STEEPEST-ASCENT CONSTRAINED APPROACH TO MAXIMUM ENTROPY

G. P. Beretta
Universita di Brescia
Brescia, Italy, and

Massachusetts Institute of Technology
Cambridge, Massachusetts

HTD-Vol. 80

Second Law Analysis of Heat Transfer in Energy Systems

presented at
THE WINTER ANNUAL MEETING OF
THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS
BOSTON, MASSACHUSETTS
DECEMBER 13–18, 1987

sponsored by
THE HEAT TRANSFER DIVISION, ASME

edited by
R. F. BOEHM
UNIVERSITY OF UTAH

N. LIOR
UNIVERSITY OF PENNSYLVANIA

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

STEEPEST-ASCENT CONSTRAINED APPROACH TO MAXIMUM ENTROPY

G. P. Beretta
Universita di Brescia
Brescia, Italy, and
Massachusetts Institute of Technology
Cambridge, Massachusetts

ABSTRACT

A rate equation for a discrete probability distribution is discussed as a route to describe smooth relaxation towards the maximum entropy distribution compatible at all times with one or more linear constraints. The entropy follows the path of steepest ascent compatible with the constraints. The rate equation is consistent with the Onsager theorem of reciprocity and the fluctuation-dissipation theorem. The mathematical formalism was originally developed to obtain a quantum theoretical unification of mechanics and thermodynamics. It is presented here as part of an effort to develop tools for the treatment of nonequilibrium problems with engineering applications.

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 [1-2]. 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 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 it follows the path of steepest entropy ascent compatible with the constraints. We also discuss a generalization to treat constraints with specified time-dependent magnitudes.

The formalism presented here has features of great generality, and adaptability to different applications. It was originally developed to obtain a quantum theoretical unification of mechanics and thermodynamics [3-4]. But here it has been abstracted from its original physics purpose and is presented to the thermal engineering community as a mathematical tool in an attempt to stimulate discussion on the subject. The hope is that the discussion will help to identify engineering applications in which the treatment of nonequilibrium problems could use the powerful formalism we present.

Because the audience is likely to be familiar with the so-called maximum entropy formalism [1], and with the constrained-equilibrium method for the treatment of nonequilibrium in chemical engineering [2], we present our rate equation is this framework. We have already presented elsewhere the generalization to a wider class of constrained extremum problems [5].

2. NONEQUILIBRIUM PROBLEMS

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

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

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

$$\sum_i p_i A_{ki} = \langle A_k\rangle \qquad k = 0,1,\ldots,n \tag{1b}$$

where A_{ki} is the magnitude of the k-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 $\mathbf{p}^*$ can be written as

$$p_i^* = Q^{-1} \exp(-\sum_{k=1}^{n} \lambda_k A_{ki}) \tag{2a}$$

$$Q = \sum_i \exp(-\sum_{k=1}^{n} \lambda_k A_{ki}) \tag{2b}$$

where the Lagrange multipliers λ_1, ..., λ_n are determined by the values $\langle A_1\rangle$, ..., $\langle A_n\rangle$ of the constraints.

The extension of the maximum entropy problem to the nonequilibrium domain that we wish to consider is the following.

Nonequilibrium Problem 1

We seek a time-dependent probability distribution, namely, a vector function $\mathbf{p}(t) = \{p_1(t),\ldots,p_i(t),\ldots\}$, whose entropy $S(\mathbf{p}(t))$ is strictly increasing with time, and such that the magnitudes $\langle A_k\rangle$ of the constraints are time-invariant, namely,

$$\sum_i p_i(t) A_{ki} = \langle A_k\rangle \qquad k = 0,1,\ldots,n \tag{3}$$

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

$$\sum_i p_i(t) A_{ki} = \sum_i p_i(0) A_{ki} \qquad k = 0,1,\ldots,n \tag{4a}$$

and

$$dS(\mathbf{p})/dt = -d[\sum_i p_i(t) \ln p_i(t)]/dt > 0 \tag{4b}$$

A further generalization of the above nonequilibrium problem is one in which the magnitudes of the constraints are assigned a definite time-dependence.

Nonequilibrium Problem 2

Given an initial distribution $\mathbf{p}_0$, we seek a time-dependent distribution $\mathbf{p}(t)$ with $\mathbf{p}(0) = \mathbf{p}_0$ such that at all times t

$$d\langle A_k\rangle/dt = -d[\sum_i p_i(t) A_{ki}]/dt = \alpha_k(t,\mathbf{p}(t)) \qquad k = 0,1,\ldots,n \tag{5a}$$

and

$$dS(\mathbf{p})/dt = -d[\sum_i p_i(t) \ln p_i(t)]/dt > 0 \tag{5b}$$

where the rates $\alpha_k(t,\mathbf{p}(t))$ are given functions of time and of the instantaneous probability distribution. For example, Problem 2 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 2 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.

First, we discuss a way to construct a differential equation for the probability distribution $\mathbf{p}$, namely, an equation of the form

$$d\mathbf{p}/dt = \mathbf{D}(\mathbf{p}) \tag{6}$$

whose solutions are solutions of Problem 1.

Then, we discuss a way to construct a differential equation of the form

$$d\mathbf{p}/dt = \mathbf{R}_1(t,\mathbf{p}) + \ldots + \mathbf{R}_n(t,\mathbf{p}) + \mathbf{D}(\mathbf{p}) \tag{7}$$

whose solutions are solutions of Problem 2.

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. STEEPEST-ASCENT INCREASE OF ENTROPY

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

$$d\mathbf{x}/dt = \tau(\mathbf{p})^{-1}\,[\mathbf{f} - (\mathbf{f})_{L(\mathbf{g}_0,\mathbf{g}_1,\ldots,\mathbf{g}_n)}] \tag{8}$$

where vectors $\mathbf{g}_0$, $\mathbf{g}_1$, ..., $\mathbf{g}_n$, and $\mathbf{f}$ are defined by Relations A7 and A8, the vector

$$(\mathbf{f})_{L(\mathbf{g}_0,\mathbf{g}_1,\ldots,\mathbf{g}_n)} \tag{9}$$

is the orthogonal projection of $\mathbf{f}$ onto the linear span of vectors $\mathbf{g}_0$, $\mathbf{g}_1$, ..., $\mathbf{g}_n$, and $\tau(\mathbf{p})$ may be any strictly positive functional of the probability distribution $\mathbf{p}$ including a constant (with the dimensions of time).

Using Relations A2 and A17, and some procedure to eliminate from the set $\mathbf{g}_0$, $\mathbf{g}_1$, ..., $\mathbf{g}_n$ those vectors that are linearly dependent on the others, we may readily verify that Equation 8 induces an equation for $d\mathbf{p}/dt$ which contains only the square x_i^2 of the new variables and, therefore, is of the form of Equation 6.

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

$$dG_k/dt = \mathbf{g}_k \cdot d\mathbf{x}/dt = \tau^{-1}\,\mathbf{g}_k \cdot [\mathbf{f} - (\mathbf{f})_{L(\mathbf{g}_0,\mathbf{g}_1,\ldots,\mathbf{g}_n)}] = 0 \tag{10}$$

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

$$\begin{aligned} dS/dt = dF/dt = \mathbf{f}\cdot d\mathbf{x}/dt &= \tau^{-1}\,\mathbf{f}\cdot[\mathbf{f} - (\mathbf{f})_{L(\mathbf{g}_0,\mathbf{g}_1,\ldots,\mathbf{g}_n)}] \\ &= \tau^{-1}\,[\mathbf{f} - (\mathbf{f})_{L(\mathbf{g}_0,\mathbf{g}_1,\ldots,\mathbf{g}_n)}]\cdot[\mathbf{f} - (\mathbf{f})_{L(\mathbf{g}_0,\mathbf{g}_1,\ldots,\mathbf{g}_n)}] \\ &= \tau\; d\mathbf{x}/dt\cdot d\mathbf{x}/dt \ge 0 \end{aligned} \tag{11}$$

and the equal sign in Relation 11 applies if and only if vector $\mathbf{f}$ is in $L(\mathbf{g}_0,\mathbf{g}_1,\ldots,\mathbf{g}_n)$ and, therefore, is a linear combination of vectors $\mathbf{g}_0$, $\mathbf{g}_1$, ..., $\mathbf{g}_n$, namely, there is a set of multipliers λ_0, λ_1, ..., λ_n such that

$$\mathbf{f} = \sum_{k=0}^{n} \lambda_k \mathbf{g}_k \tag{12}$$

Using Relations A7 and A8, Condition 12 becomes

$$-x_i - x_i \ln x_i^2 = \sum_{k=0}^{n} \lambda_k x_i A_{ki} \qquad i = 1, 2, \ldots \tag{13}$$

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

$$p_i \ln p_i = -p_i\,(1 + \sum_{k=0}^{n} \lambda_k A_{ki}) \qquad i = 1, 2, \ldots \tag{14}$$

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 the distribution tends towards an equilibrium distribution 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 as long as there is some probability p_i 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 $G_k(\mathbf{x}) = \langle A_k \rangle$ for $k = 0,1,\ldots,n$ where $\langle A_k \rangle$ 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 $F(\mathbf{x}) = S$ where S varies from 0 to the maximum value compatible with the magnitudes of the constraints. Every trajectory $\mathbf{x}(t)$ 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 the next section 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. ONSAGER RECIPROCITY AND FLUCTUATION-DISSIPATION RELATIONS

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

$$\sum_i p_i A_{ki} = \langle A_k \rangle \qquad k = 0,1,\ldots,n,\ldots \tag{15}$$

where the first $n+1$ functionals coincide with the constraints, but the set is now extended to as many functionals as needed to completely specify the distribution $\mathbf{p}$. If the functionals are all linearly independent, then we need as many as there are probabilities in the distribution (minus one because of normalization). We call such a set of functionals a complete set of independent properties of the probability distribution.

We will denote by $Y_0, Y_1, \ldots, Y_k, \ldots$ a complete set of property functionals of the variables x_i^2, namely,

$$Y_k(\mathbf{x}) = \sum_i x_i^2 A_{ki} \qquad k = 0,1,\ldots \qquad (16)$$

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_k as

$$\mathbf{y}_k = \{\partial Y_k/\partial x_1,\ldots,\partial Y_k/\partial x_i,\ldots\} = \{2x_1 A_{k1},\ldots,2x_i A_{ki},\ldots\} \qquad k = 0,1,\ldots \qquad (17)$$

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

$$Y_k = \tfrac{1}{4}\, \mathbf{y}_0 \cdot \mathbf{y}_k \qquad (18)$$

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

$$Y_{km}(\mathbf{x}) = \sum_i x_i^2 A_{ki} A_{mi} - \sum_i x_i^2 A_{ki} \sum_i x_i^2 A_{mi} \qquad (19)$$

$$= \tfrac{1}{4}\, \mathbf{y}_k \cdot \mathbf{y}_m - (\tfrac{1}{4}\, \mathbf{y}_0 \cdot \mathbf{y}_k)(\tfrac{1}{4}\, \mathbf{y}_0 \cdot \mathbf{y}_m) \qquad (20)$$

which represent the covariance or codispersion of properties Y_k and Y_m. In particular, the functional Y_{kk} represents the variance or dispersion (also, fluctuation) of property Y_k. We now consider the entropy functional

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

and its gradient vector

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

$$= \{-2x_1 - 2x_1 \ln x_1^2,\ldots,-2x_i - 2x_i \ln x_i^2,\ldots\} \qquad (22)$$

and further assume that, when evaluated at a given distribution $\mathbf{x}$, the property functionals Y_k in the complete set have gradient vectors $\mathbf{y}_k$ 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 $\mathbf{f}$ can be written as

$$\mathbf{f} = \sum_k \lambda_k \mathbf{y}_k \qquad (23)$$

where the scalars $\lambda_0, \lambda_1, \ldots$ are determined by the set of equations

$$\sum_k \lambda_k A_{ki} = -1 - \ln x_i^2 \qquad (x_i \neq 0) \qquad i=1,2,\ldots \qquad (24)$$

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

The entropy functional F may also be written as

$$F(\mathbf{x}) = -\sum_i x_i^2 \ln x_i^2 = \tfrac{1}{4}(1 + \mathbf{y}_0 \cdot \mathbf{f})$$

$$= \tfrac{1}{4} + \tfrac{1}{4} \sum_k \lambda_k \mathbf{y}_0 \cdot \mathbf{y}_k = \tfrac{1}{4} + \sum_k \lambda_k Y_k \qquad (25)$$

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

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

$$d\mathbf{x}/dt = \tau^{-1} \left[\mathbf{f} - (\mathbf{f})_{L(g_0,g_1,\ldots,g_n)}\right] \qquad (26)$$

which can now also be written as

$$d\mathbf{x}/dt = \tau^{-1} \sum_m \lambda_m \left[\mathbf{y}_m - (\mathbf{y}_m)_{L(g_0,g_1,\ldots,g_n)}\right] \qquad (27)$$

where we used Equation 23 for f. The rates of change of properties Y_k are then given by

$$dY_k/dt = \mathbf{y}_k \cdot d\mathbf{x}/dt = \sum_m \lambda_m L_{km} \qquad (28)$$

where we defined the functionals

$$L_{km} = \tau^{-1}\, \mathbf{y}_k \cdot \left[\mathbf{y}_m - (\mathbf{y}_m)_{L(g_0,g_1,\ldots,g_n)}\right] \qquad (29a)$$

$$= \tau^{-1}\left[\mathbf{y}_k - (\mathbf{y}_k)_{L(g_0,g_1,\ldots,g_n)}\right] \cdot \left[\mathbf{y}_m - (\mathbf{y}_m)_{L(g_0,g_1,\ldots,g_n)}\right] \qquad (29b)$$

and in writing Equation 29b we have used Equation A13 with $\mathbf{g} = (\mathbf{y}_k)_{L(g_0,g_1,\ldots,g_n)}$ and subtracted a zero from Equation 29a. Relation 28 shows that the rates of change (or generalized "fluxes") dY_k/dt and the affinities (or generalized "forces") λ_m are linearly interrelated by the coefficients (or generalized "conductivities") L_{km}.

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

$$L_{km} = \tau^{-1} \frac{\begin{vmatrix} \mathbf{y}_k\cdot\mathbf{y}_m & \mathbf{y}_k\cdot\mathbf{h}_1 & \cdots & \mathbf{y}_k\cdot\mathbf{h}_r \\ \mathbf{y}_m\cdot\mathbf{h}_1 & \mathbf{h}_1\cdot\mathbf{h}_1 & \cdots & \mathbf{h}_r\cdot\mathbf{h}_1 \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{y}_m\cdot\mathbf{h}_r & \mathbf{h}_1\cdot\mathbf{h}_r & \cdots & \mathbf{h}_r\cdot\mathbf{h}_r \end{vmatrix}}{\begin{vmatrix} \mathbf{h}_1\cdot\mathbf{h}_1 & \cdots & \mathbf{h}_r\cdot\mathbf{h}_1 \\ \cdots & \cdots & \cdots \\ \mathbf{h}_1\cdot\mathbf{h}_r & \cdots & \mathbf{h}_r\cdot\mathbf{h}_r \end{vmatrix}} \qquad (30)$$

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

$$L_{km} = L_{mk} \qquad (31)$$

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

$$[\mathbf{L}] = \begin{bmatrix} L_{00} & L_{01} & \cdots & L_{0m} & \cdots \\ L_{10} & L_{11} & \cdots & L_{1m} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ L_{k0} & L_{k1} & \cdots & L_{km} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \qquad (32)$$

is a Gram matrix and as such it is nonnegative. Matrix $[\mathbf{L}]$ is strictly positive only if the vectors $\mathbf{y}_k-(\mathbf{y}_k)_{L(\mathbf{g}_0,\mathbf{g}_1,\ldots,\mathbf{g}_n)}$ are all linearly independent, in which case the set of Equations 28 may be solved to yield

$$\lambda_m = \sum_k ([\mathbf{L}]^{-1})_{mk}\, dY_k/dt \qquad (33)$$

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

$$dS/dt = dF/dt = \mathbf{f}\cdot d\mathbf{x}/dt$$

$$= \sum_k \lambda_k \mathbf{y}_k\cdot d\mathbf{x}/dt = \sum_k \lambda_k\, dY_k/dt \qquad (34a)$$

$$= \tau\; d\mathbf{x}/dt\cdot d\mathbf{x}/dt = \sum_k\sum_m \lambda_k L_{km} \lambda_m \qquad (34b)$$

and, if $[\mathbf{L}]$ is strictly positive,

$$= \sum_k\sum_m dY_m/dt\; ([\mathbf{L}]^{-1})_{mk}\, dY_k/dt \qquad (35)$$

Finally, comparing Relation 20, for the codispersion Y_{km} of properties Y_k and Y_m, and Relation 30, for the generalized conductivities L_{km}, we see that there is a relation between L_{km}, Y_{km} and all the codispersions of the constraints, and properties Y_k and Y_m. We may greatly simplify these relations if, for a given distribution $\mathbf{x}$, we further restrict the choice of the complete set of linearly independent property functionals Y_k so that $\mathbf{y}_0 = \mathbf{h}_1$, $\mathbf{y}_1 = \mathbf{h}_2$, ..., $\mathbf{y}_{r-1} = \mathbf{h}_r$ where $\mathbf{h}_1$, ..., $\mathbf{h}_r$ are linearly independent vectors spanning the manifold $L(\mathbf{g}_0,\mathbf{g}_1,\ldots,\mathbf{g}_n)$ generated by the constraints and, moreover, we select the functionals for $k > r$ so that the codispersions Y_{k0}, Y_{k1}, ..., $Y_{k(r-1)}$ (Equation 20) are all equal to zero (notice that $Y_{k0} = 0$ implies $Y_k = 0$). For this particular choice, by studying Relation 30 for the generalized conductivities, we may readily verify that

$$L_{km} = 0 \qquad \text{whenever } k < r \text{ or } m < r \qquad (36a)$$

$$L_{km} = 4\, Y_{km}/\tau \qquad \text{for } k \geq r \text{ and } m \geq r \qquad (36b)$$

and, therefore, we find a direct relation between the covariance and the generalized conductivity of the pair of properties Y_k and Y_m. In particular, for $k = m > r$ we find $L_{kk} = 4Y_{kk}/\tau$ which is a relation between the variance (or fluctuation) and the direct conductivity (or dissipation) of property Y_k.

5. TIME-DEPENDENT CONSTRAINTS

In terms of the notation already introduced, let us consider the differential equation

$$d\mathbf{x}/dt = \gamma_k^{-1}\,[\mathbf{g}_k - (\mathbf{g}_k)_{L(\mathbf{g}_0,\mathbf{g}_1,\ldots,\mathbf{g}_{k-1},\mathbf{g}_{k+1},\ldots,\mathbf{g}_n)}] \qquad (37a)$$

where

$$\gamma_k = \beta_k(t,\mathbf{x}(t))^{-1}\, \mathbf{g}_k\cdot[\mathbf{g}_k - (\mathbf{g}_k)_{L(\mathbf{g}_0,\mathbf{g}_1,\ldots,\mathbf{g}_{k-1},\mathbf{g}_{k+1},\ldots,\mathbf{g}_n)}] \qquad (37b)$$

and $\beta_k(t,\mathbf{x}(t)) = \alpha_k(t,\mathbf{p}(t))$ as specified by Equation 5a of

Problem 2. Again, we may readily verify that Equation 37 induces an equation for dp/dt which contains only the square x_i^2 of the new variables and, therefore, is of the form

$$dp/dt = R_k(t,p) \quad (38)$$

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

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

Clearly, when two or more constraints have a specified rate of change, then Equation 7 yields many orthogonal contributions to dx/dt. The terms $R_1,\ldots,R_n$ (each with structure similar to that given by Equation 37) cause the shifting constraints to follow the specified rates of change $\alpha_1,\ldots,\alpha_n$. The contribution of these terms to the rate of entropy change does not have a definite sign. The term **D** (as given by Equation 8) gives instead a positive definite contribution to the rate of entropy change and tends to attract the distribution **x** towards a path of steepest entropy ascent compatible with the instantaneous values of the constraints.

We may finally note that by substituting $L(g_0,g_1,\ldots,g_{k-1},g_{k+1},\ldots,g_n)$ in Equation 37 with $L(f,g_0,g_1,\ldots,g_{k-1},g_{k+1},\ldots,g_n)$ we would obtain a rate equation causing the k-th constraint to follow the specified rate α_k 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 k-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 [1,6].

APPENDIX

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

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

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

$$p_i = x_i^2 \qquad i = 1, \ldots \quad (A2)$$

We now rewrite Problem 2 as follows

$$dG_k(x(t))/dt = \beta_k(t,x(t)) \quad (A3a)$$

and

$$dF(x(t))/dt > 0 \quad (A3b)$$

where

$$x = \{x_1,\ldots,x_i,\ldots\} \quad (A4)$$

$$G_k(x) = \sum_i x_i^2 A_{ki} \qquad k = 0,1,\ldots,n \quad (A5)$$

$$F(x) = -\sum_i x_i^2 \ln x_i^2 \quad (A6)$$

Next, we define vectors representing gradients of the constraints $G_k(x)$ and of the entropy functional $F(x)$,

$$g_k = \{\partial G_k/\partial x_1,\ldots,\partial G_k/\partial x_i,\ldots\} = \{2x_1A_{k1},\ldots,2x_iA_{ki},\ldots\}$$
$$k = 0,1,\ldots, n \quad (A7)$$

$$f = \{\partial F/\partial x_1,\ldots,\partial F/\partial x_i,\ldots\}$$
$$= \{-2x_1-2x_1\ln x_1^2,\ldots,-2x_i-2x_i\ln x_i^2,\ldots\} \quad (A8)$$

so that the rates of change of the functions $G_0, G_1, \ldots, G_n$, and F are given by

$$dG_k/dt = g_k \cdot dx/dt \quad (A9)$$

$$dF/dt = \mathbf{f} \cdot d\mathbf{x}/dt \qquad \text{(A10)}$$

where, clearly, $d\mathbf{x}/dt = \{dx_1/dt, \ldots, dx_i/dt, \ldots\}$ and the dot product has the obvious meaning (for example, $\mathbf{f} \cdot d\mathbf{x}/dt = f_1 dx_1/dt + \ldots + f_i dx_i/dt + \ldots$).

Given a set of vectors $\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n$, the symbol

$$L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n) \qquad \text{(A11)}$$

will denote their linear span, i.e., the linear manifold containing all the vectors that are linear combinations of $\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n$. Given another vector $\mathbf{b}$, the symbol

$$(\mathbf{b})_{L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n)} \qquad \text{(A12)}$$

will denote the orthogonal projection of $\mathbf{b}$ onto the linear manifold $L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n)$, namely, the unique vector in $L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n)$ such that its dot product with any other vector $\mathbf{g}$ in $L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n)$ equals the dot product of $\mathbf{b}$ with $\mathbf{g}$, i.e.,

$$\mathbf{g} \cdot (\mathbf{b})_{L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n)} = \mathbf{g} \cdot \mathbf{b} \qquad \text{(A13)}$$

for every $\mathbf{g}$ in $L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n)$.

In terms of a set of linearly independent vectors $\mathbf{h}_1, \ldots, \mathbf{h}_r$ spanning the manifold $L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n)$, where clearly $r \le n$, we can write two equivalent explicit expressions for the projection $(\mathbf{b})_{L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n)}$ of vector $\mathbf{b}$ onto $L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n)$. The first is

$$(\mathbf{b})_{L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n)} = \sum_{k=1}^{r} \sum_{m=1}^{r} (\mathbf{b} \cdot \mathbf{h}_k) \left[M(\mathbf{h}_1, \ldots, \mathbf{h}_r)^{-1}\right]_{km} \mathbf{h}_m \qquad \text{(A14)}$$

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

$$M(\mathbf{h}_1, \ldots, \mathbf{h}_r) = \begin{bmatrix} \mathbf{h}_1 \cdot \mathbf{h}_1 & \cdots & \mathbf{h}_r \cdot \mathbf{h}_1 \\ \cdots & \cdots & \cdots \\ \mathbf{h}_1 \cdot \mathbf{h}_r & \cdots & \mathbf{h}_r \cdot \mathbf{h}_r \end{bmatrix} \qquad \text{(A15)}$$

The second expression is a ratio of two determinants

$$(\mathbf{b})_{L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n)} = - \frac{\begin{vmatrix} 0 & \mathbf{h}_1 & \cdots & \mathbf{h}_r \\ \mathbf{f} \cdot \mathbf{h}_1 & \mathbf{h}_1 \cdot \mathbf{h}_1 & \cdots & \mathbf{h}_r \cdot \mathbf{h}_1 \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{f} \cdot \mathbf{h}_r & \mathbf{h}_1 \cdot \mathbf{h}_r & \cdots & \mathbf{h}_r \cdot \mathbf{h}_r \end{vmatrix}}{\begin{vmatrix} \mathbf{h}_1 \cdot \mathbf{h}_1 & \cdots & \mathbf{h}_r \cdot \mathbf{h}_1 \\ \cdots & \cdots & \cdots \\ \mathbf{h}_1 \cdot \mathbf{h}_r & \cdots & \mathbf{h}_r \cdot \mathbf{h}_r \end{vmatrix}} \qquad \text{(A16)}$$

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

$$\mathbf{b} - (\mathbf{b})_{L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n)} = \frac{\begin{vmatrix} \mathbf{f} & \mathbf{h}_1 & \cdots & \mathbf{h}_r \\ \mathbf{f} \cdot \mathbf{h}_1 & \mathbf{h}_1 \cdot \mathbf{h}_1 & \cdots & \mathbf{h}_r \cdot \mathbf{h}_1 \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{f} \cdot \mathbf{h}_r & \mathbf{h}_1 \cdot \mathbf{h}_r & \cdots & \mathbf{h}_r \cdot \mathbf{h}_r \end{vmatrix}}{\begin{vmatrix} \mathbf{h}_1 \cdot \mathbf{h}_1 & \cdots & \mathbf{h}_r \cdot \mathbf{h}_1 \\ \cdots & \cdots & \cdots \\ \mathbf{h}_1 \cdot \mathbf{h}_r & \cdots & \mathbf{h}_r \cdot \mathbf{h}_r \end{vmatrix}} \qquad \text{(A17)}$$

where in writing Equation A17 we used Equation A16.

The vector represented by Equation A17 has the relevant property

$$\mathbf{g}_k \cdot [\mathbf{b} - (\mathbf{b})_{L(\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_n)}] = 0 \qquad k = 0, 1, \ldots, n \qquad \text{(A18)}$$

which follows directly from Relation A13, i.e., the vector $\mathbf{b} - (\mathbf{b})_{L(g_0,g_1,...,g_n)}$ is orthogonal to manifold $L(g_0,g_1,...,g_n)$. Moreover, we have the other relevant property

$$\mathbf{b}\cdot[\mathbf{b} - (\mathbf{b})_{L(g_0,g_1,...,g_n)}]$$
$$= [\mathbf{b} - (\mathbf{b})_{L(g_0,g_1,...,g_n)}]\cdot[\mathbf{b} - (\mathbf{b})_{L(g_0,g_1,...,g_n)}] \geq 0 \quad \text{(A19)}$$

where the strict inequality applies whenever $\mathbf{b}$ is not in $L(g_0,g_1,...,g_n)$.

In the paper, we make extensive use of the notation and relations just discussed [5,7].

REFERENCES

1. 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. 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, Book H0341C, Vol. 3, p. 135, WA/AES (1986).

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

4. 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. G.P. Beretta, ASME Paper, Book H0341C, Vol. 3, p. 129, WA/AES (1986).

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

7. An earlier version of the present paper has been presented at the IV International Symposium on Second Law Analysis of Thermal Systems in Rome, Italy, May 1987.

ACKNOWLEDGMENT

The author is indebted with Dr. G.N. Hatsopoulos and the Thermo Electron Corporation for constant support of my attempts to develop methods for nonequilibrium thermodynamic analysis.