LETTER TO THE EDITOR

Maximum entropy production and the fluctuation theorem

R C Dewar

Unité EPHYSE, INRA Centre de Bordeaux-Aquitaine, BP 81, 33883 Villenave d'Ornon Cedex, France

E-mail: dewar@bordeaux.inra.fr

Abstract
Recently the author used an information theoretical formulation of non-equilibrium statistical mechanics (MaxEnt) to derive the fluctuation theorem (FT) concerning the probability of second law violating phase-space paths. A less rigorous argument leading to the variational principle of maximum entropy production (MEP) was also given. Here a more rigorous and general mathematical derivation of MEP from MaxEnt is presented, and the relationship between MEP and the FT is thereby clarified. Specifically, it is shown that the FT allows a general orthogonality property of maximum information entropy to be extended to entropy production itself, from which MEP then follows. The new derivation highlights MEP and the FT as generic properties of MaxEnt probability distributions involving anti-symmetric constraints, independently of any physical interpretation. Physically, MEP applies to the entropy production of those macroscopic fluxes that are free to vary under the imposed constraints, and corresponds to selection of the most probable macroscopic flux configuration. In special cases MaxEnt also leads to various upper bound transport principles. The relationship between MaxEnt and previous theories of irreversible processes due to Onsager, Prigogine and Ziegler is also clarified in the light of these results.

PACS numbers: 05.70.Ln, 65.40.Gr, 89.70.+c

The constrained maximization of Shannon information entropy (MaxEnt) is an algorithm for constructing probability distributions from partial information, which resides naturally within the framework of Bayesian probability theory [1]. It is a statistical inference tool of considerable generality. In particular, Jaynes [2] has proposed MaxEnt as a universal method for constructing the microscopic probability distributions of equilibrium and non-equilibrium statistical mechanics. Recently the author applied MaxEnt to the information entropy of microscopic phase-space paths, and derived three generic non-equilibrium properties from the resulting MaxEnt path distribution: the fluctuation theorem (FT), maximum entropy production

(MEP) and self-organized criticality for slowly flux-driven systems [3]. In this letter we present a more rigorous and general mathematical derivation of MEP from MaxEnt which clarifies the relationship between MEP and the FT. This new derivation of MEP also sheds light on the relationship between MaxEnt and previous theories of irreversible processes due to Onsager [4], Prigogine [5] and Ziegler [6].

Let us begin by recalling some general mathematical properties of MaxEnt distributions which hold irrespective of any physical assumptions [1]. The purpose here is to show that MEP and the FT are generic properties of a certain class of MaxEnt distributions. The physical interpretation of MaxEnt and MEP will be addressed subsequently. We consider a situation with n potential outcomes and $m < n$ constraints in the form of known values F_k of certain functions f_k ($1 \leqslant k \leqslant m$). We wish to assign Bayesian probabilities p_i ($1 \leqslant i \leqslant n$) so as to quantify our partial information about the outcomes. In the MaxEnt approach we choose the distribution p_i that maximizes the Shannon information entropy

$$H = -\sum_{i=1}^{n} p_i \log p_i \tag{1}$$

subject to the constraints

$$\langle f_k \rangle \equiv \sum_{i=1}^{n} p_i f_k(i) = F_k, \qquad 1 \leqslant k \leqslant m \tag{2}$$

$$\sum_{i=1}^{n} p_i = 1. \tag{3}$$

This procedure gives the least-biased probability distribution consistent with the available information (i.e. the m constraints, the n outcomes and the mn values $f_k(i)$) because, by construction, p_i contains this information alone. Equation (2) does not mean that the constraints F_k must take the form of sample averages (although this case is not excluded); F_k can be any data whatsoever, including single measurements. Equation (2) simply means that p_i must agree with the available data in the sense that F_k is recoverable from p_i as the expectation value $\langle f_k \rangle$, the rationale being that $\langle f_k \rangle$ is the estimator of f_k with minimum expected square error [7]. Implicit here is the Bayesian interpretation of p_i as a measure of our state of knowledge about the real world rather than an inherent property of the real world (a sampling frequency) [1].

By introducing the vector of Lagrange multipliers $(\lambda_1, \ldots, \lambda_m) \equiv \boldsymbol{\lambda}$ and defining the partition function

$$Z(\boldsymbol{\lambda}) \equiv \sum_{i=1}^{n} \exp \left\{ \sum_{k=1}^{m} \lambda_k f_k(i) \right\} \tag{4}$$

the MaxEnt distribution is obtained as

$$p_i = \frac{1}{Z(\boldsymbol{\lambda})} \exp \left\{ \sum_{k=1}^{m} \lambda_k f_k(i) \right\} \equiv p(i \mid \boldsymbol{F}C) \tag{5}$$

where the notation $p(i \mid \boldsymbol{F}C)$ (the probability of i given $\boldsymbol{F}$ and C) reminds us that p_i is conditional on the available information consisting of the constraint vector $\boldsymbol{F} \equiv (F_1, \ldots, F_m)$ together with the prior information C (e.g. microscopic physics in the case of statistical mechanics [8]) that determines the set of allowed outcomes and corresponding function values $f_k(i)$. The Lagrange multipliers $\boldsymbol{\lambda}$ may be expressed in terms of the constraints $\boldsymbol{F}$ by solving the relations

$$F_k = \langle f_k \rangle = \frac{\partial \log Z(\boldsymbol{\lambda})}{\partial \lambda_k}, \qquad 1 \leqslant k \leqslant m, \tag{6}$$

leading directly to the reciprocity relations and fluctuation formulae

$$B_{jk} \equiv \frac{\partial F_k}{\partial \lambda_j} = \frac{\partial F_j}{\partial \lambda_k} = \frac{\partial^2 \log Z(\boldsymbol{\lambda})}{\partial \lambda_j \partial \lambda_k} = \langle f_j f_k \rangle - \langle f_j \rangle \langle f_k \rangle, \qquad 1 \leqslant j, k \leqslant m. \tag{7}$$

The maximum value of the information entropy is

$$H_{\max} = \log Z(\boldsymbol{\lambda}) - \sum_{k=1}^{m} \lambda_k F_k. \tag{8}$$

The solution of equation (6) for $\boldsymbol{\lambda}$, when substituted into equation (8), gives $H_{\max}$ as a function of the constraint vector $\boldsymbol{F}$ alone, a function which will be denoted by $S(\boldsymbol{F})$. In order to keep the notation simple in the following analysis, we temporarily suppress the dependence of S on the prior information C, although it will be introduced later when required (equation (23)). Equation (8) describes a Legendre transformation between $S(\boldsymbol{F})$ and $\log Z(\boldsymbol{\lambda})$. By considering small changes in the constraints $\boldsymbol{F}$, it is straightforward to show that

$$\frac{\partial S(\boldsymbol{F})}{\partial F_k} = -\lambda_k, \qquad 1 \leqslant k \leqslant m. \tag{9}$$

Equation (9) is a general orthogonality property of MaxEnt distributions. It states that the vector of Lagrange multipliers $\boldsymbol{\lambda}$ is everywhere locally normal to the contours of S in F-space (and points in the direction of decreasing S). Thus $\boldsymbol{\lambda}$ specifies not only the MaxEnt distribution corresponding to a given constraint vector $\boldsymbol{F}$, but also how S will vary under small changes in the constraints in the neighbourhood of $\boldsymbol{F}$. Further differentiation of equation (9) gives the reciprocity relations in terms of S as

$$A_{jk} \equiv \frac{\partial \lambda_k}{\partial F_j} = \frac{\partial \lambda_j}{\partial F_k} = -\frac{\partial^2 S(\boldsymbol{F})}{\partial F_j \partial F_k}, \qquad 1 \leqslant j, \quad k \leqslant m. \tag{10}$$

A and B are symmetric, positive definite matrices. Furthermore A (the entropy curvature matrix) is the inverse of B (the covariance matrix), provided we are dealing with linearly independent constraints (i.e. $\partial F_j / \partial F_k = \delta_{jk}$).

Next we consider those situations in which the potential outcomes can be grouped in pairs $(i+, i-)$ with respect to which each function f_k is anti-symmetric,

$$f_k(i-) = -f_k(i+), \qquad 1 \leqslant k \leqslant m. \tag{11}$$

Then the MaxEnt distribution satisfies the generic 'fluctuation theorem' (FT)

$$\frac{p_{i+}}{p_{i-}} = \exp \left\{ 2 \sum_{k=1}^{m} \lambda_k f_k(i+) \right\}. \tag{12}$$

A variety of relationships of this general form have been derived for stationary and non-stationary physical systems, in terms of the ratio of probabilities of pairs of microscopic phase-space trajectories related by path reversal [9]. It was recognized in [10] that a common explanation for these relationships lies in the hypothesis that the trajectories have a Gibbs-type probability distribution. MaxEnt provides the natural formalism in which Gibbs-type distributions emerge, whether or not they refer to physical systems. Thus the FT is not confined to physical systems alone but arises in a (potentially large) class of statistical inference problems involving constraint functions which are anti-symmetric (*sensu* equation (11)).

The generic FT (equation (12)) has important implications for the functional form of the relationship between $\boldsymbol{\lambda}$ and $\boldsymbol{F}$. This may be seen explicitly by expressing the FT in terms of $p(\boldsymbol{f})$, the p.d.f. of $\boldsymbol{f} \equiv (f_1, \ldots, f_m)$,

$$\frac{p(\boldsymbol{f})}{p(-\boldsymbol{f})} = \exp \left(2 \sum_{k=1}^{m} \lambda_k f_k \right) \tag{13}$$

and invoking the quadratic (steepest descent) approximation to $p(\boldsymbol{f})$ around $\boldsymbol{f} = \boldsymbol{F}$,

$$p(\boldsymbol{f}) \propto \exp\left\{-\frac{1}{2}\sum_{j,k=1}^{m}(f_j - F_j)A_{jk}(f_k - F_k)\right\}. \tag{14}$$

It is easily verified that equations (13) and (14) are satisfied simultaneously for all $\boldsymbol{f}$ if and only if

$$\lambda_k = \sum_{j=1}^{m}A_{jk}(\boldsymbol{F})F_j \tag{15}$$

which, by inversion, implies

$$F_k = \sum_{j=1}^{m}B_{jk}(\boldsymbol{\lambda})\lambda_j. \tag{16}$$

Here we emphasize that, in general, the 'constitutive relations' between $\boldsymbol{\lambda}$ and $\boldsymbol{F}$ described by equations (15) and (16) are nonlinear due to the dependences, respectively, of A on $\boldsymbol{F}$ and B on $\boldsymbol{\lambda}$.

These consequences of the FT may be re-expressed as gradient properties of the 'dissipation function' (so-called for reasons that will become apparent in equation (27)), defined by

$$d \equiv 2\sum_{k=1}^{m}\lambda_k f_k. \tag{17}$$

Substitution of $A_{jk} = \partial\lambda_j/\partial F_k$ into equation (15) leads straightforwardly to an orthogonality condition in F-space satisfied by the mean dissipation $D \equiv \langle d \rangle$,

$$\frac{\partial D(\boldsymbol{F})}{\partial F_k} = 4\lambda_k, \qquad 1 \leqslant k \leqslant m. \tag{18}$$

Thus the FT extends the orthogonality property of S (equation (9)) to the mean dissipation function: the vector of Lagrange multipliers $\boldsymbol{\lambda}$ indicates the direction in F-space of steepest *descent* on the S-surface and of steepest *ascent* on the mean dissipation surface $D(\boldsymbol{F})$. In other words, the D-surface is essentially an inverted copy of the S-surface. Similarly, from $B_{jk} = \partial F_j/\partial\lambda_k$ and equation (16), we obtain the dual orthogonality condition in λ-space:

$$\frac{\partial D(\boldsymbol{\lambda})}{\partial\lambda_k} = 4F_k, \qquad 1 \leqslant k \leqslant m. \tag{19}$$

The constraint vector $\boldsymbol{F}$ indicates the direction in λ-space of steepest ascent on the dissipation surface $D(\boldsymbol{\lambda})$. Equations (15) and (16) also yield the relations

$$D = 2\sum_{j,k=1}^{m}A_{jk}(\boldsymbol{F})F_j F_k = 2\sum_{j,k=1}^{m}B_{jk}(\boldsymbol{\lambda})\lambda_j\lambda_k = \frac{1}{2}\Delta^2 \tag{20}$$

where $\Delta^2 \equiv \langle d^2 \rangle - D^2$ is the variance of d. As implied by the arguments of A and B, in general D is not a quadratic function of $\boldsymbol{F}$ or $\boldsymbol{\lambda}$.

As we now show, the orthogonality conditions (equations (18) and (19)) derived from the FT imply that D adopts its maximum value allowed under the imposed constraints. Let us consider equation (19) first. A given set of constraints defines a fixed vector $\boldsymbol{F}$ which, to be specific, we denote by $\boldsymbol{F}^*$. The MaxEnt value of the Lagrange multiplier vector, $\boldsymbol{\lambda}^*$, corresponding to the prescribed $\boldsymbol{F}^*$ must then satisfy the orthogonality condition

$\partial D(\boldsymbol{\lambda})/\partial\lambda_k|_{\lambda=\lambda*} = 4F_k^*$. This coincides with the value of $\boldsymbol{\lambda}$ that maximizes $D(\boldsymbol{\lambda})$ subject to the constraint

$$D(\boldsymbol{\lambda}) = 2\sum_{k=1}^{m}\lambda_k F_k^* \tag{21}$$

as may be verified by forming the Lagrangian function

$$\Psi(\boldsymbol{\lambda}) \equiv D(\boldsymbol{\lambda}) + \gamma\left\{D(\boldsymbol{\lambda}) - 2\sum_{k=1}^{m}\lambda_k F_k^*\right\} \tag{22}$$

in which γ is a Lagrange multiplier; setting $\partial\Psi(\boldsymbol{\lambda})/\partial\lambda_k|_{\lambda=\lambda*} = 0$ reproduces the required orthogonality condition for $\gamma = -2$. That this solution is a maximum follows from $\partial^2\Psi/\partial\lambda_j\partial\lambda_k = (1+\gamma)\partial^2 D/\partial\lambda_j\partial\lambda_k = -4B_{jk}$ and the positive definiteness of B. Geometrically, the equation $2\sum_{k=1}^{m}\lambda_k F_k^* = c$ defines a plane $\pi(c)$ in λ-space whose normal lies in the direction of the prescribed $\boldsymbol{F}^*$, while the equation $D(\boldsymbol{\lambda}) = c$ defines a contour of the mean dissipation function in λ-space. Equation (21) then implies that the values of c allowed by the constraints $\boldsymbol{F}^*$ are those for which the plane $\pi(c)$ intersects the contour $D(\boldsymbol{\lambda}) = c$. From among these allowed values of c the orthogonality condition selects the value c^* for which $\pi(c^*)$ intersects the contour $D(\boldsymbol{\lambda}) = c^*$ tangentially, at a single point $\boldsymbol{\lambda}^*$ corresponding to the maximum allowed value of c.

In contrast to this result, we note that $D(\boldsymbol{\lambda})$ has a *minimum* at $\boldsymbol{\lambda}^*$ with respect to variations in $\boldsymbol{\lambda}$ which are restricted to the plane $\pi(c^*)$, as may be verified by forming the Lagrangian function

$$\Phi(\boldsymbol{\lambda}) \equiv D(\boldsymbol{\lambda}) + \gamma\left\{2\sum_{k=1}^{m}\lambda_k F_k^* - c*\right\};$$

setting $\partial\Phi(\boldsymbol{\lambda})/\partial\lambda_k|_{\lambda=\lambda*} = 0$ gives the orthogonality condition at $\boldsymbol{\lambda}^*$ for $\gamma = -2$, and $\partial^2\Phi/\partial\lambda_j\partial\lambda_k = \partial^2 D/\partial\lambda_j\partial\lambda_k = 4B_{jk}$ implies that this stationary point is a minimum. Anticipating the physical interpretation of D as entropy production (see equation (28)), Prigogine's principle of minimum entropy production [5] may be seen as a special case of this latter type of constrained variation in which one or more components of $\boldsymbol{F}^*$ is zero (say, $F_k^* = 0$ for $1 \leqslant k \leqslant s < m$). Then within the plane $\pi(c^*)$ the complementary components of $\boldsymbol{\lambda}$ (i.e. λ_k for $s < k \leqslant m$) are fixed, and $D(\boldsymbol{\lambda})$ has a minimum with respect to variation in the free components of $\boldsymbol{\lambda}$ (i.e. λ_k for $1 \leqslant k \leqslant s$). This last statement is in fact somewhat more general than Prigogine's result, because it does not require the usual near-equilibrium assumption of linear constitutive relations.

In the MaxEnt distribution, the information concerning the constraints $\boldsymbol{F}$ is encoded in the vector of Lagrange multipliers $\boldsymbol{\lambda}$ via equation (6). It is then a matter of convenience whether we choose $\boldsymbol{F}$ or $\boldsymbol{\lambda}$ to describe the imposed constraints. By an argument identical to that leading to maximum dissipation in λ-space, the orthogonality condition in F-space (equation (18)) is equivalent to maximizing $D(\boldsymbol{F})$ subject to the constraint $D(\boldsymbol{F}) = 2\sum_{k=1}^{m}\lambda_k^* F_k$ in which $\boldsymbol{\lambda}^*$ is prescribed, the solution being a maximum on account of the positive definiteness of A.

We now consider the less restrictive problem in which the prior information C represents all of the available information, and F_k ($1 \leqslant k \leqslant m$) represent free (linearly independent) variables to be determined from C. Here C includes not only the prior information (e.g. microscopic physics) determining the set of allowed outcomes $1 \leqslant i \leqslant n$, but also the conditions of the problem (e.g. steady-state conditions, critical thresholds) determining the set of variables F_k ($1 \leqslant k \leqslant m$) that are linearly independent and their domain of allowed values in F-space. We wish to estimate the F_k from knowledge of C alone. A fundamental property of any such estimate, $\boldsymbol{F}_C$, is its redundancy in the sense that knowing both $\boldsymbol{F}_C$ and C cannot alter our uncertainty about the outcomes based on knowledge of C alone. We may express this

mathematically through the redundancy condition

$$S(\boldsymbol{F}_C C) = S(C) \qquad\qquad\qquad (23)$$

where $S(\boldsymbol{F}_C C)$ and $S(C)$ are the MaxEnt information entropies of the conditional probabilities $p(i \mid \boldsymbol{F}_C C)$ and $p(i \mid C)$ respectively. The distribution $p(i \mid \boldsymbol{F}_C C)$ is given by equation (5) with $\boldsymbol{F} = \boldsymbol{F}_C$, and we now indicate explicitly the dependence of the corresponding information entropy, $S(\boldsymbol{F}_C C)$, on the prior information C. As before, we consider the case where the functions f_k are anti-symmetric over the outcomes (*sensu* equation (11)). In that case a dissipation function $D(\boldsymbol{F})$ can be defined, as before, through $p(i \mid \boldsymbol{F} C)$. Suppose now that $D(\boldsymbol{F})$ has an upper bound $D_{\max}(C)$ within the domain of F-space compatible with C. It may be shown from the orthogonality property of $D(\boldsymbol{F})$ established above that the only estimate $\boldsymbol{F}_C$ satisfying both the redundancy condition (equation (23)) and the upper bound constraint on $D(\boldsymbol{F})$ is that for which $D(\boldsymbol{F})$ attains its upper bound value $D_{\max}(C)$ (see the appendix). Physically, this result corresponds to MEP (see below). Furthermore, in the special case where C admits only one independent variable F (i.e. $m = 1$) and where F itself has an upper bound $F_{\max}(C)$, it may be shown similarly (see the appendix) that the only consistent estimate of F from C is $F_{\max}(C)$. This latter result corresponds to various maximum flux principles in physics (see below). Other than the existence of the upper bounds $D_{\max}(C)$ and $F_{\max}(C)$, these results are independent of the specific nature of the constraint C. The property of maximum dissipation (and maximum flux) is generic to MaxEnt estimates of anti-symmetric functions.

So far we have deliberately avoided any physical interpretation in the derivation of the above results, in order to expose them as generic properties of MaxEnt which apply to a certain class of problems involving statistical inference from partial information. We now ask: if MaxEnt is fundamentally an algorithm of Bayesian statistical inference, why should we expect it to work as a description of nature? The answer is that, in its application to statistical mechanics, MaxEnt predicts that behaviour which is selected *reproducibly* by nature under the imposed constraints [2, 3]. The latter consist of the applied experimental conditions (constraints $\boldsymbol{F}$ or, equivalently, $\boldsymbol{\lambda}$) together with the known or hypothesized microscopic physics defining the allowed microstates or phase-space trajectories (prior information C). Evidently, knowledge of $\boldsymbol{F} C$ alone must be sufficient to predict any result that nature reproduces under $\boldsymbol{F} C$. This is precisely the information from which the MaxEnt distribution $p(i \mid \boldsymbol{F} C)$ is constructed, and from which all physical observables are predicted through expectation values taken over $p(i \mid \boldsymbol{F} C)$. In the less restrictive problem of estimating $\boldsymbol{F}$ from C, the redundancy condition of equation (23) is interpreted in physical terms as the requirement that $\boldsymbol{F}_C$ is reproducibly selected under C. In statistical terms, reproducible behaviour simply means the most probable behaviour: it is reproducible precisely because it is characteristic of each of the overwhelming majority of microscopic states or paths consistent with the imposed constraints [2, 3].

To summarize, MaxEnt is a statistically based physical selection principle which extends Boltzmann's insight to non-equilibrium systems. It predicts the reproducible (i.e. most probable) behaviour selected under given constraints. However, the predictive success of MaxEnt (like that of equilibrium statistical mechanics) hinges on having correctly identified the constraints that actually apply in nature. In that respect MaxEnt (and hence the principles of maximum dissipation and maximum flux) remains essentially a trial-and-error procedure; its failures inform us of new constraints (new physics).

We now discuss the physical interpretation of the dissipation function D as entropy production. Recently the author applied MaxEnt to the non-equilibrium stationary behaviour of a finite system (volume V, boundary Ω) exchanging energy and matter with its environment [3]. There the possible outcomes consisted of microscopic phase-space paths Γ over a finite time

interval τ. The constraints F_k took the general form of initial values (at $t = 0$) of some scalar density at each point $x \in V$, denoted here by $\rho(x, 0)$, together with the time-averaged values of the outward normal component of the corresponding flux density at each point $x \in \Omega$, denoted by $\bar{F}^n(x)$, where the overbar indicates a time average over the interval $t \in (0, \tau)$. The construction of the MaxEnt path distribution p_Γ then proceeded as described above: the sum over k in equation (5) passes over, in the continuum limit, to spatial integrals over V and Ω, giving

$$\sum_{k=1}^m \lambda_k f_k(i) \rightarrow \int_V \alpha(x)\rho(x, 0)_\Gamma \, \mathrm{d}V + \int_\Omega \eta(x)\bar{F}^n(x)_\Gamma \, \mathrm{d}A \qquad (24)$$

where $\alpha(x)$ and $\eta(x)$ are Lagrange multipliers associated with $\rho(x, 0)$ and $\bar{F}^n(x)$, respectively. Contributions like equation (24) arising from different scalar densities ρ (e.g. internal energy, mass) combine additively. As shown in [3], $\alpha(x)$ on the boundary is related to $\eta(x)$ through the local continuity equation that applies to each path Γ, $\partial\rho(x, t)_\Gamma/\partial t = -\nabla \cdot F(x, t)_\Gamma + Q(x, t)_\Gamma$, in which Q is a local source; specifically we find $\eta(x) = -\tau\alpha(x)/2$. Therefore the exponent (path action) of the MaxEnt path distribution becomes

$$\sum_{k=1}^m \lambda_k f_k(i) \rightarrow \int_V \alpha(x)\rho(x, 0)_\Gamma \, \mathrm{d}V - \frac{\tau}{2}\int_\Omega \alpha(x)\bar{F}^n(x)_\Gamma \, \mathrm{d}A. \qquad (25)$$

The inclusion of the local density constraints $\rho(x, 0)$, embodied in the first term of equation (25), leads to a physical interpretation of the Lagrange multiplier $\alpha(x)$ and hence of the second term in equation (25). Specifically, the requirement that p_Γ corresponds to the Gibbs grand-canonical distribution in the equilibrium limit, $\bar{F}^n(x) = 0$, implies $\alpha = -1/k_\mathrm{B}T$ when ρ is the internal energy density and $\alpha = \mu/k_\mathrm{B}T$ when ρ is the mass density of some chemical species, where T is temperature, μ is the species chemical potential and k_B is Boltzmann's constant.

Retaining these identifications, we may consider the MaxEnt distribution constructed from the surface flux constraints alone, whose path action is the second term in equation (25),

$$\sum_{k=1}^m \lambda_k f_k(i) \rightarrow -\frac{\tau}{2}\int_\Omega \alpha(x)\bar{F}^n(x)_\Gamma \, \mathrm{d}A. \qquad (26)$$

The point here is that now all the constraint functions are anti-symmetric (*sensu* equation (11)): the allowed microscopic paths Γ (whether they are governed by deterministic or stochastic equations of motion) can be grouped in pairs related by path reversal, under which $\bar{F}^n(x)_\Gamma$ is anti-symmetric. Therefore maximum dissipation applies as described above. When $\bar{F}^n(x)$ is the internal energy flux density ($\alpha = -1/k_\mathrm{B}T$), for example, the mean dissipation is

$$D = 2\sum_{k=1}^m \lambda_k F_k \rightarrow \frac{\tau}{k_\mathrm{B}}\int_\Omega \frac{\bar{F}^n(x)}{T(x)} \, \mathrm{d}A \qquad (27)$$

which may be identified with the (dimensionless) thermodynamic entropy exported across Ω during interval τ by heat flow; analogous contributions to D from mass flows across Ω combine additively. In the steady state, equation (27) may also be written as the volume integral

$$D = \frac{\tau}{k_\mathrm{B}}\int_V \left\{ \bar{F}(x) \cdot \nabla\left(\frac{1}{T(x)}\right) + \frac{\bar{Q}(x)}{T(x)} \right\} \, \mathrm{d}V \quad \text{(steady-state)} \qquad (28)$$

which is the corresponding thermodynamic entropy production within V, the two terms in the integrand representing thermal and frictional dissipation, respectively. In physical terms, therefore, maximum dissipation translates to maximum entropy export, or equivalently in the

steady state, maximum entropy production (MEP). Note, however, that stationarity is not a necessary condition for maximum entropy export; in equation (27), $\bar{F}^n(\boldsymbol{x})$ could represent the surface flux at $\boldsymbol{x}$ averaged over a small time interval τ during the (non-stationary) macroscopic evolution of the system.

We illustrate these general results with two physical examples. First, in Rayleigh–Bénard convection a fixed temperature gradient is imposed across a fluid layer enclosed between two horizontal conducting surfaces, corresponding to the case of prescribed $\boldsymbol{\lambda}^*$ (inverse temperature gradient). In this one-dimensional problem, MEP then implies that the selected steady-state vertical heat flux across the fluid ($\boldsymbol{F}$) is given by the maximum value of $\boldsymbol{F}$ allowed under $\boldsymbol{\lambda}^*$; this result may also be seen more directly as an application of the maximum flux principle. An analogous result holds for the momentum flux across a fluid layer subjected to a fixed shear. These upper bound transport principles have been shown to reproduce several key features of the observed vertical profiles of temperature and velocity in steady-state thermal and shear turbulence [11]. Alternatively, imposing a fixed heat input (prescribed $\boldsymbol{F}^*$) leads to selection of the maximum (inverse) temperature gradient ($\boldsymbol{\lambda}$) allowed by $\boldsymbol{F}^*$.

Secondly, within a simple 10-box zonal model of Earth's climate, Paltridge [12] showed that MEP reproduces the observed zonal distributions of meridional heat transport, surface temperature and cloud fraction with remarkable accuracy. This application corresponds to the less restrictive problem where the variables F_k (horizontal heat flux between boxes k and $k + 1$) and λ_k (inverse temperature gradient between boxes k and $k + 1$) are free variables to be selected by the imposed constraints C (steady-state energy balance under prescribed solar radiation inputs at the top of the atmosphere). Paltridge's analysis also included the supplementary hypothesis that within each zone the surface-to-atmosphere convective transport of sensible and latent heat is maximal for a given configuration of horizontal heat fluxes F_k in the atmosphere and oceans. This additional 'convection hypothesis'—which predicts the homeostatic regulation of surface temperature by clouds [13]—is another example of the maximum flux principle obtained here from MaxEnt, in which it is assumed that adjustment of vertical convection occurs over a much shorter timescale than adjustment of the F_k. These and other applications of MEP are discussed in [14]. Except in the simplest cases, most physical applications of MEP correspond to the less restrictive problem of variable $\boldsymbol{F}$ and $\boldsymbol{\lambda}$ rather than prescribed $\boldsymbol{F}^*$ or $\boldsymbol{\lambda}^*$.

The present derivation of MEP from MaxEnt refines and generalizes the less rigorous argument advanced previously [3]. It demonstrates explicitly the relationship between MEP and the FT via the orthogonality property of the dissipation function D, and furthermore shows how MEP, the FT and orthogonality of D are generic properties of MaxEnt distributions under anti-symmetric constraints. It also makes clear that in physical applications MEP applies to the entropy production of those linearly independent fluxes that are free to vary under the imposed constraints.

Finally, we comment briefly on the relationship between MaxEnt and previous theories of irreversible processes due to Onsager [4], Prigogine [5] and Ziegler [6]. Jaynes [1, 2] already noted that Onsager-type reciprocity relations (equation (7)) emerge naturally as generic properties of MaxEnt distributions, independently of any underlying physical assumptions. In addition, Županović *et al* [15] showed recently for the case of linear flux–force relationships that MEP is equivalent to the Onsager–Rayleigh principle of 'least dissipation of energy' (despite the name), and that MEP manifests itself as Kirchoff's loop law in linear electric networks and their analogues in chemical reaction networks [16]. In this work, the constitutive relations between $\boldsymbol{F}$ and $\boldsymbol{\lambda}$ derived from the FT (equations (15) and (16)) provide nonlinear generalizations of Onsager's linear flux–force relations, and imply that the Onsager–Rayleigh principle (MEP) may be extended to systems far from equilibrium.

As noted after equation (22), Prigogine's principle of minimum entropy production [5] can be interpreted within the MaxEnt formalism as a special case of the behaviour of the dissipation function $D(\boldsymbol{\lambda})$ under variations in $\boldsymbol{\lambda}$ that are restricted to the plane $\pi(c^*)$. Maximum dissipation is a more general result which describes the behaviour of $D(\boldsymbol{\lambda})$ under all possible variations in $\boldsymbol{\lambda}$ permitted by the imposed constraints and thus represents a physical selection principle under given constraints. MEP applies both close to and far from equilibrium.

In Ziegler's formalism [6], a dissipation function $\Phi = \Sigma_k A_k^{(d)} v_k$ was defined in terms of dissipative forces $A_k^{(d)}$ and corresponding velocities v_k, with Φ being considered the fundamental quantity from which the dissipative forces were to be derived. Assuming Φ is given as a function of the v_k alone, Ziegler derived the orthogonality condition $A_k^{(d)} \propto \partial \Phi / \partial v_k$ on the grounds that it is the only possible vectorial relation between $A_k^{(d)}$ and Φ. Then the principle of maximum dissipation followed as above. Ziegler's formalism is discussed further in the context of Earth's climate in [13]. MaxEnt provides a more fundamental statistical basis for the orthogonality condition. In MaxEnt the fundamental quantity is not the dissipation function D but the information entropy H, from whose maximization D emerges as a derived quantity. Via the FT, MaxEnt establishes MEP as a physical selection principle for the most probable configuration of macroscopic fluxes under given constraints.

Acknowledgments

Appendix. Maximum dissipation and maximum flux principles under arbitrary constraints

Given the prior information C, let there be m linearly independent unknowns F_k ($1 \leqslant k \leqslant m$). We wish to estimate the values of F_k from knowledge of C alone. This estimate, denoted by $\boldsymbol{F}_C$, must satisfy the redundancy condition, equation (23), as well as the upper bound constraint on the dissipation function, $D(\boldsymbol{F}) \leqslant D_{\max}(C)$, within the domain of F-space compatible with C. The nature of the constraint C is otherwise arbitrary. The distribution $p(i \mid C)$ from which $\boldsymbol{F}_C$ is to be estimated is obtained from MaxEnt by maximizing the information entropy

$$H(C) = -\sum_{i=1}^{n} p(i \mid C) \log p(i \mid C) \tag{A.1}$$

subject to the three constraints

$$H(C) = S(\boldsymbol{F}_C C) \tag{A.2}$$

$$\langle D \rangle \equiv \sum_{i=1}^{n} p(i \mid C) D(\boldsymbol{f}(i)) \leqslant D_{\max}(C) \tag{A.3}$$

$$\sum_{i=1}^{n} p(i \mid C) = 1. \tag{A.4}$$

In equation (A.2), $S(\boldsymbol{F}_C C)$ is the MaxEnt information entropy of the conditional probability $p(i \mid \boldsymbol{F}_C C)$, and the estimate $\boldsymbol{F}_C$ is to be recovered from $p(i \mid C)$ in the usual way (equation (2)),

$$\boldsymbol{F}_C = \sum_{i=1}^{n} p(i \mid C) \boldsymbol{f}(i). \tag{A.5}$$

Inequality constraints like equation (A.3) can be incorporated into MaxEnt in a similar way to equality constraints [17]. We introduce the Lagrangian function

$$\psi = H(C) + \alpha(H(C) - S(\boldsymbol{F}_C C)) + \beta(\langle D \rangle - D_{\max}(C)) + \gamma \left(\sum_{i=1}^{n} p(i \mid C) - 1 \right) \tag{A.6}$$

in which α, β and γ are Lagrange multipliers. The only difference with regard to treatment of the inequality constraint on $\langle D \rangle$ is that now we have two possibilities: either $\beta = 0$ and $\langle D \rangle$ is strictly less than $D_{\max}(C)$ (the constraint is inactive) or $\beta \neq 0$ and $\langle D \rangle = D_{\max}(C)$ (the constraint is active, i.e. maximum dissipation). Setting $\partial \psi / \partial p(i \mid C) = 0$ yields the solution

$$p(i \mid C) = \frac{1}{Z(\alpha, \beta)} \exp \left\{ \frac{\beta}{1+\alpha} D(\boldsymbol{f}(i)) + \frac{\alpha}{1+\alpha} \sum_{k=1}^{m} \lambda_k(\boldsymbol{F}_C) f_k(i) \right\} \tag{A.7}$$

in which λ_k arises from differentiating $S(\boldsymbol{F}_C C)$ with respect to F_k as in equation (9), and the partition function $Z(\alpha, \beta)$ is a normalization factor arising from equation (A.4).

For the case $\beta = 0$ (constraint on $\langle D \rangle$ inactive), we have

$$p(i \mid C) = \frac{1}{Z(\alpha, 0)} \exp \left\{ \frac{\alpha}{1+\alpha} \sum_{k=1}^{m} \lambda_k(\boldsymbol{F}_C) f_k(i) \right\}. \tag{A.8}$$

Comparing equation (A.8) with equation (5), we see that in this case $p(i \mid C)$ coincides with the MaxEnt distribution $p(i \mid \tilde{\boldsymbol{F}} C)$ specified by the Lagrange multiplier $\tilde{\boldsymbol{\lambda}} = \alpha \boldsymbol{\lambda}(\boldsymbol{F}_C)/(1+\alpha)$, and its information entropy is therefore given by $S(C) = S(\tilde{\boldsymbol{F}} C)$. As established in the main text, the value $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{F}_C)$ maximizes the dissipation function $D(\boldsymbol{\lambda})$ subject to the constraint $D(\boldsymbol{\lambda}) = 2\boldsymbol{\lambda} \cdot \boldsymbol{F}_C$, and consequently $D(\boldsymbol{\lambda}(\boldsymbol{F}_C)) > D(\tilde{\boldsymbol{\lambda}})$ for any finite α. Because the D-surface is essentially an inverted copy of the S-surface (see the main text), this last inequality implies that $S(\boldsymbol{F}_C C) < S(\tilde{\boldsymbol{F}} C) = S(C)$ in violation of the redundancy condition (equation (A.2)). The case $\beta = 0$ is thereby excluded.

The redundancy condition can only be satisfied when $\beta \neq 0$, i.e. the inequality constraint must be a strict equality: $\langle D \rangle = D_{\max}(C)$. In many practical applications (e.g. physical systems with a large number of microscopic degrees of freedom) we can ignore fluctuations in $\boldsymbol{f}$ around its mean value $\boldsymbol{F}_C$ (i.e. $\langle D \rangle = D(\boldsymbol{F}_C)$), and we therefore have $D(\boldsymbol{F}_C) = D_{\max}(C)$. The MaxEnt estimate of $\boldsymbol{F}$ from the constraint C then coincides with the value of $\boldsymbol{F}$ that maximizes the dissipation function under C.

In the special case where C admits only one unknown F and where F itself has an upper bound $F_{\max}(C)$, the same analysis with equation (A.3) replaced by

$$\langle F \rangle \equiv \boldsymbol{F}_C = \sum_{i=1}^{n} p(i \mid C) f(i) \leqslant F_{\max}(C) \tag{A.9}$$

shows that the only estimate of F from C satisfying both the redundancy condition (equation (A.2)) and the upper bound constraint on F is $F = F_{\max}(C)$. In physical applications this result corresponds to various maximum flux principles (see the main text).

References

[1]   Jaynes E T 2003 *Probability Theory: the Logic of Science* ed G L Bretthorst (Cambridge: Cambridge University Press)
[2]   Jaynes E T 1957 *Phys. Rev.* **106** 620
      Jaynes E T 1957 *Phys. Rev.* **108** 171
      Jaynes E T 1979 *The Maximum Entropy Principle* ed R D Levine and M Tribus (Cambridge, MA: MIT Press) p 15
      Jaynes E T 1980 *Ann. Rev. Phys. Chem.* **31** 579
[3]   Dewar R C 2003 *J. Phys. A: Math. Gen.* **36** 631
      Dewar R C 2004 *Non-Equilibrium Thermodynamics and the Production of Entropy* ed A Kleidon and R Lorenz (Berlin: Springer) p 41
[4]   Onsager L 1931 *Phys. Rev.* **37** 237
      Onsager L 1931 *Phys. Rev.* **37** 2265
[5]   Prigogine I 1962 *Introduction to Non-Equilibrium Thermodynamics* (New York: Wiley-Interscience)
[6]   Ziegler H 1983 *An Introduction to Thermomechanics* (Amsterdam: North-Holland)
[7]   Jaynes E T 1985 *Maximum-Entropy and Bayesian Methods in Inverse Problems* ed C R Smith and W T Grandy (Dordrecht: Reidel) p 21
[8]   Evans R M L 2005 *J. Phys. A: Math. Gen.* **38** 293
[9]   Evans D J and Searles D J 2002 *Adv. Phys.* **51** 1529
[10] Maes C 1999 *J. Stat. Phys.* **95** 333
[11] Kerswell R R 2002 *J. Fluid Mech.* **461** 239 and references therein
[12] Paltridge G W 1978 *Q. J. Met. Soc.* **104** 927
[13] O'Brien D M and Stephens G L 1995 *Q. J. Met. Soc.* **121** 1773
[14] Ozawa H, Ohmura A, Lorenz R D and Pujol T 2003 *Rev. Geophys.* **41** 1018
      Kleidon A and Lorenz R (ed) 2004 *Non-Equilibrium Thermodynamics and the Production of Entropy* (Berlin: Springer)
[15] Županović P, Juretić D and Botrić S 2005 *Fizika* A at press
[16] Županović P, Juretić D and Botrić S 2004 *Phys. Rev.* E **70** 056108
      Županović P and Juretić D 2004 *Croat. Chem. Acta* **77** 561
[17] Ishwar P and Moulin P 2003 *IEEE Trans. Signal Process.* **51** 698