

Article

## Modeling Non-Equilibrium Dynamics of a Discrete Probability Distribution: General Rate Equation for Maximal Entropy Generation in a Maximum-Entropy Landscape with Time-Dependent Constraints

Gian Paolo Beretta

Università di Brescia, via Branze 38, Brescia, I-25123, Italy, E-mail: beretta@ing.unibs.it

Temporary address: Massachusetts Institute of Technology, Room 3-237, Cambridge, MA 01239, USA

Received: 17 December 2007; in revised form: 30 July 2008 / Accepted: 30 July 2008 / Published: 14 August 2008

---

**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 resulting dynamics follows the path of steepest entropy 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 in a general, non-quantal formulation as a part of an effort to develop tools for the phenomenological treatment of non-equilibrium problems with applications in engineering, biology, sociology, and economics. The rate equation is also extended to include the case of assigned time-dependences of the constraints and the entropy, such as for modeling non-equilibrium energy and entropy exchanges.

**Keywords:** maximum entropy formalism, maximum entropy production, steepest entropy ascent, nonequilibrium dynamics, irreversible thermodynamics, Onsager reciprocity, fluctuation-dissipation relations, entropy generation by irreversibility, nonequilibrium temperature, constrained equilibrium, nonlinear dynamical systems

---

## 1. Foreword

The term *entropy* is used in many different contexts wherever the relevant *state description* is in terms of a probability distribution over some set of possible events which characterize the *system description*. Depending on the context, such events may be *microstates*, or *eigenstates*, or *configurations*, or *trajectories*, or *transitions*, or *mutations*, and so on.

Mathematically, entropy is functional of the state probabilities which quantifies their *spread* according to some set of defining axioms discussed by several authors during this “Facets of Entropy” Conference. In this sense, the use of a common name for all the possible different state functionals that share such broad defining features, may have some unifying advantage from the conceptual point of view (in the same sense as the term “house” unifies in one concept an infinite variety of very different realizations), and may be useful to understand analogies and suggest inter-breeding developments between very different fields of research that share similar probabilistic descriptions.

On the other hand, from a physics point of view, entropy is very precise, single property of matter that can be measured in every laboratory by means of well defined, if not standard measurement procedures (the most general and rigorous axiomatic definition currently available, and valid also for non-equilibrium states, is found in Ref. [1], see also [2]). The first instance of the use of the term *entropy* for this property is in a 1865 paper by the German professor of mathematical physics, Rudolf Clausius [3]. Since then, the term remained restricted to the domain of physics and thermodynamics until in 1929 von Neumann [4] pioneered the reasoning that was subsequently followed by Shannon [5] and spread thereafter in various directions, many of which unrelated to physics considerations.

In fields outside physics and thermodynamics the characterization of probability distributions by means of some entropy functional has generally been beneficial and provided new insights. But the inflation of different meanings all assigned to the single word entropy has in turn originated confusion and misunderstandings about the physical (thermodynamic) entropy. The analogies between probabilistic mathematical theories in different fields is not sufficient to make inferences about interpretations of the meaning of the physical entropy. A physical theory [6] is not just a set of “formal links” between concepts (i.e., a mathematical theory) but also of “empirical links” between such concepts and the *empirical world*, the *plane of perceptions*, the empirical data, the measurement procedures. Hence, we should beware of the possible abuses and superficial or false conclusions that may arise from making inferences from incomplete analogies between theories.

Having said so, below I hope I will not add to the confusion by proposing some general mathematics which does originate from a physical theory, but (with the due caution about meanings and interpretation) might find use in several other fields, where the maximum entropy formalism has already found an application and provides therefore an existing relevant background, or “landscape”. We are interested in “non-equilibrium probability distributions” and their time evolution towards the maximal entropy distribution. We develop a general family of dynamical equations with which we can describe for example a constrained, maximal entropy generation probability redistribution dynamics, with either “natural constraints” such as the constants of the motion of the system, or externally assigned, (rate-controlled) time dependent constraints and/or entropy changes.

## 2. 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 [7–11]. 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 an arbitrary non-equilibrium probability distribution towards the maximum entropy distribution. The main form of the nonlinear rate equation we propose for the probability distribution has the feature that it keeps the constraints constant at their initial values (or varies them in time according to an externally provided law) 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 non-equilibrium 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, which may have specified constant values (thus representing constants of the “motion”) or may have a specified time-dependence (as for example in rate-controlled constrained equilibrium modeling of chemical kinetics [10, 11] or other dynamical system modeling).

The formalism presented here has features of great generality, and is presented in several forms that are adaptable to different applications. It was originally “designed” by the author in 1981 [12] and subsequently developed [13–18] to obtain an equation of motion for a quantum theoretical unification of mechanics and thermodynamics [19, 20] whereby the most general known statement of second law of thermodynamics, due to Hatsopoulos and Keenan [21], emerges as a fundamental dynamical theorem about the stability of the equilibrium states of a microscopic dynamics which generalizes the Schrödinger equation. Recently, the original equation has been partially rediscovered in the same context [22]. The idea of steepest-entropy-ascent time evolution of a probability distribution has been recently rediscovered in [23] but with important differences we discuss in [18].

In the present paper, the formalism is abstracted from its original physics purpose. It is presented as a mathematical tool that may find phenomenological modeling applications in other fields. This abstracted version, originally developed by the author in 1986 [24, 25], has also been recently rediscovered in a maximum entropy formalism context [26].

Because the so-called maximum entropy formalism [7–9] is well known, and the constrained-equilibrium method for the treatment of non-equilibrium in chemical engineering is easily understood [10, 11], we illustrate our rate equation in this framework, assuming for simplicity linear constraints. The generalization to nonlinear constraints has also been developed and is straightforward, but will be discussed elsewhere.

### 3. Maximum Entropy Landscape

The maximum entropy problem which sets our context (landscape, to use the terminology of nonlinear optimization) is that of seeking a probability distribution, namely, a probability vector  $\mathbf{p} = p_1, \dots, p_i, \dots$ , whose entropy

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

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

$$\sum_i p_i A_{ki} = \langle A_k \rangle \quad k = 0, 1, \dots, n \tag{2}$$

where  $A_{ki}$  is the magnitude of the  $k$ -th constraint in base state  $i$ . 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

$$p_i^* = \frac{\exp(-\sum_{k=1}^n \lambda_k A_{ki})}{Q}, \text{ with } Q = \sum_i \exp\left(-\sum_{k=1}^n \lambda_k A_{ki}\right) \tag{3}$$

where the Lagrange multipliers  $\lambda_1, \dots, \lambda_n$  are determined by the values  $\langle A_1 \rangle, \dots, \langle A_n \rangle$  of the constraints.

Given the above landscape, the extensions to the non-equilibrium domain that we wish to consider are the following.

### 4. Non-equilibrium Problem A: Constant Values of the Constraints

We seek a time-dependent probability distribution, namely, a vector function (one-parameter family)  $\mathbf{p}(t) = \{p_1(t), \dots, p_i(t), \dots\}$ , whose entropy  $S(\mathbf{p}(t))$  is non-decreasing with time  $t$ ,

$$\frac{dS(\mathbf{p})}{dt} = -\frac{d}{dt} \left[ \sum_i p_i(t) \ln p_i(t) \right] \geq 0 \tag{4}$$

and such that the magnitudes  $\langle A_k \rangle$  of the constraints are time-invariant, namely,

$$\sum_i p_i(t) A_{ki} = \langle A_k \rangle \quad k = 0, 1, \dots, n \tag{5}$$

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  $t$

$$\sum_i p_i(t) A_{ki} = \sum_i p_i(0) A_{ki} \quad k = 0, 1, \dots, n \tag{6}$$

### 5. Non-equilibrium Problem B: Time-Dependent Values of the Constraints

A further generalization of the preceding non-equilibrium problem is one in which the magnitudes  $\langle A_k \rangle$  of the constraints are assigned a definite time-dependence.

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

$$\frac{d\langle A_k \rangle}{dt} = \frac{d}{dt} \left[ \sum_i p_i(t) A_{ki} \right] = \alpha_k(t, \mathbf{p}(t)) \quad k = 0, 1, \dots, n \quad (7)$$

and

$$\frac{dS(\mathbf{p})}{dt} = -\frac{d}{dt} \left[ \sum_i p_i(t) \ln p_i(t) \right] \geq 0 \quad (8)$$

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

For example, Problem B can be applied in the context of the constrained-equilibrium method for chemical kinetics [10, 11]. 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 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 B 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

$$\frac{d\mathbf{p}}{dt} = D(\mathbf{p}) \quad (9)$$

whose solutions are solutions of Problem A.

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

$$\frac{d\mathbf{p}}{dt} = \mathbf{R}_1^s(t, \mathbf{p}) + \dots + \mathbf{R}_n^s(t, \mathbf{p}) + \mathbf{D}(\mathbf{p}) \quad (10)$$

whose solutions are solutions of Problem B.

These differential equations and their main properties are presented in terms of the notation introduced in the next two sections.

## 6. Reformulation in terms of square-root probabilities

Because the variables  $p_1, \dots, p_i, \dots$  represent probabilities, both non-equilibrium problems A and B just defined are subject to the additional set of inequality constraints

$$p_i \geq 0 \quad \text{or} \quad p_i(t) \geq 0 \quad (11)$$

It is therefore convenient to change variables to the set of square-root probabilities  $x = \{x_1, \dots, x_i, \dots\}$  from which the probabilities may be computed according to the relations

$$p_i = x_i^2 \quad i = 1, \dots \quad (12)$$

Were it only to cope with the nonnegativity constraints 11, we could choose as new variables any function of the probabilities whose inverse is even. The main reason for choosing  $x_i = \sqrt{p_i}$  is geometrical and part of the steepest entropy ascent assumption, therefore we discuss it at the beginning of Section 8. We note here that for example in Ref. [23], where a steepest entropy ascent dynamics is derived without switching to the square root representation, the entropy gradient is not well-defined on the entire domain (it diverges whenever one of the probabilities is zero) and as a result the dynamics has unphysical effects and does not conserve positivity when solved backwards in time.

The normalization constraint may be written as

$$G_0(\mathbf{x}) = \sum_i x_i^2 = \mathbf{x} \cdot \mathbf{x} = \mathbf{x} \cdot \mathbf{I} \cdot \mathbf{x} = 1 \tag{13}$$

where  $\mathbf{I} = \text{diag}(1)$  and the dot product has the obvious meaning (for example,  $\mathbf{f} \cdot \mathbf{g} = f_1 g_1 + \dots + f_i g_i + \dots$ ). Hence, in square-root probability space the set of possible probability distributions maps to the unit sphere.

In terms of the new variables, we rewrite Problem B as follows (Problem A is a particular case)

$$\frac{dG_k(\mathbf{x}(t))}{dt} = \dot{G}_k(t, \mathbf{x}(t)) = \alpha(t, \mathbf{p}(t)) \tag{14}$$

and

$$\frac{dF(\mathbf{x}(t))}{dt} > 0 \tag{15}$$

where our state vector is now

$$\mathbf{x} = \{x_1, \dots, x_i, \dots\} \tag{16}$$

our constraints are

$$G_k(\mathbf{x}) = \sum_i x_i^2 A_{ki} = \mathbf{x} \cdot \mathbf{A}_k \cdot \mathbf{x} \quad k = 0, 1, \dots, n \tag{17}$$

where  $\mathbf{A}_k = \text{diag}(A_{ki})$  and the entropy functional is

$$F(\mathbf{x}) = - \sum_i x_i^2 \ln x_i^2 = \mathbf{x} \cdot \mathbf{S} \cdot \mathbf{x} \tag{18}$$

where  $\mathbf{S} = \text{diag}(-\ln x_i^2)$ . Clearly  $F(\mathbf{x})$  coincides with the entropy functional  $S(\mathbf{p})$  in view of Equations 12.

### 7. Gradients of Entropy and Constraints with respect to the Square-Root Probabilities

Next, we define vectors representing “gradients” of the constraints  $G_k(\mathbf{x})$  and of the entropy functional  $F(\mathbf{x})$ , with respect to the state vector  $\mathbf{x}$ ,

$$\mathbf{g}_k = \frac{\partial G_k}{\partial \mathbf{x}} = \left\{ \frac{\partial G_k}{\partial x_1}, \dots, \frac{\partial G_k}{\partial x_i}, \dots \right\} = \{2x_1 A_{k1}, \dots, 2x_i A_{ki}, \dots\} = 2\mathbf{A}_k \cdot \mathbf{x} \quad k = 0, 1, \dots, n \tag{19}$$

$$\mathbf{f} = \frac{\partial F}{\partial \mathbf{x}} = \left\{ \frac{\partial F}{\partial x_1}, \dots, \frac{\partial F}{\partial x_i}, \dots \right\} = \{-2x_1 - 2x_1 \ln x_1^2, \dots, -2x_i - 2x_i \ln x_i^2, \dots\} = 2(\mathbf{S} - \mathbf{I}) \cdot \mathbf{x} \tag{20}$$

so that the rates of change of the functionals  $G_0, G_1, \dots, G_n$ , and  $F$  induced by the time dependence  $\mathbf{x}(t)$  of the state vector, are given by

$$\frac{dG_k}{dt} = \mathbf{g}_k \cdot \dot{\mathbf{x}} \tag{21}$$

$$\frac{dF}{dt} = \mathbf{f} \cdot \dot{\mathbf{x}} \tag{22}$$

where  $\dot{\mathbf{x}} = d\mathbf{x}(t)/dt = \{dx_1(t)/dt, \dots, dx_i(t)/dt, \dots\}$ .

### 8. Time Evolution Along the Path of Steepest Entropy Ascent

In order to introduce the geometrical notion of steepest entropy ascent, we need to define what we mean by “distance” between two probability distributions, and by “length” of a curve in probability space defined by a trajectory, a time evolution. The proper unique natural metric for this purpose is known in statistics as the Fisher-Rao metric (see e.g. [27–29]). For a one-parameter family of discrete distributions,  $\mathbf{p}(t)$ , where  $t$  is the parameter, the distance between distributions  $\mathbf{p}(t + dt)$  and  $\mathbf{p}(t)$  is

$$\begin{aligned} d\ell &= \frac{1}{2} \sqrt{\sum_i p_i \left(\frac{d \ln p_i}{dt}\right)^2} dt = \frac{1}{2} \sqrt{\sum_i \frac{1}{p_i} \left(\frac{dp_i}{dt}\right)^2} dt = \sqrt{\sum_i \left(\frac{d\sqrt{p_i}}{dt}\right)^2} dt \\ &= \sqrt{\sum_i (\dot{x}_i)^2} dt = \sqrt{\dot{\mathbf{x}} \cdot \dot{\mathbf{x}}} dt \end{aligned} \tag{23}$$

Thus, square-root probabilities  $x_i = \sqrt{p_i}$  are the most natural variables in that:

- the space becomes the unit sphere,  $\mathbf{x} \cdot \mathbf{x} = 1$  ( $\sum_i p_i = 1$ );
- the Fisher-Rao metric simplifies to  $d\ell = \sqrt{\dot{\mathbf{x}} \cdot \dot{\mathbf{x}}} dt$ , or equivalently  $d\ell^2 = d\mathbf{x} \cdot d\mathbf{x}$ ;
- the distance between any two distributions is the angle  $d(\mathbf{x}_1, \mathbf{x}_2) = \cos^{-1}(\mathbf{x}_1 \cdot \mathbf{x}_2)$ .

We now focus on Problem A, and write  $\dot{\mathbf{x}} = \dot{\mathbf{x}}_L + \dot{\mathbf{x}}_{\perp L}$  where  $L = L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)$  is the linear manifold spanned by the gradient vectors associated with the constraints,  $dG_k/dt = 0$ , see Eq. 21. In order to satisfy all the constraints, it is clear that  $\dot{\mathbf{x}}$  must be orthogonal to  $L$ , i.e.,  $\dot{\mathbf{x}} = \dot{\mathbf{x}}_{\perp L}$ . Noting that also  $\mathbf{f}$  may be written as  $\mathbf{f} = \mathbf{f}_L + \mathbf{f}_{\perp L}$ , we may therefore rewrite Eq. 22 as

$$\frac{dF}{dt} = \mathbf{f}_{\perp L} \cdot \dot{\mathbf{x}}_{\perp L} \tag{24}$$

It is now apparent that the maximal increase of  $F$  obtains when  $\dot{\mathbf{x}}_{\perp L}$  is in the same direction as  $\mathbf{f}_{\perp L}$ .

Hence, the rate equation that we propose is expressed by the differential equation

$$\dot{\mathbf{x}} = \frac{1}{\tau(\mathbf{p})} \mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n) \tag{25}$$

where  $\tau(\mathbf{p})$  is some strictly positive functional of the probability distribution  $\mathbf{p}$  (possibly a constant) with dimensions of time, vectors  $\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n$ , and  $\mathbf{f}$  are defined by Relations 19 and 20, and the vector

$$\mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n) \tag{26}$$

is the orthogonal complement of the projection of  $\mathbf{f}$  onto the linear span of vectors  $\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n$ , for which the Appendix gives explicit expressions using well-known but seldom encountered linear algebra notations.

The notation and more explicit relations are reviewed in the Appendix. Using Relations 12 and 91, and some procedure to eliminate from the set  $\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n$  those vectors (if any) that are linearly dependent on the others, we may readily verify that Equation 25 induces a differential equation for the probabilities  $\mathbf{p}$  which contains only the square  $x_i^2$  of the new variables and, therefore, is of the form of Equation 9.

By virtue of Relations 21 and 92, we verify that Equation 25 maintains the magnitude of each constraint invariant, i.e.,

$$\frac{dG_k}{dt} = \mathbf{g}_k \cdot \dot{\mathbf{x}} = \frac{1}{\tau} \mathbf{g}_k \cdot \mathbf{f}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} = 0 \tag{27}$$

whereas, by virtue of Relations 22 and 94 we verify that the value of the entropy functional is non-decreasing under Equation 25, i.e., (later we obtain other important expressions, e.g., Eq. 54)

$$\frac{dS}{dt} = \frac{dF}{dt} = \mathbf{f} \cdot \dot{\mathbf{x}} = \frac{1}{\tau} \mathbf{f} \cdot \mathbf{f}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} = \frac{1}{\tau} \mathbf{f}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} \cdot \mathbf{f}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} = \tau \dot{\mathbf{x}} \cdot \dot{\mathbf{x}} \geq 0 \tag{28}$$

where the equal sign (no entropy generation) applies if and only if vector  $\mathbf{f}$  is in  $L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)$  and, therefore, is a linear combination of vectors  $\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n$ , namely, there is a set of multipliers  $\lambda_0, \lambda_1, \dots, \lambda_n$  such that

$$\mathbf{f} = \sum_{k=0}^n \lambda_k \mathbf{g}_k \tag{29}$$

Using Relations 19 and 20, Condition 29 becomes

$$-x_i - x_i \ln x_i^2 = \sum_{k=0}^n \lambda_k x_i A_{ki} \quad i = 1, 2, \dots \tag{30}$$

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

$$p_i \ln p_i = -p_i \left( 1 + \sum_{k=0}^n \lambda_k A_{ki} \right) \quad i = 1, 2, \dots \tag{31}$$

To be satisfied for each  $i$  this requires either  $p_i = 0$  or  $\ln p_i = -(1 + \sum_{k=0}^n \lambda_k A_{ki})$  which we may express in compact form by

$$p_i = \frac{\delta_i \exp(-\sum_{k=1}^n \lambda_k A_{ki})}{\mathcal{Q}}, \text{ with } \mathcal{Q} = \sum_i \delta_i \exp\left(-\sum_{k=1}^n \lambda_k A_{ki}\right) \tag{32}$$

where  $\delta_i$  is a boolean variable that may be only either 0 or 1, and we evaluated  $\lambda_0$  by the normalization condition.

It is noteworthy that Equation 25 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, a feature that guarantees that each probability  $p_i$ , if positive at some instance must be positive at all times, for the dynamics prohibits its ever crossing zero. This feature also guarantees that the solutions are well defined not only in forward time but backwards

in time as well. From the present state, we can solve for future states but we can also reconstruct the past. Notice that this mathematical reversibility (existence of inverse of the dynamical map) has nothing to do with the “physical” irreversibility whereby the entropy increases in forward time and decreases in backward time.

From Relations 28 and 31 it follows that the effect of Equation 25 is to smoothly rearrange – without violating any constraint – the nonzero probabilities in the distribution, towards higher entropy distributions until a partially canonical distribution defined by Equation 32 is approached, in which the initially zero probabilities are still equal to zero whereas the initially nonzero probabilities are distributed canonically.

Clearly, such partially canonical equilibrium distribution maximizes the entropy functional when restricted to the subset of distributions that have equal to zero all the initially zero probabilities. Such distributions are equilibrium, i.e., time invariant, but are unstable whenever there is some probability  $p_i$  equal to zero. We can see this by noting that a minor perturbation of the distribution which changes a zero probability to an arbitrarily small nonzero value, would proceed away towards a quite different equilibrium of higher entropy. If initially all the probabilities in the distribution have nonzero values, then Equation 25 takes it directly towards the unique stable equilibrium distribution compatible with the initial values of the constraints, which is the maximum entropy distribution given by Equation 3.

Geometrically, we can visualize the effect of Equation 25 as follows. Consider the surface defined by  $G_k(\mathbf{x}) = \langle A_k \rangle$  for  $k = 0, 1, \dots, n$  where  $\langle A_k \rangle$  are the magnitudes of the constraints fixed by the initial distribution. On this surface we can identify contour curves of constant entropy, generated by intersecting it with the constant-entropy surfaces  $F(\mathbf{x}) = S$  where  $S$  varies from 0 to the maximum value compatible with the given magnitudes of the constraints. Every trajectory  $\mathbf{x}(t)$  generated by Equation 25 lies on the constant- $G_k$  surface 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.

We finally notice that, using Relations 28, Equation 25 may also be rewritten as

$$\dot{\mathbf{x}} = \dot{S} \frac{\mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)}{\mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n) \cdot \mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} \tag{33}$$

This form may be useful in variants of Problem A whereby one needs to specify explicitly the rate of entropy change  $\dot{S}$  along the steepest entropy ascent (or descent, if  $\dot{S}$  is for some reason specified negative and hence the inequality in Eq. 4 reversed) together with the condition that the probability distribution must maintain invariant all the constants of the motion  $G_k$ 's. Notice, however, that this form is undefined at any stable maximum entropy equilibrium state except if  $\dot{S}$  is a function of and goes to zero faster than  $\mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n) \cdot \mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)$ .

Moreover, going back to the Fisher-Rao metric, Relations 28 allow us to write the following equivalent expressions

$$d\ell = \sqrt{\dot{\mathbf{x}} \cdot \dot{\mathbf{x}}} dt = \sqrt{\frac{1}{\tau} \frac{dS}{dt}} dt = \frac{1}{\tau} \frac{dS}{d\ell} dt \tag{34}$$

from which the relaxation time  $\tau$  may be interpreted as the characteristic time of the rate  $d\ell/dt$  at which the state vector follows the steepest entropy ascent trajectory, modulated by the rate  $dS/d\ell$  at which the

entropy functional increases along the trajectory,

$$\frac{1}{\tau} = \frac{d\ell/dt}{dS/d\ell} \tag{35}$$

or, alternatively, we see that when time is measured in units of the 'local' (or instantaneous) value of  $\tau$ , the 'speed' along the steepest entropy ascent trajectory coincides with the 'local' (or instantaneous) rate of entropy increase along the trajectory,

$$\frac{d\ell}{d(t/\tau)} = \frac{dS}{d\ell} \tag{36}$$

which justifies the interpretation of  $\tau$  as an intrinsic dynamical time.

In the next section we discuss two further properties of Equation 25 related to Onsanger's reciprocity and the fluctuation-dissipation theorem. As already mentioned, the mathematical structure of Equation 25 was originally developed by the author within the context of a unified theory of mechanics and thermodynamics that we call quantum thermodynamics [3, 13, 14, 16–18]. A useful application in that framework could be to compute transport properties in the far non-equilibrium domain.

### 9. Onsanger 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

$$\sum_i p_i A_{ki} = \langle A_k \rangle \quad k = 0, 1, \dots, n, \dots \tag{37}$$

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. Let us denote by  $Y_0, Y_1, \dots, Y_k, \dots$  one such complete set of property functionals of the variables  $x_i^2$ , namely,

$$Y_k(\mathbf{x}) = \sum_i x_i^2 A_{ki} \quad k = 0, 1, \dots \tag{38}$$

such that if the values of all these functionals are given, then the values of all the  $x_i^2$  are determined.

For simplicity, let us further assume that functional  $Y_0$  is the normalization constraint, i.e.,  $A_{0i} = 1$  for each  $i$ . Next we define the gradient vector of each functional  $Y_k$  as

$$\mathbf{y}_k = \{\partial Y_k/\partial x_1, \dots, \partial Y_k/\partial x_i, \dots\} = \{2x_1 A_{k1}, \dots, 2x_i A_{ki}, \dots\} \quad k = 0, 1, \dots \tag{39}$$

Note that in terms of the gradient vectors,  $\mathbf{y}_0 = 2\mathbf{x}$  and the "mean" functionals  $Y_k$  may be written as

$$Y_k = \frac{1}{4} \mathbf{y}_0 \cdot \mathbf{y}_k \quad Y_0 = \frac{1}{4} \mathbf{y}_0 \cdot \mathbf{y}_0 = 1 \tag{40}$$

and we may define the "deviation" vectors

$$\Delta \mathbf{y}_k = \frac{1}{2} (\mathbf{y}_k - \mathbf{y}_0 Y_k) \tag{41}$$

so that clearly  $\Delta \mathbf{y}_k \cdot \mathbf{y}_0 = 0$ . By noting that  $2 \sum_k \dot{p}_k A_{ik} A_{jk} = \dot{\mathbf{y}}_i \cdot \mathbf{y}_j = \mathbf{y}_i \cdot \dot{\mathbf{y}}_j$ , it is also easy to check the following interesting relations

$$\Delta \dot{\mathbf{y}}_i \cdot \Delta \mathbf{y}_j = \Delta \mathbf{y}_i \cdot \Delta \dot{\mathbf{y}}_j \tag{42}$$

which may be interpreted as a kind of “kinematic reversibility” inherent in our gradient formalism, and holds independently of the steepest ascent dynamical assumption.

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

$$\begin{aligned} \Delta Y_k \Delta Y_m(\mathbf{x}) &= \Delta \mathbf{y}_k \cdot \Delta \mathbf{y}_m = \sum_i x_i^2 A_{ki} A_{mi} - \sum_i x_i^2 A_{ki} \sum_j x_j^2 A_{mj} \\ &= \frac{1}{4} \mathbf{y}_k \cdot \mathbf{y}_m - \left(\frac{1}{4} \mathbf{y}_0 \cdot \mathbf{y}_k\right) \left(\frac{1}{4} \mathbf{y}_0 \cdot \mathbf{y}_m\right) \end{aligned} \tag{43}$$

which represent the “covariance” or “codispersion” of properties  $Y_k$  and  $Y_m$ . In particular, the functional  $\Delta Y_k \Delta Y_k$  represents the variance or dispersion (also, fluctuation) of property  $Y_k$ . In the notation of the Appendix,

$$\Delta Y_k \Delta Y_m(\mathbf{x}) = \Delta \mathbf{y}_k \cdot \Delta \mathbf{y}_m = \frac{1}{16} \begin{vmatrix} \mathbf{y}_k \cdot \mathbf{y}_m & \mathbf{y}_k \cdot \mathbf{y}_0 \\ \mathbf{y}_0 \cdot \mathbf{y}_m & \mathbf{y}_0 \cdot \mathbf{y}_0 \end{vmatrix} \tag{44}$$

and

$$\Delta Y_k \Delta Y_k(\mathbf{x}) = \Delta \mathbf{y}_k \cdot \Delta \mathbf{y}_k = \frac{1}{16} \det M(\mathbf{y}_k, \mathbf{y}_0) \tag{45}$$

We now consider the entropy functional

$$F(\mathbf{x}) = - \sum_i x_i^2 \ln x_i^2 \tag{46}$$

and its gradient vector

$$\mathbf{f} = \{\partial F/\partial x_1, \dots, \partial F/\partial x_i, \dots\} = \{-2x_1 - 2x_1 \ln x_1^2, \dots, -2x_i - 2x_i \ln x_i^2, \dots\} \tag{47}$$

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 that have zero entries corresponding to the zero  $x_i$ 's, so that there is a unique set of scalars  $\lambda_0, \lambda_1, \dots$  such that the vector  $\mathbf{f}$  can be written as

$$\mathbf{f} = \sum_k \lambda_k \mathbf{y}_k \tag{48}$$

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

$$\sum_k \lambda_k A_{ki} = -1 - \ln x_i^2, \text{ with the index } i \text{ restricted to the set for which } x_i \neq 0. \tag{49}$$

The entropy functional  $F$  may also be written as

$$F(\mathbf{x}) = - \sum_i x_i^2 \ln x_i^2 = 1 + \frac{1}{4} \mathbf{y}_0 \cdot \mathbf{f} = 1 + \frac{1}{4} \sum_k \lambda_k \mathbf{y}_0 \cdot \mathbf{y}_k = 1 + \sum_k \lambda_k Y_k \tag{50}$$

where we used Equations 40 and 48.

We see from Relation 50 that the scalar  $\lambda_k$  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  $\mathbf{x}$ , i.e.,  $\lambda_k = \partial F / \partial Y_k$ .

Also here we may define the “deviation” vector

$$\Delta \mathbf{f} = \frac{1}{2}(\mathbf{f} - \mathbf{y}_0(F - 1)) \tag{51}$$

so that clearly  $\Delta \mathbf{f} \cdot \mathbf{y}_0 = 0$ .

In the notation of the Appendix, we may write the codispersion of entropy  $F$  and property  $Y_m$  as

$$\Delta F \Delta Y_m(\mathbf{x}) = \Delta \mathbf{f} \cdot \Delta \mathbf{y}_m = \frac{1}{16} \begin{vmatrix} \mathbf{f} \cdot \mathbf{y}_m & \mathbf{f} \cdot \mathbf{y}_0 \\ \mathbf{y}_0 \cdot \mathbf{y}_m & \mathbf{y}_0 \cdot \mathbf{y}_0 \end{vmatrix} \tag{52}$$

and the entropy fluctuations as

$$\Delta F \Delta F(\mathbf{x}) = \Delta \mathbf{f} \cdot \Delta \mathbf{f} = \frac{1}{16} \det M(\mathbf{f}, \mathbf{y}_0) \tag{53}$$

In terms of this notation, when vectors  $\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n$  are linearly independent and  $\mathbf{g}_0 = 2\mathbf{x}$  is the gradient associated with the normalization constraint, then the rate of entropy generation (Eq. 28) may be written as a ratio of Gram determinants as follows (see Relations 97 and 98 in the Appendix)

$$\frac{dS}{dt} = \frac{1}{\tau} \frac{\det M(\Delta \mathbf{f}, \Delta \mathbf{g}_1, \dots, \Delta \mathbf{g}_n)}{\det M(\Delta \mathbf{g}_1, \dots, \Delta \mathbf{g}_n)} \geq \frac{1}{\tau} \Delta \mathbf{f} \cdot \Delta \mathbf{f} \tag{54}$$

This last inequality may be interpreted as a “time-entropy” uncertainty relation [30] in the following sense

$$\tau_S \sqrt{\Delta \mathbf{f} \cdot \Delta \mathbf{f}} \geq \tau \tag{55}$$

where the characteristic time of entropy change is defined by the relation

$$\frac{1}{\tau_S} = \frac{dS/dt}{\sqrt{\Delta \mathbf{f} \cdot \Delta \mathbf{f}}} \tag{56}$$

We are now ready to consider the time dependence of the properties  $Y_k$  and the entropy  $F$  as induced by the Equation 25 for the probability distribution, i.e., by the rate equation

$$\dot{\mathbf{x}} = \tau^{-1} \mathbf{f}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} = \tau^{-1} \sum_m \lambda_m \mathbf{y}_{m \perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} \tag{57}$$

where we made use of Equation 48 for  $\mathbf{f}$ .

The rates of change of properties  $Y_k$  are then given by

$$\dot{Y}_k = \mathbf{y}_k \cdot \dot{\mathbf{x}} = \dot{\mathbf{y}}_k \cdot \mathbf{x} = \sum_m L_{km} \lambda_m \tag{58}$$

The rate of entropy production by

$$\dot{S} = \dot{F} = \tau \dot{\mathbf{x}} \cdot \dot{\mathbf{x}} = \tau^{-1} \sum_k \sum_m \lambda_k L_{km} \lambda_m \tag{59}$$

where we define the functionals

$$L_{km} = \tau^{-1} \mathbf{y}_k \cdot \mathbf{y}_{m \perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} \tag{60}$$

Relation 58 shows that the rates of change (or generalized “fluxes”)  $\dot{Y}_k$  and the affinities (or generalized “forces”)  $\lambda_m$  are linearly interrelated by the coefficients (or generalized “conductivities”)  $L_{km}$ . The linear dependence of each rate from every affinity, is a fundamental physical assumption in the so-called “linear” theory of irreversible processes, a well-known theory [31–39] that deeply entangles physics and engineering by regularizing and unifying a wealth of experimental evidence of “direct” and “cross” effects. Direct effects (like Fick’s law of mass diffusion; Fourier’s law of heat conduction; Navier’s law of momentum diffusion; Ohm’s law of charge diffusion; and some aspects of chemical kinetics) arise from the linear relation between the rate of change (or the local flux vector) of one property and *its* conjugated affinity (or the spatial gradient of *its* conjugated driving force), such as between the heat flux vector and the temperature gradient in Fourier’s law. Cross effects (like Dufour’s, Soret’s and other thermodiffusion effects; Knudsen’s and other thermo-osmotic effects; electro-osmotic and streaming electro-kinetic effects; Seebeck’s, Peltier’s and other thermoelectric effects) arise from the linear relation between the rate of change (or the local flux vector) of one property and the conjugated affinity (or the spatial gradient of the conjugated driving force) of *another* property, such as between the heat flux vector and the species concentration gradient in Dufour’s thermodiffusion law.

Using Equation 95 (Appendix) with  $\mathbf{a} = \mathbf{y}_k$  and  $\mathbf{b} = \mathbf{y}_m$  we obtain

$$L_{km} = \tau^{-1} \mathbf{y}_{k \perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} \cdot \mathbf{y}_{m \perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} \tag{61}$$

which shows that the conductivities  $L_{km}$  satisfy the reciprocity relations

$$L_{km} = L_{mk} \tag{62}$$

Using Equation 91 (Appendix) we also find the following explicit expression for the generalized conductivities  $L_{km}$ ,

$$L_{km} = \frac{1}{\tau} \frac{\begin{vmatrix} \mathbf{y}_k \cdot \mathbf{y}_m & \mathbf{y}_k \cdot \mathbf{h}_1 & \cdots & \mathbf{y}_k \cdot \mathbf{h}_r \\ \mathbf{y}_m \cdot \mathbf{h}_1 & \mathbf{h}_1 \cdot \mathbf{h}_1 & \cdots & \mathbf{h}_r \cdot \mathbf{h}_1 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_m \cdot \mathbf{h}_r & \mathbf{h}_1 \cdot \mathbf{h}_r & \cdots & \mathbf{h}_r \cdot \mathbf{h}_r \end{vmatrix}}{\begin{vmatrix} \mathbf{h}_1 \cdot \mathbf{h}_1 & \cdots & \mathbf{h}_r \cdot \mathbf{h}_1 \\ \vdots & \ddots & \vdots \\ \mathbf{h}_1 \cdot \mathbf{h}_r & \cdots & \mathbf{h}_r \cdot \mathbf{h}_r \end{vmatrix}} \tag{63}$$

where vectors  $\mathbf{h}_1, \dots, \mathbf{h}_r$  may be any linearly independent set spanning the manifold  $L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)$ .

Moreover, it follows from Relation 60 that the matrix of generalized conductivities

$$[\mathbf{L}] = \begin{bmatrix} L_{00} & L_{01} & \cdots & L_{0m} & \cdots \\ L_{10} & L_{11} & \cdots & L_{1m} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ L_{k0} & L_{k1} & \cdots & L_{km} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{bmatrix} \tag{64}$$

is a Gram matrix and as such it is nonnegative definite.

Notice that the above proof of Onsager reciprocity requires the steepest entropy ascent assumption. The “kinematic reversibility” feature expressed by Eqs. 42, which we noted is intrinsic in our gradient formalism, is in the background, but it is not sufficient to entail Eqs. 62, nor does it appear to play any role in our proof. On the other hand, other than the steepest entropy ascent assumption, the derivation of Eqs. 62 required neither any additional dynamical assumptions, such as other forms of “microscopic reversibility”, nor any *ad hoc* approximations, such as linearizations that would limit their validity only to the near equilibrium regime. Therefore, our result holds in the entire domain, i.e., even in the arbitrarily far nonequilibrium regime.

The literature on Onsager reciprocity for physical systems originates from the work of Onsager [31], Casimir [32], Callen [33], Kubo [34], and others [35–39]. Onsager’s reciprocal relations are important universal features of irreversible processes. In different ways, the cited pioneers have shown that the reciprocal relations can only be derived by complementing the known (reversible) dynamical principles with some reasonable additional principle (e.g., the so-called “principle of microscopic reversibility”), assumption (e.g., that of sufficiently small perturbations, to ensure a linear behavior of the system), or approximation (e.g., that of sufficiently small deviations from thermodynamic equilibrium; see e.g. [40] for a derivation in the near-chemical-equilibrium chemical kinetics domain). Levine [7] has also shown that the reciprocal relations can be derived and extended to a higher nonequilibrium domain by adopting the assumption (or approximation) that systems proceed along a path continuously maximizing the entropy functional subject to some (possibly unknown) set of linear time-dependent constraints. This approach is closely related to the cited “rate controlled constrained equilibrium method”.

Matrix  $[L]$  is positive definite only if the vectors  $y_k \perp L(g_0, g_1, \dots, g_n)$  are all linearly independent, in which case the set of Equations 58 may be solved to yield

$$\lambda_m = \sum_k ([L]^{-1})_{mk} \dot{Y}_k \tag{65}$$

Therefore, in this case ( $[L]$  positive definite), the rate of entropy generation (Equation 28) may be rewritten in the following several forms

$$\dot{S} = \dot{F} = \sum_k \lambda_k y_k \cdot \dot{x} = \sum_k \lambda_k \dot{Y}_k = \tau \dot{x} \cdot \dot{x} = \sum_k \sum_m \lambda_k L_{km} \lambda_m = \sum_k \sum_m \dot{Y}_m ([L]^{-1})_{mk} \dot{Y}_k \tag{66}$$

In other words, the rate of entropy generation is a quadratic form in the affinities or generalized forces, it is also a quadratic form in the rates of change of the complete set of base observables, and it is the sum of the products of each such rate of change and its conjugated affinity. We emphasize again that each element of the generalized conductivity matrix  $L$  and of its inverse, the generalized resistance matrix  $L^{-1}$ , are nonlinear functional of the probability distribution only, not of the square-root probabilities, which therefore have played only an auxiliary role in the derivation.

Finally, comparing Relation 43, for the codispersion  $\Delta Y_k \Delta Y_m$  of properties  $Y_k$  and  $Y_m$ , and Relation 63, for the generalized conductivities  $L_{km}$ , we see that there is a relation between  $L_{km}$ ,  $\Delta Y_k \Delta Y_m$  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  $x$ , we further restrict the choice of the complete set of linearly independent property functionals  $Y_k$  so that  $y_0 = h_1, y_1 = h_2, \dots, y_{r-1} = h_r$  where  $h_1, \dots, h_r$  are

linearly independent vectors spanning the manifold  $L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)$  generated by the constraints and, moreover, we select the functionals  $Y_k$  for  $k > r$  so that the codispersions  $\Delta Y_k \Delta Y_0, \Delta Y_k \Delta Y_1, \dots, \Delta Y_k \Delta Y_{r-1}$  (Equation 43) are all equal to zero. For this particular choice, by studying Relation 63 for the generalized conductivities, we may readily verify that

$$L_{km} = 0 \quad \text{whenever } k < r \text{ or } m < r \tag{67}$$

$$L_{km} = 4\Delta Y_k \Delta Y_m / \tau \quad \text{for } k \geq r \text{ and } m \geq r \tag{68}$$

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

### 10. Time-Dependent Constraints

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

$$\dot{\mathbf{x}} = \dot{G}_k \frac{\mathbf{g}^{k \perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k-1}, \mathbf{g}_{k+1}, \dots, \mathbf{g}_n)}}{\mathbf{g}^{k \perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k-1}, \mathbf{g}_{k+1}, \dots, \mathbf{g}_n)} \cdot \mathbf{g}^{k \perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k-1}, \mathbf{g}_{k+1}, \dots, \mathbf{g}_n)}} \tag{69}$$

where  $\dot{G}_k = \alpha_k(t, \mathbf{p}(t))$  are given functions of time and of the instantaneous probability distribution. Clearly, the rates of change cannot be completely arbitrary, as Equation 69 is undefined unless  $\dot{G}_k$  is a function of and goes to zero faster than  $\mathbf{g}^{k \perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k-1}, \mathbf{g}_{k+1}, \dots, \mathbf{g}_n)} \cdot \mathbf{g}^{k \perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k-1}, \mathbf{g}_{k+1}, \dots, \mathbf{g}_n)}$ . Once the rate  $\dot{G}_k$  is specified, the probability distribution changes along the steepest  $G_k$  ascent (or descent, if  $\dot{G}_k$  is negative) compatible with maintaining invariant the other constants of the motion.

Again, we also readily verify that Equation 69 induces an equation for  $\dot{\mathbf{p}}$  which contains only the square  $x_i^2$  of the new variables and, therefore, is of the form

$$\dot{\mathbf{p}} = \mathbf{R}_k(t, \mathbf{p}) \tag{70}$$

We may readily verify that Equation 69 induces an evolution of the probability distribution  $\mathbf{p}$  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  $\dot{G}_k(t, \mathbf{p}(t))$ .

Geometrically, we could visualize the effect of Equation 69 as follows. We consider the surface defined by  $G_m(\mathbf{x}) = \langle A_m \rangle$  for  $m = 0, 1, \dots, k-1, k+1, \dots, n$  where  $\langle A_m \rangle$  are the magnitudes of the constraints (except the  $k$ -th) fixed by the initial distribution. On this surface, we can identify contour lines along which the  $k$ -th constraint is constant, generated by intersecting the surface just defined with the surface  $G_k(\mathbf{x}) = \langle A_k \rangle$  where  $\langle A_k \rangle$  varies over a feasible range of values. Every trajectory  $\mathbf{p}(t)$  generated by Equation 69 lies on the surface 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 all other constraints. In this sense, Equation 69 determines the minimal change in  $\mathbf{x}$  that is necessary in order to change the  $k$ -th constraint at the specified rate  $\dot{G}_k$ .

With this equation, however, the rate of entropy change does not have a definite sign. So, we finally note that by substituting  $L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k-1}, \mathbf{g}_{k+1}, \dots, \mathbf{g}_n)$  in Equation 69 with

$$L(\mathbf{f}, \mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k-1}, \mathbf{g}_{k+1}, \dots, \mathbf{g}_n) \tag{71}$$

we obtain a rate equation causing the  $k$ -th constraints to follow the specified rate  $\dot{G}_k$  while maintaining a zero change for the other constraints, as well as a zero rate of change of the entropy. In other words, this would describe an isoentropic change of the magnitude of the  $k$ -th constraints. We therefore denote by

$$\dot{\mathbf{p}} = \mathbf{R}_k^s(t, \mathbf{p}) \tag{72}$$

the differential equation for the probabilities generated by the following differential equation for the square-root probabilities

$$\dot{\mathbf{x}} = \dot{G}_k \frac{\mathbf{g}^{k \perp L}(\mathbf{f}, \mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k-1}, \mathbf{g}_{k+1}, \dots, \mathbf{g}_n)}{\mathbf{g}^{k \perp L}(\mathbf{f}, \mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k-1}, \mathbf{g}_{k+1}, \dots, \mathbf{g}_n) \cdot \mathbf{g}^{k \perp L}(\mathbf{f}, \mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k-1}, \mathbf{g}_{k+1}, \dots, \mathbf{g}_n)} \tag{73}$$

This form obtains the specified rate  $\dot{G}_k$  while the probability distribution changes along the steepest  $G_k$  ascent (or descent, if  $\dot{G}_k$  is negative) compatible with maintaining invariant all other constants of the motion and the entropy.

Clearly, if we want two or more constraints to have a specified rate of change, then Equation 10,

$$\frac{d\mathbf{p}}{dt} = \mathbf{R}_1^s(t, \mathbf{p}) + \dots + \mathbf{R}_n^s(t, \mathbf{p}) + \mathbf{D}(\mathbf{p}) \tag{74}$$

yields many orthogonal contributions to  $\dot{\mathbf{x}}$ . The terms  $\mathbf{R}_1^s, \dots, \mathbf{R}_n^s$  (each with structure similar to that given by Equation 69) cause the shifting constraints to follow the specified rates of change  $\dot{G}_1, \dots, \dot{G}_n$ . The term  $\mathbf{D}$  (as given by Equation 25) gives instead a non-negative definite contribution to the rate of entropy change and tends to attract the distribution  $\mathbf{x}$  towards a path of steepest entropy ascent compatible with the instantaneous values of the constraints.

### 11. Two Thermodynamic Examples

A less general but more clear illustration of the above formalism obtains when we assume that in addition to the normalization constraint  $G_0$  there is only one constraint  $G_1$ , the energy.

#### 11.1. Modeling the Extraction of the Adiabatic Availability of a System in a Non-Equilibrium State

For example, suppose the task is to extract the adiabatic availability of the system (see Ref. [1] for the precise definition of this fundamental thermodynamic concept; briefly: the maximum amount of energy that can be extracted from the system in the form of mechanical work without leaving any other net effects in its environment). Then the rate equation

$$\dot{\mathbf{x}} = \dot{G}_1 \frac{\mathbf{g}^{1 \perp L}(\mathbf{f}, \mathbf{g}_0)}{\mathbf{g}^{1 \perp L}(\mathbf{f}, \mathbf{g}_0) \cdot \mathbf{g}^{1 \perp L}(\mathbf{f}, \mathbf{g}_0)} \tag{75}$$

tells us, for a given non-equilibrium distribution  $\mathbf{x}$  of square-root probabilities (in general representing some non-equilibrium state), the rate at which we must change each probability in the distribution in order to reduce the energy  $\langle G_1 \rangle$  of the system at a given rate  $\dot{G}_1$  while maintaining the entropy constant. Of course, once we reach the minimum energy compatible with the initial entropy, the state is a canonically distributed stable equilibrium state (i.e., we reached a maximum entropy distribution for the current energy) and  $\mathbf{g}^{1 \perp L}(\mathbf{f}, \mathbf{g}_0) = 0$ , no more energy can be extracted isoentropically.

Recall that the explicit expression of  $\mathbf{g}^{1 \perp L}(\mathbf{f}, \mathbf{g}_0)$  can be easily obtained using the notation developed in the Appendix.

11.2. Modeling a Non-Work Interaction between a System in a Non-Equilibrium State and a Thermal Reservoir at Temperature  $T_R$

Also pertaining to thermodynamics is the modeling of a non-work interaction (again see Ref. [1] for rigorous definitions), i.e., an interaction where in addition to energy exchange between the interacting systems, there is also entropy exchange. Suppose further that the exchange occurs between our system in a non-equilibrium state described by an arbitrary probability distribution, and a thermal reservoir at temperature  $T_R$ . Then, the ratio of the energy to the entropy exchanged is equal to  $T_R$ . So, the energy and entropy balance equations in this case are (recall that in our notation here, energy is  $G_1$ )

$$\dot{G}_1 = \dot{G}_1^{\leftarrow} \tag{76}$$

and

$$\dot{S} = \frac{\dot{G}_1^{\leftarrow}}{T_R} + \dot{S}_{\text{gen}} \tag{77}$$

where by  $\dot{G}_1^{\leftarrow}$  and  $\dot{G}_1^{\leftarrow}/T_R$  we denote the rates at which respectively energy and entropy are exchanged via the heat interaction (assumed positive if into the system, negative if out of the system).

Then the rate equation

$$\dot{\mathbf{x}} = \dot{G}_1 \frac{\mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0)}{\mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0) \cdot \mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0)} + \left( \frac{\dot{G}_1}{T_R} + \dot{S}_{\text{gen}} \right) \frac{\mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1)}{\mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1) \cdot \mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1)} \tag{78}$$

tells us, for any given non-equilibrium distribution  $\mathbf{x}$  of square-root probabilities (in general representing some non-equilibrium state), the rate at which we must change each probability in the distribution in order to obtain the given energy and entropy exchanges, as well as the given amount of entropy generation by irreversibility. We can even set  $\dot{S}_{\text{gen}} = 0$  and see therefore what the changes would be if the non-work interaction occurs for example much faster than the typical time scale of entropy generation.

Finally, because form 78 may be somewhat misleading if one forgets the conditions mentioned above on the dependences of  $\dot{G}_1$  and  $\dot{S}$  near a canonical distribution, we prefer to rewrite it in terms of characteristic times of energy exchange

$$\dot{G}_1 = \frac{\text{sign}(\dot{G}_1^{\leftarrow})}{\tau_{\dot{G}_1^{\leftarrow}}} \mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0) \cdot \mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0) \tag{79}$$

and of entropy generation by irreversibility

$$\dot{S}_{\text{gen}} = \frac{1}{\tau_{\dot{S}_{\text{gen}}}} \mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1) \cdot \mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1) \tag{80}$$

so that the rate equation takes the form

$$\dot{\mathbf{x}} = \frac{\text{sign}(\dot{G}_1^{\leftarrow})}{\tau_{\dot{G}_1^{\leftarrow}}} \left( \mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0) + \frac{1}{T_R} \frac{\mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0) \cdot \mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0)}{\mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1) \cdot \mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1)} \mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1) \right) + \frac{1}{\tau_{\dot{S}_{\text{gen}}}} \mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1) \tag{81}$$

from which we see that the characteristic time of entropy exchange is related to that of energy exchange and to the temperature of the thermal reservoir as follows

$$\tau_{\dot{S}^{\leftarrow}} = \frac{\mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1) \cdot \mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1)}{\mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0) \cdot \mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0)} T_R \tau_{\dot{G}_1^{\leftarrow}} \tag{82}$$

or, equivalently, we may see that the following state functional of the system has to match the temperature of the reservoir

$$\frac{\mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0) \cdot \mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0)}{\mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1) \cdot \mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1)} = \frac{\tau_{\dot{S}^+}}{\tau_{\dot{G}_1^-}} \tag{83}$$

and, at least in this restricted sense, we could enforce an interpretation of this state functional as a “non-equilibrium temperature” of the system. The question of how and whether the concept of temperature should be extended to the nonequilibrium domain is currently a frontier topic in physics and engineering (see, e.g., [41]).

Notice that, using the explicit forms developed in the Appendix and in Section 9, we may rewrite the above expressions in terms of the ratio of energy to entropy fluctuations, by virtue the identity

$$\frac{\mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0) \cdot \mathbf{g}_{1\perp L}(\mathbf{f}, \mathbf{g}_0)}{\mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1) \cdot \mathbf{f}_{\perp L}(\mathbf{g}_0, \mathbf{g}_1)} = \frac{\Delta G_1 \Delta G_1}{\Delta F \Delta F} \tag{84}$$

## 12. Conclusions

The structure of the family of rate equations we discuss in this paper represents a flexible framework in which to cast non-equilibrium time-dependent 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 25 is more general than any equation assuming that the probability distribution always maximizes the entropy functional subject to some “possibly unknown” set of constraints [7, 8, 28].

Generalizations of the present formalism to non-extensive entropy functionals or other convex functionals of the probability distribution, as well as to nonlinear constraints, have been developed and will be presented elsewhere.

## Acknowledgements

I am personally deeply indebted to Dr. George N. Hatsopoulos for his support and his constant enthusiasm and inspiration in my continuous attempts to understand the fundamental physical roots of thermodynamics, and to develop methods for non-equilibrium thermodynamic analysis. Likewise, I am grateful to Prof. Elias P. Gyftopoulos for sharing with me his enthusiasm, passion, sometimes frustration, and certainly very long hours of hard joint work behind each dense page of Ref. [1]. Finally, I thank the MIT Mechanical Engineering Department for hosting me during the 2007 Fall term when this paper was prepared, and the Università di Brescia for supporting this visit.

## Appendix. Orthogonal Decomposition of a Vector with respect to a Linear Manifold

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

$$L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n) \tag{85}$$

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

$$\mathbf{b}_{L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} \tag{86}$$

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

$$\mathbf{g} \cdot \mathbf{b}_{L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} = \mathbf{g} \cdot \mathbf{b} \tag{87}$$

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

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

$$\mathbf{b}_{L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} = \sum_{k=1}^r \sum_{m=1}^r (\mathbf{b} \cdot \mathbf{h}_k) [M(\mathbf{h}_1, \dots, \mathbf{h}_r)^{-1}]_{km} \mathbf{h}_m \tag{88}$$

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

$$M(\mathbf{h}_1, \dots, \mathbf{h}_r) = \begin{bmatrix} \mathbf{h}_1 \cdot \mathbf{h}_1 & \cdots & \mathbf{h}_r \cdot \mathbf{h}_1 \\ \vdots & \ddots & \vdots \\ \mathbf{h}_1 \cdot \mathbf{h}_r & \cdots & \mathbf{h}_r \cdot \mathbf{h}_r \end{bmatrix} \tag{89}$$

The second expression is a ratio of two determinants

$$\mathbf{b}_{L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} = - \frac{\begin{vmatrix} 0 & \mathbf{h}_1 & \cdots & \mathbf{h}_r \\ \mathbf{b} \cdot \mathbf{h}_1 & \mathbf{h}_1 \cdot \mathbf{h}_1 & \cdots & \mathbf{h}_r \cdot \mathbf{h}_1 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b} \cdot \mathbf{h}_r & \mathbf{h}_1 \cdot \mathbf{h}_r & \cdots & \mathbf{h}_r \cdot \mathbf{h}_r \end{vmatrix}}{\begin{vmatrix} \mathbf{h}_1 \cdot \mathbf{h}_1 & \cdots & \mathbf{h}_r \cdot \mathbf{h}_1 \\ \vdots & \ddots & \vdots \\ \mathbf{h}_1 \cdot \mathbf{h}_r & \cdots & \mathbf{h}_r \cdot \mathbf{h}_r \end{vmatrix}} \tag{90}$$

where the determinant at the denominator, also given by  $\det M(\mathbf{h}_1, \dots, \mathbf{h}_r)$ , is always strictly positive because the vectors  $\mathbf{h}_1, \dots, \mathbf{h}_r$  are linearly independent.

In the paper, our rate equations are expressed in terms of vectors of the form

$$\mathbf{b}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} = \mathbf{b} - \mathbf{b}_{L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} = \frac{\begin{vmatrix} \mathbf{b} & \mathbf{h}_1 & \cdots & \mathbf{h}_r \\ \mathbf{b} \cdot \mathbf{h}_1 & \mathbf{h}_1 \cdot \mathbf{h}_1 & \cdots & \mathbf{h}_r \cdot \mathbf{h}_1 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b} \cdot \mathbf{h}_r & \mathbf{h}_1 \cdot \mathbf{h}_r & \cdots & \mathbf{h}_r \cdot \mathbf{h}_r \end{vmatrix}}{\begin{vmatrix} \mathbf{h}_1 \cdot \mathbf{h}_1 & \cdots & \mathbf{h}_r \cdot \mathbf{h}_1 \\ \vdots & \ddots & \vdots \\ \mathbf{h}_1 \cdot \mathbf{h}_r & \cdots & \mathbf{h}_r \cdot \mathbf{h}_r \end{vmatrix}} \tag{91}$$

where in writing Equation 91 we make use of Equation 90. The vector  $\mathbf{b}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)}$  is orthogonal to manifold  $L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)$ ; indeed, the vector represented by Equation 91 has the relevant property

$$\mathbf{g}_k \cdot \mathbf{b}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} = 0 \quad k = 0, 1, \dots, n \tag{92}$$

which follows directly from Relation 87, and hence the relation

$$\mathbf{b} = \mathbf{b}_{L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} + \mathbf{b}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} \tag{93}$$

represents the unique orthogonal decomposition of vector  $\mathbf{b}$  with respect to manifold  $L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)$ .

Moreover, we have the other obvious, but relevant properties

$$\mathbf{b} \cdot \mathbf{b}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} = \mathbf{b}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} \cdot \mathbf{b}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} \geq 0 \tag{94}$$

where the strict inequality applies whenever  $\mathbf{b}$  is not in  $L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)$ , and for any  $\mathbf{a}$  and  $\mathbf{b}$

$$\mathbf{a} \cdot \mathbf{b}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} = \mathbf{a}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} \cdot \mathbf{b}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} \geq 0 \tag{95}$$

An important formula which derives from Eq. 91 and the usual properties of determinants, is

$$\mathbf{b}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} \cdot \mathbf{b}_{\perp L(\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_n)} = \frac{\det M(\mathbf{b}, \mathbf{h}_1, \dots, \mathbf{h}_n)}{\det M(\mathbf{h}_1, \dots, \mathbf{h}_n)} \tag{96}$$

Moreover, choosing the set of linearly independent vectors  $\mathbf{h}_1, \dots, \mathbf{h}_n$  so that  $\mathbf{h}_n = 2\mathbf{x}$  is the gradient of the normalization constraint, and using the notation introduced in Section 9, i.e., defining the “mean” functionals  $H_j = \mathbf{h}_j \cdot \mathbf{h}_n/4$ ,  $B = \mathbf{b} \cdot \mathbf{h}_n/4$ , and the “deviation” vectors  $\Delta\mathbf{h}_j = (\mathbf{h}_j - \mathbf{h}_n H_j)/2$ ,  $\Delta\mathbf{b} = (\mathbf{b} - \mathbf{h}_n B)/2$ , it is easy to show that

$$\frac{\det M(\mathbf{b}, \mathbf{h}_1, \dots, \mathbf{h}_n)}{\det M(\mathbf{h}_1, \dots, \mathbf{h}_n)} = \frac{\det M(\Delta\mathbf{b}, \Delta\mathbf{h}_1, \dots, \Delta\mathbf{h}_{n-1})}{\det M(\Delta\mathbf{h}_1, \dots, \Delta\mathbf{h}_{n-1})} \tag{97}$$

As a final remark, we write the following generalized form of the Cauchy-Schwarz inequality

$$\det M(\Delta\mathbf{b}, \Delta\mathbf{h}_1, \dots, \Delta\mathbf{h}_{n-1}) \leq \det M(\Delta\mathbf{h}_1, \dots, \Delta\mathbf{h}_{n-1}) \Delta\mathbf{b} \cdot \Delta\mathbf{b} \tag{98}$$

In the paper, we make extensive use of the notation and relations discussed in this appendix [24, 25].

### References

1. Gyftopoulos, E.P.; Beretta, G.P., *Thermodynamics. Foundations and Applications*; Dover: Mineola, 2005.
2. Zanchini, E.; Beretta, G.P., Rigorous Axiomatic Definition of Entropy Valid Also For Non-Equilibrium States. In *Meeting the Entropy Challenge*; Beretta, G.P.; Ghoniem A.; Hatsopoulos, G.N., Ed.; AIP Book, CP1033, in press (2008).
3. Clausius, R., Über Verschiedene für die Anwendungen Bequeme Formen der Hauptgleichungen der Mechanischen Wärmetheorie. Memoir read at the Philos. Soc. Zürich on April 24; *Pogg. Ann.*, **1865**, 125, 353-400. English translation in: *The Second Law of Thermodynamics*; Kestin, J. Ed.; Dowden, Hutchinson and Ros: Stroudsburg, 1976. pp. 162193.

4. von Neumann, J., Beweiss des Ergodensatzes und des H-Theorems in der neuen Mechanik. *Zeitschrift fr Physik*, **1929**, 57, 30-70.
5. Shannon, C.E., The Mathematical Theory of Communication. *Bell System Technical J.*, **1948**, 27, 379423; and **1948**, 27, 623656.
6. Margenau, H., *The Nature of Physical Reality*; McGraw-Hill, 1950.
7. Levine, R.D.; Tribus, M., Ed., *The Maximum Entropy Formalism*; The M.I.T. Press: Cambridge, Mass., 1979.
8. Alhassid, Y.; Agmon, N.; Levine, R.D., Collision Experiments with Partial Resolution of Final States: Maximum Entropy Procedure and Surprisal Analysis. *Phys. Rev. C*, **1978**, 79, 1775-1788.
9. Levine, R.D., Entropy and Macroscopic Disequilibrium. II. The Information Theoretic Characterization of Markovian Relaxation Processes. *J. Chem. Phys.* **1976**, 65, 3302-3315.
10. Keck, J.C., Rate-Controlled Constrained-Equilibrium Method for Treating Reactions in Complex Systems. In *The Maximum Entropy Formalism*, Levine, R.D.; Tribus, M., Ed.; The M.I.T. Press: Cambridge, Mass., 1979. pp. 219-245.
11. Beretta, G.P.; Keck, J.C., The Constrained Equilibrium Approach to Nonequilibrium Dynamics. In *Computer-Aided Engineering of Energy Systems, Second Law Analysis and Modeling*, Gaggioli, R.A., Ed.; ASME Book H0341C-AES, 1986; Vol. 3, pp. 135-139.
12. Beretta, G.P.; On the General Equation of Motion of Quantum Thermodynamics and the Distinction between Quantal and Nonquantal Uncertainties, Sc.D. thesis, MIT, Cambridge, MA, 1981; arXiv:quant-ph/0509116.
13. Beretta, G.P.; Gyftopoulos, E.P.; Park, J.L.; Hatsopoulos, G.N., Quantum Thermodynamics. A New Equation of Motion for a Single Constituent of Matter. *Nuovo Cimento B*, **1984**, 82, 169-191.
14. Beretta, G.P.; Gyftopoulos, E.P.; Park, J.L., Quantum Thermodynamics. A New Equation of Motion for a General Quantum System. *Nuovo Cimento B*, **1985**, 87, 77-97. 77-97.
15. Beretta, G.P.; Steepest Entropy Ascent in Quantum Thermodynamics, in *The Physics of Phase Space*, Edited by Y.S. Kim and W.W. Zachary, Lecture Notes in Physics, Vol. 278, Springer-Verlag, pp. 441-443 (1986).
16. Beretta, G.P., A General Nonlinear Evolution Equation for Irreversible Conservative Approach to Stable Equilibrium. In *Frontiers of Nonequilibrium Statistical Physics*, Moore, G.T.; Scully, M.O., Ed.; Plenum Press: New York, 1986. pp 193-204.
17. Beretta, G.P., Quantum Thermodynamics of Nonequilibrium. Onsager Reciprocity and Dispersion-Dissipation Relations. *Found. Phys.*, **1987**, 17, 365-381.
18. Beretta, G.P., Nonlinear Model Dynamics for Closed-System, Constrained, Maximal-Entropy-Generation Relaxation by Energy Redistribution. *Phys. Rev. E*, **2006**, 73, 026113.
19. Hatsopoulos, G.N.; Gyftopoulos, E.P., A Unified Quantum Theory of Mechanics and Thermodynamics. *Found. Phys.*, **1976**, 6, 15-31; **1976**, 6, 127-141; **1976**, 6, 439-455; **1976**, 6, 561-570.
20. Maddox, J.; Uniting Mechanics and Statistics, *Nature*, **1985**, 316, 11.
21. Hatsopoulos G.N.; Keenan, J.H.; *Principles of General Thermodynamics*, Wiley, 1965.
22. S. Gheorghiu-Svirschevski, Nonlinear Quantum Evolution with Maximal Entropy Production, *Phys. Rev. A*, **2001**, 63, 022105; and Addendum, *Phys. Rev. A*, **2001**, 63, 054102.

23. Lemanska, M.; Jaeger, Z.; A Nonlinear Model for Relaxation in Excited Closed Physical Systems, *Physica D*, **2002**, *170*, 72-86.
24. Beretta, G.P.; A New Approach to Constrained-Maximization Nonequilibrium Problems, in *Computer-Aided Engineering of Energy Systems: Second Law Analysis and Modeling*, Edited by R.A. Gaggioli, ASME Book H0341C-AES, Vol. 3, pp. 129-134 (1986).
25. Beretta, G.P.; Dynamics of Smooth Constrained Approach to Maximum Entropy, in *Second Law Analysis of Thermal Systems*, Edited by M.J. Moran and E. Sciubba, ASME Book I00236, pp. 17-24 (1987).
26. Caticha, A.; Change, Time and Information Geometry, *AIP Conf. Proc.*, **2001**, *568*, 72.
27. Wootters, W.K.; Statistical Distance and Hilbert Space, *Phys. Rev. D*, **1981**, *23*, 357-362.
28. Salamon, P.; Nulton, J.D.; Berry, R.S., Length in Statistical Thermodynamics, *J. Chem. Phys.*, **1985**, *82*, 2433-2436.
29. Braunstein, S.L.; Caves, C.M.; Statistical Distance and the Geometry of Quantum States, *Phys. Rev. Lett.*, **1994**, *72*, 3439-3443.
30. Beretta, G.P.; Time-Energy and Time-Entropy Uncertainty Relations in Dissipative Quantum Systems, arXiv:quant-ph/0511091 (2005).
31. Onsager, L.; Reciprocal Relations in Irreversible Processes I, *Phys. Rev.*, **1931**, *37*, 405; Reciprocal Relations in Irreversible Processes II, *38*, 2265 (1931).
32. Casimir, H.B.G.; On Onsager's Principle of Microscopic Reversibility, *Rev. Mod. Phys.*, **1945**, *17*, 343.
33. Callen, H.B.; Phenomenological Theory of Irreversible Processes, Thesis, MIT, **1947**, unpublished.
34. Kubo, R.; Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems, *J. Phys. Soc. Jpn.*, **1957**, *12*, 570-586.
35. Callen, H.B.; Barasch, M.L.; Jackson, L.L.; Statistical Mechanics of Irreversibility, *Phys. Rev.*, **1952**, *88*, 1382.
36. Callen, H.B.; Welton, T.A.; Irreversibility and Generalized Noise, *Phys. Rev.*, **1951**, *83*, 34; Greene, R.F.; Callen, H.B.; On the Formalism of Thermodynamic Fluctuation Theory, *Phys. Rev.*, **1951**, *83*, 1231; Callen, H.B.; Greene, R.F.; On a Theorem of Irreversible Thermodynamics, *Phys. Rev.*, **1952**, *86*, 702; Greene, R.F.; Callen, H.B.; On a Theorem of Irreversible Thermodynamics. II, *Phys. Rev.*, **1952**, *88*, 1387.
37. Onsager, L.; Machlup, S.; Fluctuations and Irreversible Processes, *Phys. Rev.*, **1953**, *91*, 1505.
38. Prigogine, I.; *Introduction to Thermodynamics of Irreversible Processes*; C.C. Thomas, Springfield, Illinois, 1965.
39. De Groot, S.R.; Mazur, P.; *Non-Equilibrium Thermodynamics*; Dover, Mineola, 1984.
40. Beretta, G.P.; Gyftopoulos, E.P.; Thermodynamic Derivations of Conditions for Chemical Equilibrium and of Onsager Reciprocal Relations for Chemical Reactors, *J. Chem. Phys.*, **2004**, *121*, 2718-2728.
41. Casas-Vazquez J.; Jou, D.; Temperature in Non-equilibrium States: a Review of Open Problems and Current Proposals, *Rep. Progr. Phys.*, **2003**, *66*, 1937.

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).