DYNAMICS OF SMOOTH CONSTRAINED APPROACH TO MAXIMUM ENTROPY

G. P. Beretta
Dipartimento di Energetica
Politecnico di Milano, Italy
and
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts

SECOND LAW
ANALYSIS OF
THERMAL SYSTEMS

presented at

THE FOURTH INTERNATIONAL SYMPOSIUM ON SECOND LAW ANALYSIS OF THERMAL SYSTEMS
ROME, ITALY
MAY 25–29, 1987

sponsored by

THE ADVANCED ENERGY SYSTEMS DIVISION, ASME
ASSOCIAZONE TERMOTECNICA ITALIANA (ATI), SEZIONE LAZIALE
UNIVERSITA DI ROMA I "LA SAPIENZA"

edited by

M.J. MORAN
OHIO STATE UNIVERSITY

E. SCIUBBA
UNIVERSITA DI ROMA I "LA SAPIENZA"

ASME Book I00236, pp. 17-24 (1987)

THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS
United Engineering Center • 345 East 47th Street • New York, N.Y. 10017

DYNAMICS OF SMOOTH CONSTRAINED APPROACH TO MAXIMUM ENTROPY

G. P. Beretta
Dipartimento di Energetica
Politecnico di Milano, Italy
and
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts

ABSTRACT

A rate equation for a discrete probability distribution is discussed as a route to describe a smooth relaxation towards the maximum entropy distribution compatible at all times with one or more linear constraints.

1. INTRODUCTION

The determination of a probability distribution of maximum entropy subject to a set of linear constraints has applications in many areas of engineering, physics, chemistry, and information theory [1]. The maximum entropy distribution typically represents an equilibrium or a constrained-equilibrium state of the system under study. This paper addresses a generalization of the maximum entropy problem to the nonequilibrium domain, by discussing a general rate equation for the description of smooth constrained relaxation of nonequilibrium probability distributions towards the maximum entropy distribution.

The nonlinear rate equation for the probability distribution has the feature that it keeps the constraints constant at their initial values and increases the entropy value until the probabilities converge to the maximum entropy distribution. The rate equation is also consistent with an Onsager reciprocity theorem and a fluctuation-dissipation theorem, both extended to the entire nonequilibrium domain.

Geometrically, every trajectory generated by the rate equation in state space has the property that its image on the intersection of the entropy surface with the constraint hyperplane is at each point orthogonal to the constant entropy contours, i.e., on such intersection the image of the trajectory follows a path of steepest entropy ascent compatible with the constraints.

We also discuss a generalized rate equation to treat constraints with specified time-dependent magnitudes.

2. NONEQUILIBRIUM PROBLEM

The maximum entropy problem which sets our context is that of seeking a probability distribution, namely, a probability vector $\underline{p} = \{p_1,\ldots,p_i,\ldots\}$, whose entropy $S(\underline{p})$

$$S(\underline{p}) = -\textstyle\sum_i p_i \ln p_i \tag{1a}$$

is maximal subject to given magnitudes $\langle A_r\rangle$ of one or more constraints

$$\textstyle\sum_i p_i A_{ri} = \langle A_r\rangle \qquad r = 0,1,\ldots,m \tag{1b}$$

where A_{ri} is the magnitude of the r-th constraint in state i, namely, a state represented by a probability distribution with $p_i = 1$ and $p_{j\neq i} = 0$. We will assume that the first constraint is the normalization condition, so that $A_{0i} = 1$ for each i and $\langle A_0\rangle = 1$.

The maximizing distribution $\underline{p}^*$ can be written as

$$p_i^* = \frac{1}{Q}\exp\Big(-\sum_{r=1}^{m}\lambda_r A_{ir}\Big) \tag{2a}$$

$$Q = \textstyle\sum_i \exp\Big(-\sum_{r=1}^{m}\lambda_r A_{ir}\Big) \tag{2b}$$

where the Lagrange multipliers $\lambda_1, \ldots, \lambda_m$ are determined by the values $\langle A_1\rangle,\ldots,\langle A_m\rangle$ of the constraints.

The extension of the maximum entropy problem to the nonequilibrium domain that we wish to consider is the following. We seek a time-dependent probability distribution, namely, a vector function $\underline{p}(t) = \{p_1(t),\ldots,p_i(t),\ldots\}$, whose entropy $S(\underline{p}(t))$ is strictly increasing with time, and such that the magnitudes $\langle A_r\rangle$ of the constraints are time-invariant, namely,

$$\textstyle\sum_i p_i(t) A_{ri} = \langle A_r\rangle \qquad r = 0,1,\ldots,m \tag{3}$$

for all times t. Alternatively, given an initial distribution $\underline{p}_0$ we seek a time-dependent distribution $\underline{p}(t)$ with $\underline{p}(0) = \underline{p}_0$ such that at all times t

$$\sum_i p_i(t)A_{ri} = \sum_i p_i(0)A_{ri} \qquad r = 0,1,\ldots,m \quad (4a)$$

and

$$-\frac{d}{dt}\sum_i p_i(t)\ln p_i(t) > 0 \qquad (4b)$$

A further generalization of the above nonequilibrium problem is one in which the magnitudes of the constraints are assigned a definite time-dependence. Thus, given an initial distribution $\underline{p}_0$ we seek a time-dependent distribution $\underline{p}(t)$ with $\underline{p}(0) = \underline{p}_0$ such that at all times t

$$\frac{d}{dt}\sum_i p_i(t)A_{ri} = \alpha_r(t,\underline{p}(t)) \qquad r = 1,\ldots,m \quad (5a)$$

and

$$-\frac{d}{dt}\sum_i p_i(t)\ln p_i(t) > 0 \qquad (5b)$$

where the rates $\alpha_r(t,\underline{p}(t))$ are given functions of time and of the instantaneous probability distribution.

For example, Problem 5 can be applied in the context of the constrained-equilibrium method for chemical kinetics [2]. According to this method, the chemical composition of a complex reacting system is assumed at all times to be that of a constrained-equilibrium state of maximum entropy subject to the usual normalization, energy, and stoichiometry constraints, plus an additional set of constraints each representing a class of rate-controlling reactions. The magnitudes of these additional constraints are continuously updated according to a kinetic model for the rates of the controlling reactions. Problem 5 represents a generalization of the constrained-equilibrium method where, instead of assuming instantaneous entropy maximization immediately after each update of the rate-controlling constraints, we assume a smooth approach to maximum entropy continuously compatible with the shifting magnitudes of the constraints.

In Section 3, we discuss a way to construct a differential equation for the probability distribution $\underline{p}$, namely, an equation of the form

$$\dot{\underline{p}} = \underline{D}(\underline{p}) \qquad (6)$$

whose solutions are solutions of Problem 4.

In Section 5, we discuss a way to construct a differential equation of the form

$$\dot{\underline{p}} = \underline{R}_1(t,\underline{p}) + \ldots + \underline{R}_m(t,\underline{p}) + \underline{D}(\underline{p}) \qquad (7)$$

whose solutions are solutions of Problem 5.

These differential equations and their main properties are presented in terms of the notation introduced in the Appendix. From here on, we assume familiarity with the useful and nontrivial notation in the Appendix.

3. CONSTRAINED INCREASE OF ENTROPY

In terms of the notation defined in the Appendix, we propose to consider the differential equation

$$\dot{\underline{x}} = \frac{1}{\tau}\,[\underline{f} - (\underline{f})_{L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m)}] \qquad (8)$$

where vectors $\underline{g}_0$, $\underline{g}_1$, ..., $\underline{g}_m$, and $\underline{f}$ are defined by Relations A7 and A8. Using Relations A2 and A17, and some procedure to eliminate from the set $\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m$ those vectors that are linearly dependent on the others, we may readily verify that Equation 8 induces an equation for $\dot{\underline{p}}$ which contains only the square x_i^2 of the new variables and, therefore, is of the form

$$\dot{\underline{p}} = \frac{1}{\tau}\,\underline{D}(\underline{p}) \qquad (9)$$

Here, τ can be either a positive constant or a positive functional of the probabilities $\underline{p}$.

By virtue of Relations A9 and A19, we conclude that the magnitude of each constraint is invariant under Equation 8, i.e.,

$$\dot{G}_r = \underline{g}_r\cdot\dot{\underline{x}} = \frac{1}{\tau}\,\underline{g}_r\cdot[\underline{f} - (\underline{f})_{L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m)}] = 0 \qquad (10)$$

By virtue of Relation A10 and A19, we conclude that the value of the entropy functional is nondecreasing under Equation 8, i.e.,

$$\dot{S} = \dot{F} = \underline{f}\cdot\dot{\underline{x}} = \frac{1}{\tau}\,\underline{f}\cdot[\underline{f} - (\underline{f})_{L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m)}] \qquad (11a)$$

$$= \frac{1}{\tau}\,[\underline{f}-(\underline{f})_{L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m)}]\cdot[\underline{f}-(\underline{f})_{L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m)}] \qquad (11b)$$

$$= \tau\,\dot{\underline{x}}\cdot\dot{\underline{x}} \geq 0 \qquad (11c)$$

and the equal sign in Relation 11 applies if and only if vector $\underline{f}$ is in $L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m)$ and, therefore, is a linear combination of vectors $\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m$, namely, there is a set of multipliers $\lambda_0,\lambda_1,\ldots,\lambda_m$ such that

$$\underline{f} = \sum_{r=1}^{m}\lambda_r\underline{g}_r \qquad (12)$$

Using Relations A7 and A8, Condition 12 becomes

$$-x_i - x_i\ln x_i^2 = \sum_{r=0}^{m}\lambda_r x_i A_{ri} \qquad i = 1,\ldots \qquad (13)$$

or, multiplying it by x_i and using $p_i = x_i^2$,

$$p_i \ln p_i = -p_i(1 + \sum_{r=0}^{m}\lambda_r A_{ri}) \qquad (14)$$

It is noteworthy that Equation 8 cannot alter the value of a p_i or x_i which is initially zero, namely, an initially zero probability remains zero at all times. Thus, from Relations 11 and 14 it follows that the effect of Equation 8 is to smoothly rearrange -- without violating any constraint -- the nonzero probabilities in the distribution towards higher entropy distributions until as time tends to infinity the distribution tends towards an equilibrium defined by Equation 14 in which the initially zero probabilities are still equal to zero whereas the initially nonzero probabilities are distributed canonically. Clearly, such equilibrium distribution would be unstable until some probability is equal to zero, because a minor perturbation of the distribution which sets this probability to an arbitrarily small nonzero value would proceed away towards a different equilibrium of higher entropy. If initially all the probabilities in the distribution have nonzero values, then Equation 8 takes the distribution directly towards the unique stable equilibrium distribution compatible with the initial values of the constraints and given by Equations 2.

Geometrically, we could visualize the effect of Equation 8 as follows. Consider the hyperplane defined by $G_r(\underline{x}) = \langle A_r \rangle$ for $r = 0,1,\ldots,m$ where $\langle A_r \rangle$ are the magnitudes of the constraints fixed by the initial distribution. On this hyperplane we can identify contour curves of constant entropy, generated by intersecting the hyperplane with the constant entropy surfaces $F(\underline{x}) = S$ where S varies from 0 to the maximum value compatible with the magnitudes of the constraints. Every trajectory $\underline{x}(t)$ generated by Equation 8 lies on the hyperplane and is at each point orthogonal to the constant entropy contour passing through that point. In this sense, the trajectory follows a path of steepest entropy ascent compatible with the constraints.

In Section 4 we discuss two further properties of Equation 8 related to Onsager's reciprocity and the fluctuation-dissipation theorem.

The mathematical structure of Equation 8 was originally developed by the author within the context of a unified theory of mechanics and thermodynamics that we call quantum thermodynamics [3-4].

4. RECIPROCITY AND FLUCTUATION-DISSIPATION RELATIONS

An indirect way to specify a probability distribution $\underline{p}$ is to specify the mean values of a sufficient number of independent linear functionals of of the distribution such as

$$\sum_i p_i A_{ri} = \langle A_r \rangle \qquad r = 0,1,\ldots,m,\ldots \tag{15}$$

where the first m+1 functionals coincide with the constraints, but the set is now extended to as many functionals as needed to completely specify the distribution $\underline{p}$. If the functionals are all linearly independent, then we need as many as there are probabilities in the distribution. We call such a set of functionals a complete set of independent properties of the probability distribution. We will denote by $Y_0, Y_1, \ldots, Y_r, \ldots$ a complete set of property functionals of the variables x_i^2, namely,

$$Y_r(\underline{x}) = \sum_i x_i^2 A_{ri} \qquad r = 0,1,\ldots \tag{16}$$

such that if the values of all these functionals are given then the values of all the x_i^2 are determined. For simplicity, we shall further assume that functional Y_0 is the normalization constraint, i.e., $A_{0i} = 1$ for each i.

We then define the gradient vectors of the functionals Y_r

$$\underline{y}_r = \{\partial Y_r/\partial x_1, \ldots, \partial Y_r/\partial x_i, \ldots\}$$

$$= \{2x_1 A_{r1}, \ldots, 2x_i A_{ri}, \ldots\} \qquad r = 0,1,\ldots \tag{17}$$

In terms of the gradient vectors, the functionals Y_r may be written as

$$Y_r = \frac{1}{4}\, \underline{y}_0 \cdot \underline{y}_r \tag{18}$$

In terms of functionals Y_r we may also form the following useful nonlinear functionals

$$Y_{rs}(\underline{x}) = \sum_i x_i^2 A_{ri} A_{si} - \sum_i x_i^2 A_{ri} \sum_j x_j^2 A_{sj} \tag{19}$$

$$= \frac{1}{4}\, \underline{y}_r \cdot \underline{y}_s - \frac{1}{16}\, (\underline{y}_0 \cdot \underline{y}_r)(\underline{y}_0 \cdot \underline{y}_s) \tag{20}$$

which represent the covariance or codispersion of properties Y_r and Y_s. In particular, the functional Y_{rr} represents the variance or dispersion (also, fluctuation) of property Y_r.

We now consider the entropy functional

$$F(\underline{x}) = -\sum_i x_i^2 \ln x_i^2 \tag{21}$$

and its gradient vector

$$\underline{f} = \{-2x_1 - 2x_1 \ln x_1^2, \ldots, -2x_i - 2x_i \ln x_i^2, \ldots\} \tag{22}$$

and further assume that, when evaluated at a given distribution $\underline{x}$, the property functionals Y_r in the complete set have gradient vectors $\underline{y}_r$ that are all linearly independent and span the entire set of vectors with zero entries corresponding to the zero x_i's, so that there is a unique set of scalars $\lambda_0, \lambda_1, \ldots$ such that the vector $\underline{f}$ can be written as

$$\underline{f} = \sum_r \lambda_r \underline{y}_r \tag{23}$$

where the scalars λ_0, $\lambda_1, \ldots$ are determined by the set of equations

$$\sum_r \lambda_r A_{ri} = -1 - \ln x_i^2 \qquad (x_i \neq 0) \tag{24}$$

with the index i restricted to the set of nonzero x_i's.

The entropy functional F may also be written as

$$F = -\sum_i x_i^2 \ln x_i^2 = \frac{1}{4}(1 + \underline{y}_0 \cdot \underline{f}) \tag{25a}$$

$$= \frac{1}{4} + \frac{1}{4} \sum_r \lambda_r \underline{y}_0 \cdot \underline{y}_r \tag{25b}$$

$$= \frac{1}{4} + \sum_r \lambda_r Y_r \tag{25c}$$

where we used Equations 23 and 18. We see from Relation 25c that the scalar λ_r can be interpreted as an affinity or generalized "force" representing the marginal impact of property Y_r onto the value of the entropy about a given distribution $\underline{x}$.

We are now ready to consider the time dependence of the properties Y_r and the entropy F as induced by the Equation 8 for the probability distribution, i.e., by the rate equation

$$\dot{\underline{x}} = \frac{1}{\tau}\,[\underline{f} - (\underline{f})_{L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m)}] \tag{26}$$

which can now also be written as

$$\dot{\underline{x}} = \frac{1}{\tau}\,\sum_s \lambda_s [\underline{y}_s - (\underline{y}_s)_{L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m)}] \tag{27}$$

where we used Equation 23 for $\underline{f}$. The rates of change of properties Y_r are then given by

$$\dot{Y}_r = \underline{y}_r \cdot \dot{\underline{x}} = \sum_s \lambda_s L_{rs} \tag{28}$$

where we defined the functionals

$$L_{rs} = \frac{1}{\tau}\,\underline{y}_r \cdot [\underline{y}_s - (\underline{y}_s)_{L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m)}] \tag{29a}$$

$$= \frac{1}{\tau}\,[\underline{y}_r - (\underline{y}_r)_{L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m)}] \cdot [\underline{y}_s - (\underline{y}_s)_{L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m)}] \tag{29b}$$

and in writing Equation 29b we have used Equation A13 with $\underline{a} = (\underline{y}_r)_L$ and subtracted a zero from Equation 29a.

Relation 28 shows that the rates of change (or generalized "fluxes") $\dot{Y}_r$ and the affinities (or generalized "forces") λ_s are linearly interrelated by the coefficients (or generalized "conductivities") L_{rs}.

If we use Equation A17 to write an explicit expression for the generalized conductivities L_{rs}, we find

$$L_{rs} = \frac{1}{\tau}\,\frac{\begin{vmatrix} \underline{y}_r\cdot\underline{y}_s & \underline{y}_r\cdot\underline{h}_1 & \cdots & \underline{y}_r\cdot\underline{h}_k \\ \underline{y}_s\cdot\underline{h}_1 & \underline{h}_1\cdot\underline{h}_1 & \cdots & \underline{h}_k\cdot\underline{h}_1 \\ \vdots & \vdots & \ddots & \vdots \\ \underline{y}_s\cdot\underline{h}_k & \underline{h}_1\cdot\underline{h}_k & \cdots & \underline{h}_k\cdot\underline{h}_k \end{vmatrix}}{\begin{vmatrix} \underline{h}_1\cdot\underline{h}_1 & \cdots & \underline{h}_k\cdot\underline{h}_1 \\ \vdots & \ddots & \vdots \\ \underline{h}_1\cdot\underline{h}_k & \cdots & \underline{h}_k\cdot\underline{h}_k \end{vmatrix}} \tag{30}$$

and, because determinants are invariant under transposition, we find that the conductivities L_{rs} satisfy the reciprocity relations

$$L_{rs} = L_{sr} \tag{31}$$

Moreover, it follows from Relation 29b that the matrix of generalized conductivities

$$[L] = \begin{bmatrix} L_{00} & L_{01} & \cdots & L_{0s} & \cdots \\ L_{10} & L_{11} & \cdots & L_{1s} & \cdots \\ \vdots & \vdots & & \vdots & \\ L_{r0} & L_{r1} & \cdots & L_{rs} & \cdots \\ \vdots & \vdots & & \vdots & \end{bmatrix} \tag{32}$$

is a Gram matrix and as such it is nonnegative. Matrix [L] is strictly positive only if the vectors $\underline{y}_r - (\underline{y}_r)_L$ are all linearly independent, in which case the set of Equations 28 may be solved to yield

$$\lambda_s = \sum_r \dot{Y}_r ([L]^{-1})_{sr} \tag{33}$$

The rate of entropy increase (Equation 11) may be rewritten in the following several ways

$$\dot{S} = \dot{F} = \underline{f}\cdot\dot{\underline{x}} = \sum_r \lambda_r \underline{y}_r \cdot \dot{\underline{x}} = \sum_r \lambda_r \dot{Y}_r \tag{34a}$$

$$= \tau\,\dot{\underline{x}}\cdot\dot{\underline{x}} = \sum_r \sum_s \lambda_r \lambda_s L_{rs} \tag{34b}$$

and, if [L] is strictly positive,

$$= \sum_r \sum_s \dot{Y}_r \dot{Y}_s ([L]^{-1})_{rs} \tag{35}$$

Finally, comparing Relation 20, for the codispersion Y_{rs} of properties Y_r and Y_s, and Relation 30, for the generalized conductivities L_{rs}, we see that there is a relation between L_{rs}, Y_{rs} and all the codispersions of the constraints and properties Y_r and Y_s. We may greatly simplify these relations if, for a given distribution $\underline{x}$, we further restrict the choice of the complete set of linearly independent property functionals Y_r so that $\underline{y}_0 = \underline{h}_1$, $\underline{y}_1 = \underline{h}_2$, ..., $\underline{y}_{k-1} = \underline{h}_k$ where $\underline{h}_1,\ldots,\underline{h}_k$ are linearly independent vectors spanning the manifold $L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_m)$ generated by the constraints and, moreover, we select the functionals for $r \geq k$ so that the codispersions Y_{r0}, Y_{r1}, ..., $Y_{r(k-1)}$ (Equation 20) are all equal to zero (notice that $Y_{r0} = 0$ implies $Y_r = 0$). For this particular choice, by studying Relation 30 for the generalized conductivities, we may readily verify that

$$L_{rs} = 0 \qquad \text{whenever} \quad r < k \quad \text{or} \quad s < k \tag{36a}$$

$$L_{rs} = \frac{4}{\tau}\,Y_{rs} \qquad \text{for} \quad r \geq k \quad \text{and} \quad s \geq k \tag{36b}$$

and, therefore, we find a direct relation between the covariance and the generalized conductivity of the pair of properties Y_r and Y_s. In particular, for r =

$s \geq k$ we find $L_{rr} = 4Y_{rr}/\tau$ which is a relation between the variance (or fluctuation) and the direct conductivity (or dissipation) of property Y_r.

5. TIME DEPENDENT CONSTRAINTS

In terms of the notation introduced in Section 3, let us consider the differential equation

$$\dot{\underline{x}} = \frac{1}{\gamma_r}[\underline{g}_r - (\underline{g}_r)_{L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_{r-1},\underline{g}_{r+1},\ldots,\underline{g}_m)}] \tag{37a}$$

where

$$\gamma_r = \frac{1}{\alpha_r(t,\underline{x})}\ \underline{g}_r\cdot[\underline{g}_r - (\underline{g}_r)_{L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_{r-1},\underline{g}_{r+1},\underline{g}_m)}] \tag{37b}$$

and $\alpha_r(t,\underline{x}) = \alpha_r(t,\underline{p})$ as specified by Equation 5a of Problem 5. Again, we may readily verify that Equation 37 induces an equation for $\dot{\underline{p}}$ which contains only the square x_i^2 of the new variables and, therefore, is of the form

$$\dot{\underline{p}} = \underline{R}_r(t,\underline{p}) \tag{38}$$

We may also readily verify that Equation 37 induces an evolution of the probability distribution $\underline{p}$ along which the magnitudes of all the constraints except the r-th are time-invariant, whereas the magnitude of the r-th constraint varies with a rate of change equal to $\alpha_r(t,\underline{p})$.

Geometrically, we could visualize the effect of Equation 37 as follows. We consider the hyperplane defined by $G_s(\underline{x}) = \langle A_s\rangle$ for $s = 0,1,\ldots,r-1,r+1,\ldots,m$ where $\langle A_s\rangle$ are the magnitudes of the constraints (except the r-th) fixed by the initial distribution. On this hyperplane we can identify contour lines along which the r-th constraint is constant, generated by intersecting the hyperplane just defined with the hyperplane $G_r(\underline{x}) = \langle A_r\rangle$ where $\langle A_r\rangle$ varies over a feasible range of values. Every trajectory $\underline{p}(t)$ generated by Equation 37 lies on the hyperplane of the fixed constraints and is at each point orthogonal to the constant-G_r contour line passing through that point. In this sense, the trajectory follows a path along the gradient of G_r compatible with the other constraints. In this sense, Equation 37 determines the minimal change in $\underline{x}$ that is necessary in order to change the r-th constraint at the specified rate α_r.

Clearly, when two or more constraints have a specified rate of change, then Equation 7 yields many orthogonal contributions to $\dot{\underline{x}}$. The terms $\underline{R}_1,\ldots,\underline{R}_m$ (with structure given by Equation 37) cause the shifting constraints to follow the specified rates of change $\alpha_1,\ldots,\alpha_m$. The contribution of these terms to the rate of entropy change does not have a definite sign. The term $\underline{D}$ (as given by Equation 8) gives instead a positive definite contribution to the rate of entropy change and tends to attract the distribution $\underline{x}$ towards a path of steepest entropy ascent compatible with the instantaneous values of the constraints.

We may finally note that by substituting $L(\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_{r-1},\underline{g}_{r+1},\ldots,\underline{g}_m)$ in Equation 37 with $L(\underline{f},\underline{g}_0,\underline{g}_1,\ldots,\underline{g}_{r-1},\underline{g}_{r+1},\ldots,\underline{g}_m)$ we would obtain a rate equation causing the r-th constraint to follow the specified rate α_r while maintaining a zero change for the other constraints, and also a zero rate of change of the entropy. In other words, this would describe an isoentropic change of the magnitude of the r-th constraint.

We conclude that the notation introduced in the Appendix and the structure of the rate equations discussed in this paper represent a flexible framework in which to cast nonequilibrium problems where it is necessary to describe a smooth constrained approach to a maximum entropy distribution with or without varying magnitudes of the constraints.

It is noteworthy that the time evolution generated by our rate Equation 8 is more general than any equation assuming that the probability distribution always maximizes the entropy functional subject to some "possibly unknown" set of constraints [1,5].

APPENDIX

Because the variables $p_1,\ldots,p_i,\ldots$ represent probabilities, all the problems defined in Section 2 are subject to the additional set of inequality constraints

$$p_i \geq 0 \quad \text{or} \quad p_i(t) \geq 0 \tag{A1}$$

For this reason, it is convenient to change variables to a new set $\underline{x} = \{x_1,\ldots,x_i,\ldots\}$ from which probabilities may be computed according to the relations

$$p_i = x_i^2 \qquad i = 1,\ldots \tag{A2}$$

We now rewrite Problem 5 as follows

$$\frac{d}{dt}G_r(\underline{x}(t)) = \alpha_r(t,\underline{x}(t)) \tag{A3a}$$

and

$$\frac{d}{dt}F(\underline{x}(t)) > 0 \tag{A3b}$$

where

$$\underline{x} = \{x_1,\ldots,x_i,\ldots\} \tag{A4}$$

$$G_r(\underline{x}) = \textstyle\sum_i x_i^2 A_{ri} \qquad r = 0,1,\ldots,m \tag{A5}$$

$$F(\underline{x}) = -\textstyle\sum_i x_i^2 \ln x_i^2 \tag{A6}$$

Next, we define vectors representing gradients of the constraints and of the entropy functional

$$\underline{g}_r = \{\partial G_r/\partial x_1,\ldots,\partial G_r/\partial x_i,\ldots\}$$
$$= \{2x_1A_{r1},\ldots,2x_iA_{ri},\ldots\} \quad r = 0,1,\ldots,m \tag{A7}$$

$$\underline{f} = \{\partial F/\partial x_1, \ldots, \partial F/\partial x_i, \ldots\}$$

$$= \{-2x_1 - 2x_1 \ln x_1^2, \ldots, -2x_i - 2x_i \ln x_i^2, \ldots\} \quad \text{(A8)}$$

so that the rates of change of the functions G_0, G_1, ..., G_m, and F are given by

$$\dot{G}_r = \underline{g}_r \cdot \underline{\dot{x}} \quad \text{(A9)}$$

$$\dot{F} = \underline{f} \cdot \underline{\dot{x}} \quad \text{(A10)}$$

where, clearly, $\underline{\dot{x}} = \{\dot{x}_1, \ldots, \dot{x}_i, \ldots\}$ and the dot product has the obvious meaning (for example, $\underline{f} \cdot \underline{\dot{x}} = f_1 \cdot \dot{x}_1 + \ldots + f_i \cdot \dot{x}_i + \ldots$).

Given a set of vectors $\underline{a}_1, \ldots, \underline{a}_n$ (for example, in our context, the set of vectors $\underline{g}_0, \underline{g}_1, \ldots, \underline{g}_m$ just defined), the symbol

$$L(\underline{a}_1, \ldots, \underline{a}_n) \quad \text{(A11)}$$

will denote their linear span, i.e., the linear manifold containing all the vectors that are linear combinations of $\underline{a}_1, \ldots, \underline{a}_n$. Given another vector $\underline{b}$, the symbol

$$(\underline{b})_{L(\underline{a}_1, \ldots, \underline{a}_n)} \quad \text{(A12)}$$

will denote the orthogonal projection of $\underline{b}$ onto the linear manifold $L(\underline{a}_1, \ldots, \underline{a}_n)$, namely, the unique vector in $L(\underline{a}_1, \ldots, \underline{a}_n)$ such that its dot product with any other vector $\underline{a}$ in $L(\underline{a}_1, \ldots, \underline{a}_n)$ equals the dot product of $\underline{b}$ with $\underline{a}$, i.e.,

$$\underline{a} \cdot (\underline{b})_{L(\underline{a}_1, \ldots, \underline{a}_n)} = \underline{a} \cdot \underline{b} \quad \text{(A13)}$$

for every $\underline{a}$ in $L(\underline{a}_1, \ldots, \underline{a}_n)$.

In terms of a set of linearly independent vectors $\underline{h}_1, \ldots, \underline{h}_k$ spanning the manifold $L(\underline{a}_1, \ldots, \underline{a}_n)$, where clearly $k \le n$, we can write two equivalent explicit expressions for the projection A12 of vector $\underline{b}$ onto $L(\underline{a}_1, \ldots, \underline{a}_n)$. The first is

$$(\underline{b})_{L(\underline{a}_1, \ldots, \underline{a}_n)} = \sum_{p=1}^{k} \sum_{q=1}^{k} (\underline{b} \cdot \underline{h}_p)[M(\underline{h}_1, \ldots, \underline{h}_k)^{-1}]_{pq} \underline{h}_q \quad \text{(A14)}$$

where $M(\underline{h}_1, \ldots, \underline{h}_k)^{-1}$ is the inverse of the Gram matrix

$$M(\underline{h}_1, \ldots, \underline{h}_k) = \begin{bmatrix} \underline{h}_1 \cdot \underline{h}_1 & \cdots & \underline{h}_k \cdot \underline{h}_1 \\ \vdots & \ddots & \vdots \\ \underline{h}_1 \cdot \underline{h}_k & \cdots & \underline{h}_k \cdot \underline{h}_k \end{bmatrix} \quad \text{(A15)}$$

The second expression is a ratio of two determinants

$$(\underline{b})_{L(\underline{a}_1, \ldots, \underline{a}_n)} = - \frac{\begin{vmatrix} 0 & \underline{h}_1 & \cdots & \underline{h}_k \\ \underline{f} \cdot \underline{h}_1 & \underline{h}_1 \cdot \underline{h}_1 & \cdots & \underline{h}_k \cdot \underline{h}_1 \\ \vdots & \vdots & \ddots & \vdots \\ \underline{f} \cdot \underline{h}_k & \underline{h}_1 \cdot \underline{h}_k & \cdots & \underline{h}_k \cdot \underline{h}_k \end{vmatrix}}{\begin{vmatrix} \underline{h}_1 \cdot \underline{h}_1 & \cdots & \underline{h}_k \cdot \underline{h}_1 \\ \vdots & \ddots & \vdots \\ \underline{h}_1 \cdot \underline{h}_k & \cdots & \underline{h}_k \cdot \underline{h}_k \end{vmatrix}} \quad \text{(A16)}$$

where the determinant at the denominator is always strictly positive because the vectors $\underline{h}_1, \ldots, \underline{h}_k$ are linearly independent. In the paper, we often make use of vector differences such as

$$\underline{b} - (\underline{b})_{L(\underline{a}_1, \ldots, \underline{a}_n)} = \frac{\begin{vmatrix} \underline{f} & \underline{h}_1 & \cdots & \underline{h}_k \\ \underline{f} \cdot \underline{h}_1 & \underline{h}_1 \cdot \underline{h}_1 & \cdots & \underline{h}_k \cdot \underline{h}_1 \\ \vdots & \vdots & \ddots & \vdots \\ \underline{f} \cdot \underline{h}_k & \underline{h}_1 \cdot \underline{h}_k & \cdots & \underline{h}_k \cdot \underline{h}_k \end{vmatrix}}{\begin{vmatrix} \underline{h}_1 \cdot \underline{h}_1 & \cdots & \underline{h}_k \cdot \underline{h}_1 \\ \vdots & \ddots & \vdots \\ \underline{h}_1 \cdot \underline{h}_k & \cdots & \underline{h}_k \cdot \underline{h}_k \end{vmatrix}} \quad \text{(A17)}$$

where in writing Equation A17 we used Equation A16.

The vector represented by Equation A17 has the relevant property

$$\underline{a}_r \cdot (\underline{b} - (\underline{b})_{L(\underline{a}_1, \ldots, \underline{a}_n)}) = 0 \qquad r = 1, \ldots, n \quad \text{(A18)}$$

which follows directly from Relation A13, i.e., the vector $\underline{b} - (\underline{b})_L$ is orthogonal to manifold L. Moreover, we have the other relevant property

$$\underline{b} \cdot (\underline{b} - (\underline{b})_L) = (\underline{b} - (\underline{b})_L) \cdot (\underline{b} - (\underline{b})_L) \ge 0 \quad \text{(A19)}$$

where the strict inequality applies whenever $\underline{b}$ is not in L.

In the paper, we make extensive use of the notation and relations just discussed [6].

REFERENCES

1. See, for example, R.D. Levine and M. Tribus, Editors, The Maximum Entropy Formalism, The M.I.T. Press, Cambridge, Mass., 1979; Y. Alhassid, N. Agmon and R.D. Levine, Chem. Phys. Lett., Vol. 53, 22 (1978); R.D. Levine, J. Chem. Phys., Vol. 65, 3302 (1976) and references therein.

2. See, for example, J.C. Keck, in The Maximum Entropy Formalism, R.D. Levine and M. Tribus, Editors, The M.I.T. Press, Cambridge, Mass., 1979; G.P. Beretta and J.C. Keck, ASME Paper, WA/AES - Vol. 3, p. 135, (1986).

3. G.N. Hatsopoulos and E.P. Gyftopoulos, Foundations of Physics, Vol. 6, 15, 127, 436, 561 (1976).

4. See, for example, G.P. Beretta, E.P. Gyftopoulos, J.L. Park and G.N. Hatsopoulos, Nuovo Cimento B, Vol. 82, 169 (1984); G.P. Beretta, E.P. Gyftopoulos and J.L. Park, Nuovo Cimento B, Vol. 87, 77 (1985); G.P. Beretta, in Frontiers of Nonequilibrium Statistical Physics, G.T. Moore and M.O. Scully, Editors, Plenum Press, New York, 193 (1986); G.P. Beretta, Foundations of Physics, Vol. 17, 365 (1987).

5. P. Salamon, J.D. Nulton and R.S. Berry, J. Chem. Phys., Vol. 82, 2433 (1985).

6. See also G.P. Beretta, ASME Book H0341C, Vol. 3, 129 (1986).