATION ENTROPY

Consider a spin density matrix ρ which is a Hermitian, $\nu\times\nu$ positive matrix with unit trace. The rank of ρ may also be given and it satisfies

$$\text{rank}(\rho)=r\leq\nu. \tag{1}$$

For any such matrix one defines the information-entropy functional[2] $E[\rho]$,

$$E[\rho]=-\text{Tr}(\rho\ln\rho)\,. \tag{2}$$

From the Hermiticity and positivity of ρ and from the trace condition it follows that its eigenvalues ρ_α, $\alpha=1,\ldots,\nu$, are real and obey

$$0\leq\rho_\alpha\leq 1\,. \tag{3}$$

It is convenient to arrange the eigenvalues ρ_α in a descending order as follows:

$$\rho_1\geq\rho_2\geq\cdots\geq\rho_\nu\,. \tag{4}$$

Expressed in terms of the eigenvalues, the definition of the information entropy reads

$$E[\rho]=-\sum_{\alpha=1}^{\nu}\rho_\alpha\ln\rho_\alpha\,. \tag{5}$$

When the rank of ρ is strictly smaller than ν we have

$$\rho_r>0\,,\quad \rho_{r+1}=\cdots=\rho_\nu=0\,. \tag{6}$$

But the function $x\ln x$ tends to zero when x tends to zero from above, so that the sum defining $E[\rho]$ is altered to contain contributions only from strictly positive eigenvalues and can finally be written as

$$E[\rho]=-\sum_{\alpha=1}^{r}\rho_\alpha\ln\rho_\alpha\,. \tag{7}$$

From the fact that the positive ρ_α are not greater than 1 it follows that

$$E[\rho]\geq 0\,, \tag{8}$$

while from the trace condition and the concavity[4] of the function $-x\ln x$ it can be shown that

$$E[\rho]\leq\ln r\,. \tag{9}$$

Having defined the information entropy $E[\rho]$ of a spin density matrix ρ we now consider the mathematical problem of maximizing $E[\rho]$ subject to the constraints imposed by the experimental data. To understand the nature of the problem we start by noting that the number f, of independent real numbers necessary to specify a Hermitian $\nu\times\nu$ matrix ρ of rank r and unit trace is given by

$$f=\nu^2-(\nu-r)^2-1\,. \tag{10}$$

The known experimental information we shall consider is given in the form of mean values and the error matrix of certain linear combinations of matrix elements of ρ. For example, one may determine experimentally the moments of the angular distribution of a two-body decay of a resonant state. Now, any real linear combination of the elements of a Hermitian matrix ρ can be written in the form $\text{Tr}(\rho M)$ with M a $\nu\times\nu$ Hermitian matrix. Therefore the experimental data are given in the form of average values (usually moments) m_j and a corresponding error matrix E_{ij} where

$$\text{Tr}(\rho M_j)=m_j\,,\quad j=1,\ \ldots\,,k\,. \tag{11}$$

The k matrices M_j are Hermitian numerical matrices

$$M_j=M_j^\dagger\,,\quad j=1,\ldots,k\,, \tag{12}$$

and the $k+1$ matrices I, $M_1,\ldots,M_k$ are assumed to be linearly independent. The number k is typically smaller than the number f of the independent real numbers necessary to specify the density matrix. If that is indeed the case, it is impossible to reconstruct ρ uniquely.

Among all possible spin density matrices which are consistent with experimental data we show in the next section how to reconstruct the spin density matrix corresponding to maximum entropy. We do not intend to give any special meaning to this particular matrix, but merely to use it as a starting point for our subsequent search procedure. The maximization of entropy applied here should be considered just as a tool for constraining an otherwise underconstrained problem and finding a solution. However, it is still true that the density matrix constructed this way has the property of being statistically least biased. Moreover, if we adopt the attitude of Doncel, Michel, and Minnaert[5] the maximum-entropy density matrix corresponds

to the measured point in the observed polarization domain.

III. MAXIMIZATION OF THE INFORMATION ENTROPY

We maximize the information entropy $E[\rho]$ subject to the subsidiary conditions

$$\mathrm{Tr}(\rho) = 1 , \tag{13}$$

$$\mathrm{Tr}(\rho M_j) = m_j , \quad j = 1, \ldots, k . \tag{14}$$

To take the subsidiary conditions into account we introduce the Lagrange multipliers $\lambda^0, \lambda^1, \ldots, \lambda^k$ and consider the functional[2]

$$I(\rho; \lambda_0, \lambda) = -\mathrm{Tr}(\rho \ln \rho) - \lambda^0[\mathrm{Tr}(\rho) - 1] - \sum_{j=1}^{k} \lambda^j [\mathrm{Tr}(\rho M_j) - m_j] . \tag{15}$$

A vanishing first variation of the functional $I(\rho; \lambda)$ with respect to its independent variables yields the conditions for a local extremum. As shown in Appendices A and C this leads to the following results. One constructs the matrix $M(\lambda)$ given by

$$M(\lambda) = \sum_{j=1}^{k} \lambda^j M_j . \tag{16}$$

This is a Hermitian matrix and thus can be diagonalized and represented in the form[6]

$$M(\lambda) = \sum_{\alpha=1}^{\nu} \mu_\alpha(\lambda) P_\alpha(\lambda) , \tag{17}$$

where $\mu_\alpha(\lambda)$ are the eigenvalues of $M(\lambda)$ and $P_\alpha(\lambda)$ are Hermitian orthonormal projectors with unit trace. For almost every point λ the eigenvalues of $M(\lambda)$ are nondegenerate and we arrange them in increasing order:

$$\mu_1 < \mu_2 < \cdots < \mu_\nu .$$

We now choose the first r eigenvalues and construct the function $\Gamma_r(\lambda)$,

$$\Gamma_r(\lambda) = \ln \sum_{\alpha=1}^{r} \exp[-\mu_\alpha(\lambda)] + \sum_{j=1}^{k} \lambda^j m_j \tag{18}$$

and a corresponding density matrix

$$\rho_r(\lambda) = \left[\sum_{\alpha=1}^{r} \exp[-\mu_\alpha(\lambda)]\right]^{-1} \times \sum_{\beta=1}^{r} \exp[-\mu_\beta(\lambda)] P_\beta(\lambda) . \tag{19}$$

We show in Appendix C that

$$\frac{\partial \Gamma_r(\lambda)}{\partial \lambda^i} = -\mathrm{Tr}[\rho_r(\lambda) M_i] + m_i , \quad i = 1, \ldots, k , \tag{20}$$

which suggests defining the function $\chi^2(\lambda)$ as

$$\chi_r^2(\lambda) = \sum_{i,j=1}^{k} \frac{\partial \Gamma_r(\lambda)}{\partial \lambda^i} (E^{-1})^{ij} \frac{\partial \Gamma_r(\lambda)}{\partial \lambda^j} , \tag{21}$$

where $(E^{-1})^{ij}$ are the elements of the inverse of the error matrix corresponding to the measured m_i. We show in Appendix C that $\Gamma_r(\lambda)$ is a convex function[4] and therefore if it has a local minimum at a point $\bar{\lambda}$ where it is differentiable, its gradient vanishes at $\bar{\lambda}$ and from Eq. (20) we get

$$\mathrm{Tr}[\rho_r(\bar{\lambda}) M_i] = m_i , \quad i = 1, \ldots, k . \tag{22}$$

Furthermore, by construction $\rho_r(\bar{\lambda})$ is Hermitian, positive, and has unit trace and rank r. Thus the matrix $\rho_r(\bar{\lambda})$ is the spin density with maximum information entropy given by

$$E[\rho_r(\bar{\lambda})] = \Gamma_r(\bar{\lambda}) . \tag{23}$$

It is also clear from the definition Eq. (21) that $\chi_r^2(\bar{\lambda})$ vanishes.

When the rank of ρ is equal to its dimension we have

$$\Gamma_\nu(\lambda) = \ln \sum_{\alpha=1}^{\nu} \exp[-\mu_\alpha(\lambda)] + \sum_{j=1}^{k} \lambda^j m_j = \ln \mathrm{Tr}\left[\exp\left(-\sum_{i=1}^{k} \lambda^i M_i\right)\right] + \sum_{j=1}^{k} \lambda^j m_j . \tag{24}$$

Let $\bar{\lambda}$ be a minimum of $\Gamma_\nu(\lambda)$; then

$$\rho_\nu(\bar{\lambda}) = \left\{\mathrm{Tr}\left[\exp\left(-\sum_{i=1}^{k} \bar{\lambda}^i M_i\right)\right]\right\}^{-1} \exp\left(-\sum_{j=1}^{k} \bar{\lambda}^j M_j\right) . \tag{25}$$

We show in Appendix C that $\Gamma_r(\lambda)$ is in general a convex function while $\Gamma_\nu(\lambda)$ is a strictly convex function.[4] The analysis of $\Gamma_\nu(\lambda)$ is thus fairly simple. If the function $\Gamma_\nu(\lambda)$ is bounded from below it has a unique local minimum which is also its global minimum.[4] If this is the case then Eq. (23) implies that $\Gamma_\nu(\lambda) \geq 0$ for any λ. Thus a point λ where $\Gamma_\nu(\lambda)$ is negative is a clear signal that $\Gamma_\nu(\lambda)$ is unbounded from below. But if $\Gamma_\nu(\lambda)$ is unbounded from below then it has no local minimum.

The boundedness of $\Gamma_\nu(\lambda)$ is determined by its behavior on a large enough sphere around the origin in the k-dimensional λ space. If we check the behavior of $\Gamma_\nu(\lambda)$ on a radial ray emanating from the origin we see that $\Gamma_\nu(\lambda)$ is bounded from below if for any unit vector $\hat{\lambda}$

$$\mu_1(\hat{\lambda}) \leq \sum_{i=1}^{k} \hat{\lambda}^i m_i \leq \mu_\nu(\hat{\lambda}) . \tag{26}$$

But for a positive density matrix this condition is always true. Thus if $\Gamma_\nu(\lambda)$ is unbounded from below the measured average values m_i violate the positivity of ρ.

The situation is more complicated if the rank of ρ is strictly smaller than its dimension. In that case there are several possibilities.

If $\Gamma_r(\lambda)$ is unbounded from below then there is

no density matrix of rank r which maximizes the entropy and reproduces the experimental average values. If that is the case one should construct $\Gamma_{r+1}(\lambda)$.

If $\Gamma_r(\lambda)$ is bounded below and attains a local minimum at a point in which it is differentiable, one can construct the density matrix according to Eq. (19).

If $\Gamma_r(\lambda)$ is bounded below and attains a local minimum at a point in which it is not differentiable, we must have at that point $\bar{\lambda}$ a degeneracy of the eigenvalues

$$\mu_r(\bar{\lambda}) = \mu_{r+1}(\bar{\lambda}) . \tag{27}$$

We assume that this is the only degeneracy at $\bar{\lambda}$. In this case the projection operator $P_r(\bar{\lambda})$ is not unique but is a member of a one-parameter family of projectors $P_r(\bar{\lambda}, \theta)$. We then construct $\chi^2(\bar{\lambda}, \theta)$ and minimize it with respect to θ for fixed $\bar{\lambda}$, thus determining θ and the corresponding projector. The density matrix is once again reconstructed according to Eq. (19).

IV. A SIMPLE EXAMPLE

In order to illustrate the procedure of maximization of the information entropy let us consider the following simple example. Let $\nu = r = 2$ so that $f = 3$ and let $k = 1$. As $k = 1$ we drop the corresponding summation indices and write simply m, M, λ. Suppose further that $2\rho_{12} = m$ so that

$$M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .$$

The spectral representation of λM is unique for $\lambda \neq 0$ and given by

$$\lambda M = \lambda \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} - \lambda \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} . \tag{28}$$

Thus

$$\begin{aligned} \Gamma_2(\lambda) &= \ln[\exp(\lambda) + \exp(-\lambda)] + \lambda m \\ &= \ln 2 + \ln \cosh\lambda + \lambda m . \end{aligned} \tag{29}$$

For any finite m this is a strictly convex function. Therefore, if it is bounded from below it has a unique local minimum which is its global minimum. If it is unbounded from below then it has no local minimum. Which of these two possibilities obtains may be determined by examining the behavior of $\Gamma_2(\lambda)$ for large values of $|\lambda|$. Indeed we have

$$\begin{aligned} \Gamma_2(\lambda) &\underset{\lambda\to\infty}{\sim} \lambda(1+m) , \\ \Gamma_2(\lambda) &\underset{\lambda\to-\infty}{\sim} -\lambda(1-m) . \end{aligned} \tag{30}$$

Therefore if $-1 \leq m \leq 1$, $\Gamma_2(\lambda)$ is bounded from below in both directions of increasing $|\lambda|$, while if $1 < m$ or $m < -1$ it is unbounded in one direction. Since the eigenvalues of M are $+1$ and -1 any density matrix ρ must satisfy

$$-1 \leq \mathrm{Tr}(\rho M) \leq 1 . \tag{31}$$

Thus an experimental value for m outside the closed interval $[-1, 1]$ is unphysical. Such a value necessarily violates the positivity of ρ. Given a value of m in the physical range we determine the critical value of λ by equating the first derivative of $\Gamma_2(\lambda)$ to zero,

$$\left.\frac{\partial\Gamma_2}{\partial\lambda}\right|_{\bar{\lambda}} = \tanh\lambda + m \Big|_{\bar{\lambda}} = 0 , \tag{32}$$

$$\bar{\lambda} = -\tanh^{-1} m . \tag{33}$$

Furthermore we have

$$\left.\frac{\partial^2\Gamma_2}{\partial\lambda^2}\right|_{\bar{\lambda}} = 1 - \tanh^2\lambda \Big|_{\bar{\lambda}} = 1 - m^2 , \tag{34}$$

so that for $-1 < m < 1$ we indeed have a minimum. For $m = \mp 1$ the critical point $\bar{\lambda}$ wanders away to $\pm\infty$ and loses the character of a simple minimum. Constructing the density matrix we get

$$\begin{aligned} \rho_2(\bar{\lambda}) &= \frac{1}{2\cosh\bar{\lambda}} \left[\cosh\bar{\lambda} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \sinh\bar{\lambda} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right] \\ &= \frac{1}{2} \begin{pmatrix} 1 & m \\ m & 1 \end{pmatrix} . \end{aligned} \tag{35}$$

Finally

$$\begin{aligned} E[\rho_2(\bar{\lambda})] &= \Gamma_2(\bar{\lambda}) \\ &= \ln 2 + \ln \frac{1}{(1-m^2)^{1/2}} - m \tanh^{-1} m \\ &= -\frac{1-m}{2} \ln\left(\frac{1-m}{2}\right) - \frac{1+m}{2} \ln\left(\frac{1+m}{2}\right) . \end{aligned} \tag{36}$$

We note that $E[\rho]$ attains its *maximum* on the density matrix $\rho_2(\bar{\lambda})$ corresponding to a *minimum* of $\Gamma_2(\lambda)$.

Let us now discuss the case $r = 1$ so that $f = 2$. The function $\Gamma_1(\lambda)$ is well defined everywhere except at $\lambda = 0$ and is given by

$$\Gamma_1(\lambda) = |\lambda| + m\lambda . \tag{37}$$

We extend its definition to $\lambda = 0$ by continuity so that $\Gamma_1(0) = 0$. For m in the open interval $(-1, 1)$ the function $\Gamma_1(\lambda)$ has a global minimum at $\lambda = 0$; however, it is not differentiable at $\lambda = 0$. The reason for this is that at $\lambda = 0$ the two eigenvalues of λM are degenerate and the spectral representation ceases to be unique. Instead we have a one-parameter family of projectors

$$P_1(\theta)=\begin{pmatrix}\cos^2\theta & \sin\theta\cos\theta\\ \sin\theta\cos\theta & \sin^2\theta\end{pmatrix},$$
$$P_2(\theta)=\begin{pmatrix}\sin^2\theta & -\sin\theta\cos\theta\\ -\sin\theta\cos\theta & \cos^2\theta\end{pmatrix}. \quad (38)$$

Let us arbitrarily choose P_1 as the projection operator to be used in Eq. (19) for ρ_1. We then have

$$\rho_1 = P_1 \quad (39)$$

and

$$\mathrm{Tr}(\rho M) = 2\sin\theta\cos\theta = m\,. \quad (40)$$

Therefore

$$\cos^2\theta = \tfrac{1}{2}[1 \pm (1-m^2)^{1/2}], \quad (41)$$

which yields two solutions for ρ_1:

$$\rho_{1^+} = \frac{1}{2}\begin{pmatrix}1+(1-m^2)^{1/2} & m\\ m & 1-(1-m^2)^{1/2}\end{pmatrix} \quad (42)$$

and

$$\rho_{1^-} = \frac{1}{2}\begin{pmatrix}1-(1-m^2)^{1/2} & m\\ m & 1+(1-m^2)^{1/2}\end{pmatrix}. \quad (43)$$

We get the same solutions for ρ with the choice $\rho = P_2$ because

$$P_1\left(\theta+\frac{\pi}{2}\right) = P_2(\theta)\,. \quad (44)$$

Both solutions are matrices of rank 1 and we therefore get

$$E[\rho_{1^\pm}] = \Gamma_1(0) = 0\,. \quad (45)$$

All the details of this simple example have a clear geometric interpretation. The 2×2 spin density matrices can be parametrized by three real numbers x, y, z as follows:

$$\rho = \frac{1}{2}\begin{pmatrix}1+z & x-iy\\ x+iy & 1-z\end{pmatrix}. \quad (46)$$

The positivity of ρ implies

$$x^2+y^2+z^2 \le 1\,. \quad (47)$$

We thus get a correspondence between the points of the unit ball in three dimensions and the 2×2 density matrices. The result of the measurement of $\mathrm{Tr}(\rho M)$, namely $m\pm\sigma$, places the unknown density matrix in the slice of the ball bounded by the two planes $x = m-\sigma$ and $x = m+\sigma$.

The maximum-entropy procedure places the unknown density matrix on the intersection of the slice with the plane $y = 0$. In particular, if we look for a rank-2 maximum-entropy matrix we get the segment denoted in Fig. 1 by ρ_2, while if we look for a rank-1 matrix we get the two arcs denoted by ρ_{1^+} and ρ_{1^-}.

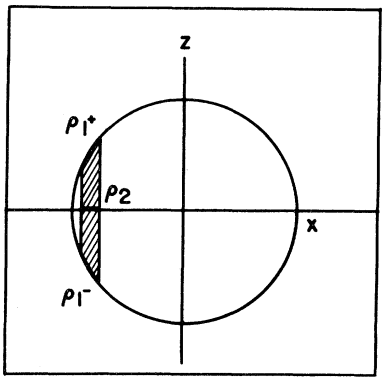

FIG. 1. A geometric interpretation of the evaluation of a 2×2 spin density matrix. The figure shows the intersection of the unit ball in three dimensions with the plane $y = 0$.

An alternative way to describe this is to say that the measurements should be interpreted as giving rise to a probability distribution on the three-dimensional unit ball. The maximum-entropy procedure of doing this is to assume that the desired distribution is uniform in the variables corresponding to the missing measurements. In our example this probability distribution is thus uniform in y and z for given x, while its dependence on x is peaked at m with a width σ.

V. PARAMETRIZATION OF THE DENSITY MATRIX

In our study we adopt a parity-eigenstate basis for the density matrix. In this basis the density matrix takes the form of a block diagonal matrix with two block submatrices on the diagonal.[7] This basis gets rid of the redundancy in the matrix elements due to parity invariance and simplifies the parametrization which we use. To be more specific, in the case which we study as a test of the method, namely, the $\pi^-\pi^+$ system in the f-mass region, the density matrix can be represented by Hermitian submatrices ${}^{(+)}\rho$ and ${}^{(-)}\rho$ (using the notation of Ref. 7) with dimensions six and three, respectively, which obey the trace condition:

$$\mathrm{trace}({}^{(+)}\rho) + \mathrm{trace}({}^{(-)}\rho) = 1\,. \quad (48)$$

A full description of the matrix can be found in Ref. 1. Since only the real part of the density matrix is related to the angular distribution, which we use as a test of our method, ${}^{(+)}\rho$ and ${}^{(-)}\rho$ are restricted to be real symmetric matrices. It is easy to see that the whole density matrix in this case can be described by 26 real parameters.

Each one of the submatrices is by definition positive semidefinite. We are using a parametri-

zation which takes care of this feature and was suggested by Daboul.[3] For the sake of completeness we outline briefly the parametrization of a real symmetric positive-semidefinite matrix with unit trace. Further details and mathematical proofs can be found elsewhere.[3,7]

Consider a real $n\times n$ symmetric positive-semidefinite matrix A. All its elements can be expressed by means of the scalar products of n vectors $y^{(i)}$, $i=1,\ldots,n$, in n-dimensional space:

$$A_{ik}=\sum_{g=1}^{n} y_g^{(i)} y_g^{(k)} . \tag{49}$$

One can always choose a basis such that the matrix $y_j^{(i)}$ is superdiagonal:

$$y_j^{(i)}=0 \quad \text{for } j>i . \tag{50}$$

The other m elements of $y_j^{(i)}$, where

$$m=\frac{n(n+1)}{2} , \tag{51}$$

can be regarded as parameters which fully describe the matrix A. The redundancy, which is caused by the fact that the parameters appear in products only, is eliminated by the requirement that n properly chosen parameters are non-negative, e.g.,

$$y_k^{(k)} \geq 0 \quad k=1,\ldots,n . \tag{52}$$

The ordered m parameters of y can be regarded as an m-dimensional vector V:

$$(V_1 V_2 V_3 V_4 \cdots V_m)=(y_1^{(1)} y_1^{(2)} y_2^{(2)} y_1^{(3)} \cdots y_n^{(n)}) . \tag{53}$$

The condition trace$(A)=1$ requires V to be a unit vector. The elements of this vector, expressed in hyperspherical coordinates in m-dimensional space, can be written as

$$\begin{aligned} V_1 &= \cos\alpha_1 , \\ V_2 &= \cos\alpha_2 \sin\alpha_1 , \\ &\cdots , \\ V_k &= \cos\alpha_k \prod_{j=1}^{k-1} \sin\alpha_j , \\ &\cdots \\ V_m &= \prod_{j=1}^{m-1} \sin\alpha_j . \end{aligned} \tag{54}$$

The $m-1$ angles α_j can finally be regarded as parameters, some of which take values in the interval $[0,\pi]$ and some in the interval $[0,\pi/2]$ so that Eq. (52) is satisfied. If the rank of the matrix A is bounded to be equal at most to r $(r<n)$, then all elements $y_j^{(k)}$ (where $j>r$) vanish and appropriate changes are to be made in the parameters α_j.

Following this prescription we parametrized the six-dimensional matrix ${}^{(+)}\rho$ by a 21-dimensional unit vector V with 20 angle parameters, and the three-dimensional matrix ${}^{(-)}\rho$ by a six-dimensional unit vector W with five angle parameters. The trace condition was fulfilled by multiplying the unit vectors V and W by $\cos\gamma$ and $\sin\gamma$, respectively, where γ is an additional angle parameter in the interval $[0,\pi/2]$.

The rank of the decay-density matrix for $f\to\pi^-\pi^+$ produced in the reaction

$$\pi+N\to f+N \tag{55}$$

which we use later, is limited to 4 if the initial nucleon N is unpolarized and the polarization of the final nucleon is not measured. It has been shown[7] that in general the rank of each of the submatrices is limited to half of the above limit; hence the rank r of either ${}^{(+)}\rho$ or ${}^{(-)}\rho$ cannot exceed $r=2$. On the other hand, since we are using only the real part of the density matrix, this limit has to be doubled. The ${}^{(-)}\rho$ matrix, being three-dimensional, is not affected by this limitation. The ${}^{(+)}\rho$ matrix, being six-dimensional, is affected so that the rank of its real part cannot exceed 4 and the number of parameters can be reduced by 3.

VI. APPLICATION TO EXPERIMENTAL DATA

The test of the method on experimental data was done by using the moments of the $\pi^-\pi^+$ angular distribution in the f-meson mass region from the reaction

$$\pi^- p\to\pi^-\pi^+ n \tag{56}$$

at 12 GeV/c incident momentum,[8] by restricting the mixture of spin states of the $\pi^-\pi^+$ system to $J=0$, 1, and 2.

First the three low-$|t|$ bins of the data were analyzed without applying any restrictions on the rank of the density matrix. The maximization of the entropy was achieved by minimizing the function $\Gamma(\lambda)$, Eq. (18), by means of the minimization routines MINUIT.[9] The minimum turns out to be extremely shallow. The density matrices obtained from this procedure, which reproduce exactly the moments used (i.e., give $\chi^2=0$), were parametrized by the hyperspherical coordinates described in Sec. V. Starting from these matrices a step-by-step exploration was made in the space defined by the hyperspherical coordinates in which the consistency with the experimental data was achieved by constraining the χ^2 to be at most equal to unity. During the exploration, the density-matrix elements were reconstructed from the values of the coordinates and the minimal and maximal

TABLE I. Example of s-channel density-matrix elements of the $\pi^+\pi^-$ system in the f-mass region from reaction $\pi^- p \to \pi^-\pi^+ n$ at 12 GeV/c, obtained by maximization of the information entropy. The notation and the bounds on the elements are taken from Ref. 8.

	$0.02 \leq -t < 0.05$ GeV^2				$0.10 \leq -t < 0.20$ GeV^2			
	Bounds		Maximum-entropy matrix elements		Bounds		Maximum-entropy matrix elements	
Density-matrix elements	Low	High	Unlimited rank	Rank $^{(+)}\rho \leq 4$	Low	High	Unlimited rank	Rank $^{(+)}\rho \leq 4$
$U(2,0)+U(2,1)+U(2,2)$	0.64	0.75	0.678	0.679	0.61	0.70	0.615	0.616
$N(2,1)+N(2,2)$	0.00	0.07	0.016	0.017	0.02	0.08	0.043	0.042
$U(2,0)=\rho_{00}^{22}$	0.57	0.65	0.609	0.609	0.31	0.38	0.337	0.337
$U(2,1)$	0.03	0.12	0.055	0.054	0.24	0.30	0.238	0.239
$U(2,2)$	0.00	0.04	0.015	0.015	0.03	0.07	0.040	0.041
$N(2,1)$	0.00	0.07	0.013	0.014	0.00	0.06	0.032	0.031
$N(2,2)$	0.00	0.03	0.003	0.004	0.00	0.04	0.010	0.011
ρ_{01}^{22}	0.09	0.13	0.109	0.109	0.16	0.19	0.190	0.190
ρ_{00}^{02}	0.29	0.39	0.326	0.323	0.21	0.28	0.221	0.218

values of each one of the matrix elements was recorded. The ranges of the density-matrix elements obtained that way are identical to those obtained by using a different method.[8]

We repeated the procedure on the same data by additionally restricting the rank of the submatrix $^{(+)}\rho$ to 4, as explained in the previous section. The maximum-entropy density matrix obeying this rank condition is nearly the same as the one that is not restricted by the rank condition (see Table I). Also the ranges of the density-matrix elements obtained that way are indistinguishable from those of the matrix with restricted rank. This result shows that the experimental moments are consistent with the rank condition. It is interesting to note that among the eigenvalues obtained by maximization of the entropy there are very small ones (see Table II), which suggests that the actual rank of the matrices could be smaller.

The maximization of the entropy was unsuccessful for the last (fourth) t bin of the data, indicating that the experimental moments there are inconsistent with positivity of the density matrix. A nonpositive-semidefinite density matrix may give rise to the negative values in the angular distribution. It is interesting to note that by reconstructing the angular distribution from the experimental moments, we found regions with negative values, which were missing in the lower-$|t|$ bins.

We found that the calculation of the maximum-entropy density matrix takes up only about one percent of the computer time used. Practically all of the computer time is taken up by the multidimensional exploration which determines the ranges of the density-matrix elements.

By comparing the time of computation using a different method,[8] we found that our method is faster by a factor of about 1.5. The use of rank limitation in our case reduces the number of parameters from 26 to 23. Since these numbers do not differ much among them, the gain in computer time due to the rank limitation only was not substantial.

For higher-spin states, however, the situation should change drastically. For instance, if spin $J=3$ is added in a $\pi\pi$ system the number of parameters is reduced from 75 to 51. Since the time of the multidimensional exploration is proportional to the square of the number of parameters (which is equivalent to the square of the dimension), the application of the rank limitation is expected to contribute an additional factor of 2 in the speed of computation. Combined with the factor of 1.5

TABLE II. Example of the eigenvalues of the density-matrix submatrices for the $\pi^-\pi^+$ system in the f-mass region from the reaction $\pi^- p \to \pi^-\pi^+ n$ at 12 GeV/c obtained by maximation of the information entropy. Results denoted by (a) are obtained without rank restrictions, while those denoted by (b) are obtained with a restriction of the rank of the submatrix $^{(+)}\rho$ to 4.

$-t$ bin (GeV^2)		$^{(-)}\rho$			$^{(+)}\rho$					
0.02–0.05	(a)	0.019	0.005	0.001	0.833	0.102	0.026	0.009	0.004	0.001
	(b)	0.019	0.005	0.002	0.830	0.107	0.027	0.010	0.000	0.000
0.10–0.20	(a)	0.042	0.025	0.003	0.734	0.132	0.041	0.020	0.002	0.001
	(b)	0.041	0.025	0.003	0.731	0.137	0.040	0.023	0.000	0.000

due to the use of maximum-entropy procedure, one obtains a factor of 3 in computer time gain. This factor should increase even faster for higher-spin states.

ACKNOWLEDGMENTS

The work of S.D. was supported by the Israel Academy of Science and Humanities, the Israel Commission for Basic Research. The work of Y.D. was supported in part by the Israel Commission for Basic Research and the U.S.-Israel Bi-National Science Foundation (BSF).

APPENDIX A: THE MAXIMIZATION CONDITIONS

In this appendix we fill in the mathematical details that were left out of the main text.

We defined a ν-dimensional spin density matrix ρ as any $\nu\times\nu$ matrix which fulfills the following conditions:

(a) Hermiticity,

$$\rho=\rho^{\dagger}; \tag{A1}$$

(b) Positivity,

$$x^{\dagger}\rho x\geq 0 \quad \forall x\in \mathbf{C}^{\nu}; \tag{A2}$$

(c) Normalization,

$$\mathrm{Tr}(\rho)=1. \tag{A3}$$

The Hermiticity implies a spectral representation[6] for ρ

$$\rho=\sum_{\alpha=1}^{\nu}\rho_{\alpha}P_{\alpha}, \tag{A4}$$

where the eigenvalues ρ_{α} are real numbers which we arrange in a descending order as follows:

$$\rho_1\geq\rho_2\geq\cdots\geq\rho_{\nu}. \tag{A5}$$

The positivity and trace conditions restrict the eigenvalues further to the unit interval

$$0\leq\rho_{\alpha}\leq 1, \quad \alpha=1,\ldots,\nu. \tag{A6}$$

The number of eigenvalues which are strictly positive gives then the rank of ρ denoted by r. Thus when r is strictly smaller than ν we have

$$\rho_r>0, \quad \rho_{r+1}=\cdots=\rho_{\nu}=0 \tag{A7}$$

and

$$\rho=\sum_{\alpha=1}^{r}\rho_{\alpha}P_{\alpha}. \tag{A8}$$

The projection operators P_{α} are Hermitian, orthonormal, and have unit trace

$$P_{\alpha}=P_{\alpha}^{\dagger}, \quad \alpha=1,\ldots,\nu, \tag{A9}$$

$$P_{\alpha}P_{\beta}=\delta_{\alpha\beta}P_{\beta}, \quad \alpha,\beta=1,\ldots,\nu, \tag{A10}$$

$$\mathrm{Tr}(P_{\alpha})=1, \quad \alpha=1,\ldots,\nu. \tag{A11}$$

The information entropy is defined on all spin density matrices as follows:

$$E[\rho]=-\mathrm{Tr}(\rho\ln\rho)=-\sum_{\alpha=1}^{r}\rho_{\alpha}\ln\rho_{\alpha}. \tag{A12}$$

The restriction (A6) on the ρ_{α} gives

$$E[\rho]\geq 0, \tag{A13}$$

while the concavity[4] of the function $-x\ln x$ gives

$$-\left(\frac{1}{r}\sum_{\alpha=1}^{r}\rho_{\alpha}\right)\ln\left(\frac{1}{r}\sum_{\alpha=1}^{r}\rho_{\alpha}\right)\geq-\sum_{\alpha=1}^{r}\frac{1}{r}\rho_{\alpha}\ln\rho_{\alpha}. \tag{A14}$$

From the normalization condition we now get

$$E[\rho]\leq\ln r. \tag{A15}$$

Let us now maximize $E[\rho]$ subject to the subsidiary conditions

$$\mathrm{Tr}(\rho)=1, \tag{A16}$$

$$\mathrm{Tr}(\rho M_j)=m_j, \quad j=1,\ldots,k. \tag{A17}$$

Introduce the corresponding Lagrange multipliers $\lambda^0,\lambda^1,\ldots,\lambda^k$ and consider the functional

$$\begin{aligned}I[\rho;\lambda]&=-\mathrm{Tr}(\rho\ln\rho)-\lambda^0[\mathrm{Tr}(\rho)-1]-\sum_{j=1}^{k}\lambda^j[\mathrm{Tr}(\rho M_j)-m_j]\\&=-\sum_{\alpha=1}^{r}\rho_{\alpha}\ln\rho_{\alpha}-\lambda^0\left(\sum_{\alpha=1}^{r}\rho_{\alpha}-1\right)-\sum_{j=1}^{k}\lambda^j\left\{\left[\sum_{\alpha=1}^{r}\rho_{\alpha}\mathrm{Tr}(P_{\alpha}M_j)\right]-m_j\right\}.\end{aligned} \tag{A18}$$

To find a local extremum of the functional I we vary ρ_{α}, P_{α}, λ^0, λ^j and equate the resulting first variation of I to zero. However, care should be exercised in the variation in two respects. First, if r, the rank of ρ, is given and is strictly less than ν (the dimension of ρ) then the $\nu-r$ vanishing eigenvalues of ρ should stay fixed at the value zero during the variation. Second, in order not to spoil the Hermiticity and orthonormality of the projectors P_{α} their infinitesimal variations should be of the form

$$\delta P_{\alpha}=i[\delta H,P_{\alpha}],$$

where δH is an arbitrary infinitesimal Hermitian matrix. From this relation we see that

$$\mathrm{Tr}(\delta P_\alpha) = 0 .$$

Computing the variation of I we get

$$\delta I = -\sum_{\alpha=1}^{r} \delta\rho_\alpha (1+\ln\rho_\alpha) - \delta\lambda^0 \left(\sum_{\alpha=1}^{r} \rho_\alpha - 1\right) - \lambda^0 \sum_{\alpha=1}^{r} \delta\rho_\alpha - \sum_{j=1}^{k} \delta\lambda^j \left[\sum_{\alpha=1}^{r} \rho_\alpha \mathrm{Tr}(P_\alpha M_j) - m_j\right]$$
$$- \sum_{j=1}^{k} \lambda^j \sum_{\alpha=1}^{r} \delta\rho_\alpha \mathrm{Tr}(P_\alpha M_j) - \sum_{j=1}^{k} \lambda^j \sum_{\alpha=1}^{r} \rho_\alpha \mathrm{Tr}(\delta H i[P_\alpha, M_j]) . \quad \text{(A19)}$$

To get the last term in the expression we used the equality

$$\mathrm{Tr}(i[\delta H, P_\alpha]M_j) = \mathrm{Tr}(\delta H i[P_\alpha, M_j])$$

which follows from the cyclic property of the trace. At a local extremum δI vanishes, and since the variations $\delta\rho_\alpha$, δH, $\delta\lambda^0$, $\delta\lambda^j$ are all independent we get the following conditions:

$$\ln\rho_\alpha + 1 + \lambda^0 + \sum_{j=1}^{k} \lambda^j \mathrm{Tr}(P_\alpha M_j) = 0 , \quad \text{(A20)}$$

$$\sum_{j=1}^{k} \lambda^j \sum_{\alpha=1}^{r} \rho_\alpha [P_\alpha, M_j] = 0 , \quad \text{(A21)}$$

$$\sum_{\alpha=1}^{r} \rho_\alpha - 1 = 0 , \quad \text{(A22)}$$

$$\sum_{\alpha=1}^{r} \rho_\alpha \mathrm{Tr}(P_\alpha M_j) - m_j = 0 , \quad j = 1, \ldots, k . \quad \text{(A23)}$$

To derive Eq. (A21) one notes that if a Hermitian matrix B satisfies the relation $\mathrm{Tr}(AB) = 0$ for an arbitrary Hermitian matrix A, then B must vanish. Now from the Hermiticity of P_α and M_j it follows that $i[P_\alpha, M]$ is Hermitian, while from the Hermiticity and arbitrariness of δH the desired Eq. (A21) follows.

We now want to solve these conditions and determine the critical values of ρ_α, P_α, λ^0, λ^j. To this end we write the second condition Eq. (A21) in the form

$$[\rho, M(\lambda)] = 0 , \quad \text{(A24)}$$

where ρ is given in Eq. (A8) and $M(\lambda)$ is the Hermitian matrix

$$M(\lambda) = \sum_{j=1}^{k} \lambda^j M_j . \quad \text{(A25)}$$

Note that although a matrix $M(\lambda)$ can be defined over the whole k-dimensional λ space, at the critical value of λ it commutes with the critical density matrix ρ. Therefore at the critical point $M(\lambda)$ and ρ can be diagonalized simultaneously or in other words they have common projectors P_α. Thus at the critical point we have the following spectral representation for M

$$M(\lambda) = \sum_{\alpha=1}^{\nu} \mu_\alpha P_{\pi(\alpha)} , \quad \text{(A26)}$$

where the eigenvalues μ_α are arranged in ascending order

$$\mu_1 \leq \mu_2 \leq \cdots \leq \mu_\nu \quad \text{(A27)}$$

while $\pi(1), \ldots, \pi(\nu)$ is an as yet unspecific permutation of $1, \ldots, \nu$. From Eq. (A26) we get

$$\mu_\alpha = \mathrm{Tr}[P_{\pi(\alpha)} M(\lambda)] = \sum_{j=1}^{k} \lambda^j \mathrm{Tr}(P_{\pi(\alpha)} M_j) . \quad \text{(A28)}$$

Substituting in the first condition, Eq. (A20), we get

$$\rho_\alpha = \exp[-(1+\lambda^0)] \exp[-\mu_{\pi^{-1}(\alpha)}] , \quad \alpha = 1, \ldots, r . \quad \text{(A29)}$$

The third condition Eq. (A22) now determines λ^0 and we have

$$\rho_\alpha = \left(\sum_{\beta=1}^{r} \exp(-\mu_{\pi^{-1}(\beta)})\right)^{-1} \exp(-\mu_{\pi^{-1}(\alpha)}) \quad \text{(A30)}$$

and

$$\rho = \left(\sum_{\beta=1}^{r} \exp(-\mu_{\pi^{-1}(\beta)})\right)^{-1} \sum_{\alpha=1}^{r} \exp(-\mu_{\pi^{-1}(\alpha)}) P_\alpha . \quad \text{(A31)}$$

Now by definition the ρ_α are arranged in a descending order. Therefore, $\mu_{\pi^{-1}(1)}, \ldots, \mu_{\pi^{-1}(r)}$ are arranged in an ascending order, and by the definition of the μ_α we must have

$$1 \leq \pi^{-1}(1) < \pi^{-1}(2) < \cdots < \pi^{-1}(r) \leq \nu . \quad \text{(A32)}$$

To determine the critical point λ we must solve the fourth condition, Eq. (A23). Substituting Eq. (A30) we get

$$\left(\sum_{\beta=1}^{r} \exp(-\mu_{\pi^{-1}(\beta)})\right)^{-1} \sum_{\alpha=1}^{r} \exp(-\mu_{\pi^{-1}(\alpha)}) \times \mathrm{Tr}(P_\alpha M_j) - m_j = 0 . \quad \text{(A33)}$$

In these equations the $\mu_{\pi^{-1}(\alpha)}$ and the P_α are dependent on λ in a nonlinear way. In order to make any progress we must now digress and investigate this dependence. This is done in Appendix B. The reader who is interested in the solution of Eq. (A33) may turn directly to Appendix C.

APPENDIX B: THE PARAMETRIC DEPENDENCE OF μ_α AND P_α: THE FEYNMAN-HELLMANN THEOREM

Define the Hermitian matrix $M(\lambda)$ over the k-dimensional λ space as

$$M(\lambda)=\sum_{j=1}^{k}\lambda^j M_j . \tag{B1}$$

For almost every point λ its eigenvalues are nondegenerate and the spectral representation of $M(\lambda)$ is given as

$$M(\lambda)=\sum_{\alpha=1}^{\nu}\mu_\alpha(\lambda)P_{\pi(\alpha)}(\lambda), \tag{B2}$$

where the eigenvalues μ_α are real and the projectors $P_{\pi(\alpha)}(\lambda)$ are Hermitian, orthonormal, one-dimensional projection operators. From the orthonormality of the $P_{\pi(\alpha)}(\lambda)$ we have

$$\left(\sum_{j=1}^{k}\lambda^j M_j\right)P_{\pi(\beta)}(\lambda)=\mu_\beta(\lambda)P_{\pi(\beta)}(\lambda) \tag{B3}$$

and

$$P_{\pi(\beta)}(\lambda)\sum_{j=1}^{k}\lambda^j M_j=\mu_\beta(\lambda)P_{\pi(\beta)}(\lambda). \tag{B4}$$

Calculating a partial derivative of Eq. (B3) with respect to λ^i we get

$$M_i P_{\pi(\beta)}(\lambda)+\left(\sum_{j=1}^{k}\lambda^j M_j\right)\frac{\partial}{\partial\lambda^i}P_{\pi(\beta)}(\lambda)$$
$$=\frac{\partial\mu_\beta}{\partial\lambda^i}P_{\pi(\beta)}(\lambda)+\mu_\beta(\lambda)\frac{\partial}{\partial\lambda^i}P_{\pi(\beta)}(\lambda). \tag{B5}$$

Multiplying on the left by $P_{\pi(\beta)}(\lambda)$ gives

$$P_{\pi(\beta)}(\lambda)M_i P_{\pi(\beta)}(\lambda)=\frac{\partial\mu_\beta(\lambda)}{\partial\lambda^i}P_{\pi(\beta)}(\lambda). \tag{B6}$$

So taking the trace of both sides yields

$$\frac{\partial\mu_\beta(\lambda)}{\partial\lambda^i}=\mathrm{Tr}(P_{\pi(\beta)}(\lambda)M_i). \tag{B7}$$

This is the Feynman-Hellmann theorem.[10] From it we get the matrices of second derivatives

$$\frac{\partial^2\mu_\beta(\lambda)}{\partial\lambda^i\partial\lambda^j}=\mathrm{Tr}\left(\frac{\partial}{\partial\lambda^j}P_{\pi(\beta)}(\lambda)M_i\right). \tag{B8}$$

Our aim now is to express the right-hand side in a way that does not involve derivatives. Since we have

$$I=\sum_{\alpha=1}^{\nu}P_{\pi(\alpha)}(\lambda) \tag{B9}$$

we may write

$$2\frac{\partial P_{\pi(\beta)}}{\partial\lambda^j}=\sum_{\alpha=1}^{\nu}\left(P_{\pi(\alpha)}\frac{\partial P_{\pi(\beta)}}{\partial\lambda^j}+\frac{\partial P_{\pi(\beta)}}{\partial\lambda^j}P_{\pi(\alpha)}\right). \tag{B10}$$

But being a projector $P_{\pi(\beta)}$ satisfies

$$(P_{\pi(\beta)})^2=P_{\pi(\beta)}, \tag{B11}$$

and differentiating with respect to λ^j gives

$$\frac{\partial P_{\pi(\beta)}}{\partial\lambda^j}=P_{\pi(\beta)}\frac{\partial P_{\pi(\beta)}}{\partial\lambda^j}+\frac{\partial P_{\pi(\beta)}}{\partial\lambda^j}P_{\pi(\beta)}. \tag{B12}$$

Subtracting this line from Eq. (B10) we get

$$\frac{\partial P_{\pi(\beta)}}{\partial\lambda^j}=\sum_{\substack{\alpha=1\\ \alpha\neq\beta}}^{\nu}\left(P_{\pi(\alpha)}\frac{\partial P_{\pi(\beta)}}{\partial\lambda^j}+\frac{\partial P_{\pi(\beta)}}{\partial\lambda^j}P_{\pi(\alpha)}\right). \tag{B13}$$

Multiplying Eq. (B5) on the left by $P_{\pi(\alpha)}$ for $\alpha\neq\beta$ we get

$$P_{\pi(\alpha)}\frac{\partial P_{\pi(\beta)}}{\partial\lambda^j}=\frac{1}{\mu_\beta-\mu_\alpha}P_{\pi(\alpha)}M_j P_{\pi(\beta)}. \tag{B14}$$

We assumed that for $\alpha\neq\beta$ the eigenvalues μ_α and μ_β are nondegenerate. Substituting Eq. (B14) and its Hermitian conjugate in Eq. (B13) we get

$$\frac{\partial P_{\pi(\beta)}}{\partial\lambda^j}=\sum_{\substack{\alpha=1\\ \alpha\neq\beta}}^{\nu}\frac{1}{\mu_\beta-\mu_\alpha}(P_{\pi(\alpha)}M_j P_{\pi(\beta)}$$
$$+P_{\pi(\beta)}M_j P_{\pi(\alpha)}). \tag{B15}$$

When substituted in Eq. (B8) this formula yields the desired result

$$\frac{\partial^2\mu_\beta}{\partial\lambda^i\partial\lambda^j}=\sum_{\substack{\alpha=1\\ \alpha\neq\beta}}^{\nu}\frac{1}{\mu_\beta-\mu_\alpha}[\mathrm{Tr}(P_{\pi(\alpha)}M_j P_{\pi(\beta)}M_i)$$
$$+\mathrm{Tr}(P_{\pi(\beta)}M_j P_{\pi(\alpha)}M_i)]. \tag{B16}$$

APPENDIX C: THE FUNCTION $\Gamma(\lambda)$ AND ITS HESSIAN

The Feynman-Hellmann theorem Eq. (B7) gives a slightly different form of Eq. (A33), namely,

$$\left[\sum_{\beta=1}^{r}\exp(-\mu_{\pi^{-1}(\beta)})\right]^{-1}\sum_{\alpha=1}^{r}\exp(-\mu_{\pi^{-1}(\alpha)})$$
$$\times\frac{\partial\mu_{\pi^{-1}(\alpha)}}{\partial\lambda^j}-m_j=0. \tag{C1}$$

This equation is just the statement that the critical values of λ for the function I are also critical for the function $\Gamma_r^\pi(\lambda)$ defined by

$$\Gamma_r^\pi(\lambda)=\ln\sum_{\alpha=1}^{r}\exp(-\mu_{\pi^{-1}(\alpha)})+\sum_{j=1}^{k}\lambda^j m_j. \tag{C2}$$

The function $\Gamma_r^\pi(\lambda)$ is actually the result of substituting in $I(\rho;\lambda^0,\lambda)$ the expression (A31) for ρ, now interpreted as defined through Eq. (B2) over the k-dimensional λ space, apart from the set of

points of measure zero where $M(\lambda)$ has degenerate eigenvalues.

Since the function $\Gamma_r^{\pi}(\lambda)$ involves in its definition only the values $\pi^{-1}(1),\ldots,\pi^{-1}(r)$ it has $\binom{\nu}{r}$ branches differing from each other by the ordered list $\pi^{-1}(1),\ldots,\pi^{-1}(r)$. In particular there is one branch for which

$$\pi^{-1}(\alpha)=\alpha,\quad \alpha=1,\ldots,r. \tag{C3}$$

We shall refer to this branch as the principal branch and denote it simply by $\Gamma_r(\lambda)$. For $r=\nu$ there is only one branch which is of course the principal branch and $\Gamma_\nu(\lambda)$ is given by

$$\Gamma_\nu(\lambda)=\ln\sum_{\alpha=1}^{\nu}\exp(-\mu_\alpha)+\sum_{j=1}^{k}\lambda^j m_j$$

$$=\ln \mathrm{Tr}\exp\left(-\sum_{j=1}^{k}\lambda^j M_j\right)+\sum_{j=1}^{k}\lambda^j m_j. \tag{C4}$$

To establish the existence and nature of the extremal points of $\Gamma_r^{\pi}(\lambda)$ we consider the Hessian[4] (the matrix of the second derivatives) of $\Gamma_r^{\pi}(\lambda)$:

$$\frac{\partial^2\Gamma_r^{\pi}(\lambda)}{\partial\lambda^i\partial\lambda^j}=\left[\sum_{\alpha=1}^{r}\exp(-\mu_{\pi^{-1}(\alpha)})\right]^{-1}\sum_{\beta=1}^{r}\exp(-\mu_{\pi^{-1}(\beta)})\frac{\partial\mu_{\pi^{-1}(\beta)}}{\partial\lambda^i}\frac{\partial\mu_{\pi^{-1}(\beta)}}{\partial\lambda^j}$$
$$-\left[\sum_{\alpha=1}^{r}\exp(-\mu_{\pi^{-1}(\alpha)})\right]^{-2}\sum_{\beta=1}^{r}\exp(-\mu_{\pi^{-1}(\beta)})\frac{\partial\mu_{\pi^{-1}(\beta)}}{\partial\lambda^i}\sum_{\gamma=1}^{r}\exp(-\mu_{\pi^{-1}(\gamma)})\frac{\partial\mu_{\pi^{-1}(\gamma)}}{\partial\lambda^j}$$
$$-\left[\sum_{\alpha=1}^{r}\exp(-\mu_{\pi^{-1}(\alpha)})\right]^{-1}\sum_{\beta=1}^{r}\exp(-\mu_{\pi^{-1}(\beta)})\frac{\partial^2\mu_{\pi^{-1}(\beta)}}{\partial\lambda^i\partial\lambda^j}. \tag{C5}$$

For an arbitrary real k-dimensional vector with components a^i $i=1,\ldots,k$ define

$$a_{\pi^{-1}(\beta)}=\sum_{i=1}^{k}a^i\frac{\partial\mu_{\pi^{-1}(\beta)}}{\partial\lambda^i} \tag{C6}$$

and

$$\bar{a}=\sum_{\beta=1}^{r}\rho_\beta a_{\pi^{-1}(\beta)}=\left[\sum_{\alpha=1}^{r}\exp(-\mu_{\pi^{-1}(\alpha)})\right]^{-1}\sum_{\beta=1}^{r}\exp(-\mu_{\pi^{-1}(\beta)})a_{\pi^{-1}(\beta)}. \tag{C7}$$

With these definitions we can write

$$\sum_{i,j=1}^{k}a^ia^j\frac{\partial^2\Gamma_r^{\pi}}{\partial\lambda^i\partial\lambda^j}=\sum_{\beta=1}^{r}\rho_\beta(a_{\pi^{-1}(\beta)}-\bar{a})^2-\sum_{\beta=1}^{r}\rho_\beta\sum_{i,j=1}^{k}a^ia^j\frac{\partial^2\mu_{\pi^{-1}(\beta)}}{\partial\lambda^i\partial\lambda^j}. \tag{C8}$$

Equation (B16) allows us to rewrite the last term in the form

$$\sum_{\beta=1}^{r}\rho_\beta\sum_{i,j=1}^{k}a^ia^j\frac{\partial^2\mu_{\pi^{-1}(\beta)}}{\partial\lambda^i\partial\lambda^j}=2\sum_{\beta=1}^{r}\rho_\beta\sum_{\substack{\alpha=1\\ \alpha\neq\beta}}^{\nu}\frac{1}{\mu_{\pi^{-1}(\beta)}-\mu_{\pi^{-1}(\alpha)}}\mathrm{Tr}[P_\alpha(\lambda)M(a)P_\beta(\lambda)M(a)], \tag{C9}$$

where

$$M(a)=\sum_{j=1}^{k}a^jM_j \tag{C10}$$

and the projectors $P_\alpha(\lambda)$, $P_\beta(\lambda)$ are projectors on eigenvectors of $M(\lambda)$. The expression $\mathrm{Tr}[P_\alpha(\lambda)M(a)P_\beta(\lambda)M(a)]$ is symmetric in α and β. It is therefore convenient to separate the right-hand side of Eq. (C9) into two parts:

$$\sum_{\beta=1}^{r}\rho_\beta\sum_{i,j=1}^{k}a^ia^j\frac{\partial^2\mu_{\pi^{-1}(\beta)}}{\partial\lambda^i\partial\lambda^j}=2\sum_{\alpha,\beta=1}^{k}\rho_\beta\frac{1}{\mu_{\pi^{-1}(\beta)}-\mu_{\pi^{-1}(\alpha)}}\mathrm{Tr}[P_\alpha(\lambda)M(a)P_\beta(\lambda)M(a)]$$
$$+2\sum_{\beta=1}^{r}\sum_{\alpha=r+1}^{\nu}\rho_\beta\frac{1}{\mu_{\pi^{-1}(\beta)}-\mu_{\pi^{-1}(\alpha)}}\mathrm{Tr}[P_\alpha(\lambda)M(a)P_\beta(\lambda)M(a)]$$
$$=\left[\sum_{\gamma=1}^{r}\exp(-\mu_{\pi^{-1}(\gamma)})\right]^{-1}\sum_{\substack{\alpha,\beta=1\\ \alpha\neq\beta}}^{r}\left[\frac{\exp(-\mu_{\pi^{-1}(\alpha)})}{\mu_{\pi^{-1}(\beta)}-\mu_{\pi^{-1}(\alpha)}}+\frac{\exp(-\mu_{\pi^{-1}(\alpha)})}{\mu_{\pi^{-1}(\alpha)}-\mu_{\pi^{-1}(\beta)}}\right]\mathrm{Tr}[P_\alpha(\lambda)M(a)P_\beta(\lambda)M(a)]$$
$$+\left[\sum_{\gamma=1}^{r}\exp(-\mu_{\pi^{-1}(\gamma)})\right]^{-1}2\sum_{\beta=1}^{r}\sum_{\alpha=r+1}^{\nu}\exp(-\mu_{\pi^{-1}(\beta)})\frac{1}{\mu_{\pi^{-1}(\beta)}-\mu_{\pi^{-1}(\alpha)}}\mathrm{Tr}[P_\alpha(\lambda)M(a)P_\beta(\lambda)M(a)]. \tag{C11}$$

But

$$\frac{\exp(-\mu_{\pi^{-1}(\beta)})}{\mu_{\pi^{-1}(\beta)}-\mu_{\pi^{-1}(\alpha)}}+\frac{\exp(-\mu_{\pi^{-1}(\alpha)})}{\mu_{\pi^{-1}(\alpha)}-\mu_{\pi^{-1}(\beta)}}=\exp\left[-\tfrac{1}{2}(\mu_{\pi^{-1}(\beta)}+\mu_{\pi^{-1}(\alpha)})\right]\frac{\sinh\left[\frac{1}{2}(\mu_{\pi^{-1}(\beta)}-\mu_{\pi^{-1}(\alpha)})\right]}{\frac{1}{2}(\mu_{\pi^{-1}(\beta)}-\mu_{\pi^{-1}(\alpha)})}, \tag{C12}$$

so that finally we have the desired result

$$\sum_{i,j=1}^{k} a^i a^j \frac{\partial^2 \Gamma_r^{\pi}}{\partial\lambda^i \partial\lambda^j}=\sum_{\beta=1}^{r}\rho_r(a_{\pi^{-1}(\beta)}-\bar{a})^2$$

$$+\left[\sum_{\gamma=1}^{r}\exp(-\mu_{\pi^{-1}(\gamma)})\right]^{-1}\sum_{\substack{\alpha,\beta=1\\ \alpha\neq\beta}}^{r}\exp\left[-\tfrac{1}{2}(\mu_{\pi^{-1}(\beta)}+\mu_{\pi^{-1}(\alpha)})\right]\frac{\sinh\left[\frac{1}{2}(\mu_{\pi^{-1}(\beta)}-\mu_{\pi^{-1}(\alpha)})\right]}{\frac{1}{2}(\mu_{\pi^{-1}(\beta)}-\mu_{\pi^{-1}(\alpha)})}\mathrm{Tr}[P_\alpha M(a) P_\beta M(a)]$$

$$+2\sum_{\beta=1}^{r}\sum_{\alpha=r+1}^{\nu}\rho_\beta\frac{1}{\mu_{\pi^{-1}(\beta)}-\mu_{\pi^{-1}(\alpha)}}\mathrm{Tr}\left[P_\alpha(\lambda)M(a)P_\beta(\lambda)M(a)\right]. \tag{C13}$$

The first two terms in this expression are non-negative. This follows from the equality

$$\mathrm{Tr}(P_\alpha M P_\beta M)=\mathrm{Tr}[(P_\alpha M P_\beta)(P_\alpha M P_\beta)^\dagger] \tag{C14}$$

and the positivity of the function $\sinh x/x$. The sign of the contribution of the last term depends on the branch of the function $\Gamma_r^{\pi}(\lambda)$ we are analyzing. In particular if it is true that

$$\mu_{\pi^{-1}(\alpha)}-\mu_{\pi^{-1}(\beta)}>0,\quad \alpha=r+1,\ldots,\nu\quad \beta=1,\ldots,r \tag{C15}$$

then this last term is also non-negative. This happens if $\mu_{\pi^{-1}(\beta)}$ for $\beta=1,\ldots,r$ are the r smallest eigenvalues and this in turn is the definition of the principal branch $\Gamma_r(\lambda)$. Thus the principal branch $\Gamma_r(\lambda)$ is a convex function. Restricting ourselves now to the principal branch we see that the right-hand side of Eq. (C13) vanishes if the following conditions hold simultaneously:

$$a_\beta=\bar{a},\quad \beta=1,\ldots,r, \tag{C16}$$

$$\mathrm{Tr}[P_\alpha(\lambda)M(a)P_\beta(\lambda)M(a)]=0,\quad \alpha,\beta=1,\ldots,r,\ \alpha\neq\beta, \tag{C17}$$

$$\mathrm{Tr}[P_\alpha(\lambda)M(a)P_\beta(\lambda)M(a)]=0,$$

$$\alpha=r+1,\ldots,\nu,\ \beta=1,\ldots,r. \tag{C18}$$

The first equation (C16) can be rewritten with the help of Eqs. (C6), (C7), and (B7) as

$$\mathrm{Tr}\{P_\beta(\lambda)[M(a)-\bar{a}P(\lambda)]\}=0,\quad \beta=1,\ldots,r, \tag{C19}$$

where

$$P(\lambda)=\sum_{\beta=1}^{r}P_\beta(\lambda). \tag{C20}$$

Together with the other two equations (C17) and (C18) this gives

$$[M(a)-\bar{a}I]P(\lambda)=0 \tag{C21}$$

or

$$\sum_{i=1}^{k}a^i M_i P(\lambda)-\bar{a}P(\lambda)=0. \tag{C22}$$

This then implies that the $k+1$ matrices $P(\lambda)$, $M_1P(\lambda),\ldots,M_kP(\lambda)$ are linearly dependent. Let us apply this result to the case $r=\nu$. In this case $P=I$, but then the $k+1$ matrices $I, M_1,\ldots,M_k$ were assumed to be linearly independent so Eq. (C22) cannot be fulfilled with nonvanishing $\bar{a}$, $a^1,\ldots,a^k$. Thus the function $\Gamma_\nu(\lambda)$ is strictly convex. It is interesting that although we started from maximizing a concave functional $E[\rho]$ we ended up with a convex function $\Gamma(\lambda)$. In Appendix D we show that this is not an accident.

APPENDIX D: THE ORIGIN OF THE CONVEXITY OF $\Gamma_r(\lambda)$

To understand the origin of the convexity of $\Gamma_r(\lambda)$ we discuss now the general set up in which it happens. Consider then a real function $g(x_1,\ldots,x_f)$ of f variables. The function $g(x)$ is supposed to be strictly concave so that the Hessian matrix $\partial^2 g(x)/\partial x_\alpha \partial x_\beta$ is supposed to be negative definite. We now maximize $g(x)$ subject to the k *linear* constraints

$$h_i(x_1,\ldots,x_f),\quad i=1,\ldots,k,\quad k<f. \tag{D1}$$

To find the maximum we introduce k Lagrange multipliers $\lambda^1,\ldots,\lambda^k$ and form the function $G(x;\lambda)$ given by

$$G(x;\lambda)=g(x)-\sum_{i=1}^{k}\lambda^i h_i(x). \tag{D2}$$

To find the constrained maximum of $g(x)$ we look for the extremum of $G(x;\lambda)$ in the $f+k$-dimensional space. The conditions for an extremum are

$$\frac{\partial G}{\partial x_\alpha}=\frac{\partial g}{\partial x_\alpha}-\sum_{i=1}^{k}\lambda^i\frac{\partial h_i}{\partial x_\alpha}=0,\quad \alpha=1,\ldots,f, \tag{D3}$$

$$\frac{\partial G}{\partial \lambda^i}=h_i(x)=0,\quad i=1,\ldots,k. \tag{D4}$$

The concavity of $g(x)$ and the linearity of the $h_i(x)$ imply that the determinant of the matrix $\partial^2 G/\partial x_\alpha \partial x_\beta$ is nonvanishing. This results in our ability to solve locally the f equations (D3) for the f unknowns $x_1, \ldots, x_f$ and express them as functions of the k Lagrange multipliers $\lambda^1, \ldots, \lambda^k$. We thus define k functions

$$x_\alpha = x_\alpha(\lambda), \quad \alpha = 1, \ldots, f,$$

that fulfill

$$\frac{\partial G}{\partial x_\alpha}(x(\lambda); \lambda) = 0 \tag{D5}$$

identically in the λ. Substituting $x_\alpha(\lambda)$ in the function G we now get a new function $F(\lambda)$ of k variables

$$F(\lambda) = G(x(\lambda); \lambda). \tag{D6}$$

Let us calculate its gradient

$$\frac{\partial F}{\partial \lambda^i} = \sum_{\alpha=1}^{f} \frac{\partial G}{\partial x_\alpha}(x(\lambda); \lambda) \frac{\partial x_\alpha}{\partial \lambda^i} + \frac{\partial G}{\partial \lambda^i}(x(\lambda); \lambda)$$
$$= h_i(x(\lambda)). \tag{D7}$$

Critical points of $F(\lambda)$ are points $\bar{\lambda}$ where the gradient $\partial F/\partial \lambda^i$ vanishes. Therefore the functions $x_\alpha(\lambda)$ map a critical point of $F(\lambda)$ on a critical point of the function $G(x; \lambda)$,

$$\bar{\lambda} \to (x(\bar{\lambda}); \bar{\lambda}), \tag{D8}$$

with the useful relation

$$F(\bar{\lambda}) = G(x(\bar{\lambda}); \bar{\lambda}) = g(x(\bar{\lambda})). \tag{D9}$$

The Hessian of $F(\lambda)$ is given by

$$\frac{\partial^2 F}{\partial \lambda^i \partial \lambda^j} = \sum_{\alpha=1}^{f} \frac{\partial h_i}{\partial x_\alpha}(x(\lambda)) \frac{\partial x_\alpha}{\partial \lambda^j}. \tag{D10}$$

Since Eqs. (D5) are f identities in λ, we get by differentiating them with respect to λ,

$$\sum_{\beta=1}^{f} \frac{\partial^2 G}{\partial x_\alpha \partial x_\beta}(x(\lambda); \lambda) \frac{\partial x_\beta}{\partial \lambda^i} + \frac{\partial^2 G}{\partial x_\alpha \partial \lambda^i}(x(\lambda); \lambda) = 0. \tag{D11}$$

The definition of $G(x; \lambda)$ and the linearity of the constraints $h_i(x)$ allow us to rewrite this as

$$\sum_{\beta=1}^{f} \frac{\partial^2 g}{\partial x_\alpha \partial x_\beta}(x(\lambda)) \frac{\partial x_\beta}{\partial \lambda^i} + \frac{\partial h_i}{\partial x_\alpha}(x(\lambda)) = 0. \tag{D12}$$

Substituting in Eq. (D10) we get

$$\frac{\partial^2 F}{\partial \lambda^i \partial \lambda^j} = -\sum_{\alpha,\beta=1}^{f} \frac{\partial^2 g}{\partial x_\alpha \partial x_\beta}(x(\lambda)) \frac{\partial x_\beta}{\partial \lambda^i} \frac{\partial x_\alpha}{\partial \lambda^j}. \tag{D13}$$

This shows that if g is strictly concave so that the matrix $\partial^2 g/\partial x_\alpha \partial x_\beta$ is negative definite, then the matrix $\partial^2 F/\partial \lambda^i \partial \lambda^j$ is positive semidefinite. In other words the functions $x(\lambda)$ map the minimum of $F(\lambda)$, if it exists, on the constrained maximum of $g(x)$.

The result we get is directly applicable to the analysis of the function $\Gamma_r(\lambda)$ because the constraints (A17) on the density matrix ρ are linear. Indeed we started from the information entropy which is strictly concave and formed $I[\rho, \lambda]$. We formed the functions $\Gamma_r^\pi(\lambda)$ in Eq. (C2) by substituting ρ in terms of λ. Our analysis in the present Appendix shows that the function formed this way should be convex. The analysis in Appendix C then shows that the branch of $\Gamma_r^\pi(\lambda)$ that fulfills this condition is the principal branch. This finally completes our proof of the results in Sec. III.

[1] G. Alexander, S. Dagan, and J. Grunhaus, Nucl. Phys. B157, 425 (1979).

[2] E. H. Wychmann, J. Math. Phys. 4, 884 (1963). See also Amnon Katz, *Principles of Statistical Mechanics, (The Information Theory Approach)* (Freeman, San Francisco, 1967), especially pp. 45–51. Most of our proofs are adaptations of the proofs of the book to the case in which a rank condition is to be satisfied.

[3] J. Daboul, Nucl. Phys. B4, 180 (1967).

[4] Joseph Stoer and Christoph Witzgall, *Convexity and Optimization in Finite Dimensions I* (Springer, Berlin, 1970), Chap. 4; R. T. Rockafellar, *Convex Analysis* (Princeton Univ., Princeton, N. J., 1970), especially Chap. 27.

[5] M. G. Doncel, L. Michel, and P. Minnaert, Nucl. Phys. B38, 477 (1972). For a slightly different attitude see, e. g., Y. Alhassid and R. D. Levine, Phys. Rev. A 18, 89 (1978).

[6] Paul R. Halmos, *Finite Dimensional Vector Spaces*, 2nd ed. (Van Nostrand Reinhold, New York, 1958), pp. 153–158.

[7] S. U. Chung and T. L. Trueman, Phys. Rev. D 11, 633 (1975).

[8] G. Alexander *et al.*, Nucl. Phys. B131, 365 (1977).

[9] F. James and M. Roos, Comput. Phys. Commun. 10, 343 (1975).

[10] R. P. Feynman, Phys. Rev. 56, 340 (1939); H. Hellmann, *Einführung in die Quantenchemie* (Deuticke, Leipzig, 1937).