Progress of Theoretical Physics, Vol. 103, No. 1, January 2000

Thermodynamic Irreversibility from High-Dimensional Hamiltonian Chaos

Shin-ichi SASA*) and Teruhisa S. KOMATSU**)

Department of Pure and Applied Sciences, University of Tokyo
Tokyo 153-8901, Japan

(Received November 9, 1999)

This paper discusses the thermodynamic irreversibility realized in high-dimensional Hamiltonian systems with a time-dependent parameter. A new quantity, the irreversible information loss, is defined from the Lyapunov analysis so as to characterize the thermodynamic irreversibility. It is proved that this new quantity satisfies an inequality associated with the second law of thermodynamics. Based on the assumption that these systems possess the mixing property and certain large deviation properties in the thermodynamic limit, it is argued reasonably that the most probable value of the irreversible information loss is equal to the change of the Boltzmann entropy in statistical mechanics, and that it is always a non-negative value. The consistency of our argument is confirmed by numerical experiments with the aid of the definition of a quantity we refer to as the excess information loss.

§1. Introduction

Thermodynamics formalizes a fundamental limitation of possible processes between equilibrium states. In particular, when a thermodynamic system is enclosed by adiabatic walls, the limitation is represented by, for example, a fact that, given a system in some initial state, it is not possible to lower the system's energy by first changing some of its other extensive variables and then returning them to their original values. Contrastingly, the energy of the system can be increased by the similar change of the other extensive variables. These two facts make clear the special nature of energy as an extensive variable. This asymmetry is the basis of *thermodynamic irreversibility*.

Thermodynamics is one of the most elegant theories being based on only a few fundamental principles.[1)] However, one may wonder how its principles emerge out of purely mechanical systems. Thermodynamic systems consist of many molecules, whose dynamics are described by Hamiltonian equations. Thus, in the idealized limit of adiabatic walls, a thermodynamic system can be regarded as a Hamiltonian system that is connected to some mechanical apparatus, but does not contact a heat reservoir. With this in mind, it may be natural to expect that thermodynamic irreversibility can be formalized in Hamiltonian systems.

Thermodynamic entropy plays a central role in the description of thermodynamic irreversibility, and the thermodynamic entropy is generally thought to be given by the logarithm of the number of micro-states. This relation, the Boltzmann formula, seems well-established as far as the calculation of equilibrium values is concerned.

*) E-mail: sasa@jiro.c.u-tokyo.ac.jp
**) E-mail: komatsu@jiro.c.u-tokyo.ac.jp

However, it has not yet been shown that the Boltzmann formula provides a complete account of irreversibility. [2)]

In this paper, we discuss thermodynamic irreversibility based on the nature of high-dimensional Hamiltonian chaos. As our most notable result, we find a new quantity which satisfies an inequality associated with thermodynamic irreversibility. We define this quantity, the "irreversible information loss", from dynamical system considerations. Furthermore, we argue reasonably that the irreversible information loss is related to the change of the Boltzmann entropy, and this leads us to conclude that the Boltzmann entropy does not decrease for any processes in the thermodynamic limit.

1.1. *Related studies*

This paper provides theoretical arguments for numerical results reported in a previously published paper, [3)] and contains a detailed description of the numerical experiment.

The present study has been carried out under the influence of several related studies. First, the attempt to construct steady state thermodynamics by Oono and Paniconi has provided the direction of the present study. [4)] They have proposed an operational method to obtain non-equilibrium thermodynamic functions. In addition, from a more general viewpoint, they emphasize the importance of theory concerning the relation between two different states. Following this manner of thinking, we have set out to study thermodynamic irreversibility from dynamical systems.

The stochastic energetics proposed by Sekimoto has given a nice example of the construction of thermodynamics from dynamical systems. [5)] Stochastic energetics formalizes energy transformation in Langevin dynamics with a clear distinction between heat and work. Sekimoto and Sasa have demonstrated the minimum work principle and defined the free energy from this principle. [6)] Their argument also includes a complementary relation which defines a new thermodynamic function of two state variables. Recently, Sekimoto and Oono have constructed an example of steady state thermodynamics by analyzing a Langevin dynamical model. [7)]

Jarzynski has proposed a new method to obtain the minimum work principle. [8)] In his approach, an equality is first proven, and then the inequality related to the minimum work principle is derived by using the Jensen inequality. A similar idea may be applied to other related problems. In fact, Hatano has proved a Jarzynski-type equality for the transition between steady states under a certain condition and has derived an inequality related to the steady state thermodynamics. [9)]

As discussed by Crooks, [10)] the Jarzynski equality is also related to the fluctuation theorem proposed by Evans, Cohen and Morriss. [11)] The fluctuation theorem claims a peculiar property of the probability of the finite time average of the entropy production in a non-equilibrium steady state. Gallovotti and Cohen have presented a mathematical proof of the fluctuation theorem based on the assumption that the steady state measure is given by the dynamical measure. [12)] Since that time, it has been shown that the fluctuation theorem holds even in stochastic systems. [13),14)] On general grounds, Maes has presented an argument that the fluctuation theorem can be understood by a Gibbs property of the space-time measure. [15)]

Transportation coefficients in non-equilibrium steady states have been expressed in terms of dynamical system quantities. There are two different approaches for this. In one approach, the viscosity is related to the sum of all Lyapunov exponents in a Hamiltonian system supplemented with a deterministic thermostatting force. [16),17)] The other approach applies to Hamiltonian systems with open boundary conditions. Here, the diffusion constant is related to the escape rate which is obtained in terms of the difference between the sum of the positive Lyapunov exponents and the Kolmogorov-Sinai entropy. [18)]

Finally, we mention some recent developments in the understanding of thermodynamics. Lieb and Yngvason wrote an important paper on axiomatic thermodynamics. [1)] They have given an explicit expression of the thermodynamic entropy based on a set of axioms concerning the adiabatic accessibility, and have proven the entropy principle, the second law of thermodynamics. Although their formulation is fully mathematical, the idea of the explicit expression of thermodynamic entropy can be translated into standard energetic thermodynamics. [19)-21)]

1.2. *Outline of the paper*

This paper consists of nine sections, each of which consists of several subsections. In order to give a self-contained explanation, we include a review of thermodynamics, Hamiltonian systems, Boltzmann entropy, and Lyapunov analysis. Some of these are no doubt rather well-known topics to specialists. However, there are not a large number of people who understand all of them well. Also, it was our intention to write this paper so that it can be understood by non-specialists, who have interest in the relation between thermodynamic irreversibility and dynamical systems. The organization of the paper is summarized below.

In §2 we start with a review of thermodynamics in an adiabatic environment. Thermodynamic irreversibility is precisely defined based on basic notions such as state and process. The essence of the thermodynamic entropy is described by the entropy principle. [1)] We then explain the reason why a Hamiltonian system with a time-dependent parameter provides a model for a thermodynamic system in an adiabatic environment. We assume the microcanonical measure for the initial conditions and that the systems possess the mixing property with respect to the measure. We also assume the existence of certain large deviation properties in order to establish correspondence with the extensivity of thermodynamics. [22)] Based on these assumptions, we define the equilibrium state and most probable process in Hamiltonian systems. After these preliminaries, we address a main question.

In §3 we first review the Boltzmann formula in statistical mechanics. In particular, defining the time-dependent Boltzmann entropy, we derive a simple form of the Boltzmann entropy change for general processes. Using this formula, we calculate the average of the Boltzmann entropy change for a step process, where the average is taken over the initial conditions sampled from the microcanonical ensemble on an energy surface. We show that the average value is positive in the thermodynamic limit. We further find that the average value is related to the fluctuation of the Boltzmann entropy change.

In §4 we review the Lyapunov analysis, which is a standard method to study

chaotic dynamical systems with numerical experiments. We start with the Gram-Schmidt decomposition, because it is the easiest computational technique for the Lyapunov analysis.[23] We then discuss the convergence property of an orthogonal frame. Since the orthogonal frame obtained from convergence does not satisfy the transitive property, we define Lyapunov vectors from the orthogonal frame so that this property is satisfied.[24] Based on these Lyapunov vectors, we define Lyapunov exponents, local expansion ratios, and the information loss rate. In order to recover the symmetry of unstable and stable directions, we also define contraction ratios. We then prove a relation between the expansion and contraction ratios. We also derive an expression for the weight on trajectory segments.

In §5 we discuss the reversibility of Hamiltonian systems. We relate the evolution map, Lyapunov vectors, and local expansion rates for time-reversed systems to those for the original system. The reversibility leads to a reversibility paradox.[25] In order to resolve the paradox, we need to consider the measure for a set of the initial conditions for time-reversed systems. This consideration leads to a reversibility relation expressed in terms of probability.

In §6 we begin with the definition of *irreversible information loss*. Using the reversibility relation mentioned above, we prove that the irreversible information loss averaged over the initial conditions is always non-negative. We define the most probable value of the irreversible information loss in the thermodynamic limit, and we present an argument that this most probable value is equal to the Boltzmann entropy change.

In §7 we define a quantity we call *excess information loss*, because this quantity is more tractable than the irreversible information loss. We present a relation between the Boltzmann entropy change and the excess information loss based on the assumption that the reversible part of the excess information loss is equal to the quasi-static excess information loss. This relation is identical to an equality proposed in a previous paper.[3] Furthermore, we briefly discuss a minimum excess information loss principle, which may be analogous in some sense to the minimum work principle in thermodynamics with an isothermal environment. We also explain the origin of the quasi-static excess information loss using Lyapunov analysis.

In §8 we report results of numerical experiments on a Fermi-Pasta-Ulam model[26] with a time-dependent nonlinear term. With these, we numerically check the assumptions in the arguments given in the previous sections, and numerically demonstrate several properties of certain quantities such as the Lyapunov exponents and Boltzmann entropy changes in this model. As the main numerical experiment, we confirm the relation between the Boltzmann entropy change and the excess information loss.

The final section is devoted to concluding remark.

1.3. *Remarks*

Our theoretical arguments include some non-rigorous, but intuitively reasonable statements. To as great an extent as possible, we state explicitly when the assertions we make are assumptions. There is one exception, however. We often use the expression $o(N)$ to represent a quantity of negligible magnitude compared to N in

the limit $N \to \infty$. This constitutes an order estimate valid in the case that the system satisfies an appropriate condition. However, we do not discuss what this condition is, nor do we explicitly state that an assumption is involved when we neglect such a quantity. We simply expect that the condition is satisfied unless an abnormal situation occurs.

We use the same font for numbers and vectors. We believe that the difference can be understood in the context. Also, a matrix is expressed as $\mathcal{A}$, and A_{ij} denotes the (ij)-element of this matrix.

§2. Preliminaries

In this section, we review thermodynamics and Hamiltonian systems. We clarify basic assumptions of our theory and address the main question of this paper.

2.1. *Thermodynamic irreversibility*

A thermodynamic system is characterized by the internal energy U and a set of work variables $\{X_i\}$. When the value of X_i is changed externally, an energy change is induced. An infinitely small response dU is written as

$$dU = \sum_i Y_i dX_i. \tag{2·1}$$

In thermodynamics, X_i is chosen as an extensive or intensive variable. Then, since the internal energy U is an extensive variable, Y_i is an intensive or extensive variable, respectively. The relation Eq. (2·1) is valid for the case that the system is enclosed by adiabatic walls. More formally, Eq. (2·1) should be regarded as a mathematical expression of the physical situation that the system is placed in an adiabatic environment.

The equilibrium state Σ is assumed to be realized when the system is left for a sufficiently long time after values of the work variables are fixed. This assumption provides the operational definition of the equilibrium state. Also, the equilibrium state Σ is assumed to be determined uniquely by the set of the values of $(U, \{X_i\})$. That is, the state Σ is identified with $(U, \{X_i\})$. When the value of X_i is changed externally, a transition from one equilibrium state Σ_0 to another one Σ_1 occurs. This transition, which is denoted by $\Sigma_0 \overset{\mathrm{a}}{\to} \Sigma_1$, is called a *thermodynamic process* or simply a *process*. More precisely, this process is called an adiabatic process realized in an adiabatic environment. However, in the argument below, we use the term "process" instead of "adiabatic process".

Let Σ_0 and Σ_1 be arbitrary equilibrium states. We then ask whether or not processes $\Sigma_0 \overset{\mathrm{a}}{\to} \Sigma_1$ and $\Sigma_1 \overset{\mathrm{a}}{\to} \Sigma_0$ are realizable. When both the processes are realizable, these are called *reversible processes*. When only one process $\Sigma_0 \overset{\mathrm{a}}{\to} \Sigma_1$ is realizable, this process is called an *irreversible process*. We can easily see that the process

$$(U, \{X_i\}) \overset{\mathrm{a}}{\to} (U', \{X_i\}) \tag{2·2}$$

provides an example of an irreversible process when $U' > U$.

The essence of thermodynamic entropy is described by the entropy principle:[1] There exists an extensive variable S given by a function of Σ such that the inequality

$$S(\Sigma_1) \geq S(\Sigma_0) \tag{2·3}$$

is satisfied if and only if a process $\Sigma_0 \xrightarrow{\text{a}} \Sigma_1$ is realizable. The extensive variable S is determined uniquely up to multiplicative and additive arbitrary constants.

Lieb and Yngvason have proved the entropy principle based on axioms concerning the adiabatic accessibility.[1] Also, in conventional thermodynamics based on work and heat, the entropy principle can be proved with some physical assumptions.[20),21)]

2.2. *Hamiltonian systems*

A Hamiltonian system is characterized by a Hamiltonian function $H(\Gamma)$, where Γ is a set of canonical coordinates $\{q_i\}$ and momenta $\{p_i\}$,

$$\Gamma_i = q_i, \tag{2·4}$$
$$\Gamma_{N+i} = p_i, \tag{2·5}$$

where $1 \leq i \leq N$. The equations of motion for q_i and p_i are given by

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \tag{2·6}$$
$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}. \tag{2·7}$$

These equations can be formally written as

$$\frac{d\Gamma}{dt} = -\mathcal{J}\frac{\partial H}{\partial \Gamma}, \tag{2·8}$$

where $\mathcal{J}$ is a $2N \times 2N$ anti-symmetric matrix which satisfies

$$\mathcal{J}^2 = -1. \tag{2·9}$$

Under an initial condition $\Gamma(0)$ given at $t = 0$, the phase space point at time t, $\Gamma(t)$, is determined by the equation of motion.

In this paper, we are concerned with Hamiltonian systems with a time-dependent parameter α. The energy of the system E at time t is given by

$$E(t) = H(\Gamma(t), \alpha(t)). \tag{2·10}$$

We then obtain the equality

$$\frac{dE}{dt} = \sum_{i=1}^{N}\left(\frac{\partial H}{\partial q_i}\frac{dq_i}{dt} + \frac{\partial H}{\partial p_i}\frac{dp_i}{dt}\right) + \frac{\partial H}{\partial \alpha}\frac{d\alpha}{dt} \tag{2·11}$$
$$= \frac{\partial H}{\partial \alpha}\frac{d\alpha}{dt}, \tag{2·12}$$

where we have used the equations of motion. This equality can be rewritten as

$$dE = Ad\alpha, \tag{2·13}$$

where

$$A = \frac{\partial H}{\partial \alpha}. \tag{2·14}$$

Comparing Eq. (2·13) with Eq. (2·1), we find that E and α correspond to the internal energy U and a work variable X. This suggests that a Hamiltonian system with a time-dependent parameter can be a dynamical system model for thermodynamics in an adiabatic environment. We proceed to our discussion based on this expectation and attempt to find necessary properties so as to establish consistency with thermodynamics.

Since we are particularly interested in thermodynamic processes, we assume that the value of α is changed in a finite time interval $[\tau_i, \tau_f]$; that is,

$$\frac{d\alpha(t)}{dt} = 0 \tag{2·15}$$

when $t \notin [\tau_i, \tau_f]$. In the argument below, we assume the condition

$$0 \ll \tau_i \leq \tau_f \tag{2·16}$$

without an explicit remark. Note that $\ll$ in Eq. (2·16) has been assumed for a technical reason. We also represent the protocol of the parameter change by $\alpha(\ )$.

2.3. *Measure*

We assume that the initial condition given at $t = 0$ is sampled from the microcanonical ensemble on an energy surface Σ. The measure for the ensemble is given by the microcanonical measure

$$\mu_{\rm mc}(d\Gamma; \Sigma) = \frac{1}{|\Sigma|}\frac{1}{|\nabla_\Gamma H|}d\sigma, \tag{2·17}$$

where $d\sigma$ is the Lebesgue measure on the energy surface, and $|\Sigma|$ is given by

$$|\Sigma| = \int d\sigma \frac{1}{|\nabla_\Gamma H|}. \tag{2·18}$$

The quantities $\mu_{\rm mc}(\Gamma; \Sigma)$ and $|\Sigma|$ are rewritten as

$$\mu_{\rm mc}(d\Gamma; \Sigma) = \frac{1}{|\Sigma|}d\Gamma\delta(H(\Gamma) - E), \tag{2·19}$$

$$|\Sigma| = \int d\Gamma\delta(H(\Gamma) - E), \tag{2·20}$$

where $d\Gamma$ is the $2N$-dimensional Lebesgue volume element in the phase space. We also note that Eq. (2·19) is given by

$$\mu_{\rm mc}(\Delta_\epsilon(\Gamma); \Sigma) = \lim_{\delta E \to 0} \frac{\mu_L(\Delta_\epsilon(\Gamma) \cap \Sigma \circ \delta E)}{\mu_L(\Sigma \circ \delta E)}, \tag{2·21}$$

where $\Sigma \circ \delta E$ denotes a union of energy surfaces from E to $E + \delta E$, $\Delta_\epsilon(\Gamma)$ is a region with size ϵ which includes the point Γ, and $\mu_{\rm L}$ is the $2N$ dimensional Lebesgue measure.

We also assume that the systems in question are ergodic and possess the mixing property. Here, a system is called "ergodic" with respect to the microcanonical measure when the equality

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T dt f(\Gamma(t)) = \int \mu_{\rm mc}(d\Gamma;\Sigma) f(\Gamma) \tag{2·22}$$

holds for an arbitrary measurable function f and almost all initial conditions $\Gamma(0)$ with respect to the measure. The mixing property with respect to the microcanonical measure means that the equality

$$\lim_{t\to\infty}\int \mu_{\rm mc}(d\Gamma(0);\Sigma) f(\Gamma(0)) g(\Gamma(t)) = \int \mu_{\rm mc}(d\Gamma;\Sigma) f(\Gamma)\int \mu_{\rm mc}(d\Gamma;\Sigma) g(\Gamma) \tag{2·23}$$

holds for arbitrary measurable functions f and g. It is easily proved that a mixing system possesses ergodicity.

Suppose that the initial condition at time $t = -t'$ $(t' > 0)$ is sampled from an ensemble with the measure

$$\mu_f(d\Gamma(-t');\Sigma) = \mu_{\rm mc}(d\Gamma(-t');\Sigma) f(\Gamma(-t')), \tag{2·24}$$

where f is a measurable function normalized in such a way that

$$\int_\Sigma \mu_{\rm mc}(d\Gamma(-t');\Sigma) f(\Gamma(-t')) = 1. \tag{2·25}$$

Then, the mixing property leads to

$$\lim_{t'\to\infty}\int \mu_f(d\Gamma(-t');\Sigma) g(\Gamma(0)) = \int \mu_{\rm mc}(d\Gamma;\Sigma) g(\Gamma). \tag{2·26}$$

That is, the average of $g(\Gamma(0))$ with respect to $\mu_f(d\Gamma(-t');\Sigma)$ is the same as the average of $g(\Gamma(0))$ with respect to $\mu_{\rm mc}(d\Gamma(0);\Sigma)$ when $t' \to \infty$. Using Eq. (2·26), we check numerically the validity of the mixing property and we can prepare the microcanonical ensemble at $t = 0$ in the following way.

First, we prepare a set of initial conditions at $t = -t'$ sampled from an ensemble with a measure absolutely continuous with respect to the Lebesgue measure on the energy surface Σ. (This can be done easily in numerical experiments.) Then, we take the average of a dynamical variable, for example A, at $t = 0$. We carry out two experiments for two different measures assumed at $t = -t'$. If the average values coincide, we may conclude that the system possesses the mixing property.*) Also, from Eq. (2·26), we find that this average value is the average for the microcanonical measure at $t = 0$, $\mu_{\rm mc}(d\Gamma;\Sigma)$. This implies that we can prepare the microcanonical ensemble at $t = 0$.

*) However, precisely speaking, this is nothing but a confirmation of one of the necessary conditions for the mixing property.

2.4. *Thermodynamic limit*

In thermodynamics, the internal energy U is an extensive variable, and a work variable is an extensive or intensive variable. In order to establish consistency with thermodynamics, we assume the following large deviation property,[22] which may be closely related to the extensivity of the energy: Let $\Pi_E(E_1)dE$ be the probability that the final energy after a parameter change takes a value in the region $[E_1, E_1 + dE]$. Then, Π_E can be written in the form

$$\Pi_E(E_1) \sim \exp(-N\phi_E(E_1/N)) \tag{2·27}$$

in the appropriate asymptotic limit, including $N \to \infty$.

Several remarks are made here. (i) The appropriate limit in Eq. (2·27) is called the *thermodynamic limit*. In the argument below, the limit $N \to \infty$ always implies the thermodynamic limit without an explicit remark. (ii) The probability of the final energy is induced from the measure for the ensemble of the initial conditions. (iii) ϕ_E is called a rate function and is a non-negative convex function that takes the value 0 at some point. The zero of ϕ_E, $\bar{E}_{1*}$, is called the most probable value of E_1/N.

We next discuss the extensivity or intensivity of α. We consider particularly the case that α is an intensive parameter. The variable A then turns out to be an extensive variable, which is characterized by the following large deviation property: Let $\Pi_A(A')dA$ be the probability that A takes a value in the region $[A', A' + dA]$ at $t = 0$. Then Π_A can be written in the form

$$\Pi_A(A') \sim \exp(-N\phi_A(A'/N)), \tag{2·28}$$

in the thermodynamic limit.

Note that the probability density Π_A is determined by the measure for the ensemble. Since the most probable value of A/N, $\bar{A}_*$, exists for each energy surface, we write $\bar{A}_*(\Sigma)$ when we wish to emphasize the state dependence.

2.5. *Equilibrium state*

We assume that the equilibrium state in thermodynamics corresponds uniquely to the energy surface. This is the reason we used the same symbol Σ for an energy surface in §2.3 and the equilibrium state in §2.1. Also, the energy surface is specified by the set of quantities (E, α). In the argument below, Σ denotes an equilibrium state, an energy surface, and a set of quantities (E, α).

Let us discuss the condition under which we can know whether or not the equilibrium state is realized. The term 'equilibrium' implies that nothing changes. Thus, it is natural to find a quantity which does not change at equilibrium. Although the energy E does not change when $t \geq \tau_f$, it would be strange for the equilibrium state to be realized immediately after the parameter change in question is completed. Thus, the energy cannot be used as an indicator of the equilibrium state. The next candidate of such an indicator may be the variable A. However, since the trajectory never converges to a fixed point, the value of A remains time dependent. After eliminating such seemingly natural choices, we realize that the argument for the nature of equilibrium should be developed with the thermodynamic limit.

Suppose that $\Gamma(\tau_f) \in \Sigma_1$. In general, $A(\tau_f)/N$ is not equal to $\bar{A}_*(\Sigma_1)$. However, from the large deviation and mixing properties, we can expect

$$\lim_{t\to\infty} \frac{A(t)}{N} \to \bar{A}_*(\Sigma_1) \tag{2·29}$$

in the thermodynamic limit. When $A(t)/N$ is sufficiently close to $\bar{A}_*(\Sigma_1)$ up to a certain time,*) we assume that the state is at equilibrium. There may be other important physical quantities to be checked. However, since we do not have any criteria to judge their importance, we assume that the relaxation of the variable A is enough to identify the equilibrium state.

2.6. *Most probable process*

Suppose that an equilibrium state Σ_0 is realized at $t = 0$ and that another equilibrium state is realized on an energy surface Σ_1 a sufficiently long time after $t = \tau_f$. We call this transition a "process" in analogy to thermodynamics. However, since Σ_1 depends on $\Gamma(0)$, Σ_1 is not determined uniquely when we assume the initial energy surface Σ_0 and the protocol of the parameter change $\alpha(\ )$. Here, in order to establish the correspondence with thermodynamics, we assume the large deviation property of the path of E: Let $\Pi_{E:\mathrm{path}}(\{E'(t), 0 \le t' \le \tau\}) \prod_t dE(t)$ be the path probability that $E(t)$ takes a value in the region $[E', E' + dE(t)]$ at time t. Then, $\Pi_{E:\mathrm{path}}$ can be written in the form

$$\Pi_{E:\mathrm{path}}(\{E'(t), 0 \le t' \le \tau\}) \sim \exp(-N\phi_{E:\mathrm{path}}(\{E'(t)/N, 0 \le t' \le \tau\})) \tag{2·30}$$

in the thermodynamic limit.

The probability density $\Pi_{E:\mathrm{path}}$ is determined by the measure for the ensemble of the initial conditions. The rate function $\phi_{E:\mathrm{path}}$ is a function of path segments $\{E'(t)/N, 0 \le t' \le \tau\}$, and there is a most probable path $\{\bar{E}_*(t'), 0 \le t' \le \tau\}$ that minimizes the rate function. Then, since the parameter is changed in a deterministic way, the most probable process is defined as $\{(N\bar{E}_*(t'), \alpha(t')),\ 0 \le t' \le \tau_f\}$. The most probable process is denoted by

$$\Sigma_0 \to_* \Sigma_1, \tag{2·31}$$

where $\Sigma_1 = (N\bar{E}_*(\tau_f), \alpha(\tau_f))$, and it is identified with the thermodynamic process $\Sigma_0 \xrightarrow{\mathrm{a}} \Sigma_1$.

2.7. *Main question*

Let us summarize our basic assumptions and address the main question. When one attempts to study thermodynamic irreversibility in Hamiltonian systems, there seem to be three problems. The first problem is related to the measure for the initial conditions, where, as one example, determining the condition for the most natural measure is concerned. The second problem is related to the reason why macroscopic variables behave in a deterministic way. The discussion of large deviation properties

*) Formally, we should consider the limit $N \to \infty$ before the limit $t \to \infty$. In an experiment with finite N, we should focus on an asymptotic regime up to a certain time.

is one way to consider this problem. In this paper, we do not consider these problems deeply. As mentioned above, we assume that the Hamiltonian systems in question possess the mixing property with respect to the microcanonical measure, and we also assume the large deviation properties of A and E in the thermodynamic limit. Putting off discussion of such problems, we come to the third important problem: We ask how the thermodynamic law is established. In other words, we ask whether or not we can find a state variable satisfying the entropy principle from dynamical systems. Let us write the question explicitly.

Let $\Sigma_0 \to_* \Sigma_1$ be an arbitrary most probable process. Then, find a state variable S such that

$$S(\Sigma_1) \geq S(\Sigma_0), \tag{2·32}$$

where equality holds only when the reverse process $\Sigma_1 \to_* \Sigma_0$ is realizable.

§3. Statistical mechanics

In statistical mechanics, the thermodynamic entropy is calculated as the Boltzmann entropy. For this reason we review fundamental properties of the Boltzmann entropy and discuss whether or not we can answer to the main question by using the Boltzmann entropy.

The thermodynamic entropy takes a constant value along an arbitrary quasi-static process $\Sigma_0 \overset{\text{qs}}{\to} \Sigma_1$, which is realized by an infinitely slow change of the parameter value. Then, in developing statistical mechanics, we first attempt to find such a quantity.[27] The adiabatic theorem ensures the existence of an invariant quantity along quasi-static processes. We thus start with the adiabatic theorem.

3.1. *Adiabatic theorem*

Let $\Omega(\Sigma)$ be the phase space volume enclosed by an energy surface $\Sigma = (E, \alpha)$:

$$\Omega(E, \alpha) = \int d\Gamma \theta(E - H(\Gamma, \alpha)). \tag{3·1}$$

When the value of α is changed in time, the energy of the system changes. We define the time evolution of the phase space volume as

$$\Omega(t) = \Omega(E(t), \alpha(t)), \tag{3·2}$$

where we note that $E(t)$ depends on $\Gamma(0)$. We then obtain

$$\frac{d\Omega}{dt} = \left[\frac{\partial \Omega}{\partial \alpha} + \frac{\partial \Omega}{\partial E} A\right] \frac{d\alpha}{dt}, \tag{3·3}$$

where we have used the equality

$$\frac{dE}{dt} = A \frac{d\alpha}{dt}, \tag{3·4}$$

which is given by Eq. (2·12).

By using the expression Eq. (3·1), we derive

$$\frac{\partial \Omega}{\partial E} = |\Sigma|, \tag{3·5}$$

$$\frac{\partial \Omega}{\partial \alpha} = -|\Sigma| \langle A \rangle_{\Sigma}, \tag{3·6}$$

where $|\Sigma|$ and $\langle f \rangle_{\Sigma}$ are defined as

$$|\Sigma| = \int d\Gamma \delta(E - H(\Gamma, \alpha)), \tag{3·7}$$

$$\langle f \rangle_{\Sigma} = \frac{1}{|\Sigma|} \int_{\Sigma} \frac{d\sigma}{|\nabla_{\Gamma} H|} f(\Gamma). \tag{3·8}$$

Here, $\langle f \rangle_{\Sigma}$ corresponds to the average of f over the micro-canonical ensemble on the energy surface Σ. Substituting Eqs. (3·5) and (3·6) into Eq. (3·3), we obtain

$$\frac{d\Omega}{dt} = |\Sigma| \left[A - \langle A \rangle_{\Sigma} \right] \frac{d\alpha}{dt} \tag{3·9}$$

$$= |\Sigma| \delta A \frac{d\alpha}{dt}, \tag{3·10}$$

where we have defined a new variable δA as

$$\delta A = A - \langle A \rangle_{\Sigma}. \tag{3·11}$$

We now prove the adiabatic theorem which states that the equality

$$\Omega(\Sigma_0) = \Omega(\Sigma_1) \tag{3·12}$$

holds for an arbitrary quasi-static process $\Sigma_0 \overset{\text{qs}}{\to} \Sigma_1$.

Proof First, set $\Sigma_0 = (E_0, \alpha_0)$ and $\Sigma_1 = (E_\infty, \alpha_\infty)$. Then, we decompose the quasi-static process into n quasi-static processes such that

$$(\alpha_j, E_j) \overset{\text{qs}}{\to} (\alpha_{j+1}, E_{j+1}), \tag{3·13}$$

where $0 \leq j \leq n-1$ and $\alpha_{j+1} = \alpha_j + \Delta\alpha$ with

$$\Delta\alpha = \frac{\alpha_\infty - \alpha_0}{n}. \tag{3·14}$$

Note that $\alpha_n = \alpha_\infty$ and $E_n = E_\infty$. We first assume

$$\Delta\Omega_j = \Omega(E_{j+1}, \alpha_{j+1}) - \Omega(E_j, \alpha_j) \tag{3·15}$$

$$= O((\Delta\alpha)^2) \tag{3·16}$$

for large n. We then obtain

$$\Omega(E_0, \alpha_0) - \Omega(E_\infty, \alpha_\infty) = \lim_{n\to\infty} \sum_{j=1}^{n-1} O((\Delta\alpha)^2) \tag{3·17}$$

$$= \lim_{n\to\infty} O\left(\frac{1}{n}\right) \tag{3·18}$$

$$= 0. \tag{3·19}$$

This proves the adiabatic theorem.

We next prove Eq. (3·16). Without loss of generality, we can assume that the value of α is monotonically changed from α_j to $\alpha_j + \Delta\alpha$. Defining the protocol of the parameter change $\alpha_\tau(\ )$ as

$$\alpha_\tau(t) = \alpha_j + (\Delta\alpha)\frac{t}{\tau}, \tag{3·20}$$

we calculate $\Delta\Omega_j$ from Eq. (3·10) in the following way:

$$\Delta\Omega_j = \lim_{\tau\to\infty} \int_0^\tau dt |\Sigma(t)| \delta A(t) \frac{d\alpha_\tau}{dt} \tag{3·21}$$

$$= \lim_{\tau\to\infty} \frac{(\Delta\alpha)}{\tau} \int_0^\tau dt |\Sigma(t)| \delta A(t). \tag{3·22}$$

When $\Delta\alpha$ is sufficiently small, $\Delta\Omega_j$ is evaluated as

$$\Delta\Omega_j = |\Sigma(0)| \lim_{\tau\to\infty} \frac{(\Delta\alpha)}{\tau} \int_0^\tau dt \delta A(t) + O((\Delta\alpha)^2) \tag{3·23}$$

$$= O((\Delta\alpha)^2), \tag{3·24}$$

where we have used

$$\lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau dt \delta A(t) = 0, \tag{3·25}$$

which is equivalent to

$$\lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau dt A(t) = \langle A \rangle_\Sigma . \tag{3·26}$$

This equality holds for almost all initial conditions with respect to the Lebesgue measure on the energy surface, because of the ergodicity with respect to the micro-canonical measure.

□

3.2. *Boltzmann entropy*

We define the Boltzmann entropy $S_{\rm B}$ as[*)]

$$S_{\rm B}(\Sigma) = \log \Omega(\Sigma), \tag{3·27}$$

where the Boltzmann constant is assumed to be unity. For later convenience, we define the temperature $T(\Sigma)$ as

$$T(\Sigma) = \left(\frac{\partial S_{\rm B}}{\partial E}\right)^{-1} = \frac{\Omega(\Sigma)}{|\Sigma|}. \tag{3·28}$$

[*)] See p. 371 in Boltzmann's book[28)] for an explicit presentation of the Boltzmann formula. However, the expression of Eq. (3·27) was first proposed by Gibbs as the quantity corresponding to the thermodynamic entropy. See p. 128 in Gibbs's book.[29)] The monograph by P. Ehrenfest and T. Ehrenfest is also useful for explanations of contemporary related ideas.[25)]

Although Eq. (3·27) is the formula which makes it possible for us to calculate the thermodynamic entropy for the equilibrium state Σ, we define the time evolution of $S_{\rm B}$ as*)

$$S_{\rm B}(t) = S_{\rm B}(\Sigma(t)). \tag{3·29}$$

Then, from Eqs. (3·10), (3·27) and (3·28), we obtain

$$\frac{dS_{\rm B}}{dt} = \frac{1}{T(\Sigma)}\delta A\frac{d\alpha}{dt}. \tag{3·30}$$

The integration of this equation over the time interval $[0,\tau]$ leads to

$$\Delta S_{\rm B} = \int_0^\tau dt \frac{\delta A(t)}{T(\Sigma(t))}\frac{d\alpha}{dt}. \tag{3·31}$$

Here, we note that $Ad\alpha$ is equal to the energy change ΔE during a time interval $[t, t+dt]$ and that $\langle A\rangle_\Sigma\, d\alpha$ may be interpreted as the quasi-static work, $W_{\rm qs}$, calculated under the condition that the system stays *virtually* on the energy surface. $W_{\rm qs}$ is identical to the work done in an actual process which can be realized when the system contacts a heat bath with a slowly changing temperature. The quasi-static heat $Q_{\rm qs}$ from the heat bath is then given by

$$Q_{\rm qs} = \Delta E - W_{\rm qs} = \delta A d\alpha. \tag{3·32}$$

Using $Q_{\rm qs}$, we rewrite Eq. (3·31) as

$$\Delta S_{\rm B} = \int_0^\tau dt \frac{1}{T}\frac{dQ_{\rm qs}}{dt} \tag{3·33}$$

$$= \int_0^\tau \frac{dQ_{\rm qs}}{T}. \tag{3·34}$$

This should be compared with Clausius's formula

$$\Delta S = \int \frac{d'Q}{T}, \tag{3·35}$$

where $d'Q$ is an infinitely small quasi-static heat exported from a heat bath. In this way, the Boltzmann entropy turns out to be identified with the thermodynamic entropy.

3.3. *Entropy change for step processes*

We discuss the step process given by

$$\frac{d\alpha}{dt} = \delta(t)\Delta\alpha. \tag{3·36}$$

In the argument below, $\Sigma_0 = (E_0, \alpha_0)$ and $\Sigma_1 = (E_1, \alpha_1)$ denote the initial and final states, respectively. By substituting Eq. (3·36) into Eq. (3·31), we have

$$\Delta S_{\rm B} = \frac{1}{2}\left[\frac{\delta A(0_+)}{T(\Sigma_1)} + \frac{\delta A(0_-)}{T(\Sigma_0)}\right]\Delta\alpha, \tag{3·37}$$

*) Do not confuse this with the time evolution of the H-function in the H-theorem of Boltzmann.[25),28)]

where we note that δA and T are discontinuous at $t = 0$. We consider the average over initial conditions sampled from the microcanonical ensemble on the energy surface Σ_0. This average is denoted by $\langle\ \rangle_0$. We calculate $\langle \Delta S_{\rm B}\rangle_0$ as

$$\langle \Delta S_{\rm B}\rangle_0 = \frac{(\Delta\alpha)^2}{2T_0}\left[\frac{1}{\Sigma}\frac{\partial\Sigma}{\partial E}\left\langle(\delta A)^2\right\rangle_0 + \frac{\partial}{\partial E}\left\langle(\delta A)^2\right\rangle_0\right] + o((\Delta\alpha)^2), \tag{3·38}$$

where $T_0 = T(\Sigma_0)$. (The proof is given below.) Further, since the relation

$$\frac{\partial}{\partial E}\left\langle(\delta A)^2\right\rangle_0 = o(N) \tag{3·39}$$

is expected when $N \to \infty$, we obtain

$$\langle \Delta S_{\rm B}\rangle_0 = \frac{(\Delta\alpha)^2}{2T_0^2}\left\langle(\delta A)^2\right\rangle_0 + o(N, (\Delta\alpha)^2) > 0, \tag{3·40}$$

where we have used

$$\frac{1}{\Sigma}\frac{\partial\Sigma}{\partial E} = \frac{1}{T} + O\left(\frac{1}{N}\right). \tag{3·41}$$

In the thermodynamic limit, $\langle \Delta S_{\rm B}\rangle_0$ is equal to the entropy difference $S_{\rm B}(\Sigma_1) - S_{\rm B}(\Sigma_0)$ for the most probable process $\Sigma_0 \to_* \Sigma_1$. Thus, we conclude

$$S_{\rm B}(\Sigma_1) \geq S_{\rm B}(\Sigma_0) + o(N, (\Delta\alpha)^2) \tag{3·42}$$

for the most probable step process $\Sigma_0 \to_* \Sigma_1$.

Furthermore, from Eq. (3·37), the fluctuation $\left\langle(\Delta S_{\rm B})^2\right\rangle_0$ is calculated as

$$\left\langle(\Delta S_{\rm B})^2\right\rangle_0 = \frac{\langle(\delta A)^2\rangle_0}{T_0^2}(\Delta\alpha)^2 + o((\Delta\alpha)^2). \tag{3·43}$$

Combining this result with Eq. (3·40), we obtain the equality

$$\langle \Delta S_{\rm B}\rangle_0 = \frac{1}{2}\left\langle(\Delta S_{\rm B})^2\right\rangle_0 + o((\Delta\alpha)^2, N). \tag{3·44}$$

This is the fluctuation-response relation for the entropy change.

Now, we prove Eq. (3·38).

Proof We first find

$$\left\langle\left[\frac{\delta A(0_+)}{T(\Sigma_1)} + \frac{\delta A(0_-)}{T(\Sigma_0)}\right]\right\rangle_0 = \frac{\langle\delta A(0_+)\rangle_0}{T_0} + O((\Delta\alpha)^2). \tag{3·45}$$

Then, from Eq. (3·37), we have

$$\langle \Delta S_{\rm B}\rangle_0 = \frac{\langle\delta A(0_+)\rangle_0}{2T_0}(\Delta\alpha) + O((\Delta\alpha)^3). \tag{3·46}$$

Let us evaluate $\langle\delta A(0_+)\rangle_0$ up to order $\Delta\alpha$, where

$$\langle\delta A(0_+)\rangle_0 = \left\langle\frac{\partial H}{\partial\alpha}(\Gamma(0_+), \alpha(0_+))\right\rangle_0 - \left\langle\left\langle\frac{\partial H}{\partial\alpha}\right\rangle_{\Sigma_1}\right\rangle_0. \tag{3·47}$$

Since $\Gamma(0_+) = \Gamma(0_-)$, we expand the first term on the right-hand side in such a way that

$$\frac{\partial H}{\partial \alpha}(\Gamma(0_+), \alpha(0_+)) = \frac{\partial H}{\partial \alpha}(\Gamma(0_-), \alpha(0_-)) + \frac{\partial^2 H}{\partial \alpha^2}(\Gamma(0_-), \alpha(0_-))\Delta\alpha + O((\Delta\alpha)^2). \tag{3·48}$$

Taking the average over the initial conditions, we obtain

$$\left\langle \frac{\partial H}{\partial \alpha}(\Gamma(0_+), \alpha(0_+)) \right\rangle_0 = \left\langle \frac{\partial H}{\partial \alpha} \right\rangle_0 + \left\langle \frac{\partial^2 H}{\partial \alpha^2} \right\rangle_0 \Delta\alpha + O((\Delta\alpha)^2). \tag{3·49}$$

We next evaluate the second term on the right-hand side of Eq. (3·47):

$$\left\langle \left\langle \frac{\partial H}{\partial \alpha} \right\rangle_{\Sigma_1} \right\rangle_0 = \left\langle \frac{1}{|\Sigma_1|} \int d\Gamma \frac{\partial H}{\partial \alpha} \delta(H(\Gamma, \alpha_1) - E_1) \right\rangle_0 . \tag{3·50}$$

We note that there are four terms which include $\Delta\alpha$ in the right-hand side of Eq. (3·50): (i) $\Delta\alpha$ appears in $1/|\Sigma_1|$; (ii) it appears in $\partial H/\partial\alpha$ in the integrand; (iii) it appears in $H(\Gamma, \alpha_1)$ in the Dirac delta function; and (iv) it appears in E_1 in the Dirac delta function. We extract the contribution proportional to $\Delta\alpha$ from each term.

(i)For the first term, the contribution is

$$-\frac{1}{|\Sigma|^2} \left. \frac{d|\Sigma|}{d\alpha} \right|_0 \int d\Gamma \frac{\partial H}{\partial \alpha} \delta(H(\Gamma, \alpha_0) - E_0), \tag{3·51}$$

where we have defined

$$\frac{d}{d\alpha} = \frac{\partial}{\partial \alpha} + A \frac{\partial}{\partial E}. \tag{3·52}$$

Here we note that the equality

$$\frac{d\Omega}{d\alpha} = 0 \tag{3·53}$$

holds owing to the adiabatic theorem. This equality and the relation $\Omega = |\Sigma|T$ lead to

$$\left. \frac{d|\Sigma|}{d\alpha} \right|_0 = -\frac{|\Sigma_0|}{T_0} \left. \frac{dT}{d\alpha} \right|_0 . \tag{3·54}$$

Thus, Eq. (3·51) becomes

$$\frac{1}{T_0} \left. \frac{\partial T}{\partial \alpha} \right|_0 \langle A \rangle_0 . \tag{3·55}$$

(ii) Without any calculation, the contribution from the second term is

$$\left\langle \frac{\partial^2 H}{\partial \alpha^2} \right\rangle_0 . \tag{3·56}$$

(iii, iv) In deriving the third and fourth terms, we employ the following formula:

$$\int d\Gamma \delta'(H - E) f(\Gamma) = \int dE' \int_{H=E'} \frac{d\sigma}{|\nabla_\Gamma H|} \delta'(E' - E) f(\Gamma) \tag{3·57}$$

$$= -\frac{\partial}{\partial E'}\left[\int_{H=E'}\frac{d\sigma}{|\nabla_\Gamma H|}f(\Gamma)\right]\Big|_{E'=E} \tag{3·58}$$

$$= -\frac{\partial}{\partial E'}\left[\langle f\rangle_{(E',\alpha)}\,|\Sigma(E',\alpha)|\right]\Big|_{E'=E}. \tag{3·59}$$

Owing to this formula, we can calculate the contribution from the third term:

$$\frac{1}{|\Sigma|_0}\int d\Gamma\left(\frac{\partial H}{\partial\alpha}\right)^2\delta'(H(\Gamma,\alpha_0)-E_0)$$

$$= -\frac{1}{|\Sigma|_0}\frac{\partial}{\partial E}\left[\left\langle\left(\frac{\partial H}{\partial\alpha}\right)^2\right\rangle_\Sigma|\Sigma|\right]\Big|_0 \tag{3·60}$$

$$= -\frac{\partial}{\partial E}\left\langle\left(\frac{\partial H}{\partial\alpha}\right)^2\right\rangle_\Sigma\Big|_0 - \frac{1}{|\Sigma|_0}\frac{\partial|\Sigma|}{\partial E}\Big|_0\left\langle\left(\frac{\partial H}{\partial\alpha}\right)^2\right\rangle_0. \tag{3·61}$$

Similarly, the contribution from the fourth term is obtained as

$$-\frac{1}{|\Sigma|_0}\int d\Gamma\frac{\partial H}{\partial\alpha}\delta'(H(\Gamma,\alpha_0)-E_0)\left\langle\frac{\partial H}{\partial\alpha}\right\rangle_0$$

$$= \frac{1}{|\Sigma|_0}\frac{\partial}{\partial E}\left[\left\langle\frac{\partial H}{\partial\alpha}\right\rangle|\Sigma|\right]\Big|_0\left\langle\frac{\partial H}{\partial\alpha}\right\rangle_0 \tag{3·62}$$

$$= \left\langle\frac{\partial H}{\partial\alpha}\right\rangle_0\frac{\partial}{\partial E}\left[\left\langle\frac{\partial H}{\partial\alpha}\right\rangle\right]\Big|_0 + \frac{1}{|\Sigma|_0}\frac{\partial|\Sigma|}{\partial E}\Big|_0\left\langle\frac{\partial H}{\partial\alpha}\right\rangle^2. \tag{3·63}$$

The contributions from the third and fourth terms can be combined in the form

$$-\frac{1}{\Sigma_0}\frac{\partial\Sigma}{\partial E}\Big|_0\left\langle(\delta A)^2\right\rangle_0 - \frac{\partial}{\partial E}\left\langle(\delta A)^2\right\rangle_0\Big|_0 - \langle A\rangle_0\frac{\partial\langle A\rangle_\Sigma}{\partial E}\Big|_0. \tag{3·64}$$

Then, all the contributions given by Eqs. (3·55), (3·56) and (3·64) are summarized as

$$\left\langle\left\langle\frac{\partial H}{\partial\alpha}\right\rangle_{\Sigma_1}\right\rangle_0 = \left\langle\frac{\partial H}{\partial\alpha}\right\rangle_0 + \left\langle\frac{\partial^2 H}{\partial\alpha^2}\right\rangle_0\Delta\alpha$$

$$- \frac{1}{\Sigma}\frac{\partial\Sigma}{\partial E}\left\langle(\delta A)^2\right\rangle_0\Delta\alpha - \frac{\partial}{\partial E}\left\langle(\delta A)^2\right\rangle_0\Delta\alpha, \tag{3·65}$$

where we have used the equality

$$\frac{1}{T}\frac{\partial T}{\partial\alpha} - \frac{\partial\langle A\rangle_\Sigma}{\partial E} = 0. \tag{3·66}$$

(The proof of this equality is given below.) Substitution of Eqs. (3·49) and (3·65) into Eq. (3·47) yields

$$\langle\delta A(0_+)\rangle_0 = \frac{1}{\Sigma_0}\frac{\partial\Sigma}{\partial E}\Big|_0\left\langle(\delta A)^2\right\rangle_0\Delta\alpha + \frac{\partial}{\partial E}\left\langle(\delta A)^2\right\rangle_\Sigma\Big|_0\Delta\alpha + O((\Delta\alpha)^2). \tag{3·67}$$

Recalling Eq. (3·46), we finally obtain Eq. (3·38).

□

We now give the proof of Eq. (3·66). One may find that Eq. (3·66) is equivalent to a Maxwell's relation.

Proof For simplicity, we use the abbreviation A for $\langle A \rangle_{\Sigma}$. Then, A is a function of $(S_{\rm B}, \alpha)$. Since we can write

$$A = A(S_{\rm B}, \alpha) = \left(\frac{\partial E}{\partial \alpha} \right)_{S_{\rm B}}, \tag{3·68}$$

we calculate

$$\left(\frac{\partial A}{\partial E} \right)_{\alpha} = \left(\frac{\partial A}{\partial S_{\rm B}} \right)_{\alpha} \left(\frac{\partial S_{\rm B}}{\partial E} \right)_{\alpha} \tag{3·69}$$

$$= \frac{1}{T} \frac{\partial^2 E}{\partial S_{\rm B} \partial \alpha} \tag{3·70}$$

$$= \frac{1}{T} \frac{\partial T}{\partial \alpha}. \tag{3·71}$$

□

3.4. *Remark*

In this section, we have shown that $\Delta S_{\rm B}$ is positive for most probable step processes. One may then ask whether or not $\Delta S_{\rm B}$ is positive for arbitrary processes. As one example, one may evaluate $\Delta S_{\rm B}$ near quasi-static processes based on several physical assumptions. However, if we consider this question from the definition of $\Delta S_{\rm B}$, it seems hard to obtain any general results. Nevertheless, since $S_{\rm B}$ is equivalent to the thermodynamic entropy, we expect the inequality

$$\Delta S_{\rm B} \geq o(N) \tag{3·72}$$

to hold for most probable processes in general. We discuss the validity of Eq. (3·72) in §6.2 based on Lyapunov analysis of chaotic systems. We now leave statistical mechanics and enter into discussion on Lyapunov analysis.

§4. Lyapunov analysis

One of essential features of chaotic systems is sensitivity to initial conditions. Consider a trajectory segment, $\{\Gamma(t),\ 0 \leq t \leq \infty\}$. Almost all trajectories starting from phase space points in a neighborhood at $\Gamma(0)$ separate exponentially in time from the trajectory $\Gamma(\ )$. Such behavior can be studied quantitatively by measuring the expansion of vectors in the tangent spaces around the trajectory. More generally, we can discuss the time evolution of the k-dimensional volume element, which is given by the exterior product of k independent vectors in the tangent space. (See the Appendix for the basic properties of the volume element and exterior product.) Such an argument includes Liouville's theorem as a special case ($k = 2N$), which states that the $2N$ dimensional volume element maintains its volume along the trajectory. From the observation for both the cases that $k = 1$ and $k = 2N$, we expect that the tangent space at each point can be decomposed into subspaces associated with

the expansion ratios. Indeed, it is known that the multiplicative ergodic theorem of Oseledets provides a mathematical description of this naive expectation.[30] In recent years, the analysis of tangent spaces, which is often referred to as *Lyapunov analysis*, has become a standard technique to study chaos owing to the establishment of a numerical calculation methods.[23]

In this section, we consider Hamiltonian systems without a parameter change, except for the final two subsections, and we review Lyapunov analysis, emphasizing its computational aspects.

When the value of α is not changed in time, the evolution map from $t = t_0$ to $t = t_1$ takes the form $U_{t_1-t_0}$ and satisfies

$$\Gamma(t) = U_t(\Gamma(0)). \tag{4·1}$$

The change of the trajectory at time t, $\delta\Gamma(t)$, with respect to an infinitely small change of the initial condition, $\delta\Gamma(0)$, can be written as

$$\begin{aligned} \delta\Gamma(t) &= U_t(\Gamma(0) + \delta\Gamma(0)) - U_t(\Gamma(0)) && (4·2)\\ &= \mathcal{T}(t, \Gamma(0))\delta\Gamma(0). && (4·3) \end{aligned}$$

Here, $\mathcal{T}(t, \Gamma(0))$ is called the "linearized evolution map" and is calculated by numerical integration of the linearized evolution equation. Note that the matrix $\mathcal{T}(t, \Gamma(0))$ is determined by the trajectory segment $\{\Gamma(t'), 0 \le t' \le t\}$. We analyze the matrix $\mathcal{T}(t, \Gamma(0))$ below.

4.1. *Gram-Schmidt decomposition*

Let $\{e_i,\ 1 \le i \le 2N\}$ be a set of orthogonal unit vectors given randomly in the tangent space at $\Gamma(0)$. For the time being, we use the abbreviation $\mathcal{T}$ for $\mathcal{T}(t, \Gamma(0))$. Since almost all vectors expand in the most unstable direction, the direction of the vector $\mathcal{T}e_1$ may approach the most unstable direction when t is sufficiently large. We thus define a unit vector in the tangent space at $\Gamma(t)$ as

$$f_1 = \frac{\mathcal{T}e_1}{|\mathcal{T}e_1|}. \tag{4·4}$$

The vector f_1 is expected to indicate the most unstable direction at $\Gamma(t)$ as $t \to \infty$. Similarly, we define the most unstable direction in the orthogonal co-space of f_1:

$$f_2 = \frac{\mathcal{T}e_2 - (\mathcal{T}e_2, f_1)}{|\mathcal{T}e_2 - (\mathcal{T}e_2, f_1)|}. \tag{4·5}$$

With similar consideration, we define the i-th unstable direction:

$$f_i = \frac{\mathcal{T}e_i - \sum_{j=1}^{i-1}(\mathcal{T}e_i, f_j)f_j}{|\mathcal{T}e_i - \sum_{j=1}^{i-1}(\mathcal{T}e_i, f_j)f_j|}. \tag{4·6}$$

Since $\{f_i,\ 1 \le i \le 2N\}$ is a set of orthonormal unit vectors in the tangent space at $\Gamma(t)$, we can find an orthogonal matrix $\mathcal{F}(t, \Gamma(0))$ given by

$$f_i = \mathcal{F}(t, \Gamma(0))e_i. \tag{4·7}$$

Further, from Eq. (4·6), $\mathcal{T}e_i$ can be written as

$$\mathcal{T}(t,\Gamma(0))e_i = \sum_k L_{ik}(t,\Gamma(0))\mathcal{F}(t,\Gamma(0))e_k, \tag{4·8}$$

where L_{ij} is the (i,j)-element of a lower triangle matrix $\mathcal{L}$. Equation (4·8) is the Gram-Schmidt decomposition of the matrix $\mathcal{T}$. Since the diagonal element is particularly important below, we write it explicitly as

$$L_{ii} = \left| \mathcal{T}e_i - \sum_{j=1}^{i-1} (\mathcal{T}e_i, f_j) f_j \right|. \tag{4·9}$$

4.2. *Convergence property*

As mentioned above, f_i indicates the i-th unstable direction only as $t \to \infty$. Let $e_{i*}(\Gamma(t))$ be the 'true' i-th unstable direction at $\Gamma(t)$. In order to have $e_{i*}(\Gamma(t))$ within a certain accuracy, we need to confirm

$$d(\mathcal{F}(t,\Gamma(0))e_i, e_{*i}(\Gamma(t))) \leq \epsilon, \tag{4·10}$$

where ϵ is a small number related to the accuracy we require, and $d(e,e')$ is the absolute value of the sine of the angle between two unit vectors e and e':

$$d(e,e') = \sqrt{1-(e,e')^2}. \tag{4·11}$$

However, since we do not have $e_{*i}(\Gamma(t))$ yet, we cannot confirm whether or not Eq. (4·10) is satisfied. Then, instead of Eq. (4·10), we check the condition

$$d(\mathcal{F}(t,\Gamma(0))e_i, \mathcal{F}(t,\Gamma(0))e'_i) \leq \epsilon \tag{4·12}$$

for two sets of orthogonal unit vectors $\{e_i, 1 \leq i \leq 2N\}$ and $\{e'_i, 1 \leq i \leq 2N\}$, which are chosen randomly. When Eq. (4·12) is satisfied, we assume that the true i-th unstable direction is determined by

$$e_{*i}(\Gamma(t)) \simeq \mathcal{F}(t,\Gamma(0))e_i, \tag{4·13}$$

within an accuracy we require.

When we numerically obtain $e_{*i}(\Gamma(0))$ at an arbitrary point $\Gamma(0)$, we consider a trajectory segment $\{\Gamma(t),\ -t_b \leq t \leq 0\}$ and check the condition

$$d(\mathcal{F}(t_b,\Gamma(-t_b))e_i, \mathcal{F}(t_b,\Gamma(-t_b))e'_i) \leq \epsilon, \tag{4·14}$$

for sufficiently large t_b. When Eq. (4·14) is satisfied, we assume

$$e_{*i}(\Gamma(0)) \simeq \mathcal{F}(t_b,\Gamma(-t_b))e_i, \tag{4·15}$$

within an accuracy we require.

We do not know a mathematical condition under which Eq. (4·12) is satisfied. In the argument below, we assume that Eq. (4·12) is satisfied and that a set of vectors $\{e_{*i}(\Gamma)\}$ is determined for an arbitrary point Γ. Once $e_{*i}(\Gamma(0))$ is determined, $e_{*i}(\Gamma(t))$ is calculated by

$$e_{*i}(\Gamma(t)) = \mathcal{F}(t,\Gamma(0))e_{*i}(\Gamma(0)). \tag{4·16}$$

4.3. *Lyapunov vectors*

To this point, we have stated that $e_{*i}(\Gamma(t))$ indicates the i-th unstable direction at $\Gamma(t)$. More precisely, $e_{*i}(\Gamma(t))$ indicates the most unstable direction in the orthogonal co-space of the subspace spanned by $\{e_{*j}(\Gamma(t)),\ 1 \leq j \leq i-1\}$ in the tangent space at $\Gamma(t)$. However, $e_{*i}(\Gamma(t))$ does *not* satisfy

$$\mathcal{T}(t, \Gamma(0))e_{*i}(\Gamma(0)) \propto e_{*i}(\Gamma(t)) \tag{4·17}$$

except for the case $i = 1$. This seems somewhat strange, because the unstable nature should be defined as something consistent along the trajectory. Thus, we define a set of vectors $\{\xi_i(\Gamma(t)),\ 1 \leq i \leq 2N\}$ which satisfies two conditions. The first condition is the transitivity,

$$\mathcal{T}(t, \Gamma(0))\xi_i(\Gamma(0)) \propto \xi_i(\Gamma(t)), \tag{4·18}$$

and the second condition is that the vector space generated by $\{e_{*j},\ 1 \leq j \leq i\}$ is spanned by $\{\xi_j,\ 1 \leq j \leq i\}$. The second condition is expressed by

$$\sum_j A_{ij}\xi_j(\Gamma(t)) = e_{*i}(\Gamma(t)), \tag{4·19}$$

where $A_{ij} = 0$ for $i < j$, and A_{ij} is regarded as the (ij)-element of a lower triangle matrix $\mathcal{A}$.

Now, we define the i-th expansion factor $\Lambda_i(t, \Gamma(0))$ in the i-th unstable direction,

$$\mathcal{T}(t, \Gamma(0))\xi_i(\Gamma(0)) = \Lambda_i(t, \Gamma(0))\xi_i(\Gamma(t)). \tag{4·20}$$

We call ξ_i the "i-th Lyapunov vector".[24] In order to determine uniquely the value of Λ_i, we assume the normalization condition that the volume of the i-dimensional parallelepiped formed by $\{\xi_j,\ 1 \leq j \leq i\}$ is unity. This condition is expressed by

$$|\wedge_{j=1}^{i} \xi_j| = 1 \tag{4·21}$$

for $1 \leq i \leq 2N$ (see the Appendix). We also assume that A_{ii} is positive. Under these conditions, we can prove that the i-th expansion factor $\Lambda_i(t, \Gamma(0))$ is calculated by the Gram-Schmidt decomposition Eq. (4·8) with $e_{*i}(\Gamma(0))$.

Proof From Eq. (4·8), we have

$$\mathcal{T}(t, \Gamma(0))e_{*i}(\Gamma(0)) = \sum_k L_{ik}(t, \Gamma(0))\mathcal{F}(t, \Gamma(0))e_{*k}(\Gamma(0)). \tag{4·22}$$

Using Eqs. (4·19) and (4·20), we rewrite the left-hand side of Eq. (4·22) as

$$\begin{aligned}
&\sum_j A_{ij}(\Gamma(0))\mathcal{T}(t, \Gamma(0))\xi_j(\Gamma(0)) \\
&\quad = \sum_j A_{ij}(\Gamma(0))\Lambda_j(t, \Gamma(0))\xi_j(\Gamma(t)) && (4·23)\\
&\quad = \sum_{jk} A_{ij}(\Gamma(0))\Lambda_j(t, \Gamma(0))(\mathcal{A}(\Gamma(t))^{-1})_{jk}e_{*k}(\Gamma(t)) && (4·24)\\
&\quad = \sum_{jk} A_{ij}(\Gamma(0))\Lambda_j(t, \Gamma(0))(\mathcal{A}(\Gamma(t))^{-1})_{jk}\mathcal{F}(t, \Gamma(0))e_{*k}(\Gamma(0)). && (4·25)
\end{aligned}$$

Comparing the right-hand side of Eq. (4·22) with Eq. (4·25), we find

$$L_{ik}(t,\Gamma(0)) = \sum_j A_{ij}(\Gamma(0))\Lambda_j(t,\Gamma(0))(\mathcal{A}(\Gamma(t))^{-1})_{jk}. \tag{4·26}$$

Further, from Eqs. (4·19) and (4·21), we can easily see

$$|A_{ii}| = 1 \tag{4·27}$$

for $1 \leq i \leq 2N$ (see the Appendix). Since A_{ii} is assumed to be positive, $A_{ii} = 1$. Then, Eq. (4·26) yields

$$\Lambda_i(t,\Gamma(0)) = L_{ii}(t,\Gamma(0)). \tag{4·28}$$

In this way, the i-th expansion factor can be calculated numerically.

□

4.4. *Lyapunov exponent*

The i-th expansion ratio $\lambda_i(\Gamma(t))$ at $\Gamma(t)$ is defined as

$$\frac{d\Lambda_i(t,\Gamma(0))}{dt} = \lambda_i(\Gamma(t))\Lambda_i(t,\Gamma(0)). \tag{4·29}$$

The long-time average of the i-th expansion ratio $\lambda_i(\Gamma(t))$ is called the i-th Lyapunov exponent, which is given by

$$\bar{\lambda}_i = \lim_{\tau\to\infty}\frac{1}{\tau}\int_0^\tau dt\lambda_i(\Gamma(t)) \tag{4·30}$$

$$= \lim_{\tau\to\infty}\frac{1}{\tau}\log\Lambda_i(\tau,\Gamma(0)). \tag{4·31}$$

The quantity $\bar{\lambda}_i$ is sometimes called the 'local' Lyapunov exponent, because it depends on $\Gamma(0)$. However, from the ergodic theorem, $\bar{\lambda}_i$ has the same value for almost all $\Gamma(0)$ with respect to the microcanonical measure. Since we assume the ergodicity of the microcanonical measure, we are not concerned with the local nature of the Lyapunov exponent.

As clearly seen from the method of construction of Lyapunov vectors, we find

$$\bar{\lambda}_1 \geq \bar{\lambda}_2 \cdots \geq \bar{\lambda}_{2N}. \tag{4·32}$$

In Hamiltonian systems, there are at least two zero Lyapunov exponents whose Lyapunov vectors indicate the normal direction of the energy surface and the tangential direction of the trajectory. In the argument below, we assume that there are N_p positive Lyapunov exponents. Unless the system has an additional conservation law such as momentum conservation, $N_p = N - 1$.

The *information loss rate* at $\Gamma(t)$, $h(\Gamma(t))$, is defined as the sum of the expansion ratios with positive Lyapunov exponents:

$$h(\Gamma(t)) = \sum_{\bar{\lambda}_i>0}\lambda_i(\Gamma(t)) \tag{4·33}$$

$$= \sum_{i=1}^{N_p}\lambda_i(\Gamma(t)). \tag{4·34}$$

Note that $h(\Gamma(t))$ represents the volume expansion ratio of the N_p-dimensional unstable space. That is, $h(\Gamma(t))$ can be rewritten as

$$h(\Gamma(t)) = \frac{d}{dt} \log | \wedge_{i=1}^{N-1} \mathcal{T}(t, \Gamma(0)) \xi_i(\Gamma(0))|. \tag{4·35}$$

The long-time average of the information loss rate, $\bar{h}$, has the same value for almost all initial conditions with respect to the microcanonical measure. It is known that $\bar{h}$ is identical to the Kolmogorov-Sinai entropy when the system is hyperbolic. [31)]

4.5. *Contraction ratio*

Let us recall that the expansion factor Λ_i is calculated by the Gram-Schmidt decomposition under the normalization condition Eq. (4·21). However, this normalization lacks balance between the unstable and stable directions. Since Hamiltonian systems possess time reversal symmetry, such unbalance will cause theoretical complications. In order to recover the symmetry, we introduce a new set of vectors, $\{\xi_i^{(s)},\ 1 \le i \le 2N\}$, given by

$$\xi_i^{(s)} = c_i \xi_{2N-i+1}, \tag{4·36}$$

where c_i is a positive number determined so as to satisfy

$$| \wedge_{j=1}^{i} \xi_j^{(s)}| = 1 \tag{4·37}$$

for $1 \le i \le 2N$.

We now define the contraction factor $\Lambda_i^{(s)}$ and contraction ratio $\lambda_i^{(s)}$ as

$$\mathcal{T}(t, \Gamma(0)) \xi_i^{(s)}(\Gamma(0)) = \frac{1}{\Lambda_i^{(s)}(t, \Gamma(0))} \xi_i^{(s)}(\Gamma(t)), \tag{4·38}$$

$$\lambda_i^{(s)}(\Gamma(t)) = \frac{d}{dt} \log \Lambda_i^{(s)}(t, \Gamma(0)). \tag{4·39}$$

As seen in the next section, the contraction ratio is related to the expansion ratio of the time-reversed trajectory, and this relation plays a role in simplifying arguments. In particular, the following relation is utilized.

$$\sum_{i=1}^{N_p} \lambda_{2N+1-i} = -\sum_{i=1}^{N_p} \lambda_i^{(s)} + \frac{d}{dt} o(N). \tag{4·40}$$

Here, the last term represents the time derivative of a function whose value is much smaller than N as $N \to \infty$. Note that the left-hand side and the first term on the right-hand side are of order N.

Proof Substituting Eq. (4·36) into Eq. (4·38), we have

$$c_i(\Gamma(0)) \Lambda_{2N+1-i}(t, \Gamma(0)) = \frac{1}{\Lambda_i^{(s)}(t, \Gamma(0))} c_i(\Gamma(t)). \tag{4·41}$$

Taking the time derivative of the logarithm of both sides of this equation yields

$$\lambda_{2N+1-i}(\Gamma(t)) = -\lambda_i^{(s)}(\Gamma(t)) + \frac{d}{dt} \log \left(\frac{c_i(\Gamma(t))}{c_i(\Gamma(0))} \right). \tag{4·42}$$

We thus obtain

$$\sum_{i=1}^{N_p} \lambda_{2N+1-i}(\Gamma(t)) = -\sum_{i=1}^{N_p} \lambda_i^{(s)}(\Gamma(t)) + \frac{d}{dt}\sum_{i=1}^{N_p} \log\left(\frac{c_i(\Gamma(t))}{c_i(\Gamma(0))}\right). \tag{4·43}$$

Let us evaluate the second term on the right-hand side of Eq. (4·43). We first define the 'angle' ϕ_i as

$$|\wedge_{j=1}^{2N}\xi_j| = |\wedge_{j=1}^{2N-i}\xi_j||\wedge_{j=2N+1-i}^{2N}\xi_j|\sin\phi_i, \tag{4·44}$$

where $0 \le \phi_i \le \pi/2$ (see the Appendix). By using the normalization condition of $\{\xi_i\}$, we rewrite Eq. (4·44) as

$$|\wedge_{j=2N+1-i}^{2N}\xi_j|\sin\phi_i = 1. \tag{4·45}$$

Using Eqs. (4·36) and (4·37), we obtain

$$c_1\cdots c_i = \sin\phi_i. \tag{4·46}$$

This leads to

$$\sum_{i=1}^{N_p} \log\left(\frac{c_i(\Gamma(t))}{c_i(\Gamma(0))}\right) = \log\left(\frac{\sin\phi_{N_p}(\Gamma(t))}{\sin\phi_{N_p}(\Gamma(0))}\right) = o(N), \tag{4·47}$$

where we have assumed

$$\sin(\phi_{N_p}(\Gamma)) \gg O(\exp(-N)), \tag{4·48}$$

which may be ensured by the condition that the unstable and stable manifolds intersect transversally.

□

4.6. *Weight on trajectory segments*

We consider a weight on the trajectory segment $\{\Gamma(t),\ 0 \le t \le \tau\}$. This weight, $W(\{\Gamma(t),\ 0 \le t \le \tau\})$, is a conditional probability of finding trajectory segments remaining in a small tube around $\{\Gamma(t),\ 0 \le t \le \tau\}$ when the initial condition is chosen in a small region around $\Gamma(0)$. More explicitly, the weight W is defined in the following way.

Suppose that the phase space is decomposed into small cells $\{\Delta_j\}$ with a sufficiently small size ϵ and that $\Gamma(0)$ is included in the i-th cell Δ_i. Then, let $N(\tau,\epsilon)$ be the number of cells which intersect $U_\tau(\Delta_i)$. We can choose the value of ϵ so that the region $U_\tau(\Delta_i)$ remains in the linear regime around $U_\tau(\Gamma(0))$. This condition may be given by

$$\epsilon \ll \epsilon_c(\tau), \tag{4·49}$$

where the value of $\epsilon_c(\tau)$ is determined by nonlinear properties of dynamical systems. Under this condition, $N(\tau,\epsilon)$ measures the number of distinguishable trajectory segments starting from the neighborhood of $\Gamma(0)$. Therefore, we define the weight as

$$W(\{\Gamma(t),\ 0 \le t \le \tau\}) = \frac{1}{N(\tau,\epsilon)}. \tag{4·50}$$

Then, we can show

$$W(\{\Gamma(t),\ 0 \le t \le \tau\}) = |\wedge_{i=1}^{N_p} \mathcal{T}(-\tau, \Gamma(\tau))\xi_i(\Gamma(\tau))| \tag{4·51}$$

in an appropriate limit of large τ and small ϵ. (Since $\epsilon_c(\tau) \to 0$ in the limit $\tau \to \infty$, we need to deal with the delicate problem of double limits. However, we assume simply that we can choose an appropriate asymptotic limit.) We further find that, by Eq. (4·35), Eq. (4·51) takes the simpler form

$$W(\{\Gamma(t),\ 0 \le t \le \tau\}) = \exp\left(-\int_0^\tau dt h(\Gamma(t))\right). \tag{4·52}$$

Proof Consider the time evolution of a small region Δ_i. This region expands and contracts in the unstable and stable directions, respectively. After a sufficiently long time, the region almost collapses into the N_p-dimensional unstable manifold and intersects cells in the unstable directions. Since the N_p-dimensional volume element in the unstable manifold at $\Gamma(\tau)$ can be written as $\wedge_{i=1}^{N_p}\xi_i(\Gamma(\tau))$, we expect

$$N(\tau,\epsilon) = \frac{|\wedge_{i=1}^{N_p} \xi_i(\Gamma(\tau))|}{|\wedge_{i=1}^{N_p} \mathcal{T}(-\tau, \Gamma(\tau))\xi_i(\Gamma(\tau))|} \tag{4·53}$$

$$= |\wedge_{i=1}^{N_p} \mathcal{T}(-\tau, \Gamma(\tau))\xi_i(\Gamma(\tau))|^{-1} \tag{4·54}$$

in an appropriate limit of large τ and small ϵ. Substituting this into Eq. (4·50) leads to Eq. (4·51).

□

4.7. *Time-dependent case*

To this point in this section, we have assumed that the value of α is not changed in time. In this subsection, we briefly discuss Lyapunov analysis for systems with a time-dependent parameter. When the value of α depends on t, the evolution map from $t = t_0$ to $t = t_1$ depends on the absolute time t_0 and t_1. Therefore, it takes the form U_{t_1,t_0}, and the linearized evolution map is written as $\mathcal{T}(t_1, t_0; \Gamma(t_0))$.

The Lyapunov analysis in such a case may be reconsidered carefully. However, we do not need general arguments. In the systems in question, the value of α is changed during a finite time interval $[\tau_i, \tau_f]$, where $0 \ll \tau_i \le \tau_f \ll \tau$. Therefore, for example, the i-th Lyapunov vectors at $\Gamma(0)$ and $\Gamma(\tau)$ can be defined as $\xi_i(\Gamma(0))$ and $\xi_i(\Gamma(\tau))$, respectively.

Although the expansion factors, Lyapunov exponents, and information loss rate do not make sense in general, the argument regarding the weight W is still valid. We can write

$$W(\{\Gamma(t),\ 0 \le t \le \tau\}) = |\wedge_{i=1}^{N_p} \mathcal{T}(0, \tau; \Gamma(\tau))\xi_i(\Gamma(\tau))| \tag{4·55}$$

for sufficiently large τ and small ϵ. We also define the *actual information loss rate* as the generalization of Eq. (4·35):

$$h_{\rm a}(t, \Gamma(0)) = \frac{d}{dt} \log |\wedge_{i=1}^{N_p} \mathcal{T}(t, 0; \Gamma(0))\xi_i(\Gamma(0))|. \tag{4·56}$$

4.8. *Liouville's theorem*

In this subsection, we review a proof of Liouville's theorem, which states that the $2N$-dimensional volume element maintains its volume along the trajectory. It is important to understand that Liouville's theorem holds even when the value of α is changed in time.

Proof We have the Hamiltonian equation

$$\frac{d\Gamma(t)}{dt} = -\mathcal{J}\left.\frac{\partial H(\Gamma,\alpha(t))}{\partial \Gamma}\right|_{\Gamma=\Gamma(t)}. \tag{4·57}$$

(See Eq. (2·8) in §2.) Since the linearized evolution equation is written as

$$\frac{d\delta\Gamma(t)}{dt} = -\mathcal{J}\left.\frac{\partial^2 H(\Gamma,\alpha(t))}{\partial \Gamma\partial\Gamma}\right|_{\Gamma=\Gamma(t)}\delta\Gamma(t), \tag{4·58}$$

the linearized evolution map $\mathcal{T}(t,0;\Gamma(0))$ satisfies the equation

$$\frac{d\mathcal{T}}{dt} = -\mathcal{J}\mathcal{B}\mathcal{T}, \tag{4·59}$$

where $\mathcal{B}$ is a symmetric matrix. We then obtain

$$\frac{d(\mathcal{T}^\dagger\mathcal{J}\mathcal{T})}{dt} = \frac{d\mathcal{T}^\dagger}{dt}\mathcal{J}\mathcal{T} + \mathcal{T}^\dagger\mathcal{J}\frac{d\mathcal{T}}{dt} \tag{4·60}$$

$$= -(\mathcal{T}^\dagger\mathcal{B}\mathcal{J}^\dagger\mathcal{J}\mathcal{T} + \mathcal{T}^\dagger\mathcal{J}\mathcal{J}\mathcal{B}\mathcal{T}) \tag{4·61}$$

$$= 0, \tag{4·62}$$

where we have used the equality

$$\mathcal{J}^\dagger\mathcal{J} = -\mathcal{J}\mathcal{J} = 1. \tag{4·63}$$

Since $\mathcal{T}(0,0;\Gamma(0)) = 1$, Eq. (4·62) leads to

$$\mathcal{T}^\dagger\mathcal{J}\mathcal{T} = \mathcal{J}. \tag{4·64}$$

Taking the determinant of both the sides gives

$$\det[\mathcal{T}^\dagger\mathcal{T}] = 1. \tag{4·65}$$

Let $\{e_i,\ 1 \le i \le 2N\}$ be an orthogonal set of unit vectors defined in the tangent space at $\Gamma(0)$. The time evolution of the $2N$-dimensional volume element $\wedge_{i=1}^{2N} e_i$ is given by $\wedge_{i=1}^{2N}\mathcal{T}e_i$, and its volume is calculated as

$$|\wedge_{i=1}^{2N}\mathcal{T}e_i| = \sqrt{\det\mathcal{T}\mathcal{T}^\dagger} \tag{4·66}$$

$$= 1 \tag{4·67}$$

(see the Appendix). Therefore, the $2N$-dimensional volume element maintains its volume along the trajectory.

□

Further, using Liouville's theorem, we can prove that the equality

$$\sum_{i=1}^{2N} \lambda_i(\Gamma(t)) = 0 \tag{4·68}$$

holds when the value of α is not changed in time.

Proof Since the the value of α is not changed in time, we obtain

$$\sum_{i=1}^{2N} \lambda_i(\Gamma(t)) = \frac{d}{dt} \sum_{i=1}^{2N} \log \Lambda_i(t, \Gamma(0)) \tag{4·69}$$

$$= \frac{d}{dt} \log | \wedge_{i=1}^{2N} \mathcal{T}(t, \Gamma(0)) \xi_i(\Gamma(0))| \tag{4·70}$$

$$= 0 \tag{4·71}$$

where Liouville's theorem is used to obtain the last line.

□

§5. Reversibility

5.1. *Reversibility in time evolution*

We define the matrix $\mathcal{R}$ as

$$(\mathcal{R}\Gamma)_i = q_i, \tag{5·1}$$

$$(\mathcal{R}\Gamma)_{i+N} = -p_i, \tag{5·2}$$

where $1 \leq i \leq N$. This matrix corresponds to the time-reversed operator acting on a phase space point. We assume that the Hamiltonian under consideration possesses time reversal symmetry:

$$H(\mathcal{R}\Gamma, \alpha) = H(\Gamma, \alpha). \tag{5·3}$$

Let $U_{t,0}$ and $\tilde{U}_{t,0}$ be the evolution maps for Hamiltonian equations with $\alpha(\ )$ and $\tilde{\alpha}(\ )$, respectively, where we have defined the time-reversed protocol of the parameter change $\tilde{\alpha}(\ )$ as

$$\tilde{\alpha}(t) = \alpha(-t). \tag{5·4}$$

Then, owing to the symmetry property Eq. (5·3), the identity

$$U_{t,0} = \mathcal{R}\tilde{U}_{-t,0}\mathcal{R} \tag{5·5}$$

holds.

Proof Let $\{\Gamma(t)\}$ and $\{\tilde{\Gamma}(t)\}$ be trajectories given by

$$\Gamma(t) = U_{t,0}(\Gamma(0)), \tag{5·6}$$

$$\tilde{\Gamma}(t) = \tilde{U}_{t,0}(\tilde{\Gamma}(0)), \tag{5·7}$$

where $\Gamma(0)$ and $\tilde{\Gamma}(0)$ are the initial conditions which satisfy the relation

$$\tilde{\Gamma}(0) = \mathcal{R}\Gamma(0). \tag{5·8}$$

From Eq. (5·7), we obtain

$$-\frac{d\tilde{\Gamma}(-t)}{dt} = -\mathcal{J}\left.\frac{\partial H(\Gamma,\tilde{\alpha}(-t))}{\partial\Gamma}\right|_{\Gamma=\tilde{\Gamma}(-t)} \tag{5·9}$$

$$= -\mathcal{J}\left.\frac{\partial H(\Gamma,\alpha(t))}{\partial\Gamma}\right|_{\Gamma=\tilde{\Gamma}(-t)}, \tag{5·10}$$

where we have used the equation of motion in the form Eq. (2·8) with the matrix $\mathcal{J}$ satisfying Eq. (2·9). On the other hand, Eq. (5·6) leads to

$$\mathcal{R}\frac{d\Gamma(t)}{dt} = -\mathcal{R}\mathcal{J}\left.\frac{\partial H(\Gamma,\alpha(t))}{\partial\Gamma}\right|_{\Gamma=\Gamma(t)} \tag{5·11}$$

$$= \mathcal{J}\mathcal{R}\left.\frac{\partial H(\Gamma,\alpha(t))}{\partial\Gamma}\right|_{\Gamma=\Gamma(t)} \tag{5·12}$$

$$= \mathcal{J}\left.\frac{\partial H(\Gamma,\alpha(t))}{\partial(\mathcal{R}\Gamma)}\right|_{\Gamma=\Gamma(t)} \tag{5·13}$$

$$= \mathcal{J}\left.\frac{\partial H(\mathcal{R}\Gamma,\alpha(t))}{\partial(\mathcal{R}\Gamma)}\right|_{\Gamma=\Gamma(t)} \tag{5·14}$$

$$= \mathcal{J}\left.\frac{\partial H(\Gamma,\alpha(t))}{\partial\Gamma}\right|_{\Gamma=\mathcal{R}\Gamma(t)}. \tag{5·15}$$

Here, the second line is obtained by the relation

$$\mathcal{R}\mathcal{J} + \mathcal{J}\mathcal{R} = 0, \tag{5·16}$$

the third line is derived from the relation $\mathcal{R}\mathcal{R} = 1$, and the equality of the fourth line comes from the symmetry property Eq. (5·3).

Comparing Eqs. (5·10) and (5·15), we find that $\tilde{\Gamma}(-t)$ and $\mathcal{R}\Gamma(t)$ obey the same evolution equation. Recalling the relation for the initial conditions Eq. (5·8), we conclude

$$\tilde{\Gamma}(-t) = \mathcal{R}\Gamma(t). \tag{5·17}$$

By using Eqs. (5·6) and (5·7), Eq. (5·17) can be rewritten as

$$\tilde{U}_{-t,0}(\mathcal{R}\Gamma(0)) = \mathcal{R}U_{t,0}(\Gamma(0)). \tag{5·18}$$

Since $\Gamma(0)$ is arbitrary, Eq. (5·5) holds.

□

In the argument below, $\tilde{\Gamma}(t)$ will be assumed to be given by Eq. (5·17).

5.2. *Reversibility in Lyapunov analysis*

First, from Eq. (5·5), we have

$$\mathcal{T}(t,0;\Gamma(0)) = \mathcal{R}\tilde{\mathcal{T}}(-t,0;\tilde{\Gamma}(0))\mathcal{R}, \tag{5·19}$$

where $\mathcal{T}(t,0;\Gamma(0))$ and $\tilde{\mathcal{T}}(t,0;\tilde{\Gamma}(0))$ are the linearized evolution maps around the trajectories $\Gamma(\ )$ and $\tilde{\Gamma}(\ )$. In particular, when the value of α is not changed, the

equality $\tilde{\mathcal{T}} = \mathcal{T}$ holds. We then prove the identities

$$\xi_i(\Gamma) = \mathcal{R}\xi_i^{(s)}(\mathcal{R}\Gamma), \tag{5·20}$$

$$\lambda_i(\Gamma) = \lambda_i^{(s)}(\mathcal{R}\Gamma), \tag{5·21}$$

$$h(\Gamma(t)) - h(\tilde{\Gamma}(-t)) = \frac{d}{dt}o(N). \tag{5·22}$$

Proof Using the matrices defined as

$$\mathcal{X}_{ij}(\Gamma) = (\xi_j(\Gamma))_i, \tag{5·23}$$

$$\mathcal{X}_{ij}^{(s)}(\Gamma) = (\xi_j^{(s)}(\Gamma))_i, \tag{5·24}$$

we can write $\mathcal{T}(t, \Gamma(0))$ in two forms,

$$\mathcal{T}(t, \Gamma(0)) = \mathcal{X}(\Gamma(t))\mathcal{M}(t, \Gamma(0))\mathcal{X}(\Gamma(0))^{-1}, \tag{5·25}$$

$$\mathcal{T}(t, \Gamma(0)) = \mathcal{X}^{(s)}(\Gamma(t))\mathcal{M}^{(s)}(t, \Gamma(0))\mathcal{X}^{(s)}(\Gamma(0))^{-1}, \tag{5·26}$$

where $\mathcal{M}(t, \Gamma(0))$ and $\mathcal{M}^{(s)}(t, \Gamma(0))$ are diagonal matrices whose (i, i)-elements are given by $\Lambda_i(t, \Gamma(0))$ and $\Lambda_i^{(s)}(t, \Gamma(0))^{-1}$, respectively. Using Eq. (5·26), we rewrite the right-hand side of Eq. (5·19) as

$$\begin{aligned}&\mathcal{R}\mathcal{X}^{(s)}(\tilde{\Gamma}(-t))\mathcal{M}^{(s)}(-t, \tilde{\Gamma}(0))\mathcal{X}^{(s)}(\tilde{\Gamma}(0))^{-1}\mathcal{R} \\ &\quad = \mathcal{R}\mathcal{X}^{(s)}(\mathcal{R}\Gamma(t))\mathcal{M}^{(s)}(-t, \tilde{\Gamma}(0))\mathcal{X}^{(s)}(\mathcal{R}\Gamma(0))^{-1}\mathcal{R},\end{aligned} \tag{5·27}$$

where we have used $\tilde{\mathcal{T}} = \mathcal{T}$. Comparing Eq. (5·27) with the right-hand side of Eq. (5·25), we obtain

$$\mathcal{X}(\Gamma) = \mathcal{R}\mathcal{X}^{(s)}(\mathcal{R}\Gamma), \tag{5·28}$$

$$\mathcal{M}(t, \Gamma(0)) = \mathcal{M}^{(s)}(-t, \mathcal{R}\Gamma(0)), \tag{5·29}$$

where we have used the normalization conditions of ξ_i and $\xi_i^{(s)}$ given by Eqs. (4·21) and (4·37).

Equation (5·28) is equivalent to Eq. (5·20), and Eq. (5·29) leads to Eq. (5·21), because of the equality

$$\frac{d}{dt}\log \Lambda_i^{(s)}(-t, \mathcal{R}\Gamma(0))^{-1} = \lambda_i^{(s)}(\tilde{\Gamma}(-t)) \tag{5·30}$$

$$= \lambda_i^{(s)}(\mathcal{R}\Gamma(t)). \tag{5·31}$$

Furthermore, $h(\Gamma(t))$ can be expressed in terms of $\{\lambda_i^{(s)}\}$ in the following way:

$$h(\Gamma(t)) = \sum_{i=1}^{N_p} \lambda_i(\Gamma(t)) \tag{5·32}$$

$$= -\sum_{i=1}^{N_p} \lambda_{2N+i-1}(\Gamma(t)) \tag{5·33}$$

$$= \sum_{i=1}^{N_p} \lambda_i^{(s)}(\Gamma(t)) + \frac{d}{dt}o(N), \tag{5·34}$$

where the second and third lines come from Eq. (4·68) and Eq. (4·40), respectively. On the other hand, from the symmetry property Eq. (5·21), $h(\tilde{\Gamma}(-t))$ can be written as

$$h(\tilde{\Gamma}(-t)) = \sum_{i=1}^{N_p} \lambda_i(\tilde{\Gamma}(-t)) \tag{5·35}$$

$$= \sum_{i=1}^{N_p} \lambda_i(\mathcal{R}\Gamma(t)) \tag{5·36}$$

$$= \sum_{i=1}^{N_p} \lambda_i^{(s)}(\Gamma(t)). \tag{5·37}$$

Comparing Eqs. (5·34) and (5·37), we obtain Eq. (5·22).

□

Using these identities, we can express the weight on the trajectory segment by the actual information loss of the time-reversed trajectory:

$$W(\{\Gamma(t),\ 0 \le t \le \tau\}) = \exp\left[\int_0^\tau dt \tilde{h}_a(-t, \tilde{\Gamma}(-\tau)) + o(N)\right], \tag{5·38}$$

where the actual information loss rate along the time-reversed trajectory $\tilde{h}_a$ is defined as

$$\tilde{h}_a(t, \tilde{\Gamma}(-\tau)) = \frac{d}{dt} \log |\wedge_{i=1}^{N-1} \tilde{\mathcal{T}}(t, -\tau; \Gamma(-\tau))\xi_i(\tilde{\Gamma}(-\tau))|. \tag{5·39}$$

Proof Let us recall the expression of the weight Eq. (4·55):

$$W(\{\Gamma(t),\ 0 \le t \le \tau\}) = |\wedge_{i=1}^{N_p} \mathcal{T}(0, \tau; \Gamma(\tau))\xi_i(\Gamma(\tau))|. \tag{5·40}$$

The right-hand side can be rewritten as

$$|\wedge_{i=1}^{N_p} \mathcal{R}\tilde{\mathcal{T}}(0, -\tau; \tilde{\Gamma}(-\tau))\mathcal{R}\mathcal{R}\xi_i^{(s)}(\mathcal{R}\Gamma(\tau))|$$

$$= |\wedge_{i=1}^{N_p} \tilde{\mathcal{T}}(0, -\tau; \tilde{\Gamma}(-\tau))\xi_i^{(s)}(\tilde{\Gamma}(-\tau))| \tag{5·41}$$

$$= |\wedge_{i=1}^{N_p} \tilde{\mathcal{T}}(0, -\tau; \tilde{\Gamma}(-\tau))\xi_{2N+1-i}(\tilde{\Gamma}(-\tau))| \exp(o(N)) \tag{5·42}$$

$$= |\wedge_{i=1}^{N_p} \tilde{\mathcal{T}}(0, -\tau; \tilde{\Gamma}(-\tau))\xi_i(\tilde{\Gamma}(-\tau))| \exp(o(N)) \tag{5·43}$$

$$= \exp\left[\int_0^\tau dt \tilde{h}_a(-t, \tilde{\Gamma}(-\tau)) + o(N)\right], \tag{5·44}$$

where the third line is obtained by using an argument in §4.5, and the fourth line comes from Liouville's theorem.

□

Similarly, the weight on the time-reversed trajectory segment $W(\{\tilde{\Gamma}(t),\ -\tau \le t \le 0\})$ can be written as

$$W(\{\tilde{\Gamma}(t),\ -\tau \le t \le 0\}) = \exp\left[\int_0^\tau dt h_a(t, \Gamma(0)) + o(N)\right]. \tag{5·45}$$

(See Eq. (4·56) for the definition of $h_a(t, \Gamma(0))$.)

5.3. *Reversibility paradox*

Suppose that there is a trajectory segment $\{\Gamma(t),\ 0 \le t \le \tau\}$ from an energy surface Σ_0 to Σ_1. Then, the time-reversed segment $\{\tilde{\Gamma}(t),\ -\tau \le t \le 0\}$ goes from Σ_1 to Σ_0. One may wonder how this fact is compatible with thermodynamic irreversibility. Essentially the same question was asked by Roschmidt. This is known as the reversibility paradox.[25] A standard answer may appeal to the operational impossibility of the time-reversed operation $\Gamma \to \mathcal{R}\Gamma$. If we were allowed to operate on the system by using the result of the observation of the trajectory, we could perform the time-reversed operation. This consideration is related to the Maxwell's demon problem.[32] However, the time dependence of α is given without any references to trajectories. Thus, in our problem, the time-reversed operation cannot be realized by $\alpha(\ )$, and the Maxwell's demon problem does not appear.

However, still the paradox is not resolved completely. In order to be compatible with thermodynamic irreversibility, there should be asymmetry between the trajectory segment $\{\Gamma(t),\ 0 \le t \le \tau\}$ and the time-reversed segment $\{\tilde{\Gamma}(t),\ -\tau \le t \le 0\}$. This asymmetry cannot come from purely mechanical considerations. We must consider the measure for the ensemble of the initial conditions of the time-reversed trajectory segment $\{\tilde{\Gamma}(t),\ -\tau \le t \le 0\}$. This ensemble, Υ_τ, is defined as a $(2N-1)$-dimensional set which satisfies

$$\tilde{U}_{0,-\tau}(\Upsilon_\tau) = \Sigma_0. \tag{5·46}$$

From the reversibility relation Eq. (5·5), we obtain

$$\Upsilon_\tau = \mathcal{R}U_{\tau,0}(\Sigma_0). \tag{5·47}$$

Owing to the chaotic nature, Υ_τ becomes a quite complicated set as τ becomes large. Here, we describe the set Υ_τ informally. We focus on the thermodynamic limit so that the structure of Υ_τ is clearly seen.

Suppose that the most probable processes $\Sigma_0 \to_* \Sigma_1$ and $\tilde{\Sigma}_1 \to_* \Sigma_0$ are realized by the protocols of the parameter change $\alpha(\ )$ and $\tilde{\alpha}(\ )$, respectively. Then, from Eq. (5·46), $\Upsilon_\tau \cap \tilde{\Sigma}_1$ becomes dominant in $\tilde{\Sigma}_1$ with respect to the microcanonical measure for $\tilde{\Sigma}_1$. On the other hand, from Eq. (5·47), $\Upsilon_\tau \cap \Sigma_1$ becomes dominant in Υ_τ with respect to the microcanonical measure for Σ_0. One may think that these two statements are apparently contradictory. However, we should note that the measures are different when we observe the set Υ_τ. Here, it is worthwhile noting that $\Upsilon_\tau \cap \Sigma_1$ is *not* dominant in Σ_1 with respect to the microcanonical measure for Σ_1, when $\Sigma_1 \neq \tilde{\Sigma}_1$. Therefore, we can imagine that the set Υ_τ has a fine structure on energy surfaces apart from $\tilde{\Sigma}_1$. In order to represent this heterogeneity quantitatively, we define a measure $\tilde{\mu}$ for the set Υ_τ as

$$\tilde{\mu}(\Delta_\epsilon(\tilde{\Gamma}(-\tau)); \Upsilon_\tau) = \lim_{\delta E \to 0} \frac{\mu_{\mathrm{L}}(\Delta_\epsilon(\tilde{\Gamma}(-\tau)) \cap \mathcal{R}U_{\tau,0}(\Sigma_0 \circ \delta E))}{\mu_{\mathrm{L}}(\mathcal{R}U_{\tau,0}(\Sigma_0 \circ \delta E))}, \tag{5·48}$$

where μ_{L} denotes the $2N$-dimensional Lebesgue measure, $\Delta_\epsilon(\tilde{\Gamma}(-\tau))$ is a small region with a size ϵ including $\tilde{\Gamma}(-\tau)$, and $\Sigma_0 \circ \delta E$ represents a set of energy surfaces with width δE (see §2.4). We then expect that $\tilde{\mu}(\Delta_\epsilon(\tilde{\Gamma}(-\tau)); \Upsilon_\tau)$ for $\tilde{\Gamma}(-\tau) \in \tilde{\Sigma}_1$ is much larger than $\tilde{\mu}(\Delta_\epsilon(\tilde{\Gamma}(-\tau)); \Upsilon_\tau)$ for $\tilde{\Gamma}(-\tau) \in \Sigma_1$.

5.4. *Reversibility in probability*

In spite of the asymmetry between the two sets Σ_0 and Υ_τ, from the reversibility of the time evolution there is a one-to-one correspondence between the set of trajectory segments from Σ_0 to $\mathcal{R}\Upsilon_\tau$ and the set of time-reversed trajectory segments from Υ_τ to Σ_0.

We now derive the reversibility relation coming from this correspondence. In §4.7, we discussed the weight on trajectory segments. The weight $W(\{\Gamma(t),\ 0 \le t \le \tau\})$ is a conditional probability of finding trajectory segments remaining in a small tube around $\{\Gamma(t),\ 0 \le t \le \tau\}$ when the initial condition is chosen in a small region $\Delta_\epsilon(\Gamma(0))$ around $\Gamma(0)$. Then,

$$\mu_{\rm mc}(\Delta_\epsilon(\Gamma(0)); \Sigma_0) W(\{\Gamma(t),\ 0 \le t \le \tau\}) \tag{5·49}$$

is the probability of finding a trajectory segment in a small tube around $\{\Gamma(t),\ 0 \le t \le \tau\}$ of all trajectory segments from Σ_0 to $\mathcal{R}\Upsilon_\tau$. This probability, from the one-to-one correspondence mentioned above, should be equal to the probability of finding a trajectory segment in a small tube around $\{\tilde{\Gamma}(t),\ -\tau \le t \le 0\}$ of all trajectory segments from Υ_τ to Σ_0. The latter probability is written as

$$\tilde{\mu}(\Delta_\epsilon(\tilde{\Gamma}(-\tau)); \Upsilon_\tau) W(\{\tilde{\Gamma}(t),\ -\tau \le t \le 0\}). \tag{5·50}$$

Therefore, we obtain the relation

$$\begin{aligned} &\mu_{\rm mc}(\Delta_\epsilon(\Gamma(0)); \Sigma_0) W(\{\Gamma(t),\ 0 \le t \le \tau\}) \\ &\qquad = \tilde{\mu}(\Delta_\epsilon(\tilde{\Gamma}(-\tau)); \Upsilon_\tau) W(\{\tilde{\Gamma}(t),\ -\tau \le t \le 0\}). \end{aligned} \tag{5·51}$$

This relation will lead to an important equality related to thermodynamic irreversibility.

§6. Irreversible information loss

6.1. *Definition*

Let us define the *irreversible information loss* I as

$$I(\tau, \Gamma(0)) = \log \frac{W(\{\tilde{\Gamma}(t),\ -\tau \le t \le 0\})}{W(\{\Gamma(t),\ 0 \le t \le \tau\})}, \tag{6·1}$$

where the right-hand side depends on $(\tau, \Gamma(0))$, because the trajectory is given by a solution to the deterministic evolution equation. Using Eqs. (5·38) and (5·45), we can write I as

$$I(\tau, \Gamma(0)) = \int_0^\tau dt [h_a(t, \Gamma(0)) - \tilde{h}_a(-t, \tilde{\Gamma}(-\tau))] + o(N). \tag{6·2}$$

This expression may represent the meaning of the term 'irreversible information loss'. Then, from the relation Eq. (5·51), we obtain

$$\int_{\Upsilon_\tau} \tilde{\mu}(\Delta_\epsilon(\Gamma(-\tau)); \Upsilon_\tau) = \int_{\Sigma_0} \mu_{\rm mc}(\Delta_\epsilon(\Gamma(0))) \exp(-I). \tag{6·3}$$

By the normalization condition of the probability, we have the equality

$$\langle \exp(-I) \rangle_0 = 1, \tag{6·4}$$

where $\langle\ \rangle_0$ denotes the average with respect to $\mu_{\rm mc}(d\Gamma; \Sigma_0)$. Using the Jensen inequality

$$\langle \exp(-I) \rangle_0 \leq \exp(-\langle I \rangle_0), \tag{6·5}$$

we obtain

$$\langle I \rangle_0 \geq 0. \tag{6·6}$$

This inequality suggests that the irreversible information loss I has a certain relation with thermodynamic irreversibility. One may find that this argument has some similarity with that of Jarzynski. [8)]

In order to discuss the convergence of $I(\tau, \Gamma(0))$ for $\tau \to \infty$, we evaluate the value of

$$\lim_{\tau\to\infty} \frac{\partial I}{\partial \tau} = \lim_{\tau\to\infty} [h_a(\tau, \Gamma(0)) - \tilde{h}_a(-\tau, \tilde{\Gamma}(-\tau))]. \tag{6·7}$$

Since the value of α is changed in a finite time interval, $\wedge_{i=1}^{N_p} \mathcal{T}(t, 0; \Gamma(0)) \xi_i(\Gamma(0))$ approaches the unstable manifold around $\Gamma(t)$ when $t \gg \tau_f$. We thus obtain

$$\lim_{\tau\to\infty} [h_a(\tau, \Gamma(0)) - \tilde{h}_a(-\tau, \tilde{\Gamma}(-\tau))] = \lim_{\tau\to\infty} [h(\Gamma(\tau)) - h(\tilde{\Gamma}(-\tau))] \tag{6·8}$$

$$= \frac{d}{dt} o(N), \tag{6·9}$$

where we have used Eq. (5·22). We further assume that this convergence is so fast that the time integration yields a finite value. Then, there exists a function $\bar{I}$ such that

$$\lim_{\tau\to\infty} \lim_{N\to\infty} \frac{1}{N} I(\tau, \Gamma(0)) = \bar{I}(\Gamma(0)). \tag{6·10}$$

6.2. *Most probable value*

We define the most probable value of $\bar{I}(\Gamma(0))$ based on the assumption of the large deviation property: Let $\Pi_I(\psi; \Sigma_0, \alpha(\ ))d\psi$ be the probability that $I(\tau, \Gamma(0))/N$ takes a value in $[\psi, \psi + d\psi]$ when the initial equilibrium state Σ_0 and the protocol of the parameter change $\alpha(\ )$ are given. Then, Π_I can be written in the form

$$\Pi_I(\psi; \Sigma_0, \alpha(\ )) \sim \exp(-N\phi_I(\psi; \Sigma_0, \alpha(\ ))) \tag{6·11}$$

in the thermodynamic limit.

The probability density Π_I is induced from the microcanonical measure for the initial conditions on Σ_0. The rate function ϕ_I is a convex and non-negative function, and the most probable value $\bar{I}_*$ satisfies

$$\phi_I(\bar{I}_*; \Sigma_0, \alpha(\ )) = 0. \tag{6·12}$$

Here, Eq. (6·6) leads to the inequality

$$\bar{I}_* \geq 0 \tag{6·13}$$

for an arbitrary most probable process.

We then attempt to find a relation between $\bar{I}_*$ and $\Delta S_{\rm B}$. First, let us note that $\langle I\rangle_0$ can be rewritten as

$$\langle I\rangle_0 = -\int_{\Sigma_0} \mu_{\rm mc}(\Delta_\epsilon(\Gamma(0));\Sigma_0)\log\frac{\tilde{\mu}(\Delta_\epsilon(\tilde{\Gamma}(-\tau));\Upsilon_\tau)}{\mu_{\rm mc}(\Delta_\epsilon(\Gamma(0));\Sigma_0)}, \tag{6·14}$$

where we have used Eq. (5·51). We consider the thermodynamic limit in the expression Eq. (6·14), although the argument is not completely formalized yet.

When the most probable process $\Sigma_0 \to_* \Sigma_1$ is realized, $\Upsilon_\tau \cap \Sigma_1$ is dominant in Υ_τ with respect to the microcanonical measure on Σ_0. Also, from the mixing property, $\Upsilon_\tau \cap \Sigma_1$ may be identified with Σ_1 in a coarse-graining description of the phase space. Thus, we can expect that $\tilde{\mu}(\Delta_\epsilon(\tilde{\Gamma}(-\tau));\Upsilon_\tau)$ in Eq. (6·14) may be replaced by $\mu_{\rm mc}(\Delta_\epsilon(\tilde{\Gamma}(-\tau));\Sigma_1)\exp(o(N))$, in an appropriate limit of small ϵ, large τ, and large N. When this is true, we can rewrite Eq. (6·14) as

$$\langle I\rangle_0 = -\int_{\Sigma_0} \mu_{\rm mc}(\Delta_\epsilon(\Gamma(0));\Sigma_0)\log\frac{\mu_{\rm mc}(\Delta_\epsilon(\tilde{\Gamma}(-\tau));\Sigma_1)}{\mu_{\rm mc}(\Delta_\epsilon(\Gamma(0));\Sigma_0)} + o(N) \tag{6·15}$$

$$= -\int_{\Sigma_0} \mu_{\rm mc}(\Delta_\epsilon(\Gamma(0));\Sigma_0)\log\frac{|\Sigma_0|}{|\Sigma_1|} + o(N). \tag{6·16}$$

Thus, we obtain

$$\bar{I}_* = \lim_{N\to\infty}\frac{1}{N}\log\frac{|\Sigma_1|}{|\Sigma_0|}. \tag{6·17}$$

This equality shows that the value of $\bar{I}_*$ is determined by the initial and final states, irrespective of the details of the process $\Sigma_0 \to_* \Sigma_1$. Furthermore, from Eq. (3·28) in §3.2, we can express $\bar{I}_*$ in the form

$$\bar{I}_* = \lim_{N\to\infty}\frac{1}{N}\log\frac{\Omega_1}{\Omega_0} \tag{6·18}$$

$$= \lim_{N\to\infty}\frac{1}{N}[S_{\rm B}(\Sigma_1) - S_{\rm B}(\Sigma_0)]. \tag{6·19}$$

That is, the most probable value of the irreversible information loss is equal to the change of the Boltzmann entropy per unit degree. Also, the inequality Eq. (6·13) can be read as

$$S_{\rm B}(\Sigma_1) \geq S_{\rm B}(\Sigma_0) + o(N) \tag{6·20}$$

for an arbitrary most probable process $\Sigma_0 \to_* \Sigma_1$. This implies that the Boltzmann entropy satisfies the entropy principle in thermodynamics.

6.3. *Fluctuation theorem*

Let us consider the probability $\Pi_I(\psi;\Upsilon_\tau,\tilde{\alpha}(\ ))d\psi$ that the irreversible information loss $\tilde{I}$ takes a value within $[N\psi, N(\psi+d\psi)]$ when the measure of the initial conditions is assumed to be $\tilde{\mu}$ and the time-reversed protocol $\tilde{\alpha}(\ )$ is given. $\Pi_I(\psi;\Upsilon_\tau,\tilde{\alpha}(\ ))d\psi$ can be written as

$$\tilde{\Pi}(\psi;\Upsilon_\tau,\tilde{\alpha}(\ ))d\psi = \int_{\Upsilon_\tau}\tilde{\mu}(\Delta_\epsilon(\tilde{\Gamma}(-\tau));\Upsilon_\tau)\mathbf{E}(\psi \leq \tilde{I}(\tau,\tilde{\Gamma}(-\tau))/N \leq \psi+d\psi), \tag{6·21}$$

where $\mathbf{E}(*)$ takes the value 1 when the statement $*$ is true. The right-hand side of this expression can be rewritten as

$$\int_{\Upsilon_\tau} \tilde{\mu}(\Delta_\epsilon(\tilde{\Gamma}(-\tau));\Upsilon_\tau)\mathbf{E}(-\psi - d\psi \le I(\tau,\Gamma(0))/N \le -\psi)$$
$$= \exp(N\psi)\int_{\Sigma_0} \mu_{\rm mc}(\Delta_\epsilon(\Gamma(0)))\mathbf{E}(-\psi - d\psi \le I(\tau,\Gamma(0))/N \le -\psi) \quad (6\cdot22)$$
$$= \exp(N\psi)\Pi_I(-\psi;\Sigma_0,\alpha(\ ))d\psi, \quad (6\cdot23)$$

where we have used Eq. (5·51) in order to obtain the second line. Therefore, we obtain

$$\frac{\Pi_I(\psi;\Sigma_0,\alpha(\ ))}{\tilde{\Pi}_I(-\psi;\Upsilon_\tau,\tilde{\alpha}(\ ))} = \exp(N\psi). \quad (6\cdot24)$$

This may be called the "fluctuation theorem in Hamiltonian systems with a time-dependent parameter".

We could not derive a useful expression of Eq. (6·24) in the thermodynamic limit. We now explain the reason. Suppose that the most probable process $\tilde{\Sigma}_1 \to_* \Sigma_0$ is realized by the time-reversed protocol $\tilde{\alpha}(\ )$. Then, the dominant region of Υ_τ with respect to the measure $\tilde{\mu}$, which contributes $\Pi_I(\psi;\Upsilon_\tau,\tilde{\alpha}(\ ))$, is around the energy surface $\tilde{\Sigma}_1$. However, if Υ_τ were replaced by $\tilde{\Sigma}_1$, ψ in Eq. (6·24) could not be substituted by, for example, the most probable value $\bar{I}_*$ for $\Sigma_0 \to_* \Sigma_1$. This precludes the significance of Eq. (6·24).

When we are concerned with an infinitely small step process, we can derive the fluctuation-response relation from Eq. (6·24). In such a process, Υ_τ may be replaced by Σ_0 at the lowest order approximation. Then, since Π_I may be approximated by a Gaussian distribution for large N, we can write

$$\log \tilde{\Pi}_I(-\psi;\Sigma_0,\tilde{\alpha}(\ )) = -N\frac{(\psi+\bar{I}'_*)^2}{2\sigma'^2} + o(N), \quad (6\cdot25)$$
$$\log \Pi_I(\psi;\Sigma_0,\alpha(\ )) = -N\frac{(\psi-\bar{I}_*)^2}{2\sigma^2} + o(N). \quad (6\cdot26)$$

The fluctuation theorem Eq. (6·24) leads to

$$-N\frac{(\psi-\bar{I}_*)^2}{2\sigma^2} + N\frac{(\psi+\bar{I}'_*)^2}{2\sigma'^2} = N\psi + o(N). \quad (6\cdot27)$$

Since $\bar{I}_*$ and $\bar{I}'_*$ are infinitely small, Eq. (6·27) may be valid for arbitrary ψ in a finite range including $\psi = 0$. Thus, we have

$$\bar{I}'_* = \bar{I}_*, \quad (6\cdot28)$$
$$\sigma'^2 = \sigma^2. \quad (6\cdot29)$$

Substituting these equalities into Eq. (6·27), we obtain

$$N\bar{I}_* = \frac{N}{2}\sigma^2 \quad (6\cdot30)$$
$$= \frac{N^2}{2}\left\langle (\delta\psi)^2 \right\rangle \quad (6\cdot31)$$
$$= \frac{1}{2}\left\langle (I - N\bar{I}_*)^2 \right\rangle. \quad (6\cdot32)$$

Comparing Eq. (6·32) with Eq. (3·44) in §3.3, we find that this result is consistent with Eq. (6·19).

§7. Excess information loss

At each point $\Gamma(t)$ in a trajectory segment, $\{\Gamma(t), 0 \le t \le \tau\}$, we can consider a Hamiltonian system defined on the energy surface $\Sigma(t)$ by *virtually* fixing the parameter value to $\alpha(t)$. Then, as discussed in §4.4, we can calculate the information loss rate $h(\Gamma;\Sigma(t))$ at $\Gamma \in \Sigma(t)$ in this virtual Hamiltonian system. We define the *excess information loss rate* as *)

$$h_{\rm ex}(t,\Gamma(0)) = h_{\rm a}(t,\Gamma(0)) - h(\Gamma(t);\Sigma(t)). \tag{7·1}$$

Further, the excess information loss $H_{\rm ex}$ is defined as the time integration of $h_{\rm ex}$:

$$H_{\rm ex}(\tau,\Gamma(0)) = \int_0^\tau dt h_{\rm ex}(t,\Gamma(0)). \tag{7·2}$$

Similarly, the excess information loss rate at $\tilde{\Gamma}(-t)$ in the time-reversed trajectory is given by

$$\tilde{h}_{\rm ex}(-t,\tilde{\Gamma}(-\tau)) = \tilde{h}_{\rm a}(-t,\tilde{\Gamma}(-\tau)) - h(\tilde{\Gamma}(-t);\Sigma(t)), \tag{7·3}$$

where $\tilde{\Gamma}(-t) \in \Sigma(t)$, and the excess information loss in the time-reversed trajectory is written as

$$\tilde{H}_{\rm ex}(\tau,\tilde{\Gamma}(-\tau)) = \int_0^\tau dt \tilde{h}_{\rm ex}(-t,\tilde{\Gamma}(-\tau)). \tag{7·4}$$

Using these quantities, we rewrite the irreversible information loss I as

$$I(\tau,\Gamma(0)) = \int_0^\tau dt[h_{\rm a}(t,\Gamma(0)) - \tilde{h}_{\rm a}(-t,\tilde{\Gamma}(-\tau))] + o(N) \tag{7·5}$$

$$= \int_0^\tau dt[h_{\rm ex}(t,\Gamma(0)) - \tilde{h}_{\rm ex}(-t,\tilde{\Gamma}(-\tau))] + o(N) \tag{7·6}$$

$$= H_{\rm ex}(\tau,\Gamma(0)) - \tilde{H}_{\rm ex}(\tau,\tilde{\Gamma}(-\tau)) + o(N), \tag{7·7}$$

where we have used the equality

$$h(\Gamma(t);\Sigma(t)) = h(\tilde{\Gamma}(-t);\Sigma(t)) + \frac{d}{dt}o(N) \tag{7·8}$$

(see Eq. (5·22)). Further, through the definition of $H_{\rm ex:rev}$,

$$H_{\rm ex:rev}(\tau,\Gamma(0)) = \frac{1}{2}[H_{\rm ex}(\tau,\Gamma(0)) + \tilde{H}_{\rm ex}(\tau,\tilde{\Gamma}(-\tau))], \tag{7·9}$$

Eq. (7·7) becomes

$$\frac{1}{2}I(\tau,\Gamma(0)) = H_{\rm ex}(\tau,\Gamma(0)) - H_{\rm ex:rev}(\tau,\Gamma(0)) + o(N). \tag{7·10}$$

*) We obtained the idea of the excess information loss from a paper by Oono and Paniconi,[4] where they defined the excess heat in constructing steady-state thermodynamics.

From the viewpoint of numerical calculations, the excess information loss is a more tractable quantity than the irreversible information loss, because $H_{\rm ex}(\tau, \Gamma(0))$ converges to a certain value $H_{\rm ex}(\infty, \Gamma(0))$ when $\tau \to \infty$. This convergence is expected from the fact that $h_{\rm ex}(t, \Gamma(0))$ converges to 0 when $t - \tau_f \to \infty$. Similarly, we expect that $\tilde{H}_{\rm ex}(\tau, \tilde{\Gamma}(-\tau))$ converges to a certain value $\tilde{H}_{\rm ex}(\infty, \tilde{\Gamma}(-\infty))$ when $\tau_i \to \infty$. (Note that $\tau \to \infty$ when $\tau_i \to \infty$.) Also, $H_{\rm ex:rev}(\infty, \Gamma(0))$ is determined.

Let us consider the average over the initial conditions sampled from the microcanonical ensemble on the energy surface Σ_0. From Eq. (7·10), we have

$$\frac{1}{2}N\bar{I}_* = \langle H_{\rm ex}\rangle_0 - \langle H_{\rm ex:rev}\rangle_0 + o(N) \tag{7·11}$$

in the thermodynamic limit, where $\langle H_{\rm ex}\rangle_0$ is the average of $H_{\rm ex}(\infty, \Gamma(0))$. Since $H_{\rm ex}(\infty, \Gamma(0))$ can be obtained numerically without referring to the time-reversed trajectory, $\langle H_{\rm ex}\rangle_0$ is a directly measurable quantity. Although $\langle H_{\rm ex:rev}\rangle_0$ is not easily obtained numerically, it may be expected to have a certain relation to the *quasi-static excess information loss* $H_{\rm ex:qs}$, which is defined as

$$H_{\rm ex:qs} = \int_0^\infty dt \frac{d\alpha}{dt}\Phi(\Sigma(t)), \tag{7·12}$$

where the quantity $\Phi(\Sigma(t))d\alpha$ is the excess information loss calculated under the assumption that the equilibrium state is *virtually* realized at each time t along the trajectory $\Gamma(\ )$. (Recall similar discussion below Eq. (3·31).) Note that $H_{\rm ex:qs}$ becomes the real excess information loss when the process is quasi-static.

We now discuss the relation between $H_{\rm ex:rev}$ and $H_{\rm ex:qs}$. We consider a step process realized by an infinitely small parameter change $\alpha \to \alpha + \Delta\alpha$ at $t = \tau_i$. Then, the quantity $\Phi(\Sigma_0)$ is given by

$$\langle H_{\rm ex}\rangle_0 = \Phi(\Sigma_0)\Delta\alpha + O((\Delta\alpha)^2). \tag{7·13}$$

Also, using Eqs. (3·40) and (6·19), we obtain

$$\bar{I}_* = O((\Delta\alpha)^2). \tag{7·14}$$

Therefore, we find

$$\langle H_{\rm ex:rev}\rangle_0 = \Phi(\Sigma_0)\Delta\alpha + O((\Delta\alpha)^2). \tag{7·15}$$

Since a quasi-static process can be realized by repeating an infinite number of infinitely small step processes, $\langle H_{\rm ex:rev}\rangle_0$ for a quasi-static process $\Sigma_0 \overset{\rm qs}{\to} \Sigma_1$ can be written as

$$\langle H_{\rm ex:rev}\rangle_0 = \int_{\alpha_0}^{\alpha_1} d\alpha \Phi(\Sigma_{(\alpha)}), \tag{7·16}$$

where $\Sigma_{(\alpha)}$ is the equilibrium state such that

$$\Sigma_0 = \Sigma_{(\alpha_0)} \overset{\rm qs}{\to} \Sigma_{(\alpha)}. \tag{7·17}$$

Then, Eq. (7·16) implies

$$\langle H_{\rm ex:rev}\rangle_0 = \langle H_{\rm ex:qs}\rangle_0 . \tag{7·18}$$

Note, however, that the validity of Eq. (7·18) is ensured only for quasi-static processes. Nevertheless, we assume that Eq. (7·18) holds at least near quasi-static processes. Based on this assumption, the right-hand side of Eq. (7·11) can be calculated numerically without referring to time-reversed trajectories, and from Eqs. (6·19), (7·11) and (7·18), we obtain the expression

$$\frac{1}{2}\langle \Delta S_{\rm B}\rangle_0 = \langle H_{\rm ex}\rangle_0 - \langle H_{\rm ex:qs}\rangle_0 + o(N). \tag{7·19}$$

7.1. *Minimum principle*

Applying the inequality Eq. (6·13) to the expression Eq. (7·11), we obtain

$$\langle H_{\rm ex}\rangle_0 \geq \langle H_{\rm ex:rev}\rangle_0 + o(N). \tag{7·20}$$

This inequality implies that *the excess information loss must not be less than its reversible part.* Further $H_{\rm ex:rev}$ may be related to the quasi-static excess information loss, as discussed in the previous subsection. Therefore, Eq. (7·20) has certain analogy with the minimum work principle in thermodynamics with an isothermal environment, which states that the work done by external agents must not be less than the quasi-static work. In this sense, Eq. (7·20) may be regarded as the minimum excess information loss principle. However, we do not yet understand the significance of the inequality Eq. (7·20). We expect that the analysis of subsystems may provide us further insight into the meaning of Eq. (7·20). This is a future problem.

7.2. *Expression of Φ*

In this subsection we derive an expression of Φ in terms of Lyapunov vectors. Suppose that the value of α is changed instantaneously from α_0 to $\alpha_0 + \Delta\alpha$ at time $t = 0$. The trajectory $\Gamma(\ )$ is not differentiable at $t = 0$. We consider the excess information loss for $\Gamma(\ )$,

$$H_{\rm ex}(\infty, \Gamma(0)) = \int_0^\infty dt[h_{\rm a}(t;\Gamma(0)) - h(\Gamma(t);\Sigma_1)], \tag{7·21}$$

where Σ_1 is an energy surface after the parameter change.

Let $\{\xi_i^{(0)},\ 1 \leq i \leq 2N\}$ be a set of Lyapunov vectors at $\Gamma(0)$ on the energy surface Σ_0. We then define a set of vectors $\{a_i(t),\ 1 \leq i \leq 2N\}$ in the tangent space at $\Gamma(t)$ as

$$a_i(t) = \mathcal{T}(t, 0; \Gamma(0))\xi_i^{(0)}, \tag{7·22}$$

where $\mathcal{T}(t, 0; \Gamma(0))$ is the linearized evolution map along the trajectory on the energy surface Σ_1. Note that a_i is not the Lyapunov vector at $\Gamma(t)$, because $\xi_i^{(0)}$ is *not* the Lyapunov vector on the energy surface Σ_1.

The i-th Lyapunov vector along the trajectory $\{\Gamma(t),\ 0 \leq t \leq \infty\}$ on the energy surface Σ_1 is denoted by $\xi_i^{(1)}(\Gamma(t))$. The expansion factor $\Lambda_i(t, \Gamma(0))$ satisfies

$$\mathcal{T}(t, 0; \Gamma(0))\xi_i^{(1)}(\Gamma(0)) = \Lambda_i(t, \Gamma(0))\xi_i^{(1)}(\Gamma(t)). \tag{7·23}$$

Using a_i and Λ_i, we can write Eq. (7·21) as

$$H_{\rm ex}(\infty, \Gamma(0)) = \lim_{t\to\infty} \left[\log | \wedge_{i=1}^{N_p} a_i(t)| - \sum_{i=1}^{N_p} \log \Lambda_i(t, \Gamma(0)) \right]. \tag{7·24}$$

Let us evaluate the right-hand side of Eq. (7·24). We expand $\xi_i^{(0)}$ in terms of the set of Lyapunov vectors $\{\xi_i^{(1)}(\Gamma(0)),\ 1 \le i \le 2N\}$ in such a way that

$$\xi_i^{(0)} = \sum_{j=1}^{2N} Q_{ij} \xi_j^{(1)}(\Gamma(0)). \tag{7·25}$$

Here, the matrix $\mathcal{Q}$ is defined at $\Gamma(0)$ and depends on $\Delta\alpha$. Then, from Eqs. (7·22) and (7·25), $a_i(t)$ can be expanded in the form

$$a_i(t) = \sum_{j=1}^{2N} Q_{ij} \Lambda_j(t, \Gamma(0)) \xi_j^{(1)}(\Gamma(t)). \tag{7·26}$$

Using this expression, we write $| \wedge_{i=1}^{N_p} a_i(t)|$ as

$$\left| \sum_{(j_1,\cdots,j_{N_p})} \left[\prod_{k=1}^{N_p} Q_{kj_k} \right] \left[\prod_{k=1}^{N_p} \Lambda_{j_k}(t, \Gamma(0)) \right] \left[\wedge_{k=1}^{N_p} \xi_{j_k}^{(1)}(\Gamma(t)) \right] \right|, \tag{7·27}$$

where the index j_k varies from 1 to $2N$. When t is sufficiently large, the contribution from the unstable directions becomes dominant. Thus, for sufficiently large t, we derive

$$\begin{aligned} \log | \wedge_{i=1}^{N_p} a_i(t)| \simeq \log \Bigg| & \sum_{j_1=1}^{N_p} \cdots \sum_{j_{N_p}=1}^{N_p} {\rm sgn}(j_1, \cdots, j_{N_p}) \left[\prod_{k=1}^{N_p} Q_{kj_k} \right] \\ & \cdot \left[\prod_{k=1}^{N_p} \Lambda_k(t, \Gamma(0)) \right] \left| \wedge_{i=1}^{N_p} \xi_i^{(1)}(\Gamma(t)) \right| \Bigg| \end{aligned} \tag{7·28}$$

$$\simeq \log \det{}_+ \mathcal{Q} + \sum_{i=1}^{N_p} \log \Lambda_i(t, \Gamma(0)), \tag{7·29}$$

where

$$\det{}_+ \mathcal{Q} = \sum_{j_1=1}^{N_p} \cdots \sum_{j_{N_p}=1}^{N_p} {\rm sgn}(j_1, \cdots, j_{N_p}) Q_{1j_1} \cdots Q_{N_p, j_{N_p}}, \tag{7·30}$$

and we have used the normalization condition $| \wedge_{i=1}^{N_p} \xi_i^{(1)}(\Gamma(t))| = 1$. Finally, substituting Eq. (7·29) into Eq. (7·24), we obtain

$$H_{\rm ex}(\infty, \Gamma(0)) = \log \det{}_+ \mathcal{Q}. \tag{7·31}$$

Therefore, from the definition of $\Phi(\Sigma_0)$, we have

$$\Phi(\Sigma_0) = \lim_{\Delta\alpha\to 0} \frac{\langle \log \det_+ \mathcal{Q} \rangle_0}{\Delta\alpha}. \tag{7·32}$$

This expression shows that Φ takes a non-zero value when the unstable manifold varies linearly for an infinitely small parameter change.

§8. Numerical experiments

To this point, we have only developed theoretical arguments. However, one may point out that these lack mathematical rigor. For this reason, we now present evidence from numerical experiments to support the validity of the theoretical arguments.

As a direct experimental test of our theory, we should check the relation between the irreversible information loss and the Boltzmann entropy. However, unfortunately, we have not yet completed this test, because it is difficult to measure numerically the irreversible information loss. The reasons for this difficulty are as follows.

First, the time-reversed trajectory is needed in calculation of $I(\tau, \Gamma(0))$. This fact creates a delicate problem: Suppose that we numerically obtain a trajectory segment $\{\Gamma(t),\ 0 \le t \le \tau\}$. Since the system is chaotic, this trajectory is not an approximation of the true trajectory starting from the initial condition $\Gamma(0)$. However, when a pseudo-orbit tracing property is valid in the system, there is a true trajectory which is close to the trajectory segment obtained numerically. Then, when we integrate the time-reversed equations of motion with the initial condition $\tilde{\Gamma}(-\tau)$, the trajectory obtained numerically deviates from $\{\tilde{\Gamma}(t),\ -\tau \le t \le 0\}$ due to the orbital instability. Therefore, in order to obtain the time-reversed trajectory $\{\tilde{\Gamma}(t),\ -\tau \le t \le 0\}$, we must store the data for the original trajectory.

Second, even if we obtain numerically $I(\tau, \Gamma(0))$, this quantity does not converge to a fixed value for $\tau \to \infty$, because only the extensive part of $I(\tau, \Gamma(0))$ converges. Therefore, it is not easy to choose the value of τ in numerical calculations. In principle, we have only to choose a large value of N so that the arbitrariness is reduced. However, we need much more time to study systems with larger N.

In this paper, instead of the irreversible information loss, we discuss the numerical study of the excess information loss, with particularly focusing on Eq. (7·19). Since $H_{\rm ex}(\tau, \Gamma(0))$ converges to $H_{\rm ex}(\infty, \Gamma(0))$ as $\tau \to \infty$, the numerical calculation of the excess information loss may be simpler than that of the irreversible information loss. Also, the time-reversed trajectory is not needed in the calculation of $H_{\rm ex}$.

8.1. *Model*

A system consisting of many molecules with short-range repulsive interaction may be the most realistic model to study thermodynamic irreversibility. However, since we are concerned with universal aspects of irreversibility, the choice of the system does not matter. Simpler models may be better for our purpose. This is the reason we numerically studied the Fermi-Pasta-Ulam (FPU) model.[26] The

Hamiltonian of the FPU model is given by

$$H(\{q_i\},\{p_i\};g)=\sum_{i=1}^{N}\left[\frac{1}{2}p_i^2+\frac{1}{2}(q_{i+1}-q_i)^2+\frac{g}{4}(q_{i+1}-q_i)^4\right], \tag{8·1}$$

where the value of g is changed in time. That is, $\alpha(\ )$ in previous sections is identified with $g(\ )$ in this section.

The evolution equations for $(\{q_i\},\{p_i\})$ are written as

$$\frac{dq_i}{dt}=p_i, \tag{8·2}$$

$$\frac{dp_i}{dt}=(q_{i+1}-q_i)+g(q_{i+1}-q_i)^3-(q_i-q_{i-1})-g(q_i-q_{i-1})^3. \tag{8·3}$$

We assume the periodic boundary conditions $q_0=q_N$ and $q_{N+1}=q_1$ in Eq. (8·3). With these boundary conditions, $\sum_i p_i$ is a conserved quantity. We assume that $\sum_i p_i=0$ for simplicity. Then, $\sum_i q_i$ also becomes a conserved quantity. We also assume $\sum_i q_i=0$. In the remaining part of this section, the energy surface with the condition $\sum_i p_i=\sum_i q_i=0$ is simply called "the energy surface". We numerically solved Eqs. (8·2) and (8·3) using the 4-th order symplectic integrator method [33] with a time step $\delta t=0.005$. Since we are concerned with the thermodynamic limit, we check the N dependence of our conclusions.

8.2. *Lyapunov analysis*

In this subsection, we assume that g takes a constant value, say g_0. Let E_0 be the energy. Then, when $E_0 g_0$ is sufficiently large, a system with large N exhibits high-dimensional chaos. As an example of such a parameter value set, $(E_0,g_0)=(1.0,10.0)$ is assumed.

We first check the mixing property with respect to the micro-canonical measure by discussing relaxation behavior. (See the last paragraph of §2.3.) Figure 1 shows

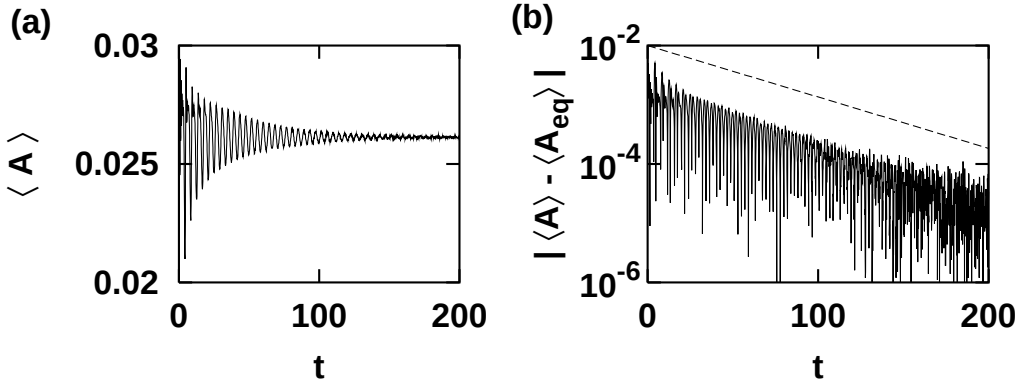

Fig. 1. Relaxation behavior of the ensemble average of A. Here $N=20$. The ensemble of initial conditions is produced arbitrarily by fixing E and A as $E=1$ and $A(0)=0.01$. (b) $\log|\langle A\rangle - A_{\rm eq}|$ as a function of t. $A_{\rm eq}$ is the equilibrium value determined by the graph (a). The dotted line represents $|\langle A\rangle - A_{\rm eq}|=\exp(-t/50)/100$.

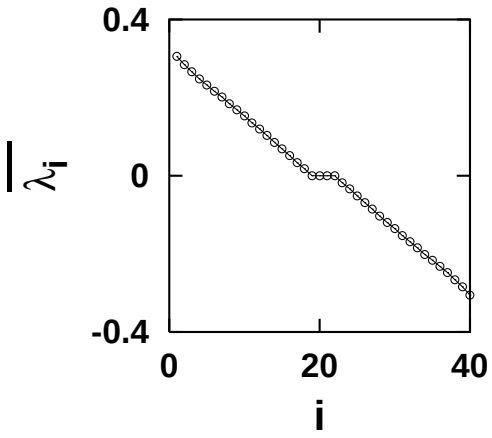

Fig. 2. Lyapunov spectrum. $N = 20$.

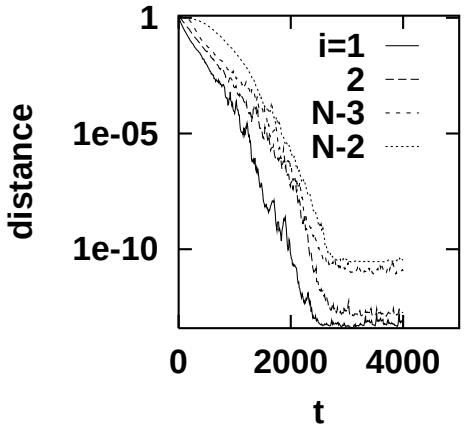

Fig. 3. Convergence of a set of orthogonal unit vectors. Here $N = 20$.

an example of how the average of A relaxes to the equilibrium value when the initial conditions are sampled from an arbitrary ensemble we assumed. As far as we checked, we observed similar relaxation behavior to the same equilibrium value for different sets of initial conditions. We thus conclude based on these numerical results that the system possesses the mixing property. Therefore, the ensemble of the initial conditions at $t = 0$ is regarded as the microcanonical ensemble with energy E_0 when the ensemble is produced by the sufficiently long time evolution of phase space points sampled from a distribution absolutely continuous with respect to the Lebesgue measure on the energy surface. Here, we remark that the relaxation curve includes an oscillatory component, while the envelop curve exhibits exponentially decreasing behavior. Both the period of oscillation and the relaxation time seem to be larger for larger N.

In order to demonstrate the chaotic nature quantitatively, we display the Lyapunov exponents in Fig. 2. Note that there are two additional zero Lyapunov exponents because of momentum conservation. That is, $N_p = N - 2$. The convergence of orthonormal frames is confirmed in the manner described in §4.2. Figure 3 shows the time evolution of the average of the distance $d(\mathcal{F}(t, \Gamma(0))e_i, \mathcal{F}(t, \Gamma(0))e'_i)$ over the initial conditions, where $i = 1, 2, N - 3$ and $N - 2$. We see that the distance decreases to a computational noise level after $t = 3000$.

8.3. *Boltzmann entropy*

The Boltzmann entropy $S_{\rm B}(E, g)$ is calculated numerically in the following way: First, according to the adiabatic theorem, which was discussed in §3.1, the phase space volume enclosed by an energy surface is conserved along quasi-static processes. Therefore, the equality

$$\Omega(E, g) = \Omega(E_*, 0) \tag{8·4}$$

holds for the quasi-static process $(E, g) \overset{\rm qs}{\to} (E_*, 0)$. Since the FPU model with $g = 0$ is reduced to the harmonic oscillator model, $\Omega(E_*, 0)$ is given by the volume of the

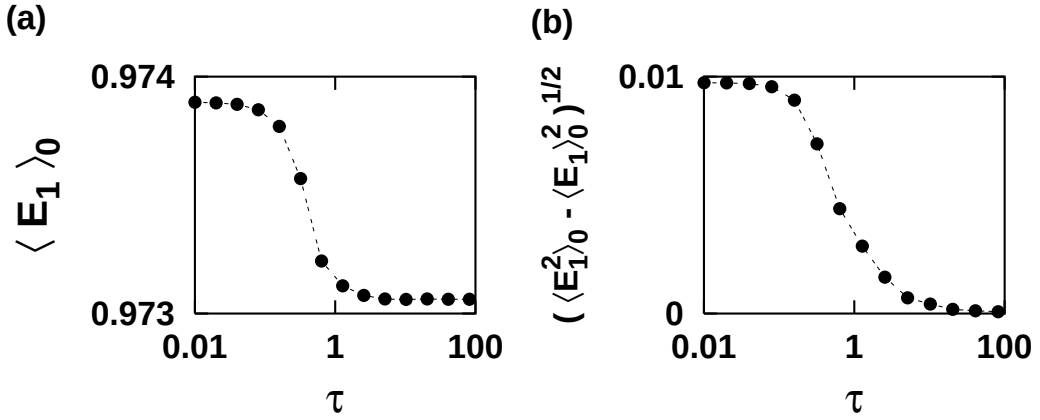

Fig. 4. (a) Average value of the final energy as a function of τ. (b) Deviation of values of the final energy as a function of τ. Here $N = 20$ and $\Delta g = -1.0$.

$2N - 2$ dimensional sphere, and it is calculated as

$$\Omega(E_*, 0) = cE_*^{N-1}, \tag{8·5}$$

where c does not depend on E_*. Thus, the Boltzmann entropy at (E, g) can be evaluated as*)

$$S_{\rm B}(E, g) = S_{\rm B}(E_*, 0) = (N - 1) \log E_*, \tag{8·6}$$

where an additive constant with respect to E_* is omitted.

In numerical experiments, a quasi-static process $(E_0, g_0) \overset{\rm qs}{\to} (E_1, g_0 + \Delta g)$ is realized by the large τ limit of the protocol $g(t) = g_0 + \Delta g t/\tau$ for $0 \le t \le \tau$. In Fig. 4, the average and deviation of E_1 are plotted as functions of τ. We find that the deviation becomes smaller for larger τ, and we may assume that quasi-static processes are realized when $\tau > 100$. The equi-entropy curve through $\Sigma_0 = (E_0, g_0)$ in Fig. 5 was obtained in this way. We express the curve by $E = E_{\rm qs}(g; \Sigma_0)$. Similarly, as shown in Fig. 5, we can draw equi-entropy curves in the (E, g) space.

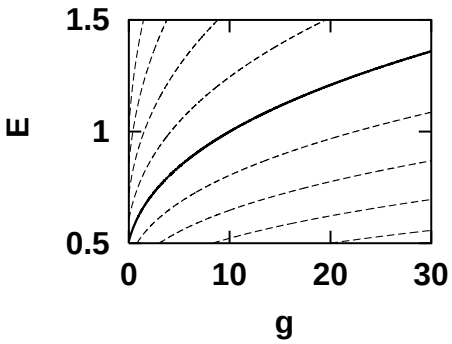

Fig. 5. Equi-entropy curve through (E_0, g_0) (solid curve) and other equi-entropy curves (dotted curves).

Now, we consider a step process realized by the instantaneous change of the value of g from g_0 to $g_1 = g_0 + \Delta g$ at $t = 0$. Then, the energy after the switching becomes E_1, whose value depends on the choice

*) In a previous paper,[3] we wrote the factor on the right-hand side as $N - 2$, not $N - 1$. This is a mistake.

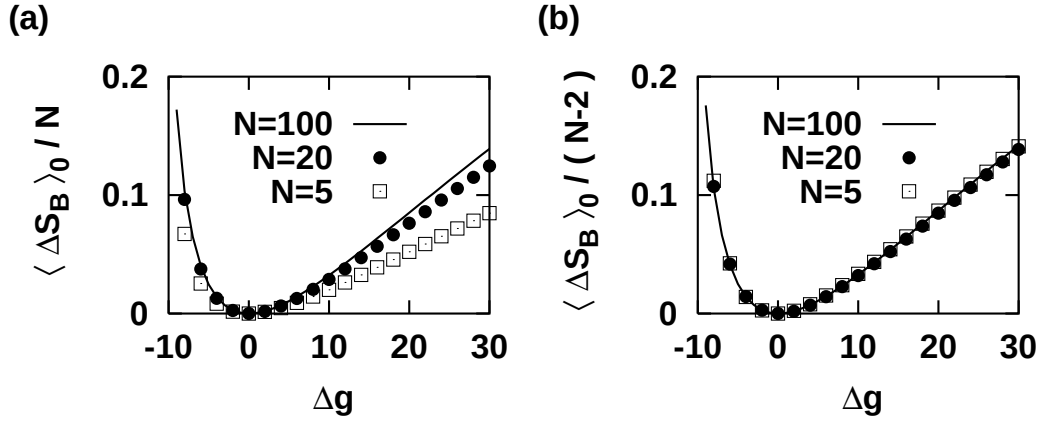

Fig. 6. Entropy difference as a function of Δg. Different normalizations for the N dependence are used in (a) and (b).

of the initial condition. The entropy difference $\Delta S_{\rm B}$ is calculated by

$$\Delta S_{\rm B} = S_{\rm B}(E_1, g_1) - S_{\rm B}(E_0, g_0), \tag{8·7}$$

with the formula Eq. (8·6). In Fig. 6, the average of the entropy difference over the initial conditions, $\langle \Delta S_{\rm B} \rangle_0$, is plotted as a function of Δg. This graph shows that $\langle \Delta S_{\rm B} \rangle_0$ is positive. Also, as shown in Fig. 7, the relative fluctuation of $\Delta S_{\rm B}$ becomes less as N is increased. This implies the existence of the large deviation property for $\Delta S_{\rm B}$.

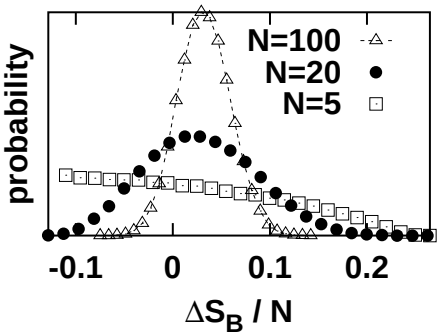

Fig. 7. The probability of the entropy difference. Here $\Delta g = 10$.

8.4. *Excess information loss*

We now give the results of numerical calculations of the excess information loss for the step processes $(E_0, g_0) \to (E_1, g_0 + \Delta g)$. In Fig. 8, $H_{\rm ex}(t, \Gamma(0))$ for four choices of the initial condition $\Gamma(0)$ are plotted as functions of t. One can see that $H_{\rm ex}(\infty, \Gamma(0))$ is clearly defined. In Fig. 9, we show the average of $H_{\rm ex}(t, \Gamma(0))$ over the initial conditions chosen from the microcanonical ensemble on the energy surface Σ_0. The quantity $\langle H_{\rm ex} \rangle_0$ is given as the value at $t = \infty$ in this graph.

In Fig. 10, $\langle H_{\rm ex} \rangle_0$ is plotted as a function of Δg. From this graph, we can evaluate the value of $\Phi(\Sigma_0)$ by the equality

$$\langle H_{\rm ex} \rangle_0 = \Phi(\Sigma_0)\Delta g + o(\Delta g) \tag{8·8}$$

for $\Delta g \to 0$. In a similar way, in principle, we can calculate the value of $\Phi(\Sigma)$ at

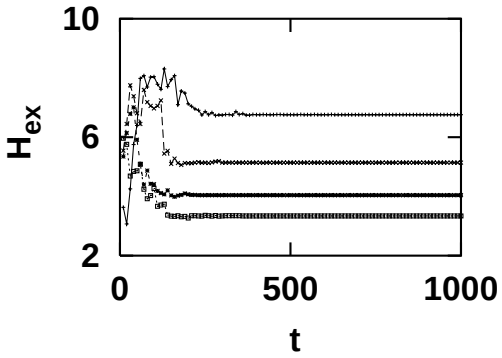

Fig. 8. $H_{\rm ex}(t, \Gamma(0))$ as functions of t for four different initial conditions. Here $N = 20$ and $\Delta g = 20$.

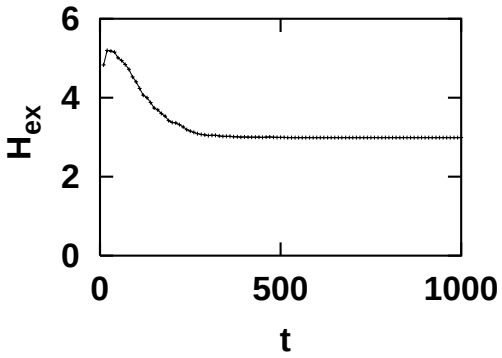

Fig. 9. Average value of $H_{\rm ex}(t, \Gamma(0))$ as a function of t. Here $N = 20$ and $\Delta g = 20$.

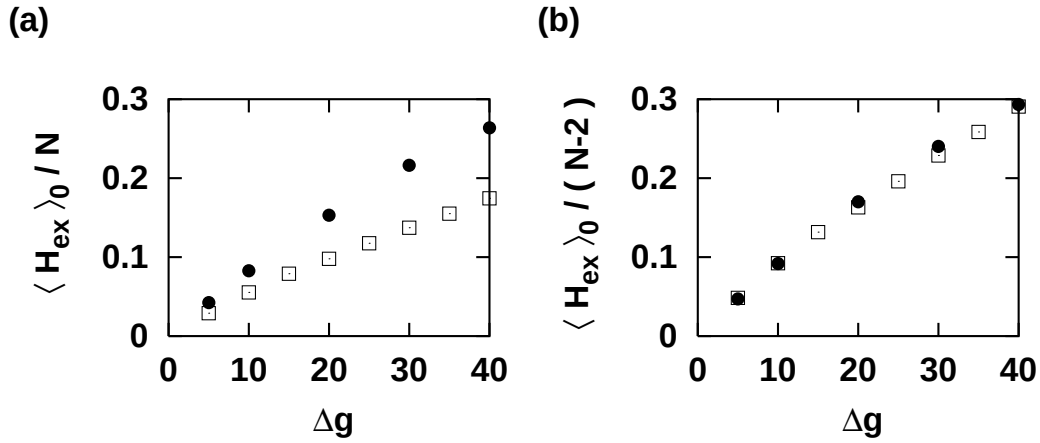

Fig. 10. $\langle H_{\rm ex} \rangle_0$ as a function of Δg. The squares and filled circles represent the data for $N = 5$ and $N = 20$, respectively. Different normalizations for the N dependence are used in (a) and (b).

each energy surface. In particular, $H_{\rm ex:qs}$ for the step process is given by

$$H_{\rm ex:qs} = \frac{1}{2}(\Phi(\Sigma_0) + \Phi(\Sigma_1))\Delta g, \tag{8·9}$$

where $\Sigma_1 = (E_1, g_1)$ (recall Eq. (7·12) for the definition of $H_{\rm ex:qs}$).

8.5. *Main experiment*

In order to check the validity of Eq. (7·19), we need to calculate $\langle H_{\rm ex:qs} \rangle_0$, the average of $H_{\rm ex:qs}$ over the initial conditions. This is written as

$$\langle H_{\rm ex:qs} \rangle_0 = \frac{1}{2}(\Phi(\Sigma_0) + \langle \Phi(\Sigma_1) \rangle_0)\Delta g. \tag{8·10}$$

However, this average cannot be calculated efficiently, because Σ_1 depends on the choice of the initial condition, and the microcanonical ensemble on Σ_1 must be

produced so as to evaluate $\Phi(\Sigma_1)$. For this reason, we find out a way to avoid the calculation of $H_{\rm ex:qs}$. We performed the reversed experiments in which the parameter g is changed from g_1 to g_0 with the initial state $\Sigma_0' = (E_{\rm qs}(g_1;\Sigma_0), g_1)$. Suppose that the process $\Sigma_0' \to \Sigma_1'$ is realized, where Σ_1' depends on the choice of the initial condition on the energy surface Σ_0'. Then, we can calculate the Boltzmann entropy change $\Delta S_{\rm B}$ and the excess information loss $H'_{\rm ex}$ for the process $\Sigma_0' \to \Sigma_1'$. These satisfy the relation

$$\frac{1}{2}\langle \Delta S_{\rm B}\rangle_0' = \left\langle H'_{\rm ex}\right\rangle_0' - \left\langle H'_{\rm ex:qs}\right\rangle_0' + o(N), \tag{8·11}$$

where $\langle\ \rangle_0'$ denotes the average over the initial conditions sampled from the microcanonical ensemble on the energy surface Σ_0', and $H'_{\rm ex:qs}$ denotes the quasi-static excess information loss for the step process $\Sigma_0' \to \Sigma_1'$. Here, we can prove the relation

$$\left\langle H_{\rm ex:qs}\right\rangle_0 + \left\langle H'_{\rm ex:qs}\right\rangle_0' = o((\Delta g)^2). \tag{8·12}$$

Proof From the definition of $H_{\rm ex:qs}$, we have

$$\left\langle H'_{\rm ex:qs}\right\rangle_0' = -\frac{1}{2}\left(\Phi(\Sigma_0') + \langle\Phi(\Sigma_1')\rangle_0'\right)\Delta g. \tag{8·13}$$

Using this equation and Eq. (8·9), we obtain

$$\left\langle H_{\rm ex:qs}\right\rangle_0 + \left\langle H'_{\rm ex:qs}\right\rangle_0' = \frac{1}{2}\left(\Phi(\Sigma_0) + \langle\Phi(\Sigma_1)\rangle_0\right)\Delta g - \frac{1}{2}\left(\Phi(\Sigma_0') + \langle\Phi(\Sigma_1')\rangle_0'\right)\Delta g. \tag{8·14}$$

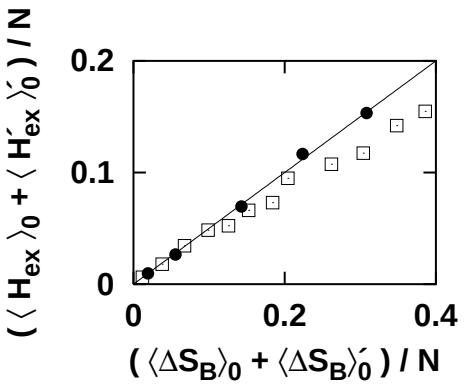

Fig. 11. $\langle H_{\rm ex}\rangle_0 + \langle H'_{\rm ex}\rangle_0'$ as a function of $\langle \Delta S_{\rm B}\rangle_0 + \langle \Delta S_{\rm B}\rangle_0'$. The squares and filled circles represent the data for $N = 5$ and $N = 20$, respectively. The solid line corresponds to $\langle H_{\rm ex}\rangle_0 + \langle H'_{\rm ex}\rangle_0' = 1/2[\langle \Delta S_{\rm B}\rangle_0 + \langle \Delta S_{\rm B}\rangle_0']$.

Recalling the energy change for step processes, we expect

$$\Sigma_0' - \Sigma_1 \sim O((\Delta g)^2), \tag{8·15}$$
$$\Sigma_0 - \Sigma_1' \sim O((\Delta g)^2). \tag{8·16}$$

Substituting these estimations into Eq. (8·14) leads to Eq. (8·12).

□

Finally, from Eqs. (7·19), (8·11) and (8·12), we obtain the equality

$$\frac{1}{2}(\langle \Delta S_{\rm B}\rangle_0 + \langle \Delta S_{\rm B}\rangle_0') = \langle H_{\rm ex}\rangle_0 + \langle H_{\rm ex}\rangle_0' + o(N, (\Delta g)^2). \tag{8·17}$$

This relation can be checked numerically. As shown in Fig. 11, Eq. (8·17) seems to be valid, and therefore our theoretical arguments turn out to be consistent with our numerical results.

§9. Concluding remark

The essence of thermodynamic irreversibility is described by the entropy principle. Therefore, when one discusses thermodynamic irreversibility in the dynamical systems, the purpose is to find a state variable satisfying the entropy principle. Our arguments are made from this natural viewpoint. In this paper, we have found that irreversible information loss leads to a state variable which satisfies the entropy principle.

We expect that our theory may be extended so as to apply to other dynamical systems without Hamiltonians. For example, in dissipative systems driven by external forces, a steady state is realized. The fluctuation properties of such systems have been discussed extensively. However, attempts at the construction of state theory have not yet been developed. Although Oono and Paniconi have proposed an operational method to construct a state variable in steady-state thermodynamics, its validity has not been confirmed.[4] We will attempt to construct non-equilibrium thermodynamic functions from dynamical system models by studying fluctuation properties.

Great variety of nonlinear dynamics is known in the contexts of fluid systems, granular systems, chemical reaction systems, and biological systems. In these systems, the concept of entropy is not self-evident, and it is difficult to connect a concept of this kind in these systems with that in thermodynamics. Even in such systems, it may be important to characterize a state in terms of state functions representing a relation between states. In particular, one of the most important questions in science regards the boundary of the living state. One may ask why we cannot restore the living state when a life is lost. This is not a problem of thermodynamic irreversibility, but rather of a sort of irreversibility in the biological world. However, the question is too general to be discussed scientifically. We should first consider more restricted phenomena related to this question. The determination of cell types in cell differentiation processes may be a good candidate. As already studied by Kaneko et al.,[34] from the dynamical system viewpoint, a cell society may be modeled by chemical networks with a variable number of cells. In these studies, it was found that a rule for the determination of cell types emerges. We expect that this rule might be formalized by state space theory, which shares common features with thermodynamics. As developed in our theory concerning thermodynamic irreversibility, we hope to find a quantity representing a type of relation associated with biological irreversibility.

Acknowledgements

The authors express special thanks to Y. Oono for thoughtful ideas on thermodynamics and enlightening discussions. They thank K. Sekimoto, H. Tasaki and S. Takesue for fruitful discussions on thermodynamic irreversibility, N. Nakagawa for helpful discussions on dynamical systems, M. Sano and T. Hatano for discussions on the fluctuation theorem, T. Chawanya, N. Ito and K. Sato for useful comments on this study, K. Kaneko for stimulating discussions on the future outlook, and P. Gas-

pard and J. R. Dorfman for their lectures at Hayama. One of the authors (T. S. K.) thanks RIKEN for their hospitality and acknowledges support as a JSPS Research Fellow.

Appendix

In this appendix we summarize basic properties of the k-dimensional volume element and k-fold exterior products of vectors.[35)]

Consider a k-dimensional surface in the n-dimensional Euclid space $\mathbf{R}^n$. The surface can be decomposed into a set of sufficiently small pieces of k-dimensional parallelepipeds. A k-dimensional parallelepiped including a point $x \in \mathbf{R}^n$ is identified to that in a tangent space at $x \in \mathbf{R}^n$.

The k-dimensional volume of the projection of the parallelepiped to a space spanned by $\{e_{i_j},\ 1 \leq j \leq k\}$ is denoted by $\omega_{i_1 i_2 \cdots i_k} dx_{i_1} \cdots dx_{i_k}$, where $dx_{i_1} \cdots dx_{i_k}$ is a k-dimensional volume measure in the space spanned by $\{e_{i_j},\ 1 \leq j \leq k\}$. The set $\{\omega_{i_1 \cdots i_k}\}$ is called the k-dimensional volume element.

Let us write the k-dimensional volume element in a coordinate-free form. We consider the parallelepiped B formed by a set of vectors $\{b_i,\ 1 \leq i \leq k\}$. We then define the k-fold exterior product $b_1 \wedge \cdots \wedge b_k$ as a map from a k-dimensional parallelepiped to its k-dimensional volume under the projection to B. Explicitly, for a k-dimensional parallelepiped formed by a set of vectors $\{a_i,\ 1 \leq i \leq k\}$, the action of the map $b_1 \wedge \cdots \wedge b_k$ is defined as

$$b_1 \wedge \cdots \wedge b_k \cdot (a_1, \cdots a_k) = \det \mathcal{G}, \tag{A·1}$$

where $G_{ij} = (b_i, a_j)$. We write $b_1 \wedge \cdots \wedge b_k$ as $\wedge_{i=1}^k b_i$.

Then, we can find an exterior product $\wedge_{i=1}^k \omega_i$ such that

$$\omega_{i_1 \cdots i_k} = \wedge_{i=1}^k \omega_i \cdot (e_{i_1}, \cdots e_{i_k}). \tag{A·2}$$

Since the exterior product is uniquely determined, the k-dimensional volume element is identified with the k-fold exterior product. (The set of vectors $\{\omega_i,\ 1 \leq i \leq k\}$ is not uniquely determined.)

The k-fold exterior product $\wedge_{i=1}^k b_i$ has two important properties, multi-linearity and skew-symmetry. The multi-linearity is constituted by the relation

$$\begin{aligned} &b_1 \wedge \cdots \wedge (c_i b_i + c_i' b_i') \wedge \cdots \wedge b_k \\ &\quad = c_i b_1 \wedge \cdots \wedge b_i \wedge \cdots \wedge b_k + c_i' b_1 \wedge \cdots \wedge b_i' \wedge \cdots \wedge b_k \end{aligned} \tag{A·3}$$

for arbitrary i, where c_i and c_i' are numbers, and the skew-symmetry is

$$\wedge_{l=1}^k b_{i_l} = \mathrm{sgn}(i_1, \cdots, i_l) \wedge_{i=1}^k b_i. \tag{A·4}$$

Here $\mathrm{sgn}(i_1, \cdots, i_k)$ takes the value 1 when the permutation $(1, \cdots, k) \to (i_1, \cdots, i_k)$ is generated by an even number of exchanges. Otherwise it takes the value -1.

Using these two properties, we can prove

$$\wedge_{i=1}^k \sum_{j=1}^k G_{ij} b_j = \det \mathcal{G} \wedge_{i=1}^k b_i. \tag{A·5}$$

Proof

$$\wedge_{i=1}^{k}\sum_{j=1}^{k}G_{ij}b_j = \sum_{(j_1,\cdots j_k)} G_{1j_1}b_{j_1}\wedge\cdots\wedge G_{kj_k}b_{j_k} \tag{A·6}$$

$$= \sum_{(j_1,\cdots j_k)} G_{1j_1}\cdots G_{kj_k}b_{j_1}\wedge\cdots\wedge b_{j_k} \tag{A·7}$$

$$= \sum_{(j_1,\cdots j_k)} \mathrm{sgn}(j_1\cdots j_k)G_{1j_1}\cdots G_{kj_k}b_1\wedge\cdots\wedge b_k \tag{A·8}$$

$$= \det\mathcal{G}b_1\wedge\cdots\wedge b_k. \tag{A·9}$$

□

The k-dimensional volume of the parallelepiped B is calculated as $\sqrt{\det\mathcal{B}}$, where $B_{ij} = (b_i, b_j)$. (We represent it by $|\wedge_{i=1}^{k} b_i|$.)

Proof We can find a set of orthogonal unit vectors $\{u_i,\ 1\le i\le k\}$ which generate the vector space spanned by $\{b_i,\ 1\le i\le k\}$. (One may construct $\{u_i,\ 1\le i\le k\}$ by employing the Gram-Schmidt procedure.) Then, the k-dimensional volume of B is given by

$$|\wedge_{i=1}^{k} b_i| = |\wedge_{i=1}^{k} b_i\cdot(u_1,\cdots,u_k)|. \tag{A·10}$$

Since b_i can be expanded in the form

$$b_i = \sum_{j=1}^{k} G_{ij}u_j, \tag{A·11}$$

we obtain

$$\wedge_{i=1}^{k} b_i = \det\mathcal{G}\wedge_{i=1}^{k} u_i, \tag{A·12}$$

where we have used Eq. (A·5). Using the identity

$$\wedge_{i=1}^{k} u_i\cdot(u_1,\cdots,u_k) = 1, \tag{A·13}$$

we can rewrite Eq. (A·10) as

$$|\wedge_{i=1}^{k} b_i| = |\det\mathcal{G}|. \tag{A·14}$$

On the other hand, since

$$(b_i,b_j) = \sum_{lm} G_{il}G_{jm}(u_l,u_m) \tag{A·15}$$

$$= \sum_{l} G_{il}G_{jl} \tag{A·16}$$

$$= \left(\mathcal{G}\mathcal{G}^{\dagger}\right)_{ij}, \tag{A·17}$$

we have

$$\mathcal{B} = \mathcal{G}\mathcal{G}^{\dagger}. \tag{A·18}$$

Therefore, the equality

$$\det\mathcal{B} = (\det\mathcal{G})^2 \tag{A·19}$$

holds. Substituting this equality into Eq. (A·14), we obtain

$$|\wedge_{i=1}^{k} b_i| = \sqrt{\det\mathcal{B}}. \tag{A·20}$$

□

Further, for arbitrary l satisfying that $1 \leq l \leq k$, the inequality

$$|b_1 \wedge \cdots \wedge b_k| \leq |b_1 \wedge \cdots \wedge b_l||b_{l+1} \wedge \cdots \wedge b_k| \tag{A·21}$$

holds. This makes it possible for us to define the 'angle' ϕ between $b_1 \wedge \cdots \wedge b_l$ and $b_{l+1} \wedge \cdots \wedge b_k$ in such a way that

$$|b_1 \wedge \cdots \wedge b_k| = |b_1 \wedge \cdots \wedge b_l||b_{l+1} \wedge \cdots \wedge b_k| \sin\phi. \tag{A·22}$$

Proof As seen in the previous proof, there exist two sets of orthogonal unit vectors $\{u_i,\ 1 \leq i \leq l\}$ and $\{u'_i,\ l+1 \leq i \leq k\}$ such that

$$b_1 \wedge \cdots \wedge b_l = |b_1 \wedge \cdots \wedge b_l| u_1 \wedge \cdots \wedge u_l, \tag{A·23}$$

$$b_{l+1} \wedge \cdots \wedge b_k = |b_{l+1} \wedge \cdots \wedge b_k| u'_{l+1} \wedge \cdots \wedge u'_k, \tag{A·24}$$

where we note that u_i is not orthogonal to u'_j. Then, we have

$$|b_1 \wedge \cdots \wedge b_k| = |b_1 \wedge \cdots \wedge b_l||b_{l+1} \wedge \cdots \wedge b_k||u_1 \wedge \cdots \wedge u_l \wedge u'_{l+1} \wedge \cdots \wedge u'_k|. \tag{A·25}$$

Now, by using the Gram-Schmidt orthogonalization, we define a new set of vectors $\{u_j,\ l+1 \leq j \leq k\}$ as

$$u_j = \frac{u'_j - \sum_{m=1}^{j-1}(u'_j, u_m)u_m}{s_j} \tag{A·26}$$

with

$$s_j = \left| u'_j - \sum_{m=1}^{j-1}(u'_j, u_m)u_m \right|. \tag{A·27}$$

Here, from the equality

$$\left| u'_j - \sum_{m=1}^{j-1}(u'_j, u_m)u_m \right|^2 + \left| \sum_{m=1}^{j-1}(u'_j, u_m)u_m \right|^2 = 1, \tag{A·28}$$

we find

$$0 \leq s_j \leq 1. \tag{A·29}$$

Then, from

$$u'_j = s_j u_j + \sum_{m=1}^{j-1}(u'_j, u_m)u_m, \tag{A·30}$$

we derive

$$|u_1 \wedge \cdots \wedge u_k \wedge u'_{k+1} \wedge \cdots \wedge u'_k| = s_{l+1} \cdots s_k. \tag{A·31}$$

Substituting Eq. (A·31) into Eq. (A·25), we finally obtain

$$|b_1 \wedge \cdots \wedge b_k| = |b_1 \wedge \cdots \wedge b_l||b_{l+1} \wedge \cdots \wedge b_k| s_{l+1} \cdots s_k \tag{A·32}$$

$$\leq |b_1 \wedge \cdots \wedge b_l||b_{l+1} \wedge \cdots \wedge b_k|. \tag{A·33}$$

□

References

1) E. H. Lieb and J. Yngvason, Phys. Rep. **310** (1999), 1.
2) J. L. Lebowitz, Physica **A194** (1993), 1.
3) S. Sasa and T. S. Komatsu, Phys. Rev. Lett. **82** (1999), 912.
4) Y. Oono and M. Paniconi, Prog. Theor. Phys. Suppl. No. 130 (1998), 29.
5) K. Sekimoto, J. Phys. Soc. Jpn. **66** (1997), 1234.
6) K. Sekimoto and S. Sasa, J. Phys. Soc. Jpn. **66** (1997), 3326.
7) K. Sekimoto, Prog. Theor. Phys. Suppl. No. 130 (1998), 17.
   K. Sekimoto and Y. Oono, private communication.
8) C. Jarzynski, Phys. Rev. **E56** (1997), 5018; Phys. Rev. Lett. **78** (1997), 2690.
9) T. Hatano, Phys. Rev. **E60** (1999), 5017.
10) G. E. Crooks, Phys. Rev. **E60** (1999), 2721.
11) D. J. Evans, E. G. D. Cohen and G. P. Morriss, Phys. Rev. Lett. **71** (1993), 2401.
12) G. Gallvotti and E. G. D. Cohen, J. Stat. Phys. **80** (1995), 931; Phys. Rev. Lett. **74** (1995), 2694.
13) J. Kurchan, J. of Phys. **A31** (1998), 3719.
14) J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95** (1999), 333.
15) C. Maes, J. Stat. Phys. **95** (1999), 367.
16) H. A. Posch and W. G. Hoover, Phys. Rev. **A38** (1988), 473; **A39** (1989), 2175.
17) D. J. Evans, E. G. D. Cohen and G. P. Morriss, Phys. Rev. **A42** (1990), 5990.
18) P. Gaspard, J. Stat. Phys. **88** (1997), 1215.
    P. Gaspard and G. Nicolis, Phys. Rev. Lett. **65** (1990), 1693.
19) K. Sato, private communication.
20) S. Sasa, *Introduction to thermodynamics* (in Japanese), (Kyoritsu, Tokyo, 2000), in press.
21) H. Tasaki, *Thermodynamics* (in Japanese), (Baifukan, Tokyo, 2000), in press.
22) Y. Oono, Prog. Theor. Phys. **99** (1989), 165, and references therein.
23) I. Shimada and T. Nagashima, Prog. Theor. Phys. **61** (1979), 1605.
24) P. Gaspard and J. R. Dorfman, Phys. Rev **E52** (1995), 3525.
25) P. and T. Ehrenfest, *The Conceptual Foundation of the Statistical Approach in Mechanics* (Dover, New York, 1990); The original work, in German, was published in 1912.
26) E. Fermi, J. Pasta and S. Ulam, in *Collected Papers of Enrico Fermi*, ed. E. Segre (University of Chicago, Chicago, 1965), vol. 2, p. 978.
27) M. Toda, R. Kubo and N. Saito, *Statistical Physics I* (2nd ed.) (Springer-Verlag, Berlin, 1992).
28) L. Boltzmann, *Lectures on Gas theory* (Dover, New York, 1995); The original work, in German, was published in 1896 and 1898.
29) J. W. Gibbs, *Elementary principles in statistical mechanics* (Dover, New York, 1960); The work was originally published in 1902.
30) V. I. Oseledets, Trans. Moskow Math. Soc. **19** (1968), 197.
31) Y. B. Pesin, Russian Math. Surveys **32:4** (1977), 55.
32) *Maxwell's Demon; Information, Entropy, Computing.* ed. H. S. Leff and A. F. Rex (Princeton U.P., Princeton, 1990).
33) H. Yoshida, Phys. Lett. **A150** (1990), 262.

34) K. Kaneko and T. Yomo, J. Theor. Biol. **199** (1999), 243.
C. Furusawa and K. Kaneko, Bull. Math. Biology **60** (1998), 659.
35) V. I. Arnold, *Mathematical methods of Classical Mechanics* (New York, Springer-Verlag, 1978).