Minimum Kullback entropy approach to the Fokker-Planck equation

A. R. Plastino,[1,*] H. G. Miller,[1,†] and A. Plastino[2,‡]
[1]*Department of Physics, University of Pretoria, Pretoria 0002, South Africa*
[2]*Departamento de Física, Universidad Nacional de La Plata, Casilla de Correo 727, 1900 La Plata, Argentina*
(Received 10 February 1997; revised manuscript received 24 April 1997)

We consider a minimum Kullback entropy approach to determine approximate, time-dependent solutions to the N-dimensional Fokker-Planck (FP) equation. It is shown that the ensuing approximate solutions to the FP equation can be derived from a variational principle. We prove that the functional relation between the time derivative of the entropy, on the one hand, and the approximate (time-dependent) distribution functions, on the other, has the same form as that corresponding to *exact* solutions to the FP equation. Other properties of the approximate Kullback solutions and some particular examples are also discussed. [S1063-651X(97)04010-5]

PACS number(s): 05.45.+b, 05.30.−d

I. INTRODUCTION

Inspired by Shannon's information theory [1], Jaynes's reformulation of statistical mechanics [2–8] greatly increased its power and scope. Jaynes provided a general prescription for the construction of a probability distribution $f(\mathbf{x})$ ($\mathbf{x}\in\mathbb{R}^N$ stands for a point in the relevant phase space), when the only available information about the system consists of the mean values of M quantities

$$\langle A_r(\mathbf{x})\rangle \equiv \int A_r(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \quad (r=1,\ldots,M). \tag{1}$$

According to Jaynes, the least biased distribution compatible with the data (1) is the one that maximizes Shannon's information measure,

$$S \equiv -\int f(\mathbf{x}) \ln f(\mathbf{x}) d\mathbf{x}, \tag{2}$$

under the constraints imposed by the mean values (1) and the appropriate normalization condition

$$\int f(\mathbf{x}) d\mathbf{x} = 1. \tag{3}$$

Jaynes's information theory approach allows us to consider more general statistical ensembles than the Gibbs microcanonical, canonical, and macrocanonical ensembles [9]. Also, it provides a natural way to treat nonequillibrium situations [9,10].

The maximum entropy principle can be applied to find approximate solutions of partial differential equations governing the time evolution of distribution functions, such as the Liouville equation [11], the Fokker-Planck equation [12,13], and the von Neumann equation in quantum mechanics [14–17]. The general idea of this approach is to follow the time evolution of just a small number of relevant mean values. With recourse to a maximum entropy ansatz for the probability distribution, a closed set of ordinary differential equations governing the evolution of these mean values is obtained [11].

The aforementioned maximum entropy scheme has been applied to a variety of physical situations. In the case of the Liouville equation (or von Neumann's in the quantum case) for Hamiltonian systems, the method has been studied in great detail [18–26]. However, the ideas of statistical mechanics can be applied to more general dynamical systems relevant in diverse areas of physics and theoretical biology. Recently, two of us showed that the maximum entropy formalism can be generalized to statistically treat the whole family of dynamical systems with divergenceless phase-space flows [11]. Hick and Stevens [12] (HS) studied Jaynes's method for obtaining approximate solutions to diffusionlike and Fokker-Planck equations. They applied such an approach to an important astrophysical problem, the cosmic ray transport equation, and obtained remarkably good numerical results. Baker-Jarvis, Racine, and Alameddine [13] have applied Jaynes's techniques to some simple examples of the Fokker-Planck equation.

A drawback of Jaynes's information theory approach for obtaining approximate solutions of differential equations is that it is very difficult to quantitatively ascertain the accuracy of the approximate results obtained. This regrettable situation arises whenever one deals with approximate treatments devised on the basis of some variational principle: Given an appropriate ansatz, the principle always provides a solution whose quality depends on the details of the particular case. HS suggested that in the case of the maximum entropy method, since the procedure implies maximizing the entropy (subject to certain constraints), one can expect it to be particularly useful in problems where the entropy of the real solution can be shown to verify an H theorem (i.e., to be a monotonically increasing function of time). This is what happens in the case of diffusionlike equations. Solutions of the more general Fokker-Planck (FP) equation, however, do not behave in this way. In order to obtain an H theorem for the FP equation [27], it is necessary to employ the relative Kullback [28] entropy between two time-dependent solutions instead of the Shannon entropy involving just one solution.

*Permanent address: Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, Casilla de Correo 727, 1900 La Plata, Argentina.
†Electronic address: hmiller@scientia.up.ac.za
‡Electronic address: plastino@venus.fisica.unlp.edu.ar

This suggests that it would be worthwhile to consider a more general maximum entropy formalism for solving approximately the FP equation, based on the extremalization of Kullback relative information entropy.

The aim of the present effort is to consider a minimum Kullback entropy scheme in order to approximately solve the general N-dimensional FP equation. Our motivation is twofold. On the one hand, the present formalism will unify and generalize some previous results that already exist in the literature. On the other hand, we will analytically prove some properties of the maximum entropy solution of the FP equation, giving insights into a method whose validity has so far come mostly from the analysis of numerical examples.

The paper is organized as follows. In Sec. II we give a brief review of some properties of the FP equation and the Kullback information measure. In Sec. III we introduce a variational principle for the solution of the FP equation, from which our approximate solutions can be derived. The main properties of the Kullback's extremalization approach to the FP equation are discussed in Sec. IV. Some important particular cases are considered in Sec. V. Our main conclusions are drawn in Sec. VI.

II. THE FOKKER-PLANCK EQUATION AND KULLBACK'S INFORMATION MEASURE

A. The Fokker-Planck operator

The N-dimensional FP equation can be cast as [27]

$$\frac{\partial W(\mathbf{x},t)}{\partial t}=L_{FP}W(\mathbf{x},t), \tag{4}$$

where $W(\mathbf{x},t)$ is a time-dependent normalized density function

$$\int W(\mathbf{x},t)d\mathbf{x}=1, \tag{5}$$

$\mathbf{x}$ is a vector belonging to $\mathbb{R}^N$, and the FP differential operator L_{FP} is given, in terms of a drift vector $\mathbf{V}_D$ with components $D_i(\mathbf{x})$ and of a diffusion tensor $\mathbf{D}$ with components $D_{ij}(\mathbf{x})$, by (Einstein convention employed)

$$L_{FP}=-\frac{\partial}{\partial x_i}D_i(\mathbf{x})+\frac{\partial^2}{\partial x_i\partial x_j}D_{ij}(\mathbf{x}). \tag{6}$$

The matrix $\{D_{ij}\}$ is assumed to be symmetric and positive definite, i.e.,

$$D_{ij}v_iv_j\geqslant 0, \tag{7}$$

for all vectors $\mathbf{v}\in\mathbb{R}^N$. The adjoint operator $L_{FP}^{\dagger}$ is

$$L_{FP}^{\dagger}=D_i(\mathbf{x})\frac{\partial}{\partial x_i}+D_{ij}(\mathbf{x})\frac{\partial^2}{\partial x_i\partial x_j}. \tag{8}$$

Given two probability distributions W_1 and W_2, the operators L_{FP} and $L_{FP}^{\dagger}$ verify

$$\int W_1L_{FP}[W_2]d\mathbf{x}=\int W_2L_{FP}^{\dagger}[W_1]d\mathbf{x}. \tag{9}$$

B. Kullback's relative information measure

The relative information measure $Q[f|f_0]$,

$$Q[f(\mathbf{x})|f_0(\mathbf{x})]\equiv\int f(\mathbf{x})\ln\left[\frac{f(\mathbf{x})}{f_0(\mathbf{x})}\right]d\mathbf{x}, \tag{10}$$

was originally introduced by Kullback [28] and can be regarded as a *statistical* distance between two probability distributions [29–31]. If we take f_0 to be the *uniform* probability distribution one immediately finds [28]

$$Q[f|f_0]=\text{const}-S, \tag{11}$$

where S stands for Shannon's measure [6]. Thus minimizing the Kullback distance is tantamount to maximizing S. In what follows it will prove convenient to work with a quantity equal to *minus* the Kullback distance. Thus we set (and shall exclusively use here from)

$$K[f_1|f_2]=-Q[f_1|f_2]. \tag{12}$$

If we have a prior bias towards the distribution $f_0(\mathbf{x})$, a reasonable prescription for statistical inference [28] is to adopt the distribution that minimizes Kullback's information (*maximizes* $K[f|f_0]$) under the constraints imposed by normalization and the relevant mean values (1). That minimum Kullback entropy distribution is given by [5]

$$f_{MK}(\mathbf{x})=\frac{1}{Z}f_0(\mathbf{x})\exp\left(-\sum_{r=1}^{M}\lambda_rA_r(\mathbf{x})\right), \tag{13}$$

where $\{\lambda_r, r=1,\ldots,M\}$ are the M Lagrange multipliers associated with the known mean values and the partition function Z is given by

$$Z\equiv\int f_0(\mathbf{x})\exp\left(-\sum_{r=1}^{M}\lambda_rA_r(\mathbf{x})\right)d\mathbf{x}. \tag{14}$$

Of course, if we have no prior information at all, Kullback's inference principle yields the distribution $f_0(\mathbf{x})$. Kullback's approach is a straightforward generalization of Jaynes's prescription to the case where an initial, nonuniform distribution $f_0(\mathbf{x})$ is available. The relevant mean values $\langle A_i\rangle$ and the concomitant Lagrange multipliers λ_i are connected by the well-known relation

$$\frac{\partial K}{\partial\langle A_i\rangle}=\lambda_i. \tag{15}$$

C. Time derivative of Kullback's relative entropy

Given two distributions W and W_0, where W is a solution of the FP equation and W_0 is an arbitrary prior distribution, the time derivative of Kullback's relative entropy is given by

$$\frac{dK}{dt}=-\int(L_{FP}W)\ln\left(\frac{W}{W_0}\right)d\mathbf{x}+\int\frac{\partial W_0}{\partial t}\left(\frac{W}{W_0}\right)d\mathbf{x}. \tag{16}$$

In the particular case where W_0 is itself a solution to the FP equation, it can be shown that [27]

$$\frac{dK}{dt}=\int \frac{W^3}{W_0^2}D_{ij}\frac{\partial}{\partial x_i}\left(\frac{W}{W_0}\right)\frac{\partial}{\partial x_j}\left(\frac{W}{W_0}\right)d\mathbf{x} \tag{17}$$

and since D_{ij} is assumed to be a positive definite matrix

$$\frac{dK}{dt}\geqslant 0, \tag{18}$$

which constitutes an H theorem for the FP equation [27].

III. VARIATIONAL PRINCIPLE

We introduce now an appropriate variational principle for the FP equation, using an auxiliary quantity $A(\mathbf{x},t)$. The variational principle should provide us with equations of motion for both W and A. The idea of introducing auxiliary quantities is commonly employed in order to formulate variational principles leading to various important evolution equations of mathematical physics [32]. We start by devising a suitable action for the FP equation,

$$I=\int_{t_1}^{t_2}dt\int d\mathbf{x}\left\{\frac{\partial W}{\partial t}-L_{FP}W\right\}A(\mathbf{x},t)+\int_{t_1}^{t_2}D[W]dt - \int W(\mathbf{x},t_2)A(\mathbf{x},t_2)d\mathbf{x}, \tag{19}$$

where the functional $D[W]$ is defined as

$$D[W]=-\int L_{FP}W\ln\left(\frac{W}{W_0}\right)d\mathbf{x}+\int\frac{\partial W_0}{\partial t}\left(\frac{W}{W_0}\right)d\mathbf{x}. \tag{20}$$

and W_0 is an arbitrary prior distribution not necessarily a solution of the FP equation. Notice that if $W(\mathbf{x},t)$ is an *exact* solution of the FP equation, we then have

$$\frac{dK}{dt}=D[W]. \tag{21}$$

Let us consider the variational principle

$$\delta I=0, \tag{22}$$

with mixed boundary conditions

$$W(\mathbf{x},t_1)=W(\mathbf{x})_{in} \tag{23}$$

and

$$A(\mathbf{x},t_2)=A(\mathbf{x})_{out}. \tag{24}$$

The above variational principle implies that for arbitrary variations $\delta W(\mathbf{x},t)$ and $\delta A(\mathbf{x},t)$, we should have

$$\int_{t_1}^{t_2}dt\int d\mathbf{x}\left\{\frac{\partial W}{\partial t}-L_{FP}W\right\}\delta A=0 \tag{25}$$

and

$$\int_{t_1}^{t_2}dt\int d\mathbf{x}\left\{\frac{\partial A}{\partial t}+L_{FP}^{\dagger}A-\frac{\delta D[W]}{\delta W}\right\}\delta W=0, \tag{26}$$

where

$$\frac{\delta D[W]}{\delta W}=-L_{FP}^{\dagger}\left\{\ln\left(\frac{W}{W_0}\right)\right\}-\frac{1}{W}(L_{FP}W)+\frac{\partial(\ln W_0)}{\partial t} \tag{27}$$

is the functional derivative of $D[W]$.

From Eqs. (25) and (26) we obtain partial differential equations for both $W(\mathbf{x},t)$ and $A(\mathbf{x},t)$. Our equations are

$$\frac{\partial W(\mathbf{x},t)}{\partial t}-L_{FP}W(\mathbf{x},t)=0, \tag{28}$$

i.e., the Fokker-Planck equation, and

$$\frac{\partial A}{\partial t}+L_{FP}^{\dagger}A-\frac{\delta D[W]}{\delta W}=0. \tag{29}$$

If we could perform an unrestricted variational procedure (VP), the exact FP solution would ensue. A *restricted* VP, on the other hand, would provide us with approximate FP solutions. We construct now a such a variational ansatz for the two functions $W(\mathbf{x},t)$ and $A(\mathbf{x},t)$, expressed in terms of M appropriate functions $A_i(\mathbf{x})$. The ansatz for A is just an (inhomogeneous) linear combination of the $A_i(\mathbf{x})$,

$$A(\mathbf{x},t)=\alpha_0(\alpha_1,\ldots,\alpha_M)+\sum_{i=1}^{M}\alpha_i(t)A_i(\mathbf{x}), \tag{30}$$

while the ansatz for W is a distribution (verifying the appropriate normalization and boundary conditions) parametrized in terms of the M mean values $\langle A_i\rangle$,

$$W(\mathbf{x},t)=W(\langle A_1\rangle,\ldots,\langle A_M\rangle). \tag{31}$$

In what follows the mean values $\langle A_i\rangle$ are regarded as the M variational parameters characterizing the ansatz for W. Hence the two functions A and W are given in terms of a set of $2M$ variational parameters

$$\{\alpha_1,\ldots,\alpha_M,\langle A_1\rangle,\ldots,\langle A_M\rangle\}. \tag{32}$$

As stated above, the procedure is an approximate one because the set (32) is not the most general one could think of.

Introduction of the ansatz for $A(\mathbf{x},t)$ into the variational equation (25) yields

$$\int_{t_1}^{t_2}dt\,\delta\alpha_0(t)\int d\mathbf{x}\left\{\frac{\partial W}{\partial t}-L_{FP}W\right\} + \sum_{i=1}^{M}\int_{t_1}^{t_2}dt\,\delta\alpha_i(t)\int d\mathbf{x}\left\{\frac{\partial W}{\partial t}-L_{FP}W\right\}A_i(\mathbf{x})=0. \tag{33}$$

Due to both the normalization and the boundary conditions on $W(\mathbf{x},t)$, the first term in Eq. (33) vanishes and since the $\delta\alpha_i(t)$ are arbitrary we obtain

$$\frac{d\langle A_i\rangle}{dt}=\langle L_{FP}^{+}A_i\rangle \quad (i=1,\ldots,M). \tag{34}$$

Furthermore, from the ansatz (31), the variational equation (26), and taking into account that the variations $\delta\langle A_i\rangle$ are arbitrary, we obtain

$$\int \left(\frac{\partial A}{\partial t}+L_{FP}^{\dagger}A\right)\frac{\partial W}{\partial\langle A_i\rangle}d\mathbf{x}=\frac{\partial D[W]}{\partial\langle A_i\rangle}\quad (i=1,\ldots,M). \tag{35}$$

Replacing $A(\mathbf{x},t)$ by the ansatz (30) yields

$$\frac{d\alpha_0}{dt}\int\frac{\partial W}{\partial\langle A_i\rangle}d\mathbf{x}+\sum_{j=1}^{M}\frac{d\alpha_j}{dt}\int A_j\frac{\partial W}{\partial\langle A_i\rangle}d\mathbf{x} + \int\left[L_{FP}^{\dagger}\left(\alpha_0+\sum_{j=1}^{M}\alpha_jA_j\right)\right]\frac{\partial W}{\partial\langle A_i\rangle}d\mathbf{x} =\frac{\partial D[W]}{\partial\langle A_i\rangle}, \tag{36}$$

while the normalization condition gives

$$\int\frac{\partial W}{\partial\langle A_i\rangle}d\mathbf{x}=\frac{\partial}{\partial\langle A_i\rangle}\int W d\mathbf{x}=0. \tag{37}$$

It is also clear that (remember that our ansatz for W is parametrized by the M mean values $\langle A_i\rangle$)

$$\frac{\partial}{\partial\langle A_i\rangle}\int A_jW d\mathbf{x}=\frac{\partial\langle A_j\rangle}{\partial\langle A_i\rangle}=\delta_{ij} \tag{38}$$

and

$$L_{FP}^{\dagger}\alpha_0=0. \tag{39}$$

Finally, Eq. (36) together with the last equations we have just derived yields equations of motion for the variational parameters α_i,

$$\frac{d\alpha_i}{dt}=-\sum_{j=1}^{M}\alpha_j\frac{\partial\langle L_{FP}^{\dagger}A_j\rangle}{\partial\langle A_i\rangle}+\frac{\partial D[W]}{\partial\langle A_i\rangle}\quad (i=1,\ldots,M), \tag{40}$$

which, together with the equations of motion (34) for the mean values of the A_i, determine the temporal evolution of both W and $A(\mathbf{x},t)$. Summing up, starting from a variational principle that leads to the FP equation, we have obtained an approximate FP solution W_{variat} in terms of the set of parameters (32).

IV. MINIMUM ENTROPY APPROACH TO THE FP EQUATION

A. The time-dependent minimum entropy ansatz

We now focus our attention upon a particular parametrization of the probability distribution W in terms of M relevant mean values. The time evolution of a set of M appropriate relevant mean values $\langle A_i\rangle$ is given by

$$\frac{d}{dt}\langle A_i\rangle=\langle L_{FP}^{\dagger}A_i\rangle\quad (i=1,\ldots,M). \tag{41}$$

We will evaluate the right-hand side of the above equations by employing the Kullback minimum entropy ansatz

$$W_{MK}(\mathbf{x},t)=\frac{1}{Z}W_0(\mathbf{x},t)\exp\left\{-\sum_{i=1}^{M}\lambda_i(t)A_i(\mathbf{x})\right\}. \tag{42}$$

It is clear that W_{MK} is determined, at each instant, by the instantaneous values adopted by the relevant mean values $\langle A_i\rangle$, which are taken as constraints in the extremalization of K. Hence the approximate minimum entropy solution $W_{MK}(\mathbf{x},t)$ is constructed in such a way as to satisfy the equations of motion (41) of the M relevant mean values. Of course, in this approximate procedure the time evolution depends on the temporal behavior of the Lagrange multipliers $\lambda_i(t)$, and we need to ascertain their functional dependence on time.

Furthermore, we have noted above that, in order to develop confidence in the approximate procedures, we need to ascertain that the temporal evolution of the Kullback entropy one constructs with the approximate FP solution behaves in the correct fashion. We tackle this question first.

B. Time derivative of Kullback's entropy

Let us now consider the time derivative of the Kullback entropy K_{MK} evaluated on the approximate minimum entropy solution $W_{MK}(\mathbf{x},t)$. The time derivative of Kullback relative entropy is given by

$$\frac{dK_{MK}}{dt}=-\int\left[\sum_{i=1}^{M}\frac{\partial W_{MK}}{\partial\langle A_i\rangle}\frac{d\langle A_i\rangle}{dt}\right]\ln\left(\frac{W_{MK}}{W_0}\right)d\mathbf{x} + \int\frac{\partial W_0}{\partial t}\left(\frac{W_{MK}}{W_0}\right)d\mathbf{x}, \tag{43}$$

which can be rewritten as

$$\frac{dK_{MK}}{dt}=\sum_{i=1}^{M}\frac{\partial K}{\partial\langle A_i\rangle}\frac{d\langle A_i\rangle}{dt}+\int\frac{\partial W_0}{\partial t}\left(\frac{W_{MK}}{W_0}\right)d\mathbf{x}, \tag{44}$$

where the partial derivatives of the relative entropy with respect to the mean values are to be taken at fixed W_0 (i.e., at a given time). By recourse to the equations of motion of the relevant mean values and the thermodynamic relations (15), Eq. (44) can be recast as

$$\frac{dK_{MK}}{dt}=\int W_{MK}\left\{L_{FP}^{\dagger}\left(\sum_{i=1}^{M}\lambda_iA_i(\mathbf{x})\right)\right\}d\mathbf{x} + \int\frac{\partial W_0}{\partial t}\left(\frac{W_{MK}}{W_0}\right)d\mathbf{x}, \tag{45}$$

which easily leads to

$$\frac{dK_{MK}}{dt}=-\int(L_{FK}W_{MK})\left[\left(-\sum_{i=1}^{M}\lambda_iA_i(\mathbf{x})\right)+\ln(Z^{-1})\right]d\mathbf{x} + \int\frac{\partial W_0}{\partial t}\left(\frac{W_{MK}}{W_0}\right)d\mathbf{x}. \tag{46}$$

From the form of the maximum entropy ansatz, it is now clear that the time derivative of the relative entropy, evaluated on the minimum entropy approximate solution W_{MK}, is given by

$$\frac{dK_{MK}}{dt}=-\int (L_{FP}W_{MK})\ln\left(\frac{W_{MK}}{W_0}\right)d\mathbf{x}+\int\frac{\partial W_0}{\partial t}\left(\frac{W_{MK}}{W_0}\right)d\mathbf{x}$$
$$=D[W_{MK}]. \tag{47}$$

We see that the functional relation between the time derivative of K and the approximate minimum entropy solution W_{MK} is exactly the same as that of the time derivative of the relative entropy in terms of an *exact* solution of the FP equation [see Eq. (16)]. In the special case of a prior distribution $W_0(\mathbf{x},t)$, which is an exact solution of the FP equation, this result has the important consequence that *our minimum entropy ansatz $W_{MK}(\mathbf{x},t)$ verifies the same H theorem satisfied by the exact solution.*

C. Equations of motion for the Lagrange multipliers

The equations of motion for the M Lagrange multipliers λ_i are

$$\frac{d\lambda_i}{dt}=\sum_{j=1}^{M}\frac{\partial\lambda_i}{\partial\langle A_j\rangle}\frac{d\langle A_j\rangle}{dt}+\frac{\partial\lambda_i}{\partial t}$$
$$=\sum_{j=1}^{M}\frac{\partial\lambda_j}{\partial\langle A_i\rangle}\frac{d\langle A_j\rangle}{dt}+\frac{\partial\lambda_i}{\partial t}=\frac{\partial}{\partial\langle A_i\rangle}\left\{\sum_{j=1}^{M}\lambda_j\frac{d\langle A_j\rangle}{dt}\right\}$$
$$-\sum_{j=1}^{M}\lambda_j\frac{\partial\langle L_{FP}^{\dagger}A_j\rangle}{\partial\langle A_i\rangle}+\frac{\partial\lambda_i}{\partial t}, \tag{48}$$

where the partial derivatives with respect to the mean values $\langle A_i\rangle$ are evaluated at fixed W_0 (i.e., at a given time), while the partial time derivatives of the Lagrange multipliers are taken at fixed values of the M moments $\langle A_i\rangle$ (i.e., they describe the changes in the λ's due to the explicit time dependence of the prior distribution $W_0(\mathbf{x},t)$; this time dependence implies that the λ's will change even if the relevant moments remain fixed).

With the same conventions for the partial derivatives we have

$$\frac{\partial D[W_{MK}]}{\partial\langle A_i\rangle}=\frac{\partial}{\partial\langle A_i\rangle}\left[\frac{dK_{MK}}{dt}\right]=\frac{\partial}{\partial\langle A_i\rangle}\left[\sum_{j=1}^{M}\lambda_j\frac{d\langle A_j\rangle}{dt}\right]+\frac{\partial\lambda_i}{\partial t}. \tag{49}$$

From the above expressions we arrive at the equations of motion for the λ's

$$\frac{d\lambda_i}{dt}=-\sum_{j=1}^{M}\lambda_j\frac{\partial\langle L_{FP}^{\dagger}A_j\rangle}{\partial\langle A_i\rangle}+\frac{\partial D[W_{MK}]}{\partial\langle A_i\rangle}. \tag{50}$$

We can see that these equations of motion for the Lagrange multipliers *coincide* with the equations of motion (40) for the parameters α_i, derived from the variational principle based on the action (19).

Summing up, we tried to build an approximate minimum entropy approach to the FP equation based upon the idea of approximately closing the set of equations (41) with the help of the minimum entropy form (42). We find that this Kullback minimum entropy ansatz for the distribution function is *identical* to the one obtained in Sec. III from the action variational principle if we identify the parameters α_i with the Lagrange multipliers λ_i.

D. Hamiltonian structure

Defining the Hamiltonian

$$\mathcal{H}(\lambda_1,\ldots,\lambda_M,\langle A_1\rangle,\ldots,\langle A_M\rangle)$$
$$=\sum_{i=1}^{M}\lambda_i\langle L_{FP}^{\dagger}A_i\rangle-D[W(\langle A_1\rangle,\ldots,\langle A_M\rangle)], \tag{51}$$

the equations of motion for the relevant mean values $\langle A_i\rangle$ and for the corresponding Lagrange multipliers λ_i can be cast into the Hamiltonian form

$$\frac{d\langle A_i\rangle}{dt}=\frac{\partial\mathcal{H}}{\partial\lambda_i} \tag{52}$$

and

$$\frac{d\lambda_i}{dt}=-\frac{\partial\mathcal{H}}{\partial\langle A_i\rangle}. \tag{53}$$

We see that the relevant mean values and the Lagrange multipliers are conjugate variables not only in Jaynes's thermodynamical sense, but also in a Hamiltonian sense. In the definition of the Hamiltonian (51), the quantities $\langle L_{FP}^{\dagger}A_i\rangle$ and $D[W]$ are regarded as functions of the M relevant mean values. Furthermore, the relevant mean values $\langle A_i\rangle$ and the Lagrange multipliers λ_i are regarded as independent variables. However, of all the possible solutions

$$\{\langle A_i\rangle(t),\lambda_i(t)\}\quad(i=1,\ldots,M), \tag{54}$$

of the Hamiltonian equations (52) and (53), only those with initial conditions satisfying the relations

$$\langle A_i\rangle(t_0)=\int A_i(\mathbf{x})Z^{-1}\exp\left(-\sum_{i=1}^{M}\lambda_i(t_0)A_i(\mathbf{x})\right)$$
$$(i=1,\ldots,M) \tag{55}$$

are relevant to our problem. The relations (55) determine an M-dimensional hypersurface embeded in the $2M$-dimensional phase space of our Hamiltonian system. It is clear that this hypersurface is an invariant set of the equations of motion (52) and (53). This means that a solution to these equations with initial conditions on this hypersurface will always remain on that subset of the phase space.

V. PARTICULAR CASES

A. Gibbs-Shannon-Jaynes entropy

If we take as the prior distribution the uniform probability distribution, the relative Kullback measure reduces (up to an additive constant) to the usual Gibbs-Shannon-Jaynes entropy. As observed above, our approach should in this case translate itself into the one provided by Jaynes's maximum entropy principle (MEP) (as applied to the FP equation).

Here we make the connection with the approximate maximum entropy scheme proposed by Hick and Stevens [12] to deal with the FP equation and, in particular, with the cosmic ray transport equation. Their approach is seen to constitute a particular instance of the one advanced in the present paper. Including convection and pitch angle scattering, the cosmic ray transport equation [12] in a (spatially) one-dimensional geometry is

$$\frac{\partial f}{\partial t}=\frac{1}{2}\nu\frac{\partial}{\partial\mu}(1-\mu^2)\frac{\partial f}{\partial\mu}-\mu v\frac{\partial f}{\partial x}, \tag{56}$$

where $f(x,\mu,t)$ is the particle distribution function, v is the particle velocity, $\mu=\hat{v}\cdot\hat{x}$ is the cosine of the pitch angle, and ν is the scattering frequency.

HS have obtained numerical results that clearly show that the maximum entropy approach can profitably be employed. They argue that the success of the method can be partially justified by the fact that, for exact solutions $f(x,\mu,t)$ of the transport equation, it can be shown that

$$\frac{dS}{dt}=\frac{1}{2}\nu\int_{-\infty}^{+\infty}dx\int_{-1}^{+1}d\mu\,\frac{1-\mu^2}{f}\left(\frac{\partial f}{\partial\mu}\right)^2>0, \tag{57}$$

so that the entropy S is a monotonically increasing function of time.

From our analysis of the time derivative of Kullback's information it follows that, in the particular case of Shannon's entropy, *the functional relation between dS/dt and the distribution function is, for the MEP solution, the same as for the exact solution*. In particular, we proved in analytical fashion that, if the exact solutions satisfy an H theorem (i.e., the entropy is a monotonically increasing function of time), then the maximum entropy solutions verifies *the same H theorem*. Concrete and tangible analytical support for the HS argument, which was based primarily on numerical evidence, is thus provided by the present considerations.

B. Generalized Liouville equation

If the diffusion tensor $\{D_{ij}(\mathbf{x})\}$ vanishes, the FP equation reduces to the Liouville equation

$$\frac{\partial W(\mathbf{x},t)}{\partial t}+\sum_{i=1}^{M}\frac{\partial}{\partial x_i}[WD_i(\mathbf{x})]=\frac{\partial W(\mathbf{x},t)}{\partial t}+\boldsymbol{\nabla}\cdot(W\mathbf{V}_D)=0, \tag{58}$$

where

$$\boldsymbol{\nabla}=\left(\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_N}\right) \tag{59}$$

stands for the N-dimensional nabla symbol. The above Liouville equation describes the time evolution of a statistical ensemble of identical dynamical systems, each evolving according to a flux in phase space given by the vector field $\mathbf{V}_D(\mathbf{x})$. The time evolution of each member of the ensemble is given by the system of ordinary differential equations

$$\frac{d\mathbf{x}}{dt}=\mathbf{V}_D(\mathbf{x}). \tag{60}$$

Note that the Liouville equation (58) is more general than the one presented in most textbooks of classical mechanics. Liouville's equation is usually associated with Hamiltonian systems, but the equation originally introduced by Liouville [33] is of the form (58), and corresponds to a more general situation and applies to arbitrary dynamical systems. Hamiltonian systems belong to the family of dynamical systems with divergenceless phase space flows

$$\boldsymbol{\nabla}\cdot\mathbf{V}_D=0, \tag{61}$$

in which case the Liouville equation adopts the form

$$\frac{\partial W(\mathbf{x},t)}{\partial t}+\mathbf{V}_D\cdot\boldsymbol{\nabla}W=0. \tag{62}$$

The study of the general Liouville equation is important in order to apply the ideas of statistical mechanics to non-Hamiltonian systems [34–37,39]. In particular, two of us have recently shown that Jaynes's approach to statistical mechanics can be applied to general dynamical systems with divergenceless flows in phase space [11]. In addition to Hamiltonian systems, this family encompasses other interesting cases, such as Nambu systems [38] and Bialynicky-Birula–Morrison dynamical systems [40]. Our present results, in the particular case of vanishing diffusion tensor, imply that the maximum entropy approach can be applied to solve approximately the Liouville equation for general dynamical systems, even if they have a nonvanishing phase flow divergence.

C. Closed linear equations of motion for the relevant mean values

Our maxent approach to the FP equation simplifies considerably in the case that the mean values $\langle L_{FP}^{\dagger}A_i\rangle$ can be expressed as a linear combination of the relevant mean values $\langle A_i\rangle$,

$$\langle L_{FP}^{\dagger}A_i\rangle=\sum_{j=1}^{M}g_{ij}\langle A_j\rangle\quad(i=1,\ldots,M). \tag{63}$$

In such cases, the relevant mean values evolve according to a closed linear system of ordinary differential equations

$$\frac{d\langle A_i\rangle}{dt}=\sum_{j=1}^{M}g_{ij}\langle A_j\rangle\quad(i=1,\ldots,M) \tag{64}$$

that can be solved *in an exact fashion*. Such a closure relation holds, for example, if we have a linear drift

$$D_i(\mathbf{x})=\sum_{j=1}^{N}a_i^j x_j+a_i^0, \tag{65}$$

a quadratic diffusion tensor

$$D_{ij}(\mathbf{x})=\sum_{k,l=1}^{N}b_{ij}^{kl}x_kx_l+\sum_{k=1}^{N}c_{ij}^{k}x_k+c_{ij}^{0}, \tag{66}$$

and we chose as relevant mean values the set of linear and quadratic moments

$$\langle x_i\rangle,\langle x_i x_j\rangle \quad (i,j=1,\ldots,N). \tag{67}$$

D. Exact solution as prior distribution

An interesting possibility is that of choosing as the prior distribution a particular known solution $W_0(\mathbf{x},t)$ of the FP equation. In that case, the time derivative of the entropy (of the approximate maxent solution W_{MK}) is

$$\frac{dK_{MK}}{dt}=-\int L_{FP}W_{MK}\ln\left(\frac{W_{MK}}{W_0}\right)d\mathbf{x}+\int L_{FP}W_0\left(\frac{W_{MK}}{W_0}\right)d\mathbf{x}, \tag{68}$$

which after some algebra can be put in the form

$$\frac{dK_{MK}}{dt}=\int \frac{W_{MK}^3}{W_0^2}D_{ij}\frac{\partial}{\partial x_i}\left(\frac{W_{MK}}{W_0}\right)\frac{\partial}{\partial x_j}\left(\frac{W_{MK}}{W_0}\right)d\mathbf{x}. \tag{69}$$

If the diffusion tensor is not zero we have

$$\frac{dK_{MK}}{dt}>0 \tag{70}$$

since D_{ij} is a positive-definite matrix. For the FP equation, a natural choice for the prior distribution is a stationary solution $W_0(\mathbf{x})$, which usually is easier to obtain than a particular time-dependent solution.

VI. CONCLUSION

In the present effort we have considered a minimum entropy approach to obtain approximate, time-dependent solutions to the N-dimensional Fokker-Planck equation, in the general context of Kullback's relative entropy. This method has been applied previously for solving the Liouville equation for Hamiltonian systems and recently generalized to the family of dynamical systems with divergenceless phase-space flow, albeit employing only the Gibbs-Shannon-Jaynes entropy. Some particular applications of Jaynes's MEP approach to the FP equation have also been discussed in the literature. In particular, the method has been successfully applied to the cosmic ray transport equation. However, this approach to the FP equation has so far been justified only on the basis of the analysis of particular numerical examples. In the present paper we have provided, within the context of a minimum Kullback information approach, some general analytical results that hold even if we employ the Gibbs-Shannon-Jaynes entropy since the latter is nothing more than a particular case of the more general Kullback one.

We have shown that the minimum Kullback approach can be derived from a variational principle and we obtained the equations of motion for the Lagrange multipliers associated with the relevant mean values. We have also shown that the equations of motion for the relevant mean values and the concomitant Lagrange multipliers exhibit a Hamiltonian structure, the latter being canonical conjugate to the former.

We proved that the functional relation between the time derivative of the Kullback entropy and the approximate minimum entropy solution is the same as in the case of the exact solutions. Hence, whenever the exact solutions $W(\mathbf{x},t)$ satisfy an H theorem [with respect to the prior distribution $W_0(\mathbf{x},t)$], the minimum entropy solutions $W_{MK}(\mathbf{x},t)$ *satisfy the same H theorem.*

In the particular case of diffusionlike equations, the Gibbs-Shannon-Jaynes entropy is a monotonically increasing function of time. If within our scheme a constant prior distribution is used, we obtain approximate (Jaynes) MEP solutions with the same property. Thus it is here asserted that (i) we analytically proved that the entropy of the approximate MEP solutions considered by Hick and Stevens for the cosmic ray transport equation increases with time and (ii) dS/dt, expressed in terms of the MEP solution $W_{ME}(\mathbf{x},t)$, is given by an expression identical to that corresponding to the exact solutions.

In the general case of the FP equation, the Gibbs-Shannon-Jaynes entropy does not (in general) behave monotonically with time. However, the time derivative of the Kullback relative entropy between two solutions of the FP equation has a definite sign (i.e., it obeys an H theorem). If within our minimum entropy scheme we adopt as a prior distribution a known solution of the FP equation, then the (Kullback) minimum entropy approximate solution will verify the same H theorem.

In the particular case of a vanishing diffusion tensor, the FP equation reduces to the general Liouville equation describing the evolution of a statistical ensemble of identical dynamical systems (with different initial conditions). In a previous work two of us showed that Jaynes's approach can be implemented in the case of dynamical systems with divergenceless flows. The present effort, within the framework of Kullback's minimum entropy principle, generalizes those results to the case of arbitrary dynamical systems.

Summing up, the present considerations unify and generalize previous work related to the information theory approach to the Liouville and FP equations. Numerical applications of this method to particular instances exist in the literature. In the case of Liouville's equation, for Hamiltonian and divergenceless dynamical systems, some analytical results are also available. We have extended all these results to the N-dimensional Fokker-Planck equation, within the more general framework of Kullback's minimum entropy prescription.

ACKNOWLEDGMENTS

The financial support of the Foundation for Research Development is gratefully acknowledged. A.P. gratefully acknowledges partial support from Argentina's National Research Council (CONICET).

[1] A. I. Khinchin, *Mathematical Foundations of Information Theory* (Dover, New York, 1957).
[2] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[3] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[4] E. T. Jaynes, in *Statistical Physics*, edited by W. K. Ford (Benjamin, New York, 1963).
[5] E. T. Jaynes in *Maximum Entropy and Bayesian Methods in Science and Engineering, Vol. 1: Foundations*, edited by G. J.

Erikson and C. Ray Smith (Kluwer, Dordrecht, 1988).
[6] A. Katz, *Statistical Mechanics* (Freeman, San Francisco, 1967).
[7] A. Hobson, *Concepts in Statistical Mechanics* (Gordon and Breach, London, 1971).
[8] R. Balian, *From Microphysics to Macrophysics* (Springer, Berlin, 1992).
[9] W. Grandy, *Foundations of Statistical Mechanics* (Reidel, Dordrecht, 1987).
[10] H. Grabert, *Projection Operator Techniques in Nonequilibrium Statistical Mechanics* (Springer, Berlin, 1982).
[11] A. R. Plastino and A. Plastino, Physica A **232**, 458 (1996).
[12] P. Hick and G. Stevens, Astron. Astrophys. **172**, 350 (1987).
[13] J. Baker-Jarvis, M. Racine, and J. Alameddine, J. Math. Phys. (N.Y.) **30**, 1459 (1989).
[14] Y. Alhassid and R. Levine, J. Chem. Phys. **67**, 4321 (1977).
[15] Y. Alhassid and R. Levine, Phys. Rev. A **18**, 89 (1978).
[16] D. Otero, A. Plastino, A. N. Proto, and G. Zannoli, Phys. Rev. A **26**, 1209 (1982).
[17] D. Otero, A. Plastino, A. N. Proto, and G. Zannoli, Z. Phys. A **316**, 323 (1984).
[18] N. Z. Tishby and R. D. Levine, Phys. Rev. A **30**, 1427 (1984).
[19] E. Duering, D. Otero, A. Plastino, and A. N. Proto, Phys. Rev. A **32**, 2455 (1985).
[20] R. Levine, D. Otero, A. Plastino, and A. N. Proto, Nucl. Phys. A **454**, 338 (1986).
[21] D. Otero, A. Plastino, A. N. Proto, and S. Misrahi, Phys. Rev. A **33**, 3446 (1986).
[22] E. Duering, D. Otero, A. Plastino, and A. N. Proto, Phys. Rev. A **35**, 2314 (1987).
[23] R. Balian and M. Veneroni, Ann. Phys. (N.Y.) **164**, 334 (1985).
[24] R. Balian and M. Veneroni, Ann. Phys. (N.Y.) **174**, 229 (1987).
[25] R. Balian and M. Veneroni, Ann. Phys. (N.Y.) **187**, 29 (1988).
[26] J. Aliaga, D. Otero, A. Plastino, and A. N. Proto, Phys. Rev. A **37**, 918 (1988).
[27] H. Risken, *The Fokker-Planck Equation* (Springer, New York, 1989).
[28] S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959); S. Kullback and R. A. Leibler, Ann. Math. Stat. **22**, 79 (1951).
[29] W. K. Wootters, Phys. Rev. D **23**, 357 (1981).
[30] S. M. Brunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).
[31] M. Ravicule, M. Casas, and A. Plastino, Phys. Rev. E (to be published).
[32] P. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[33] J. Liouville, J. Math. Pure Appl. **3**, 342 (1838).
[34] L. Andrey, Phys. Lett. **111A**, 45 (1985).
[35] W.-H. Steeb, Physica A **95**, 181 (1979).
[36] L. Andrey, Phys. Lett. **114A**, 183 (1986).
[37] E. Kerner, Phys. Lett. A **151**, 401 (1990).
[38] Y. Nambu, Phys. Rev. D **7**, 2405 (1973).
[39] S. Codriansky *et al.*, J. Phys. A **27**, 2565 (1994).
[40] I. Bialynicki-Birula and P. Morrison, Phys. Lett. A **158**, 453 (1991).