Gibbs Ensembles, Equidistribution of the Energy of Sympathetic Oscillators and Statistical Models of Thermostat

V. V. Kozlov*

V. A. Steklov Institute of Mathematics, Russian Academy of Sciences,
ul. Gubkina 8, Moscow, 119991 Russia

Received June 13, 2007; accepted October 28, 2007

Abstract—The paper develops an approach to the proof of the “zeroth” law of thermodynamics. The approach is based on the analysis of weak limits of solutions to the Liouville equation as time grows infinitely. A class of linear oscillating systems is indicated for which the average energy becomes eventually uniformly distributed among the degrees of freedom for any initial probability density functions. An example of such systems are sympathetic pendulums. Conditions are found for nonlinear Hamiltonian systems with finite number of degrees of freedom to converge in a weak sense to the state where the mean energies of the interacting subsystems are the same. Some issues related to statistical models of the thermostat are discussed.

MSC2000 numbers: 37A60, 60K35, 70H05, 82B30, 40A9

DOI: 10.1134/S1560354708030015

Key words: Hamiltonian system, sympathetic oscillators, weak convergence, thermostat

1. INTRODUCTION

The proof of the “zeroth” law of thermodynamics (an isolated system tends to the state of thermal equilibrium) is a key problem of the nonequilibrium statistical mechanics. Among other things this problem includes the analysis of equalizing of the temperature of the interacting subsystems of a single isolated system. These issues are intimately related to the study of the system’s trend to distribute uniformly the energy among the degrees of freedom. Within the realm of equilibrium statistical mechanics this property immediately follows from the canonical Gibbs distribution. Another problem from this field is the problem of thermostat that consists in proving that a system with thermostat tends to the state of statistical equilibrium in which the system’s temperature tends to that of the thermostat. Here the term thermostat should be interpreted in statistical sense, that is, a system with a great number of degrees of freedom; the thermostat interacts with the system under consideration via a “weak”-coupling mechanism.

The problem of equipartition of energy among the degrees of freedom of an oscillating system was considered in the classical work by Fermi, Pasta and Ulam [1]. They analyzed a one-dimensional array of \(N\) identical particles with forces between neighbors containing nonlinear terms. Contrary to the expectations, for \(N = 64\) the system’s energy did not show tendency towards equipartition of energy in the Fourier modes, instead the system periodically returned back very close to its original condition. However, there is nothing surprising in such a behavior of the system because it follows from Poincaré’s theorem on recurrences that the energies of distinct particles oscillate in time and of course do not tend to any fixed values as time goes to infinity. Nevertheless, Figure 9 from [1] clearly shows that the time averages of these energies irreversibly tend to their limit values.

N. N. Bogolubov ([2], Chapter IV) seems to be the first to give an example of statistical model of thermostat. Consider a system \(S\) which is an ordinary harmonic oscillator with the Hamiltonian

\[ H_S = \frac{p^2 + \omega^2 q^2}{2}. \]

*E-mail: kozlov@pran.ru
The thermostat is modeled as a system of \( N \) independent oscillators (\( N \) is large) with the Hamiltonian

\[
H_{\Sigma} = \frac{1}{2} \sum_{j=1}^{n} (p_j^2 + \omega_j^2 q_j^2).
\]

The Hamiltonian describing the interaction is assumed to be a quadratic form

\[
H_{\Sigma} = \varepsilon \sum_{j=1}^{n} \alpha_j q_j q,
\]

where \( \alpha_j = \text{const} \) and \( \varepsilon \) is a small parameter. If \( \alpha_j < 0 \), then for small \( \varepsilon > 0 \) the perturbation \( H_{\Sigma} \) can be rewritten in a more familiar manner as the potential energy of coupled oscillators

\[
\frac{\varepsilon}{2} \sum |\alpha_j| (q_j - q)^2.
\]

In so doing, however, the frequencies \( \omega \) and \( \omega_j \) (as well as \( \varepsilon \)) are changed by a small quantity.

At the initial time the state of the oscillator with Hamiltonian \( H_{\Sigma} \) is assumed to be determinate, while the oscillators of the “large” system \( \Sigma \) are governed by the Gibbs distribution with one and the same temperature \( T \). It is shown in [2] that if \( N \to \infty \) and \( \varepsilon \) is small, then eventually (under some general conditions) the system \( S \) gets arbitrarily close to the equilibrium statistical state with the same temperature \( T \).

There are quite a number of works concerned with study of dynamical models of thermostat (Gaussian thermostat, Nosé–Hoover thermostat). References to the original works and discussion can be found in the survey reports in [3]. In the context of these models, the equations of motion are augmented with non-Hamiltonian or dissipative terms so that the distribution density tends to the canonical Gibbs distribution (or the total energy of the system tends to a fixed value). Being just phenomenological and not statistical, such models are not discussed here.

2. THE ZEROTH LAW OF THERMODYNAMICS

Prior to discussion of the topics mentioned, we need a rigorous definition of statistical (thermal) equilibrium of a dynamical system. Our starting point in this is the theory of Gibbs ensembles.

Let \( \Gamma \) be the phase space of the dynamical system

\[
\dot{z} = v(z), \quad z \in \Gamma.
\]

Its phase flow \( \{g^t\} \) (\( t \) is time) preserves a measure \( d\mu \) that will be referred to as phase volume. For Hamiltonian systems \( d\mu \) is the Liouville measure. On \( \Gamma \) we also introduce a probabilistic measure \( d\nu_t = \rho_t d\mu \) whose density \( \rho_t \in L_1(\Gamma, d\mu) \) and satisfies the Liouville equation

\[
\frac{\partial \rho_t}{\partial t} + \text{div} (\rho_t v) = 0.
\]

Since the measure \( d\mu \) is preserved (\( \text{div} v = 0 \)), it follows that \( \rho_t \) is an integral of the dynamical system (1). Therefore,

\[
\rho_t(z) = \rho_0 \left( g^{-t}(z) \right),
\]

where \( \rho_0 \) is the initial probability density in \( \Gamma \).

Already Gibbs attempted to prove that as \( t \to \infty \) the density \( \rho_t \) converges (in some sense) to a stationary density \( \overline{\rho} \), thereby indicating the onset of statistical equilibrium. In the ordinary sense, the limit does not exist because in the generic case the density \( \rho_t \) as a function of \( t \) exhibits undamped oscillations.

However, the situation is not that desperate: the traditional convergence can be replaced by a stronger convergence, say, in the sense of Cesàro:

\[
\overline{\rho}(z) = \lim_{t \to \infty} \frac{1}{t} \int_0^t \rho_u(z) du = \lim_{t \to \infty} \frac{1}{t} \int_0^t \rho_0 \left( g^{-u}(z) \right) du.
\]
According to an ergodic theorem due to Birkhoff and Khinchin the function $\bar{p}$ has the following properties: 1) it is defined almost everywhere, 2) it is invariant under $\{g^t\}$, 3) $\bar{p} \in L_1(\Gamma, d\mu)$ and 4) if $\mu(\Gamma) < \infty$, then the measure $d\nu = \bar{p}d\mu$ is a probability measure, that is,

$$\int_{\Gamma} \bar{p}d\mu = \int_{\Gamma} \rho t d\mu = 1.$$  \hspace{1cm} (3)

Of course, $\mu(\Gamma) = \infty$ for Hamiltonian systems. Nevertheless, equation (3) still holds provided that the levels of constant energy are compact.

**Definition 1.** We say that a statistical dynamical system $(\Gamma, v, d\mu, \rho_t)$ converges to an equilibrium system $(\Gamma, v, d\mu, \overline{\rho})$, if the density $\overline{\rho}$ satisfies (2).

It should be noted that the value of (2) is almost everywhere the same both for $t \to +\infty$ and $t \to -\infty$. Therefore, the states of the statistical equilibrium “in the future” and “in the past” are identical. This simple observation agrees well with the property of reversibility of the dynamical system (1) and distinguishes our approach from the traditional understanding of reversibility, based on the properties of the Boltzmann kinetic equation. It is well known yet that the Boltzmann equation is incompatible with the property of reversibility of the original equations.

It should be emphasized that instead of the convergence in the sense of Cesàro (2) we can use any other linear and regular summation method that includes the Cesàro method as a partial case (e.g., the Abel summation method). In this connection the convergence of the system to a statistical equilibrium in accord with Definition 1 does not imply that one is supposed to replace the density $\rho_t$ with its average over $[0, t]$ and then let time $t$ go to infinity. Let us clarify this point.

Consider a dynamical function $\varphi: \Gamma \to \mathbb{R}$; its average at time $t$ reads

$$\int_{\Gamma} \rho_t \varphi d\mu.$$  \hspace{1cm} (4)

If $\varphi \in L_p$ (for this it suffices to assume that $\rho_0 \in L_p$), then for the integral (4) to exist it should be assumed that $\varphi \in L_q\left(\frac{1}{p} + \frac{1}{q} = 1\right)$. For $p = 1$ the function $\varphi$ must be essentially bounded. For example, let $\varphi$ be the characteristic function of a bounded measurable domain $D$ in the configuration space of a Hamiltonian system; then $\varphi$ can be extended to a measurable function defined on the whole phase space. In this case, the integral (4) gives the fraction of Hamiltonian systems from the Gibbs ensemble that lie in $D$ at time $t$.

**Definition 2.** A measure $d\nu_t = \rho_t d\mu$, ($\rho_0 \in L_p$) is said to converge weakly to a measure $d\nu = \overline{\rho}d\mu$ if

$$\int_{\Gamma} \rho_t \varphi d\mu \to \int_{\Gamma} \overline{\rho} \varphi d\mu$$  \hspace{1cm} (5)

for any function $\varphi$ from $L_q$.

In such a case we will also say that the density $\rho_t$ converges weakly to the density $\overline{\rho}$ as $t \to \infty$.

The following a priori statement holds true: if $\rho_t$ converges weakly to $\overline{\rho}$, then the function $\overline{\rho}$ is given by equation (2). For $p = q = 2$ this statement was proved in [4] and in the general case in [5].

A definition of a statistical (thermal) equilibrium in terms of weak convergence of the probabilistic measure $d\nu_t$ is given in [6]. The definition provides a natural transition from micro- to macro-description when the focus is on the evolution of the average (and thus most probable) values of dynamical functions while the evolution of the functions themselves is of no importance. Within the framework of such an approach it makes sense to discuss the convergence only to some stationary average values of dynamical functions and not to the probability density function. Summing up: the probability density function can be said to manifest itself, for the most part, in calculating averages rather than being considered “an individual entity”.

It turns out that a weak convergence takes place under some additional conditions, which are presented in [4] in terms of spectral properties of the one-parametric group of unitary operators.
generated by the phase flow of the system (1). However, some classes of dynamical systems possess weak convergence, e.g. systems with laminated flows [7]. Among these are Hamiltonian systems with homogeneous potential of degree $m \neq 2$. The proofs are based on new modifications of ergodic theorems [7, 8] and here is an example:

Suppose that $\mu(M) < \infty$ and $f_1, f_2$ lie in $L_2(\Gamma, d\mu)$; then

$$
\lim_{t \to \infty} \frac{1}{\mu(\Gamma)} \int_\Gamma \exp \left( \frac{i^2}{2\sigma^2} f_1 (g^t(z)) f_2(z) \right) d\mu dt = \int_\Gamma f_1 f_2 d\mu,
$$

(6)

where $T_1$ is the Birkhoff average value (2) of the function $f_1$. In particular, if the dynamical system (1) is ergodic (but not necessarily mixing), then the right-hand side of (1.6) can be factorized as

$$
\frac{1}{\mu(\Gamma)} \int_\Gamma f_1 d\mu \int_\Gamma f_2 d\mu.
$$

Thus, as the variance $\sigma^2$ grows infinitely the functions $f_1 (g^t(z))$ and $f_2(z)$ become (on the average) statistically independent: the integral of the product of the two functions equals the product of their integrals. Historically, ergodic theory grew out of statistical mechanics as a product of attempts to prove Boltzmann’s physical ideas on thermal equilibrium.

Unfortunately, in linear oscillating systems (the potential energy is a homogeneous quadratic form: $m = 2$) weak convergence of probabilistic measures never occurs: as a rule, the integrals (4) change harmonically and do not have a limit as $t \to \infty$.

To improve the situation, we should use a more restrictive definition of weak convergence (5) in which it is additionally required that

$$
\lim_{t \to \infty} \frac{1}{t} \int_0^t \left( \int_\Gamma \rho_s \varphi d\mu \right) ds = \int_\Gamma \overline{\rho} \varphi d\mu.
$$

(7)

**Definition 3.** We shall say that $\rho_t$ converges weakly to $\overline{\rho}$ in the sense of Cesàro, if for any trial function $\varphi$ the equality (7) holds.

**Theorem 1.** If $\rho_0 \in L_p$ ($p \geq 1$), then $\rho_t$ always converges weakly in the sense of Cesàro to $\overline{\rho} \in L_p$ defined by (2).

This theorem is a simple consequence of well-known ergodic theorems and some new ideas from [5], Theorem 1 shows that the definitions of statistical equilibrium (2) and (7) ($\rho_t$ is replaced with the stationary function $\overline{\rho}$) are equivalent. Thus, we have a blanket definition of statistical equilibrium, which, for example, applies to degenerate linear Hamiltonian systems.

Of course, for Hamiltonian systems the limit density $\overline{\rho}$ is not, generally speaking, identical with the density of the canonical Gibbs distribution. Yet there is nothing surprising in it. The Gibbs distribution contains the absolute temperature, which indeed makes some sense when the dynamics of interacting subsystems is analyzed: only in this case the temperature of various systems can be compared. A statistical derivation of the Gibbs distribution is an independent problem, which is closely connected to chaotization in an ensemble of weakly interacting Hamiltonian systems. Within such an approach, the idea of *thermodynamical limit* plays an important role. The Boltzmann–Gibbs gas is a good example. The gas is a system of $N$ identical small balls ($N$ is large); the balls collide elastically with each other and with the walls of the rectangular box. If the system is assumed to be ergodic (even not necessarily mixing) on constant energy levels, then the weak limit of the probability density (Definition 2) is a measurable function, which depends only on the total energy of the gas. Under some assumptions (of analytical rather than physical nature) on can prove that for large $N$ the velocities of the particles will (asymptotically) follow a Maxwell distribution [9].

It should be noted that the formula (2) for the density of stationary probability distribution was discussed already by N.N. Bogolubov in his famous report [10]. There he stressed that the ergodic theory of Hamiltonian systems must incorporate and extensively use the concept of thermodynamic limit in its further development. Sharing Bogolubov’s point of view, we show that systems with finite (even small) number of degrees of freedom can exhibit intriguing statistical properties.
3. EQUIDISTRIBUTION OF THE ENERGY BETWEEN SYMPATHETIC PENDULUMS

Consider two identical oscillators connected with an elastic spring (sympathetic pendulums). This is a linear Hamiltonian system with quadratic Hamiltonian function of the form

\[ H = \frac{1}{2}(p_1^2 + q_1^2) + \frac{1}{2}(p_2^2 + q_2^2) + \frac{\varepsilon}{2}(q_1 - q_2)^2, \quad \varepsilon = \text{const} > 0. \]  

(8)

For simplicity, we assume their natural frequencies to be 1. An important feature of such a system is the presence of partial solutions characterized by energy transfer between the pendulums. Here we study a statistical counterpart of this phenomenon.

Let \( \rho_0(p_1, q_1, p_2, q_2) \) be an initial probability density in the four-dimensional phase space. The only restriction imposed on \( \rho_0 \) is that it is a non-negative summable function and the energy mean value

\[ E = \int_{\mathbb{R}^4} H\rho_0 d^2pd^2q \]  

(9)

exists. If we replace \( \rho_0 \) in this formula with the solution to the Liouville equation with this initial condition, then evidently, the total energy remains unchanged. Consider the mean energies of the oscillators

\[ E_j(t) = \frac{1}{2} \int_{\mathbb{R}^4} (p_j^2 + q_j^2)\rho_t d^2pd^2q, \quad j = 1, 2. \]  

(10)

Since the integral (9) is bounded, these integrals exist for any \( t \). However, unlike the total energy (9), they are independent of time. Since the functions \( p_j^2 + q_j^2 \) are not bounded in \( \mathbb{R}^4 \), speaking formally, Theorem 1 does not apply. However, the following theorem is valid.

**Theorem 2.** The limits

\[ \lim_{t \to \infty} \frac{1}{t} \int_0^t E_j(s)ds = \overline{E}_j, \quad j = 1, 2, \]

exist and are equal; moreover, as \( \varepsilon \to 0 \) \( \overline{E}_1 = \overline{E}_2 = \frac{E}{2} \).

Thus, whatever the initial probability density, the mean energies of the oscillators are asymptotically \( (t \to \infty) \) equal. Surprisingly, the system of connected oscillators is not ergodic. Moreover, the system is integrable because there is a quadratic integral of motion functionally independent with the energy \( H \). The fact that the limit mean values of the energies coincide can be treated as equalization of the temperatures of the subsystems due to an arbitrarily weak interaction. As it will be shown in the proof, the rate of the Cesàro convergence of the functions (10) is reduced with decreasing \( \varepsilon \). It should be emphasized that the temperature equalizing occurs without any passage to the thermodynamic limit.

The proof of Theorem 2 is based on the following simple proposition valid for dynamical systems of the general type with invariant measure.

**Proposition 1.** The following formula is valid

\[ \int_{\Gamma} f(g^{-t}(z)) \varphi(z)d\mu = \int f(z)\varphi(g^t(z))d\mu. \]  

(11)

This formula follows from the theorem on change of variables in a multiple integral and the invariance of \( d\mu \) under the transformation \( g^t \).
Proof (of Theorem 2). With the help of an orthogonal transformation, we introduce “normal” coordinates \( P_j, Q_j \) instead of the canonical \( p_j, q_j \)

\[
\begin{align*}
Q_1 &= \frac{q_1 + q_2}{\sqrt{2}}, & P_1 &= \frac{p_1 + p_2}{\sqrt{2}}, & Q_2 &= \frac{q_1 - q_2}{\sqrt{2}}, & P_2 &= \frac{p_1 - p_2}{\sqrt{2}}.
\end{align*}
\]

(12)

The transformation is canonical. In terms of the new variables the Hamiltonian reads

\[
H = \frac{1}{2}(P_1^2 + Q_1^2) + \frac{1}{2}(P_2^2 + \omega^2 Q_2^2), \quad \omega^2 = 1 + 2\varepsilon > 1,
\]

(13)

and the formulas (10) now look like

\[
\begin{align*}
4E_1 &= \int_{\mathbb{R}^4} [((P_1 + P_2)^2 + (Q_1 + Q_2)^2)\rho_t(P, Q)d^2Pd^2Q, \\
4E_2 &= \int_{\mathbb{R}^4} [((P_1 - P_2)^2 + (Q_1 - Q_2)^2)\rho_t(P, Q)d^2Pd^2Q.
\end{align*}
\]

(14)

Here \( \rho_t(P, Q) \) is the probability density written in terms of the new variables. In (14) it is taken into account that the Jacobian of a canonical transformation is 1.

Now let us make use of (11). The function \( \rho_t(P, Q) \) turns into \( \rho_0(P, Q) \), while \( P, Q \) in square brackets in (14) should be replaced with the solution of the linear Hamiltonian system with the Hamiltonian (13) with the initial conditions \( P, Q \):

\[
\begin{align*}
P_1 &\mapsto P_1 \cos t - Q_1 \sin t, & Q_1 &\mapsto P_1 \sin t + Q_1 \cos t, \\
P_2 &\mapsto P_2 \cos \omega t - \omega Q_2 \sin \omega t, & Q_2 &\mapsto \frac{P_2}{\omega} \sin \omega t + Q_2 \cos \omega t.
\end{align*}
\]

Since the functions \( \sin^2 t, \sin^2 \omega t, \ldots \) converge in the sense of Cesàro as \( t \to \infty \) and the functions \( \sin t \sin \omega t, \sin t \cos \omega t, \ldots \) converge to zero (because \( \omega > 1 \)), we see that

\[
\lim_{t \to \infty} \frac{1}{t} \int_0^t 4E_1(s)ds = P_1^2 + Q_1^2 + \frac{P_2^2}{2} \left( 1 + \frac{1}{\omega^2} \right) + \frac{Q_2^2}{2} \left( 1 + \omega^2 \right).
\]

(15)

Here the bar denotes averaging with respect to the measure \( \rho_0 d^2P d^2Q \). An identical formula can be obtained for \( E_2 \). This proves the theorem. \( \square \)

As an illustrative example, suppose that the initial probability density is the product of two normal densities, that is,

\[
\rho_0 = \frac{1}{(\sqrt{2\pi kT_1})^2} e^{-\frac{(p_1^2 + q_1^2)}{2kT_1}} \frac{1}{(\sqrt{2\pi kT_2})^2} e^{-\frac{(p_2^2 + q_2^2)}{2kT_2}}.
\]

Thus, at the initial time \( t = 0 \) the states of the oscillators are assumed to be statistically independent and obey the Gibbs distribution. In particular, \( T_1 \) and \( T_2 \) can be interpreted as the absolute temperatures of these one-degree-of-freedom systems.

It is easy to verify that at the initial time the mean energies of the sympathetic oscillators are

\[
kT_1 \quad \text{and} \quad kT_2
\]

respectively, and the mean potential energy of the stretched spring reads

\[
\varepsilon k \frac{T_1 + T_2}{2}.
\]

(16)

(17)

One can find that as \( t \to \infty \)

\[
E_j(t) \to \bar{E}_j = \frac{k(T_1 + T_2)}{2} \frac{2 + 4\varepsilon + \varepsilon^2}{1 + 2\varepsilon} \quad (j = 1, 2).
\]
(in the sense of Cesàro), and the mean potential energy tends to
\[ \Pi = \frac{k(T_1 + T_2) \varepsilon (1 + \varepsilon)}{2(1 + 2\varepsilon)}. \]

Of course, the sum \(E_1^2 + E_2^2 + \Pi\) coincides with mean total energy at the initial time (the sum of the three quantities given by (16) and (17)). Obviously, as \(\varepsilon \to 0\)
\[ E_j \to kT, \quad T = \frac{T_1 + T_2}{2}. \]

So, in the presence of infinitely small interaction the temperature of each oscillator tends to the arithmetic mean of their temperatures at the initial time.

The ratio
\[ \frac{\Pi}{E_j} = \frac{2(\varepsilon^2 + \varepsilon)}{\varepsilon^2 + 4\varepsilon + 2} \]
increase monotonically from 0 to 2 when \(\varepsilon\) varies within the interval \([0, \infty)\). In particular, for asymptotically large values of the spring coefficient the mean energy of two oscillators is equal to the mean energy of the spring. In view of (15), this proposition is valid irrespective of the initial probability density.

4. MULTIDIMENSIONAL ANALOG OF A SYSTEM OF SYMPATHETIC PENDULUMS

One multidimensional generalization of sympathetic oscillators can be achieved through the use of real orthogonal \(n \times n\) matrices whose components are of the same magnitude. Such matrices exist for \(n = 2^k, k \geq 1\) and can be built inductively as follows:

\[
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1
\end{pmatrix},
\begin{pmatrix}
1 & 1 & 1 & 1
1 & -1 & 1 & -1
1 & -1 & 1 & -1
1 & -1 & 1 & -1
\end{pmatrix},
\ldots
\]

(18)

Of course, the matrix components should be each time divided by \(2^{k/2}\). The block structure of the matrices is as follows:

\[
\begin{pmatrix}
A & A \\
A & -A
\end{pmatrix},
\]

where \(A\) is the previous matrix from the sequence (18). By the way, all the matrices (18) are symmetric.

Let \(\|a_{ij}\|\) be an orthogonal \(n \times n\) matrix from the sequence (18). Using this matrix we construct the Hamiltonian function
\[
H = \frac{1}{2} \sum_{j=1}^{n} (p_j^2 + q_j^2) + \varepsilon_1 \frac{1}{2} \left( \sum_{k=1}^{n} a_{1k} q_k \right)^2 + \ldots + \varepsilon_n \frac{1}{2} \left( \sum_{k=1}^{n} a_{nk} q_k \right)^2,
\]

with non-negative real \(\varepsilon_1, \ldots, \varepsilon_n\) and thereby obtain a Hamiltonian system with \(n\) degrees of freedom.
What is the physical meaning of the Hamiltonian? Let $\varepsilon_1 = 0$ and assume $\varepsilon_2, \ldots, \varepsilon_n$ to be very small but almost identical numbers. The Hamiltonian (19) can be rewritten as

$$
H = \frac{1}{2} \sum (p_j^2 + \omega_j^2 q_j^2) + \frac{1}{2} \sum_{i<j} \varepsilon_{ij} (q_i - q_j)^2,
$$

where the frequency $\omega$ and the positive coefficients $\varepsilon_{ij}$ are known functions of $\varepsilon_2, \ldots, \varepsilon_n$. Thus, we have $n$ weakly coupled linear identical oscillators. The case $n = 2$ corresponds to classic sympathetic pendulums.

Let $\rho_0 \in L_1(\mathbb{R}^{2n}, d^n p d^n q)$ be the initial probability density function and

$$
\int_{\mathbb{R}^{2n}} H \rho_0 d^n p d^n q < \infty. \tag{20}
$$

Put

$$
E_j(t) = \frac{1}{2} \int_{\mathbb{R}^{2n}} (p_j^2 + q_j^2) \rho_t d^n p d^n q \quad (1 \leq j \leq n) \tag{21}
$$

where $\rho_t$ is the solution to the Liouville equation with $\rho_0$ as the initial condition. By virtue of (20), the mean energies $E_j$ are well defined for all $t$.

**Theorem 3.** If the numbers $\varepsilon_1, \ldots, \varepsilon_n$ are all different and the condition (20) satisfied, then

$$
\lim_{t \to \infty} \frac{1}{t} \int_0^t E_j(s) ds
$$

exist and coincide.

Conceptually, the proof parallels the proof of Theorem 2 (which in its turn is a straightforward corollary of Theorem 3). Using a linear canonical transformation, introduce new canonical coordinates $P, Q$

$$
P_j = \sum a_{jk} p_k, \quad Q_j = \sum a_{jk} q_k; \quad j = 1, \ldots, n.
$$

In terms of the new variables the Hamiltonian (19) reads

$$
H = \frac{1}{2} \sum (P_j^2 + \omega_j^2 Q_j^2), \quad \omega_j^2 = 1 + \varepsilon_j > 0. \tag{22}
$$

By assumption, all the frequencies $\omega_1, \ldots, \omega_n$ are different.

Making the change of variables $p, q \mapsto P, Q$ in (21) and using (11), we arrive at

$$
E_j(t) = \frac{1}{2} \int_{\mathbb{R}^{2n}} \left[ \left( \sum a_{kj} P_k \right)^2 + \left( \sum a_{kj} Q_k \right)^2 \right] \rho_0(P, Q) d^n P d^n Q,
$$

where $P, Q$ in the square brackets must be replaced with the solution to the Hamiltonian system (22) with $P, Q$ as the initial conditions. Because the frequencies $\omega_1, \ldots, \omega_n$ are all different, we get

$$
\lim_{t \to \infty} \frac{1}{t} \int_0^t E_j(s) ds = \frac{1}{4} \left[ \sum a_{kj}^2 P_k^2 + \sum a_{kj}^2 Q_k^2 \right].
$$

Since $a_{kj}^2 = \text{const}$ (are independent of $k$ and $j$), these Cesàro means are independent of $j$.

Do linear Hamiltonian systems with Hamiltonian function (19) have partial solutions with strongly pronounced energy transfer of the bouncing type? To answer this question, consider the solution with the initial conditions

$$
q_j(0) = 0 \quad (j \geq 1); \quad \dot{q}_1(0) = v \neq 0, \quad \dot{q}_j(0) = 0 \quad (j \geq 2),
$$

where $v$ is a constant. The mean energy $E_j$ is

$$
E_j(t) = \frac{1}{2} \left( \sum a_{kj}^2 P_k^2 + \sum a_{kj}^2 Q_k^2 \right).
$$

Since $a_{kj}^2 = \text{const}$, these Cesàro means are independent of $j$.
or in the new variables

\[ Q_j(0) = 0, \quad \dot{Q}_j(0) = \frac{v}{\sqrt{n}} \quad (j \geq 1). \]

Therefore,

\[ Q_j(t) = \frac{v}{\sqrt{n}} \sin \omega_j t \quad (j \geq 1). \]

And finally,

\[ q_i(t) = \frac{v}{\sqrt{n}} \sum_{k=1}^{n} a_{ki} \sin \omega_k t, \]
\[ \dot{q}_i(t) = \frac{v}{\sqrt{n}} \sum_{k=1}^{n} a_{ki} \cos \omega_k t, \quad (23) \]

where \( a_{ki} = \pm \frac{1}{\sqrt{n}}. \)

Consider the typical case where \( \varepsilon_1, \ldots, \varepsilon_n \) are so chosen that the frequencies \( \omega_1, \ldots, \omega_n \) are not rationally related. The trajectory

\[ t \mapsto \varphi_k(t) = \omega_k t, \quad 1 \leq k \leq n, \]

is everywhere dense (and even uniformly distributed) on the \( n \)-dimensional torus \( \mathbb{T}^n = \{ \varphi_1, \ldots, \varphi_n \mod 2\pi \}. \) Therefore, the trajectory approaches each point on the torus as close as is wished. In particular, there exist sufficiently large \( t \) so that \( \cos \varphi_k \approx \sqrt{n} a_{ki} (= \pm 1). \) In view of (23), at these times \( \dot{q}_i \approx v. \) Taking into consideration the structure of the orthogonal matrices (18), it is easy to notice that the velocities of the other particles are very small: \( \dot{q}_j \approx 0 (j \neq i). \) Moreover, if \( \varepsilon_1, \ldots, \varepsilon_n \) are small, then over a sufficiently long period of time only the \( i \)-th pendulum will oscillate while the others will remain practically immovable. This phenomenon occurs repeatedly as the weakly coupled pendulums change their roles. Unlike the case \( n = 2, \) the energy transfers between pendulums not in periodic but in quasi-periodic manner.

5. MECHANISMS OF EQUIPARTITION OF MEAN ENERGY

This section deals with some geometrical and topological considerations that promote better understanding and generalization of the analytical calculations from Sections 3 and 4.

Let \( S \) be a diffeomorphism of the phase space \( \Gamma \) that preserves the invariant measure and commutes with the phase flow transformation \( \{ g^t \}. \) If \( f: \Gamma \to \mathbb{R} \) is a measurable function, then \( f_S \) is a function such that \( z \mapsto f(Sz). \) If \( f_S = f, \) the function \( f \) will be called \( S \)-invariant.

The following trivial proposition holds true.

**Proposition 2 (symmetry principle).** Let the initial probability density function \( \rho_0 \) be \( S \)-invariant. Then the mean values

\[ \int_{\Gamma} f \rho_t d\mu \quad \text{and} \quad \int_{\Gamma} f_S \rho_t d\mu \]

are equal for all \( t. \) In particular, \( \overline{f_S} = \overline{f} \) for all \( t. \)

**Proof.** We make the change of variables \( z \mapsto Sz \) in the first of the integrals (24); in this case \( f \) transforms into \( f_S, \) while the measure \( d\mu \) and function \( \rho_t \) remain the same. The last statement can be justified as follows:

\[ \rho_t(Sz) = \rho_0(g^{-t}(Sz)) = \rho_0(S(g^{-t}(z))) = \rho_0(g^{-t}(z)) = \rho_t(z). \]

\[ \square \]
As an example, consider a Hamiltonian system with
\[ H = \sum h(p_j, q_j) + \varepsilon \sum_{i<j} V(q_i - q_j). \] (25)

Here \( h \) is a function of two variables and the potential \( V \) is an even function of one variable. The Hamiltonian (25) governs the dynamics of identical one-dimensional weakly-coupled subsystems. The function (25) is invariant under transformations of \( \mathbb{R}^{2n} \) that are permutations of the pairs of canonically conjugate variables. Obviously, these transformations preserve the Liouville measure and commute with the phase flow of the Hamiltonian system.

In the theory of Bogolubov lattices it is customary to assume that the initial density \( \rho_0 \) does not change under any permutation of the coordinates and momenta of individual particles. This invariance is sometimes interpreted as a consequence of the principle of indistinguishability of identical particles adopted in the classical statistical mechanics. However, this is an additional assumption. In this case, by Proposition 2, the mean total energies of the subsystems are equal.

Nevertheless, of greater interest is the case when \( \rho_0 \) is not symmetric at the initial time. How and under what conditions does the density \( \rho \) gain the property of symmetry? Let now \( S \) be an automorphism of the space with measure \((\Gamma, d\mu)\).

**Theorem 4.** Suppose that the dynamical system (1) has \( k < n \) independent integrals of motion \( f_1, \ldots, f_k \) such that

1) they are \( S \)-invariant,

2) the integral manifolds \( M_c = \{z \in \Gamma: f_1(z) = c_1, \ldots, f_k(z) = c_k\} \) are connected for almost all \( c = (c_1, \ldots, c_k) \in \mathbb{R}^k \),

3) for almost all \( c \in \mathbb{R}^k \) the dynamical system (1) is ergodic on \( M_c \).

Then the Birkhoff average \( \overline{\rho} \) of any initial density \( \rho_0 \in L_p(\Gamma, d\mu), p \geq 1 \) is an \( S \)-invariant function from \( L_p \).

If \( f \in L_q(\Gamma, d\mu), \frac{1}{p} + \frac{1}{q} = 1 \), then under the assumptions of Theorem 4

\[
\int_{\Gamma} f \overline{\rho} d\mu = \int_{\Gamma} f_S \overline{\rho} d\mu. \] (26)

Here we make use of the assumption that the Liouville measure is \( S \)-invariant, the \( S \)-invariance of the density \( \overline{\rho} \) (Theorem 4) and the formula for change of variables in a multiple integral.

Theorem 4 is easy to prove. Since the integrals \( f_j \) are \( S \)-invariant, then so are the surfaces \( M_c \); if \( z \in M_c \), then \( Sz \in M_c \) and vice versa. Because the system (1) is ergodic on almost all \( M_c \), the density \( \overline{\rho} \) takes one and the same value almost everywhere on each connected component \( M_c \). By virtue of the topological assumption 2, the manifolds \( M_c \) are connected. Hence, for almost all \( z \in \Gamma \) we have \( \overline{\rho}(Sz) = \overline{\rho}(z) \).

For Hamiltonian systems the assumptions of Theorem 4 are most clear in the two limiting cases: \( k = \frac{\dim \Gamma}{2} \) and \( k = 1 \). In essence, the first case is closely related to completely integrable Hamiltonian systems: if \( n = 2k \), then the \( k \) independent integrals must be in involution and, therefore, each compact connected component of their common level surface is a \( k \)-dimensional torus filled with conditionally-periodic solution curves. If additionally the system is non-degenerate, then almost all invariant tori are non-resonant and obviously, the Hamiltonian system is ergodic on such tori. Thus, when \( k = \frac{\dim \Gamma}{2} \) the only essential requirement is that \( M_c \) are connected.

For \( k = 1 \), the assumptions of Theorem 4 can be summarized as follows: 1) the \((2n-1)\)-dimensional surfaces of constant energy are connected and 2) on these surfaces the Hamiltonian...
system is ergodic. The case \( k = 1 \) plays an important role in the development of the statistical model of thermostat (Section 6).

Let us show how to obtain Theorem 2 from Theorem 4 in the typical case when the frequency \( \omega = \sqrt{1 + 2 \varepsilon} \) is irrational. The linear system with Hamiltonian (8) has two independent integral of motion

\[
H \quad \text{and} \quad F = (p_1 + p_2)^2 + (q_1 + q_2)^2.
\]

These functions do not change under the substitution

\[ p_1, q_1 \leftrightarrow p_2, q_2. \]

The two-dimensional surfaces

\[ \{ H = c_1, F = c_2 \} \tag{27} \]

are always connected. To prove this, it suffices to notice that in the normal coordinates \( P, Q \) the surfaces (27) are given by the equations

\[
P_1^2 + Q_1^2 = C_1, \quad P_2^2 + \omega^2 Q_2^2 = C_2; \quad C_1, C_2 = \text{const}.
\]

If \( C_1 \) and \( C_2 \) are different from zero, then obviously, these equations define a two-dimensional torus in \( \mathbb{R}^4 \).

Since \( \omega \) is irrational, the linear Hamiltonian system is ergodic on the tori (27). Thus, all the assumptions of Theorem 4 are satisfied and therefore the means

\[
\frac{1}{2} \int_{\mathbb{R}^4} (p_1^2 + q_1^2) \rho d^2 p d^2 q \quad \text{and} \quad \frac{1}{2} \int_{\mathbb{R}^4} (p_2^2 + q_2^2) \rho d^2 p d^2 q
\]

are equal.

Similarly, Theorem 3 can be obtained from Theorem 4 under the additional assumption that the frequencies \( \omega_j = \sqrt{1 + \varepsilon_j} \) \((j = 1, \ldots, n)\) are not rationally related. However, Theorem 3 holds under a weaker condition that there are no equal frequencies.

One final remark in conclusion. It would be erroneous to conjecture that the integrals of a Hamiltonian system with \( S \)-invariant Hamiltonian are also \( S \)-invariant. For example, for \( \varepsilon = 0 \) the system with Hamiltonian (8) admits the integral \( p_1 q_2 - p_2 q_1 \), whose sign changes with interchanging the particles.

### 6. GIBBS ENSEMBLES AND STATISTICAL MODELS OF THERMOSTAT

In this final section we will discuss statistical models of thermostat in the context of the general theory of Gibbs ensembles. The ideas presented here are mostly of preliminary nature and require a more thorough elaboration.

We start with some simple things to improve our understanding of the general idea of thermostat model. Consider a standard "natural" Hamiltonian \( H = T + \varepsilon V \), where

\[
T = \frac{1}{2} \sum_{j=1}^{n} \frac{p_j^2}{2m_j}
\]

is the kinetic energy, \( V(q) \) is the potential energy and \( \varepsilon \) is a small parameter. Then we make a crucial assumption that this Hamiltonian system is \textit{ergodic} on the surfaces of constant energy \( \{ H = \text{const} \} \). Therefore, any probability density \( \rho_t \) converges in the sense of Cesàro to the density \( \overline{\rho} \), which is a function of energy.

Thus it is natural to assume that

\[
\overline{\rho} = \frac{f(\beta H)}{\int \! f(\beta H) \, d\mu}, \tag{28}
\]

where \( f \) is a measurable function and \( \beta \) is a constant factor with the dimensions of inverse energy. The function \( f \) itself can also depend on \( \beta \).
It is easy to verify that the mean values of the summands in the expression for the kinetic energy
\[ E_j = \int_\Gamma \frac{p_j^2}{2m_j} \rho d^n p d^n q \quad (1 \leq j \leq n) \] (29)
are equal, if the density is given by (28). The proof is as follows: upon substitution \( p_j \mapsto \sqrt{m_j} \tilde{p}_j \) the value of the integral (29) no longer depends on the masses \( m_1, \ldots, m_n \). Moreover, this is so for \( \varepsilon \neq 0 \) as well.

Suppose that the mean value of the potential energy \( \varepsilon V \) tends to zero as \( \varepsilon \to 0 \). This certainly is so when the configuration space is compact and \( V \) is bounded. Then, for small \( \varepsilon \), the mean total energy
\[ \int_\Gamma H \rho_0 d\mu = \int_\Gamma H \rho d\mu \]
approximately equals \( nE \) [where \( E \) is the integral (29)] and its value does not depend on \( j \).

Now we return to the statistical model of thermostat. Under the conditions listed above, our “large” system with \( n \gg 1 \) degrees of freedom is augmented by adding \( k \) one-dimensional subsystems in such a way that the new system with \( n + k \) degrees of freedom continues to be ergodic on the manifolds of constant energy.

So, let
\[ H_{n+k} = H_n + H_k + \varepsilon W_{n,k} \]
be the Hamiltonian of the enlarged system; here \( H_n(H_k) \) is the Hamiltonian of the \( (n(k)) \)-degree-of-freedom system and \( \varepsilon W_{n,k} \) is the potential energy of interaction between these systems. The function \( \varepsilon W_{n,k} \) is assumed to be measurable and bounded. Suppose that at the initial time the systems with Hamiltonians \( H_n \) and \( H_k \) were statistically independent and in the state of thermal equilibrium. This means that the initial probability density function is the product
\[ \rho^{(n+k)} = \rho^{(n)} \rho^{(k)}, \]
where \( \rho^{(n)} \) (\( \rho^{(k)} \)) is a summable function that depends on \( H_n \) (\( H_k \)) only. We have
\begin{align*}
\int H_{n+k} \rho^{(n+k)} d^n p^{n+k} d^n q \\
= \int H_n \rho^{(n)} d^n p^n d^n q + \int H_k \rho^{(k)} d^k p^k d^k q + O(\varepsilon) \\
= nE_n^{-} + kE_k^{-} + O(\varepsilon).
\end{align*}
In the last line, the notation is self-evident.

When the whole system is at the state of statistical equilibrium, the Birkhoff mean \( \rho^{(n+k)} \) is a function of \( H_{n+k} \) (by our assumption). Moreover, the integral (30) equals to
\[ (n + k)E_{n+k}^{+} + O(\varepsilon), \]
where \( E_{n+k}^{+} \) is the mean kinetic energy per one degree of freedom of the enlarged system.

Since the total energy is preserved, for small \( \varepsilon \) we get
\[ nE_n^{-} + kE_k^{-} \approx (n + k)E_{n+k}^{+}. \] (31)
Now suppose that \( nE_n^{-} \gg kE_k^{-} \), meaning that the energy of the thermostat ("heat capacity") far exceeds the energy of the added system. Let \( n \) go to \( \infty \) while \( k \) is fixed. Then it follows from (31) that
\[ \frac{n}{n + k} E_n^{-} \to E_n^{-} \approx E_{n+k}^{+}. \]
Therefore, at the state of statistical equilibrium the mean kinetic energy per one degree of freedom of the added system equals the mean kinetic energy of the one-dimensional subsystems that make
up the thermostat. In other words, the temperature of the added system becomes equal to the temperature of the thermostat.

There is one key point left to discuss: consider a Hamiltonian system of the form (25). Under what conditions does the stationary probability density \( \varrho \) depend only on the energy of the system? It is not to be supposed that in other way or other the answer reduces only to the study of the system’s ergodicity on the manifolds \( \{ H = \text{const} \} \). The problem is essentially deeper and more intriguing: of crucial importance is the smoothness of the function \( \varrho \).

The matter is that \( \overline{\varrho} \) is an integral of motion of the Hamiltonian system. On the other hand, it is a well-known fact that in the general case, the integrals of an analytic Hamiltonian system are not analytic in the whole phase space [11]. (Of course, the integrals and the Hamiltonian are assumed to be functionally independent.)

The main difficulty, however, is that even for an analytic system of differential equations and an analytic initial density \( \varrho_0 \) the Birkhoff mean \( \overline{\varrho} \) can be discontinuous on an everywhere dense set of zero measure. A simple example is a non-degenerate integrable Hamiltonian system with compact isoenergetic manifolds: as a rule, the function \( \varrho_0 \) is discontinuous at the points on the resonant invariant tori in this sense resembling the classical Riemann function, which is continuous at irrational points and discontinuous at rational points. However, the values the function \( \varrho \) takes on a zero-measure set are of no importance. In this example, the values of \( \overline{\varrho} \) at the points on the resonant tori can be adjusted so that it becomes analytic in the whole phase space.

So, consider an analytic Hamiltonian of the form (25). Suppose that the density \( \varrho \) is analytic in \( 2n \) variables \( p, q \) and also depends on \( \varepsilon \). Recall that the density of the canonical Gibbs distribution has this property. Then, in the general case, \( \varrho \) is a function of total energy. Exact conditions are given in [11] and the proof is based on the mechanism of destruction of resonant invariant tori of the unperturbed system discovered by Poincaré. Therefore, in this case the analysis of the model of thermostat considered above can be applied.

Of course, the density \( \varrho \) can be only continuous or even essentially discontinuous (but summable) function. If so, are there any significant consequences? Unfortunately, this question has been very little explored so far, so it only remains for us to advance some hypotheses.

First, it seems likely that for large \( n \) and small \( \varepsilon \) the perturbed Hamiltonian system with Hamiltonian (25) must be transitive when the energy is fixed: there is at least one trajectory that is everywhere dense. This is one of the exact formulations of Arnold’s hypothesis about diffusion in multidimensional systems. If this is so, then even a continuous density \( \overline{\varrho} \) will be a function of the system’s total energy.

Finally, there is one more plausible conjecture: let the value of \( \varepsilon \) be fixed and sufficiently small, then, for sufficiently large \( n \) a typical Hamiltonian system of the form (25) is ergodic on isoenergetic manifolds. Unfortunately, there is very little reliable information on this subject.

It conclusion, it would be pertinent to note that, according to N. N. Bogolubov [10], the behavior of a system for no matter how large but fixed \( n \) is of little interest, instead the aim is to explore the system’s behavior on the macro-level in the thermodynamic limit, where the number of degrees of freedom \( n \) and the “room” occupied by the system both go to infinity in a prescribed manner.

REFERENCES