

Generalized Statistical Mechanics and the Onsager Relations*

PETER G. BERGMANN AND ALICE C. THOMSON

Department of Physics, Syracuse University, Syracuse, New York

(Received March 20, 1953)

The Onsager relations assert certain symmetry relations for the cross coefficients that connect rates of flux in stationary nonequilibrium processes with the driving "thermodynamic forces." These relations are usually derived by an argument borrowed from fluctuation theory, with the assumption that the rate at which a fluctuation in an equilibrium ensemble regresses equals the rate at which the ensemble average of the same quantity will change in a nonequilibrium ensemble. In this paper, we have by-passed fluctuation theory and have derived similar symmetry relations for the corresponding cross coefficients in ensembles consisting of thermally isolated systems differing but slightly from an equilibrium canonical ensemble. For the case of large deviations from equilibrium, we have found expressions for the same cross coefficients which possess no symmetry properties but which are exactly the usual correlation coefficients of fluctuation theory. Because our present work is concerned with thermally isolated systems, we cannot yet describe in a satisfactory manner stationary nonequilibrium processes. For the representation of such processes, one would need an ensemble of systems in which the interaction with the surroundings is described in terms of additional random variables.

I. INTRODUCTION

THE first major contribution to a general thermodynamics of irreversible processes was the Onsager theory of interfering linear processes. In this work the cross-coefficients describing the interference were shown to be related by a set of reciprocal relations. These relations have been demonstrated by a statistical mechanical argument using properties of the equilibrium microcanonical ensemble.^{1,2} In this proof the techniques used are restricted to processes taking place near equilibrium.

Recently considerable progress has been made toward widening the scope of statistical mechanical methods for treating irreversible processes.³⁻⁵ These investigators also use the equilibrium ensemble as a starting point. They obtain nonequilibrium results by considering suitable nonequilibrium subsets of the equilibrium ensemble. That is, specific fixed values of the interesting time-dependent macroscopic variables of the system at a particular time characterize a certain subset of members of the equilibrium ensemble, and the properties of this subset are investigated. Though this approach is not restricted to near-equilibrium processes, to date results have been obtained only under such conditions.

A second method which is based on the same viewpoint is provided the use of nonequilibrium ensembles, constructed specifically as such. We have felt that this approach might provide a better starting point for the study of more general nonequilibrium situations. One such ensemble, a sort of "generalized microcanonical ensemble," has been used for a long time in connection

with the Gibbs form of the H-theorem.⁶ We are interested in the potentialities of a second type which we have called "generalized canonical ensembles." These ensembles seem in many cases to be a more adequate statistical collection than the generalized microcanonical ensembles, and are easier to work with.

Generalized canonical ensembles were introduced by one of us (Peter G. Bergmann)⁷ in a previous paper. We shall discuss the general properties of these ensembles in greater detail, and then as an example show their use in the near-equilibrium case in deriving the Onsager relations.

II. CONSTRUCTION OF THE ENSEMBLE. ENTROPY

Thermodynamic measurements on a system in nonequilibrium consist of the determination of the values of interesting local variables in the system. For example, one might measure local temperatures and pressures in a gas. The choice of variables is dependent on the type of system and the amount of its deviation from equilibrium. For instance, temperature would be meaningless in a case where the temperature gradient (i.e., the gradient of mean energy per degree of freedom) is large over a mean free path. Furthermore, care must be exercised in nonequilibrium cases that the measured quantity actually is properly interpreted in terms of the variables of the system. For example, different types of thermometers thrust into an anisotropic beam like the sun's rays will measure different "temperatures" because of their different absorption properties.

If any suitable set of variables and measurements is chosen (it might, for instance, include fluxes, or even quantities moving with the flow), then a statistical treatment in principle can be used to obtain a thermodynamic description of the process. "Suitability" here

* This research was supported in part by the United States Air Force, monitored by the Office of Scientific Research.

¹ L. Onsager, *Phys. Rev.* **37**, 405 (1931); **38**, 2265 (1931).

² H. B. G. Casimir, *Revs. Modern Phys.* **17**, 343 (1945).

³ Callen, Barasch, and Jackson, *Phys. Rev.* **88**, 1382 (1952), and references included therein.

⁴ R. F. Greene and H. B. Callen, *Phys. Rev.* **88**, 1387 (1952), and references included therein.

⁵ M. S. Greene, *J. Chem. Phys.* **20**, 1281 (1952).

⁶ R. C. Tolman, *Principles of Statistical Mechanics* (Oxford University Press, London, 1938).

⁷ P. G. Bergmann, *Phys. Rev.* **84**, 1026 (1951), referred to as I in the following.

is taken to mean that the variables selected for observation do not fluctuate significantly during the time required to complete one observation; this, of course, is dependent on the refinement of the apparatus as well as on the extent of non-steadiness of the system with respect to those variables. And it is also necessary that those variables which represent local contributions to a variable characteristic of the whole system, like the energy, have an approximate additive property; that is, that the sum of the local contributions should approximately equal the system variable, with interaction corrections negligible. For example, the total energy should approximately equal the sum of the local energies measured. But this condition generally admits a wide class of variables. In the energy example, the interaction across the boundary between regions where measurements are made is required to contribute little to the total energy, and this is frequently true even though the atomic interactions are strong, merely because of the large number of atoms in each measured region. These requirements are in some sense equivalent to saying that meaningful measurements are indeed possible. An example of the possible use of variables moving with the flow is the use of a Brownian particle as a measuring instrument in a fluid in nonequilibrium.

One further requirement is that the set of variables selected be "complete" enough that the ensemble averages described below are defined. Thus we must avoid a specification which would lead to finite ensemble densities (i.e., densities bounded away from 0) over unbounded regions of γ -space. For instance, we cannot specify temperature measurements at each end of a rod with no measurement bounding the mean total energy of the rod.

Once a set of measurements has been made, the statistical method involves the choice of a suitable collection or ensemble of identical systems in microscopic states which either exactly, or in the ensemble average, correspond to the macroscopic state implied by the measured values. Then the prediction of the future for the actual system is made through the study of the behavior in time of averages over the ensemble.

The ensemble density which was called a "generalized microcanonical ensemble" above is constructed as follows: the μ -space of the system is conceived to be divided up into cells of size corresponding to the definition of the measured local variables. Each molecule-space in the γ -space is similarly divided. The information obtained in the measurement is now interpreted, on account of the macroscopic character of the variables, as giving an occupation number to each cell in μ -space. Then, using the principle of equal a priori probabilities, we construct an ensemble which has initially constant values over corresponding cells in γ -space.

This procedure is not always appropriate. For one thing, certain important thermodynamic measurements are made on intensive parameters, for example, local

temperatures, rather than directly on dynamical variables like local energies, and it is easier to treat these parameters meaningfully with a canonical-like ensemble. Such ensembles are really defined by the values of the parameters, whereas it is more difficult to relate such quantities to the generalized microcanonical ensembles. In addition, the physical regions in which measurements are made are not usually isolated but are rather to be considered in loose contact with their neighbors, in the sense of the requirement mentioned above of "additivity" of the selected dynamical variables. This contact suggests the usefulness of canonical-like ensembles which admit members differing slightly in the values of the interesting dynamical variables.

If the interesting variables are designated as $a_i(p_1 \cdots p_n, q_1 \cdots q_n)$, (we shall assume in the following that they are not explicitly time-dependent, although this restriction is not necessary), then the generalized canonical ensemble density is

$$\mu(p, q, t) = (Z^0)^{-1} \exp(-\sum_i \beta_i^0 a_i^0),$$

$$Z^0 = \int_{\gamma\text{-space}} \exp(-\sum_i \beta_i a_i^0) dX, \quad (1)$$

where dX is the volume element in γ -space; p, q is a symbol for the set of coordinates and momenta $p_1 \cdots p_n, q_1 \cdots q_n$; β_i^0 are constants; a_i^0 is that constant of the motion which at the time of measurement equals the value of the variable a_i , that is,

$$\partial a_i^0 / \partial t + (a_i^0, H) = 0, \quad a_i^0(p, q, t) = a_i(p_0, q_0), \quad (2)$$

where p_0, q_0 are the values of p, q at t_0 . Now in some cases it may turn out that the measured quantities are the variables a_i , in which case the β_i^0 are determined by the conditions:

$$\langle a_i \rangle_N(t_0) = \int a_i \mu(t_0) dX = A_i(t_0), \quad (3)$$

where $A_i(t_0)$ is the measured value of a_i at t_0 . In other cases the β_i^0 are the measured quantities. For example, if the a_i are local energies, the β_i^0 are proportional to the reciprocals of the initial local temperatures. In this case the conditions (3) determine the average values of the a_i at time t_0 .

The density μ satisfies Liouville's equation. In fact it is merely a way of writing that solution of Liouville's equation which has the canonical form and the correct average properties at t_0 .

In (I) a quantity called the "system entropy" was introduced which takes on its minimum value at the time of observation and has other properties suitable to entropy. This quantity is defined as follows:

$$S^* = k(\ln \mu_0 + \ln Z + \sum_i \beta_i \langle a_i \rangle_N), \quad (4)$$

where k is Boltzmann's constant, μ_0 is a constant, and

β_i and Z are defined by the requirements

$$Z^{-1} \int a_i \exp(-\sum_j \beta_j a_j) dX = \langle a_i \rangle_{Av}, \quad (5)$$

$$Z = \int \exp(-\sum_j \beta_j a_j) dX.$$

That is, β_j and Z are time-dependent quantities which are used to define a fictitious canonical ensemble which yields at all times the same averages of the a_i as the actual ensemble density μ .

The quantity S^* is a generalization of the ensemble definition of Boltzmann's quantity ($-kH$) for generalized microcanonical ensembles.⁶ S^* reduces to ($-kH$) if we consider the a_i as designations of cells in phase space. Thus if γ -space is divided into small but not infinitesimal cells as described above, then we define

$$a_i = X_i^{-1} \delta_i(q, p) = \begin{cases} 1 & \text{inside the } i\text{th cell,} \\ 0 & \text{outside the } i\text{th cell,} \end{cases} \quad (6)$$

where X_i is the volume of the i th cell. H is defined as

$$H = \sum_i \mu_i \ln \mu_i X_i, \quad (7)$$

where μ_i is the average value of μ over the i th cell. Then

$$\langle a_i \rangle_{Av} = X_i^{-1} \int \delta_i \mu(q, p) dX = \mu_i.$$

From the definition of Z and β_k ,

$$Z \langle a_i \rangle_{Av} = \int a_i \exp(-\sum_k \beta_k a_k) dX,$$

$$= X_i^{-1} \int_{i\text{th cell}} \exp(-\sum_k \beta_k a_k) dX = \exp(-\beta_i / X_i).$$

Then

$$S^* = k(\ln Z - \sum_i \ln(Z \mu_i) \mu_i X_i + \ln \mu_0) \\ = -kH + k \ln \mu_0. \quad (8)$$

In (I) it is shown that S^* has an absolute minimum at the time t_0 , both with respect to past times and future times. We conjecture that the usual behavior of this function is an initial increase continuing for very large times with perhaps small undulations, then at vast intervals large fluctuations and eventually a decrease to, but not below, the original value. This behavior is, of course, the same as that usually conjectured for the ensemble definition of ($-H$), and any entropy defined for a closed system so that it does not eventually decrease would be a violation of the Poincaré Theorem on quasi periodicity of dynamical systems. Note that the property of S^* of having a minimum at t_0 which may later be reached but never surpassed, a property previously proved for ($-H$),⁶ is quite different from the behavior of the Boltzmann μ -space definition of ($-H$), which we shall designate ($-H'$). This latter

quantity can, indeed, fluctuate below its value at t_0 . The behavior of S^* and the ensemble quantity ($-H$) is, of course, an ensemble property which is dependent on the collective behavior of all the systems, and under certain rare dynamical conditions, individual systems in the ensemble may indeed eventually have values of ($-H'$) deviating below their initially observed values. The point is that even the "average" system eventually decreases in entropy, but it does not decrease below the initial value at t_0 . We identify the ensemble property of S^* with the "thermodynamic" behavior of entropy for the physical system under observation.

Taking the time derivative of S^* in (4) we get

$$\frac{dS^*}{dt} = -k \sum_i \dot{\beta}_i \langle a_i \rangle_{Av} + k \sum_i \beta_i \dot{\langle a_i \rangle}_{Av} + k \sum_i \beta_i \frac{d}{dt} \langle a_i \rangle_{Av}.$$

By integrating by parts we can put this equation in the following form:

$$dS^*/dt = k \sum_i \beta_i \langle \dot{a}_i \rangle_{Av}. \quad (9)$$

This equation suggests the role of the $k\beta_i$ as "generalized reciprocal temperatures" and the $\langle \dot{a}_i \rangle_{Av}$ as "generalized heat flows."

III. NEAR EQUILIBRIUM. ONSAGER RELATIONS

We shall assume that one of the variables a_i , say a_0 , is chosen as the Hamiltonian H . The corresponding β_0^0 (the reciprocal of kT), is the "thermodynamic force" associated with the mean total energy of the system. Then the density μ can be written as

$$\mu = (Z^0)^{-1} \exp(-\beta_0^0 H - \sum_{i=1} \beta_i^0 a_i^0),$$

and the equilibrium ensemble is characterized by zero values of all the β_i^0 , $i \neq 0$. If, for example, we have chosen the a_i for $i > 0$ as local energies (i.e., energies of parts of the system), then the $k\beta_i^0$ for $i > 0$ will be not the initial values of the local reciprocal temperatures but the differences between reciprocal local temperatures and the reciprocal over-all temperature.

The quantities $\langle a_i \rangle_{Av}$ can be expressed as functions of the set of β_j^0 and t . The derivatives $\partial \langle a_i \rangle_{Av} / \partial \beta_j^0$ are as follows:

$$\frac{\partial \langle a_i \rangle_{Av}}{\partial \beta_j^0} = -\langle a_i a_j^0 \rangle_{Av} + \langle a_i \rangle_{Av} \langle a_j^0 \rangle_{Av}. \quad (10)$$

If the initial measurements $\langle a_i^0 \rangle_{Av}$ are close to equilibrium values and, accordingly, the β_j^0 ($j \neq 0$) small, then we can express $\langle a_i \rangle_{Av}$ by the first terms in a Taylor series expansion in the β_j^0 ($j \neq 0$). That is,

$$\langle a_i \rangle_{Av}(\beta_1^0 \cdots \beta_n^0, t) \\ = \langle a_i \rangle_{Av}(0 \cdots 0, t) + \sum_j \frac{\partial \langle a_i \rangle_{Av}}{\partial \beta_j^0}(0 \cdots 0, t) \beta_j^0 + \cdots \\ = \langle a_i \rangle_{Av}(0 \cdots 0, t) + \sum_j A_{ij} \beta_j^0, \quad (11)$$

where

$$A_{ij}(t) = -\langle a_i a_j^0 \rangle_{Av}(0 \cdots 0, t) + \langle a_i \rangle_{Av}(0 \cdots 0, t) \langle a_j^0 \rangle_{Av}(0 \cdots 0, t). \quad (12)$$

The first term in $\langle a_i \rangle_{Av}(\beta_1^0 \cdots \beta_n^0, t)$ is constant in time, being an average over the equilibrium ensemble of a quantity which is not explicitly time-dependent. We shall now show that the quantities A_{ij} have certain symmetry properties. We distinguish between those variables a_i which are even functions of the momenta of the system, and those which are odd functions of the momenta. We assume in the following that the Hamiltonian H of the system is an even function of the momenta. Consider a point \mathbf{x} in phase space which moves from an initial value \mathbf{x}^0 according to the equations of motion. We can symbolize this as $\mathbf{x} = M(\mathbf{x}^0, t)$ where $M(\mathbf{x}^0, t)$ is the mapping of \mathbf{x}^0 in phase space. Any constant of the motion, in particular an ensemble density μ , will satisfy

$$\mu(\mathbf{x}, t) = \mu(\mathbf{x}^0, 0) = \mu(M^{-1}(\mathbf{x}, t), 0).$$

If we designate by a prime the transformation reversing the signs of the \mathcal{p}_k , then the even or odd character of the a_i is given by

$$a_i(\mathbf{x}) = \epsilon_i a_i(\mathbf{x}') \quad \begin{array}{l} \epsilon_i = 1 \quad \text{for even functions,} \\ \epsilon_i = -1 \quad \text{for odd functions.} \end{array} \quad (13)$$

The assumptions on the Hamiltonian lead to the property

$$M(\mathbf{x}^0, t) = [M^{-1}(\mathbf{x}^0, t)]'. \quad (14)$$

This equation is a consequence of the time reversibility of the equations of motion.

Now we want to show that

$$A_{ik} = A_{ki} \epsilon_k \epsilon_i. \quad (15)$$

We shall show this property for the first term in expression (12); the second term is treated in exactly the same way:

$$\begin{aligned} & \langle a_i a_k^0 \rangle_{Av}(0 \cdots 0, t) \\ &= (Z^0)^{-1} \int \exp(-\beta_0^0 H) a_k(\mathbf{x}^0) a_i(M(\mathbf{x}^0, t)) dX^0, \\ &= (Z^0)^{-1} \int \exp(-\beta_0^0 H) a_k(M^{-1}(\mathbf{x}, t)) a_i(\mathbf{x}) dX, \\ &= (Z^0)^{-1} \int \exp(-\beta_0^0 H) a_k([M(\mathbf{x}', t)]') a_i(\mathbf{x}) dX, \\ &= (Z^0)^{-1} \int \exp(-\beta_0^0 H') \epsilon_k a_k(M(\mathbf{x}', t)) \epsilon_i a_i(\mathbf{x}') dX', \\ &= \langle a_k a_i^0 \rangle_{Av}(0 \cdots 0, t) \epsilon_k \epsilon_i. \end{aligned} \quad (16)$$

Equations (15) are true for all time for the near-equilibrium case, but do not hold for the general case. If we carry out a similar transformation of the first term in the general expression (10) for $\partial \langle a_i \rangle_{Av} / \partial \beta_j^0$, we get

librium case, but do not hold for the general case. If we carry out a similar transformation of the first term in the general expression (10) for $\partial \langle a_i \rangle_{Av} / \partial \beta_j^0$, we get

$$\begin{aligned} \langle a_i a_k^0 \rangle_{Av} &= (Z^0)^{-1} \int \exp(-\beta_0^0 H - \sum_j \beta_j^0 \epsilon_j a_j[M(\mathbf{x}', t)]) \\ & \quad \times \epsilon_k a_k(M(\mathbf{x}', t)) \epsilon_i a_i(\mathbf{x}') dX', \\ \langle a_k a_i^0 \rangle_{Av} &= (Z^0)^{-1} \int \exp(-\beta_0^0 H - \sum_j \beta_j^0 a_j(\mathbf{x})) \\ & \quad \times a_k(M(\mathbf{x}, t)) a_i(\mathbf{x}) dX. \end{aligned} \quad (17)$$

Now we want to relate the quantities A_{ik} to the Onsager coefficients. First note that Eq. (9) identifies the β_k as the Onsager forces.² We form the difference quotients

$$G_i = \tau^{-1} [\langle a_i \rangle_{Av}(\beta_1^0 \cdots \beta_n^0, t_0 + \tau) - \langle a_i \rangle_{Av}(\beta_1^0 \cdots \beta_n^0, t_0)] \quad (18)$$

over a time interval τ which is, on the one hand, not infinitesimal, and, on the other hand, so small that the deviation from equilibrium has not materially changed. The condition for the latter is that the generalized reciprocal temperatures β_i remain unchanged over the interval τ :

$$\dot{\beta}_i \approx 0, \quad \beta_i \approx \beta_i^0. \quad (19)$$

Then we investigate the dependence of G_i on β_k :

$$\begin{aligned} L_{ik} &= \frac{\partial G_i}{\partial \beta_k} = \tau^{-1} \left[\frac{\partial \langle a_i \rangle_{Av}}{\partial \beta_k}(t_0 + \tau) - \frac{\partial \langle a_i \rangle}{\partial \beta_k}(t_0) \right], \\ &= \tau^{-1} [A_{ik}(t_0 + \tau) - A_{ik}(t_0)]. \end{aligned} \quad (20)$$

Hence L_{ik} has the same symmetry properties as A_{ik} :

$$L_{ik} = \epsilon_i \epsilon_k L_{ki}. \quad (21)$$

These latter equations are the Onsager relations.² The quantities L_{ik} are identified with the thermodynamic coefficients in linear irreversible laws.

The requirement of a lower limit on the time interval τ is necessary because the quantity $\partial \langle a_i \rangle_{Av} / \partial \beta_k = -\langle \dot{a}_i a_k^0 \rangle_{Av} + \langle \dot{a}_i \rangle_{Av} \langle a_k^0 \rangle_{Av}$ is zero at time t_0 if a_i and a_k are either both even or both odd functions of the momenta. In some sense the ensemble fails to demonstrate the dissipative process instantaneously.

The upper limit reflects the fact that the Onsager relations apply to a system in which one portion "drives" another in an approximately steady state, i.e., a state in which the parameters like temperature and chemical potentials are constant in time but vary throughout the system. We may, for example, conceive of a system composed of three parts: two large regions which act as "reservoirs" under different thermodynamic conditions and a small connecting region through which steady-state fluxes take place. In this case the β_i referring to the small sub-system are zero

by virtue of the steady fluxes, and the β_i referring to the reservoirs are zero because of the large size of the reservoirs. The a_i of the reservoirs will steadily change. In this situation the interesting relationships are between those a_i characterizing the sub-system and the driving "temperatures" of the reservoirs. Thus we require the time τ over which the difference quotients are calculated to be at least small enough that essentially steady conditions exist over it.

This treatment in terms of difference quotients is analogous to earlier derivations.² It is perhaps just as realistic, however, to identify the thermodynamic coefficients with genuine time-derivatives $\partial \langle \dot{a}_i \rangle_{Av} / \partial \beta_k$ evaluated not at t_0 but at a time $t_0 + \tau$, where τ satisfies the above restrictions; these quantities, of course, have the same symmetry properties as the difference quotients. This second viewpoint makes clearer the manner in which the nonequilibrium ensemble gives a continued description of the thermodynamic process.

It is sometimes desirable to express the Onsager coefficients as functions of an external magnetic field. We have previously assumed a closed system, but it is nevertheless easy to treat this slightly more general case. We need only remember that if the sources of the field were included within the system, the field would change its direction under reflections of the time axis, and the Hamiltonian would be invariant under the reflection. Hence if we consider the field as external, we must reverse it whenever it would be reversed if its sources were considered internal. The modified Onsager relations are then

$$L_{ik}(\mathbf{H}) = L_{ki}(-\mathbf{H}) \epsilon_i \epsilon_k. \quad (21)$$

IV. DISCUSSION

In thermodynamics we are in the habit of distinguishing between the "thermodynamic system" and its "surroundings," including possible "reservoirs." In many arguments, foremost among them those leading to the establishment of the Second Law, it is useful to shift one's point of view, to regard a certain accumulation of matter alternately as part of the "system" or as a driving reservoir. It is to be expected that a similar elasticity of approach would serve well in statistical mechanics. In the foregoing discussion, however, we have dealt exclusively with the development of systems that from a time t_0 onward are thermally isolated; hence we have been forced to include possible reservoirs as parts of the composite system. This approach can never lead to a rigorously stationary process, because any reservoir of finite size in a nonequilibrium process will undergo secular changes. If the theory is to be capable of describing stationary nonequilibrium processes, we must introduce idealized reservoirs of infinite size; that step implies that the reservoir must be treated as distinct from the thermodynamic system proper.

On the other hand, the interaction of the reservoir with the system is stochastic, both because of the stochastic nature of the system ensemble and because of the molecular structure of the reservoir or reservoirs. We are planning to address ourselves to this problem and hope to arrive eventually at a realistic statistical-mechanical description of the stationary process. Such a description should yield the Onsager relations in a thoroughly satisfactory manner, without restrictions on the choice of the time constant τ in Eqs. (18) and (20).