Simple Proof of the Ergodic Theorem

HARALD WERGELAND

Fysisk Institutt, Norges Tekniske Høgskole, Trondheim, Norway

An attempt is made to prove the equivalence of time and space averages in statistical mechanics, without invoking the theory of sets. Apology is therefore due because of the lack in rigor, but there appear to be possibilities to show how the argument could be sharpened to mathematical precision*.

Ergodic theory is now a province of pure mathematics, but its origin is in the mechanical theory of heat. Specifically it is the assumptions made by Boltzmann and Maxwell to justify the substitution of the time average of mechanical quantities by an average over the phase space. By ergodic theorem we shall presently just mean the statement (when true), that *time average equals phase average*, and not the original hypotheses upon which this conclusion has been based. It is well known, that the latter — though sufficient to prove the theorem — did not stand the test of mathematical critique.

The first encounter with this situation will usually arouse the curiosity of the student of statistical mechanics. Now, the immediate answer to him is of course that the ergodic theorem in the sense defined above, was indeed proved by Birkhoff, Neumann and others¹ in 1932. But I think that those who look up the original papers may be a little overwhelmed by their abstractness and mathematical subtlety.

It is therefore desirable to have a proof in plainer language even if not rigorous. This is the aim of the present paper. In this spirit we shall employ only the ordinary calculus and in particular no more general concept of measure than the Riemann integral. Thus point sets of Lebesgue measure zero, which play a great role in the more mathematical formulations, will be bypassed in silence.

1. STATEMENT OF THE PROBLEM

The mechanical state of a system is determined by a point,

$$\mathbf{P} = \{q_1 q_2 \dots q_f p_1 \dots p_f\} \quad (1)$$

or eventually delimited by the cartesian volume element,

$$d\Omega = dq_1 dq_2, \dots dp_f \quad (2)$$

* Private communication from Professor Kac.

in the $2f$ -dimensional phase space. The system will be restricted by external conditions to a finite part Ω of the phase space, for example the energy shell,

$$E \leq H(\mathbf{P}) \leq E + dE \quad (3)$$

A thermodynamical variable must primarily be identified with the time average

$$f^* = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(\mathbf{P}(t)) dt \quad (4)$$

of some phase function $f(\mathbf{P})$. Now, for mathematical reasons we are only able to compute phase averages

$$\bar{f} = \frac{1}{\Omega} \int_{\Omega} f(\mathbf{P}) d\Omega \quad (5)$$

But the ergodic theorem asserts that these two kinds of average are equivalent:

$$\boxed{f^* = \bar{f}} \quad (6)$$

which is the proposition to be proved.

2. TRANSITIVITY

The decisive progress in the ergodic problem was Birkhoff's idea of transitivity. In order to explain this concept it will be convenient to survey briefly also some other notions of general dynamics:

- a) A region Ω of phase space is said to be *invariant* if for any point $\mathbf{P}$ in Ω also the total orbit through $\mathbf{P}$ belongs to Ω .
- b) Such a region is said to be *indecomposable* if it cannot be divided into parts $\Omega' + \Omega''$ which are separately invariant.
- c) If the orbit from any point $\mathbf{P}$ in Ω will come arbitrarily close to any other point $\mathbf{P}'$ in Ω , then the motion is said to be *transitive*. Indecomposability and transitivity are equivalent. It is evidently a property of the Hamiltonian.

In the ergodic problem we consider closed systems, and hence the energy as constant. Certainly one might instead conceive a thermodynamical system as an open one with a stochastically time dependent Hamiltonian. Such an interpretation has actually been worked out by Delbrück and Molière². There are, however, two drawbacks in this point of view: Firstly it deprives us of the justification of statistical mechanics for a closed system. Secondly, although ergodicity could no doubt be ensured that way, it would be — as Rosenfeld has once put it — so to speak *a fortiori*. Therefore we shall assume the Hamiltonian to be independent of time and consider the energy shell as the region to be inspected for transitivity.

A little further insight into the meaning of transitivity can be obtained from a general classification of the integrals of the motion. By the elementary theory of linear partial differential equations³ the $2f-1$ time independent integrals of the Hamiltonian system can be taken to be of the form:

$$g_i(\mathbf{P}) = c_i \quad (i = 1, 2 \dots 2f-1) \quad (7)$$

Here a characteristic alternative shows up: Such an integral may or may not define a surface in Ω . In the former case eqn. (7), considered as an equation for any one single variable p_k or q_k , has only isolated solutions. The alternative is infinitely many roots in a finite interval. The integrals defining surfaces are denoted as *univalued*.

The existence of a univalued integral will lower the dimensionality of Ω by one. In the extreme case that there are $2f-1$ of them, Ω will be reduced to the one-dimensional continuum of the orbit, which now appears as a section between surfaces c_i , $c_1 \dots c_{2f-1}$. Probably this means a closed orbit, or in other words, exact periodicity.

When the number of univalued integrals is less than $2f-1$, a phase space Ω of more than one dimension remains, and fixing the non unique integrals will not further reduce the dimensionality. It is of course still true that specification of all the constants of integration, c_i , ($i = 1 \dots 2f-1$), will determine the orbit completely, but it is only the constants fixing univalued integrals which affect the long time averages. Whatever values we choose for the others, — because of the infinite multivaluedness of any variable — the orbit must interweave the remaining phase Ω space densely. We could also say the univalued integrals decompose the phase space into invariant parts $g_1(\mathbf{P}) = c_1, \dots, g_m(\mathbf{P}) = c_m$. The intersection of these, which is of dimension $2f-1-m \dots$, is indecomposable.

A simple illustration is afforded by any completely separable many body problem *e.g.* the Conditionally Periodic Systems of the old quantum theory. For these, the action variables constitute a complete set of f univalued integrals. If the frequencies are incommensurable, a remaining f -dimensional space of the coordinates is indecomposable. A two dimensional projection of the orbit will then have the character of a Lissajou figure. It should be remembered, however, that separable systems are not ergodic in the sense of statistical thermodynamics; since they have no transfer of energy between the different degrees of freedom and therefore cannot approach thermodynamical equilibrium.

The question is how many univalued integrals are in general to be expected. At any rate the energy $E(\mathbf{P}) = E$ will always be one. It is perhaps a reasonable guess that in large systems there are essentially * no others. This would leave the widest possible region, namely the energy shell indecomposable.

Apart from the hypothesis of transitivity — which is possibly a strong one — only well established theorems of dynamics are needed in the proof, *viz.* Liouville's theorem: that the motion of the elements of phase space is an incompressible flow, and Poincaré's theorem: that any state, $\mathbf{P}$ in $\Delta\Omega$, will recur. These three premises will be used repeatedly in the following.

* Strictly speaking, also the total linear momentum and the total angular momentum are always univalued integrals, but they can be disregarded for special reasons.

Fig. 1.

Fig. 2.

3. TWO LEMMAS

In order to avoid digression in the main proof, it will be useful to derive two auxiliary formulae beforehand.

3.1 The life time of a state: $\{\mathbf{P}$ in $\Delta\Omega_i\}$

When the representative point $\mathbf{P}$ traverses a particular cell $\Delta\Omega_i$ in phase space the duration of the passage depends strongly upon how it hits the cell, whether centrally or grazingly, see Fig. 1.

The mean duration or "life time" τ_i must be equal to the mean segment $\bar{l}$ of the orbit inside $\Delta\Omega_i$ — divided by the velocity v :

$$\tau_i = \frac{\bar{l}}{v(\mathbf{P}_i)} \quad (8)$$

where

$$v = [\dot{q}_1^2 + \dot{q}_2^2 + \dots + \dot{p}_f^2]^{1/2}$$

In the course of long times the orbit will intersect $\Delta\Omega$ with a dense array of parallel chords. Introducing now by an obvious geometrical analogy the "projection" σ of $\Delta\Omega$ along v (see Fig. 1) we have for the mean chord

$$\bar{l} = \frac{\Delta\Omega}{\sigma} \quad (9)$$

and consequently:

$$\tau = \frac{\Delta\Omega}{\sigma v} \quad (10)$$

3.2. The recurrence time T of $\mathbf{P}$ in $\Delta\Omega$.

Consider now an area σ orthogonal to the velocity field v in phase space.

In a time t a phase volume $\sigma \cdot v \cdot t$ will have passed the cross section σ . At latest after a time $\Omega/\sigma v$, recurrences must have taken place.

Some points may be delayed, but others will be advanced and it is a natural guess that the mean recurrence time is just

$$T = \frac{\Omega}{\sigma \cdot v} \quad (11)$$

We shall substantiate this a little: Denote the number of phase points which have passed σ once by n_0 , twice by n_1 and so forth. The total number of recurrences in a time t will then be:

$$N(t) = 1 \cdot n_1 + 2 \cdot n_2 + \dots \quad (12)$$

while the total flow points is

$$\sigma \varrho v t = n_0 + 1 \cdot n_1 + 2 \cdot n_2 + \dots \quad (13)$$

where ϱ is a constant density to be fixed later. Hence the mean interval between two recurrences is

$$\frac{t}{N} = \frac{t}{\sigma \varrho v t - n_0} \quad (14)$$

Now again by Poincaré's theorem the class of points n_0 cannot increase *i.e.* all orbits crossing σ once must sooner or later cross it again. Therefore in the course of long time

$$\frac{t}{N} = \frac{\lambda}{\sigma \varrho v} \quad (15)$$

We must now speak more definitely about the density ϱ . It may be looked upon as representing an ensemble of $\mathcal{N} = \varrho \cdot \Omega$ systems, uniformly distributed in phase. We might therefore also write

$$\frac{t}{N} = \frac{\Omega}{\sigma v} \cdot \frac{1}{\mathcal{N}} \quad (16)$$

Now since the number of recurrences as well as the number of points crossing σ in unit time is proportional to $\mathcal{N}$, the conclusion seems certain that $\mathcal{N} = 1$ or a density $\varrho = 1/\Omega$ gives the appropriate average count of recurrences in one single system. This leads then to the formula (21).

4. THE PROOF

In the integral (4) we shall understand the representation point $\mathbf{P}$ to move according to a definite Hamiltonian, transitive in Ω , so that we have incompressibility, recurrence and indecomposability. The first step is to replace the time integral by its Riemann sum:

$$\int_0^t f(t) \, dt = \lim_n \sum f(t_n) \, \Delta t_n \quad (17)$$

Here the subdivision of the time interval is arbitrary as long as

$$t = \sum \Delta t_n = \text{constant} \quad (18)$$

We may in particular choose the time differences Δt in (18) equal to the transit times of $\mathbf{P}$ through the cells $\Delta\Omega_i$ of a fixed subdivision of Ω . Now by Poincaré's theorem every cell will be traversed many times. Let Δt_{ir} be the time spent by the representation point inside $\Delta\Omega_i$ during the r 'th passage. And let $N_i(t)$ be the number of recurrences in the time interval $(0, t)$. We have then

$$\sum_n f(t_n) \Delta t_n = \sum_{i=0}^{\Omega/\Delta\Omega} \sum_{r=0}^{N_i} f(\mathbf{P}_i) \cdot \Delta t_{ir} \quad (19)$$

In the second sum we may write

$$\sum_{r=0}^{N_i} f(\mathbf{P}_i) \Delta t_{ir} = f(\mathbf{P}_i) \cdot N_i \overline{\Delta t_i} \quad (20)$$

where $\overline{\Delta t_i}$ is the average transit time over N_i recurrences.

After these preparations we can rewrite the time average of a phase function $f(\mathbf{P})$:

$$\frac{1}{t} \int_0^t f(\mathbf{P}(t)) dt = \lim_{\Delta\Omega \rightarrow 0} \sum_{i=0}^{\Omega/\Delta\Omega} f(\mathbf{P}_i) \cdot \frac{\overline{\Delta t_i}}{t/N_i} \quad (21)$$

According to the proposition (6) we have still to consider the limit $t \rightarrow \infty$ and this must be carried out before making the cells $\Delta\Omega$ infinitesimal.

Now $\overline{\Delta t_i}$ and t/N_i must obviously approach definite limits for $t \rightarrow \infty$:

$$\begin{aligned} \lim \overline{\Delta t_i} &= \tau_i = \text{mean life time of the state: } \{\mathbf{P} \text{ in } \Delta\Omega_i\} \\ \lim t/N_i &= T_i = \text{recurrence time} \end{aligned}$$

Thus we have for the time average of f over an infinite period:

$$f^* = \lim_{\Delta\Omega \rightarrow 0} \sum f(\mathbf{P}_i) \frac{\tau_i}{T_i} \quad (22)$$

But according to the formulae (10) and (11), the ratio of the life time to the recurrence time of the state: $\mathbf{P}$ in $\Delta\Omega_i$ is

$$\frac{\tau_i}{T_i} = \frac{\Delta\Omega_i}{\Omega} \quad (23)$$

Therefore we have

$$\sum_i f(\mathbf{P}_i) \frac{\tau_i}{T_i} = \sum_i f(\mathbf{P}_i) \frac{\Delta\Omega_i}{\Omega}$$

and passing to the limit in eqn. (22)

$$\lim_{i \rightarrow \infty} \frac{1}{t} \int_0^t f(\mathbf{P}(t)) dt = \frac{1}{\Omega} \int_{\Omega} f(\mathbf{P}) d\Omega$$

as was to be proved.

REFERENCES

1. See, ter Haar, D. *Rev. Mod. Phys.* **27** (1955) 289, for full references.
2. Delbrück, M. and Molière, G. *Preuss. Akad. d. Wissensch. Phys. Math. Klasse* **1936**, No. 1.
3. For example, Forsythe, R. *Differential Equations*, 6. Ed. London 1933, p. 394.

Received February 20, 1958.