数理解析研究所講究録
1507 巻 2006 年 102-117

Fluctuation relation, Jarzynski-type equality, and the second law for general collision problems

学習院大学　田崎晴明 (Hal Tasaki)
Department of Physics, Gakushuin University

We study problems where the dynamics of a macroscopic body is intrinsically coupled with its internal microscopic degrees of freedom. We assume that the internal degrees of freedom are initially in equilibrium, and treat the deterministic dynamics of the whole system. Examples include impact phenomena of macroscopic bodies which consists of a large numbers of small particles. We summarize some formal results, including fluctuation relation, Jarzynski-type equality, and the second law, which can be easily proved for general collision problems. None of them are truly new, but it might be useful to present them in a unified manner.

Contents

1 Introduction

In the present note, we treat problems in which both deterministic Hamiltonian dynamics and equilibrium statistical mechanics play fundamental roles.

As a typical example, consider a ball (or any macroscopic body) moving with a constant velocity V in a free space. If we focus on the motion of the center of mass of the ball, this is definitely a problem of deterministic classical mechanics. The center of mass has a fixed position and velocity at each moment. Thus we can say that the degrees of freedom of the center of mass are very far from equilibrium in this situation. If we focus on the behavior of the internal degrees of freedom of the ball, on the other hand, the problem becomes that with a huge number of degrees of freedom. If the ball had been in contact with a heat bath at temperature T before the ball was thrown, the internal degrees of freedom may be best described using a statistical mechanical equilibrium distribution. It is important that the motion of the center of mass and the time evolution of the internal degrees of freedom are completely decoupled when the ball is moving in a free space.

Suppose now that the ball is reflected by a static wall (Fig. 1). The process of the reflection can in principle be described by microscopic deterministic dynamics, but may be very complicated. In particular the center of mass and the internal degrees of freedom are no longer decoupled since the translation invariance is lost. After the reflection, again the center of mass and the internal degrees of freedom decouple, and the ball moves with a new constant velocity V'. Since the center of mass and the internal degrees exchange energy during the reflection, the final velocity V' is in general different from the initial velocity V.

Note that the second law of thermodynamics requires $V' \leq V$, i.e., there can only be an elastic or an inelastic scattering. If a "super-elastic" collision with $V' > V$ is possible for a macroscopic ball, then one may construct a perpetual mobile of the second kind, which transfers thermal energy into mechanical energy.

In the present note we wish to examine among other things whether the macroscopic law $V' \leq V$ can be derived from microscopic dynamics alone. In other words we are trying to derive the second law starting from microscopic dynamics. There have been indeed some rigorous derivations of the second law of thermodynamics based on microscopic mechanics [1, 2, 3, 4, 5, 6, 7]. We stress, however, that none of them apply to the present situation of a ball reflected by a wall. The reason for the inapplicability is essential rather than technical. In the above mentioned proofs of the second law, macroscopic degrees of freedom are assumed to be completely controlled by an outside agent. The agent performs an operation (represented by a time-dependent Hamiltonian) according to a pre-fixed protocol, without allowing any feedback from the microscopic degrees of freedom[1]. In the present problem of impact, on the other hand, the macroscopic motion of the ball is determined only as a solution to the equation of motion involving all the degrees of freedom. It is therefore mandatory here to take into account feedback to the macroscopic motion from the microscopic motion. The derivation of the second law which applies to collision problems appeared in [8, 9]. In section 4, we present a new derivation of one of the main results in [9].

As a result of such a derivation, we find that the law $V' \leq V$ does not hold in the most strict sense. The second law holds in the form

$$\frac{M}{2}(V')^2 \leq \frac{M}{2}V^2 + O(kT), \tag{1.1}$$

[1]In thermodynamic settings, it is implicitly assumed that the feedback from microscopic degrees of freedom becomes negligible if the system and the means of operations become macroscopic. It is hard to judge how realistic such an idealization is.

where kT is the thermal energy for a single degree of freedom. Usually the correction term $O(kT)$ is negligible, and we do observe the usual second law. But when the velocity of the center of mass is much slower than the thermal velocity $\sqrt{kT/M}$ (where M is the total mass), the $O(kT)$ correction becomes relevant, and one may observe "super-elastic" collisions. Such a slow initial velocity is definitely unrealistic for macroscopic bodies, but may be realized for extremely small bodies of "nano-scale." We shall discuss about "super-elastic" collisions through an exact fluctuation relation in section 3, and through a toy model example in section A.

Of course the existence of "super-elastic" collisions does not imply that we can construct a perpetual mobile and violate the second law. Indeed to prepare the ball in a state with velocity less than $\sqrt{kT/M}$ is a hard task, which probably requires the use of certain cooling devices. Then the machine which makes use of the energy difference $(M/2)(V')^2 - (M/2)V^2$ will be a kind of heat engine.

2 Basic setting

2.1 General setup

Let us describe our setup in a general and abstract manner. The reader who is not familiar with our treatment in [8, 9] of collision problems is suggested to also take a look at section 2.2, where we discuss concrete examples.

We consider a classical mechanical problem with a phase space Γ which can be decomposed as $\Gamma = \Gamma_{\text{macro}} \times \Gamma_{\text{micro}}$. Here Γ_{macro} is the phase space of the macroscopic variables $X \in \Gamma_{\text{macro}}$, and Γ_{micro} is that of the microscopic variables $x \in \Gamma_{\text{micro}}$. The total mechanical variables of the system is $(X, x) \in \Gamma$.

We write that Hamiltonian of the system in the form

$$H_{\text{tot}}(X, x) = H(X) + h(x) + U(X, x), \tag{2.1}$$

where $H(X)$ is the Hamiltonian for the macroscopic part, $h(x)$ is the Hamiltonian for the microscopic part, and $U(X, x)$ is the interaction potential between the two scales.

We consider the time evolution determined by the Hamiltonian (2.1). We denote by $\hat{X}_t(\cdot,\cdot)$ and $\hat{x}_t(\cdot,\cdot)$ the corresponding time evolution maps[2].

We choose the initial condition of the system probabilistically according to the distribution

$$\rho_0(X)\,\frac{e^{-\beta\, h(x)}}{Z}, \tag{2.2}$$

where $\rho_0(X) \geq 0$ is an arbitrary distribution which is normalized as

$$\int dX\, \rho_0(X) = 1, \tag{2.3}$$

[2]Let $(X(t), x(t))$ be the solution with the initial conditions $X(0) = X_0$ and $x(0) = x_0$. By using the time evolution maps, we can write the solution as $X(t) = \hat{X}_t(X_0, x_0)$ and $x(t) = \hat{x}_t(X_0, x_0)$.

and the partition function is given by

$$Z = \int dx\, e^{-\beta\, h(x)}. \tag{2.4}$$

Here dX and dx denote the Lebesgue measures on the phase spaces Γ_{macro} and Γ_{micro}, respectively.

We choose the initial distribution (2.2) so that we have

$$U(X, x) = 0, \tag{2.5}$$

with probability one, when (X, x) is sampled according to (2.2). We also assume that there is a constant $T > 0$ such that we have

$$U(\widehat{X}_T(X, x), \hat{x}_T(X, x)) = 0, \tag{2.6}$$

again with probability one. In other words, the system does not feel the interaction in the initial state and in the final state. We fix the "final time" T throughout the present note.

For an arbitrary function $f(X, x)$ of $(X, x) \in \Gamma$, we denote its average over the initial distribution (2.2) as[3]

$$\langle\, f(X, x)\,\rangle_{\rho_0} := \int dX\, dx\, f(X, x)\, \rho_0(X)\, \frac{e^{-\beta\, h(x)}}{Z}. \tag{2.7}$$

A special but important choice of $\rho_0(X)$ is the delta function $\rho_0(X) = \delta(X - X_0)$ with a fixed $X_0 \in \Gamma_{\mathrm{macro}}$. In this case, we denote the average (2.7) as

$$\langle\, f(X, x)\,\rangle_{X_0} := \int dX\, dx\, f(X, x)\, \delta(X - X_0)\, \frac{e^{-\beta\, h(x)}}{Z} = \int dx\, f(X_0, x)\, \frac{e^{-\beta\, h(x)}}{Z}. \tag{2.8}$$

A quantity of main interest is

$$W(X, x) := H(\widehat{X}_T(X, x)) - H(X) = h(x) - h(\hat{x}_T(X, x)), \tag{2.9}$$

which is the total energy that the macroscopic part has received during the collision. The equality in (2.9) follows from (2.5), (2.6), and the energy conservation law. Note that $W(X, x) = 0$ corresponds to an elastic collision, $W(X, x) < 0$ to an inelastic collision, and $W(X, x) > 0$ to a "super-elastic" collision. Since "super-elastic" collisions seem to be inhibited by the second law of thermodynamics, we expect to have $W(X, x) \leq 0$ in a certain probabilistic sense. We shall prove some results concerning the probability distribution for the quantity $W(X, x)$ and its average.

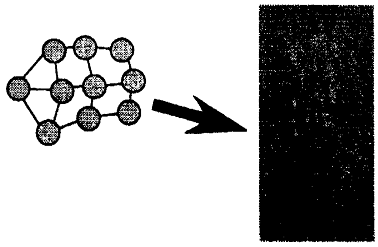

Figure 1: A "ball" consisting of N classical particles is reflected by a non-deformable wall.

2.2 Examples

2.2.1 Reflection by a potential wall

We wish to model a situation where a macroscopic ball is reflected by a wall (Fig. 1). We treat the ball as a collection of minute particles and study its motion according to microscopic dynamics. We model the wall by a fixed potential which exerts forces on the particles.

We consider a "ball" which consists of N particles. Let $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N$ be the coordinates of the particles, and $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N$ be the momenta. We assume that the total Hamiltonian is

$$H_{\text{tot}}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N) = \sum_{i=1}^{N} \frac{|\boldsymbol{p}_i|^2}{2m_i} + U_{\text{int}}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) + \sum_{i=1}^{N} U_i(\boldsymbol{r}_i). \tag{2.10}$$

Here $m_1, \ldots, m_N$ are the masses of particles. The potential $U_{\text{int}}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$ describes the forces that bind together the particles to form a stable ball. We assume that $U_{\text{int}}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$ is translationally invariant. The potential $U_i(\boldsymbol{r}_i)$ describes the force from the wall acting on the particle i. We assume that $U_i(\boldsymbol{r}_i)$ is nonvanishing only when $\boldsymbol{r}_i$ close to the surface of the wall.

Let

$$\boldsymbol{r}_{\text{cm}} := \frac{1}{M} \sum_{i=1}^{N} m_i \boldsymbol{r}_i, \quad \boldsymbol{p}_{\text{cm}} := \sum_{i=1}^{N} \boldsymbol{p}_i \tag{2.11}$$

be the position and the momentum of the center of mass, where $M := \sum_{i=1}^{N} m_i$ is the total mass. Then we set the macroscopic variable in this case as

$$X := (\boldsymbol{r}_{\text{cm}}, \boldsymbol{p}_{\text{cm}}). \tag{2.12}$$

The corresponding microscopic variable (which essentially consists of all the remaining variables) can be set as

$$x := (\tilde{\boldsymbol{r}}_1, \ldots, \tilde{\boldsymbol{r}}_{N-1}; \tilde{\boldsymbol{p}}_1, \ldots, \tilde{\boldsymbol{p}}_{N-1}), \tag{2.13}$$

where $\tilde{\boldsymbol{r}}_i := \boldsymbol{r}_i - \boldsymbol{r}_N$ is the relative coordinate, and $\tilde{\boldsymbol{p}}_i$ is its conjugate momentum.

[3] We write $A := B$ or $B =: A$ when A is defined in terms of B.

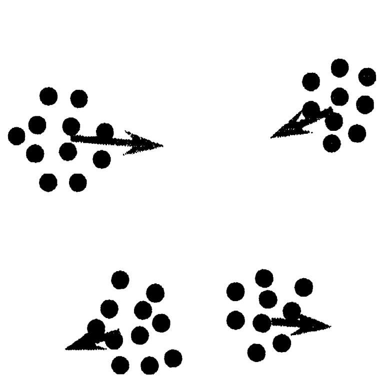

Figure 2: Two "balls" consisting of multiple classical particles undergo a collision.

The decomposition (2.1) of the total Hamiltonian is given by

$$H(X) = \frac{|\boldsymbol{p}_{\mathrm{cm}}|^2}{2M}, \quad U(X,x) = \sum_{i=1}^{N} U_i(\boldsymbol{r}_i). \tag{2.14}$$

The microscopic part $h(x)$ consists of the remainder, i.e., the kinetic energy of the internal motion and the interaction U_{int}.

Here the assumptions (2.5) and (2.6) mean that the ball is sufficiently apart from the wall surface at $t = 0$ and at $t = T$. This may be guaranteed by designing suitable potentials. See [8].

2.2.2 Collision of two particles

We can also treat the problem of two macroscopic balls colliding with each other (Fig. 2). Of course it is trivial to extend the present treatment to problems with n balls. Again the balls are treated as collections of particles. This setting is more realistic than the previous one with a static wall since all the objects are designed mechanically.

Let us consider two "balls" which consists of N_1 and N_2 particles, respectively. Let $N = N_1 + N_2$, and let $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N$ be the coordinates of the particles, and $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N$ be the momenta. We assume that the particles $1, \ldots, N_1$ form the first ball, and $N_1 + 1, \ldots, N$ form the second ball. The total Hamiltonian is

$$H_{\mathrm{tot}}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N) = \sum_{i=1}^{N} \frac{|\boldsymbol{p}_i|^2}{2m_i} + U_1(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{N_1}) + U_2(\boldsymbol{r}_{N_1+1}, \ldots, \boldsymbol{r}_N) + U_{\mathrm{int}}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N). \tag{2.15}$$

Here $m_1, \ldots, m_N$ are the masses of particles. The potentials $U_1(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{N_1})$ and $U_2(\boldsymbol{r}_{N_1+1}, \ldots, \boldsymbol{r}_N)$ describe the forces that bind together the particles to form each ball. The potential $U_{\mathrm{int}}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$ describes the interaction force between the two balls. We assume that $U_{\mathrm{int}}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$ is nonvanishing only when the two balls come close with each other. We assume that all the potentials are transitionally invariant.

Let

$$\boldsymbol{r}_{\rm cm}^{(1)} := \frac{1}{M_1}\sum_{i=1}^{N_1} m_i \boldsymbol{r}_i,\quad \boldsymbol{p}_{\rm cm}^{(1)} := \sum_{i=1}^{N_1} \boldsymbol{p}_i,\quad \boldsymbol{r}_{\rm cm}^{(2)} := \frac{1}{M_2}\sum_{i=N_1+1}^{N} m_i \boldsymbol{r}_i,\quad \boldsymbol{p}_{\rm cm}^{(2)} := \sum_{i=N_1+1}^{N} \boldsymbol{p}_i, \tag{2.16}$$

be the position and the momentum of the centers of mass of the balls, where $M_1 := \sum_{i=1}^{N_1} m_i$ and $M_2 := \sum_{i=N_1+1}^{N} m_i$ are the total masses of the balls. Then we set the macroscopic variable in this case as

$$X := (\boldsymbol{r}_{\rm cm}^{(1)}, \boldsymbol{r}_{\rm cm}^{(2)}, \boldsymbol{p}_{\rm cm}^{(1)}, \boldsymbol{p}_{\rm cm}^{(2)}). \tag{2.17}$$

The microscopic variable is again set as

$$x := (\tilde{\boldsymbol{r}}_1, \ldots, \tilde{\boldsymbol{r}}_{N-2}; \tilde{\boldsymbol{p}}_1, \ldots, \tilde{\boldsymbol{p}}_{N-2}), \tag{2.18}$$

where the variables are properly chosen as in (2.13).

The decomposition (2.1) of the total Hamiltonian is given by

$$H(X) = \frac{|\boldsymbol{p}_{\rm cm}^{(1)}|^2}{2M_1} + \frac{|\boldsymbol{p}_{\rm cm}^{(2)}|^2}{2M_2},\quad U(X,x) = U_{\rm int}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N). \tag{2.19}$$

The microscopic part $h(x)$ consists of the remainder, i.e., the kinetic energy of the internal motion and the binding potentials U_1 and U_2.

Here the assumptions (2.5) and (2.6) mean that the two balls are sufficiently apart from each other at $t = 0$ and at $t = T$. This may be guaranteed by designing suitable repulsive interaction.

3 Fluctuation relation

We shall first derive a very simple relation which is a special case of Jarzynski's "detailed fluctuation theorem" [10] (see also [11]). The relation is neat and sheds light on the nature of inelastic and "super-elastic" collisions.

Only in the present section, we assume that the Hamiltonian has a time reversal symmetry. This is satisfied in the examples of section 2.2.

Let $X_0 \in \Gamma_{\rm macro}$ be the initial state of the macroscopic part (that satisfies the requirement in section 2.1), and let $X_1 \in \Gamma_{\rm macro}$. We define

$$P(X_0 \to X_1) := \langle\, \delta(\widehat{X}_T(X,x) - X_1)\,\rangle_{X_0} = \int dX\, dx\, \delta(X - X_0)\, \delta(\widehat{X}_T(X,x) - X_1)\, \frac{e^{-\beta h(x)}}{Z}, \tag{3.1}$$

which is the probability density that the macroscopic part goes from X_0 to X_1 in the time interval T. We also define

$$\widetilde{W}(X_0 \to X_1) := H(X_1) - H(X_0), \tag{3.2}$$

which is the total energy that the macroscopic part has received during the collision, provided that X_0 and X_1 are the initial and the final states, respectively, of the macroscopic part. This is of course the same quantity as (2.9), represented in different variables.

Now from (3.1) and (3.2), we have

$$\begin{aligned}
e^{\beta\widetilde{W}(X_0\to X_1)}\,P(X_0\to X_1) &= \int dX\,dx\,\delta(X-X_0)\,\delta(\widehat{X}_T(X,x)-X_1)\,e^{\beta\{H(\widehat{X}_T(X,x))-H(X)\}}\,\frac{e^{-\beta h(x)}}{Z}\\
&= \int dX\,dx\,\delta(X-X_0)\,\delta(\widehat{X}_T(X,x)-X_1)\,e^{\beta\{h(x)-h(\hat{x}_T(X,x))\}}\,\frac{e^{-\beta h(x)}}{Z}\\
&= \int dX\,dx\,\delta(X-X_0)\,\delta(\widehat{X}_T(X,x)-X_1)\,\frac{e^{-\beta h(\hat{x}_T(X,x))}}{Z}\\
&= \int dX'\,dx'\,\delta(\widehat{X}_{-T}(X',x')-X_0)\,\delta(X'-X_1)\,\frac{e^{-\beta h(x')}}{Z},
\end{aligned} \tag{3.3}$$

where we have used the energy conservation law (2.9) to get the second equality. In the final line, we have made a change of integration variables from (X,x) to (X',x'), which are related by

$$X'=\widehat{X}_T(X,x),\quad x'=\hat{x}_T(X,x), \tag{3.4}$$

and used the Liouville theorem $dX\,dx=dX'\,dx'$.

For arbitrary $X\in\Gamma_{\text{macro}}$ and $x\in\Gamma_{\text{micro}}$, we denote by $\overline{X}\in\Gamma_{\text{macro}}$ and $\overline{x}\in\Gamma_{\text{micro}}$, respectively, the states obtained by reversing all the momenta in the original states. Then the time reversal symmetry of the Hamiltonian dynamics and (3.4) imply

$$\overline{\widehat{X}_{-T}(X',x')}=\widehat{X}_T(\overline{X'},\overline{x'}). \tag{3.5}$$

By also using $dX'\,dx'=d\overline{X'}\,d\overline{x'}$ and $h(x')=h(\overline{x'})$, we have

$$\begin{aligned}
e^{\beta\widetilde{W}(X_0\to X_1)}\,P(X_0\to X_1) &= \int d\overline{X'}\,d\overline{x'}\,\delta(\widehat{X}_T(\overline{X'},\overline{x'})-\overline{X_0})\,\delta(\overline{X'}-\overline{X_1})\,\frac{e^{-\beta h(\overline{x'})}}{Z}\\
&= \int dX\,dx\,\delta(X-\overline{X_1})\,\delta(\widehat{X}_T(X,x)-\overline{X_0})\,\frac{e^{-\beta h(x)}}{Z},
\end{aligned} \tag{3.6}$$

where, in the final line, we have simply renamed the integration variables to X and x. By comparing this with the definition (3.1), we find that

$$\boxed{e^{\beta\widetilde{W}(X_0\to X_1)}\,P(X_0\to X_1)=P(\overline{X_1}\to\overline{X_0}),} \tag{3.7}$$

which is the main result of the present section.

To see an implication of (3.7), suppose that $X_0\to X_1$ represents an inelastic collision, i.e., $\widetilde{W}(X_0\to X_1)<0$. Then since $\widetilde{W}(\overline{X_1}\to\overline{X_0})=-\widetilde{W}(X_0\to X_1)$, we find that the process

$\overline{X_1} \to \overline{X_0}$ is a "super-elastic" collision in which the internal thermal energy is converted into a mechanical energy. Therefore the relation (3.7) implies that *for any inelastic collision there exists a corresponding "super-elastic" collision.*

Of course such a conversion from thermal to mechanical energy is inhibited by the second law of thermodynamics in the macroscopic world. We indeed find from (3.7) that the "super-elastic" collision $\overline{X_1} \to \overline{X_0}$ is less unlikely by the factor $e^{\beta \widetilde{W}(X_0 \to X_1)} < 1$. In particular when the energy transfer is macroscopic and one has $-\beta \widetilde{W}(X_0 \to X_1) \gg 1$, then the factor $e^{\beta \widetilde{W}(X_0 \to X_1)}$ is essentially vanishing. We have no chance of observing a "super-elastic" collision (as we know).

When $0 < -\widetilde{W}(X_0 \to X_1) \lesssim kT$, however, the factor $e^{\beta \widetilde{W}(X_0 \to X_1)}$ is of order one. Then the symmetry (3.7) shows that it is not too unlikely to observe a "super-elastic" collision provided that an inelastic collision is possible. We expect to see "super-elastic" collisions when the energy of macroscopic part is comparable to the thermal energy kT.

The reader who is familiar with the standard fluctuation theorem might expect that our fluctuation relation (3.7) provides information about the probability distribution and the expectation value of the work $\widetilde{W}(X_0 \to X_1)$. This is unfortunately not the case here since the relation (3.7) relates a collision process with the initial state X_0 to a process with a totally different initial state $\overline{X_1}$.

Very recently Kuninaka and Hayakawa [12] made a molecular dynamics calculation of a collision of two clusters. They have observed "super-elastic collisions" when the center of mass has a very slow velocity, and found that a relation which is very close to (3.7) holds.

If we sum over X_1 in (3.7), and use (2.9), we find that

$$\langle e^{\beta W(X,x)} \rangle_{X_0} = \int d\overline{X_1}\, P(\overline{X_1} \to \overline{X_0}) \tag{3.8}$$

where the average is defined in (2.8). This reminds us of the Jarzynski equality $\langle e^{\beta W} \rangle = 1$ (which is valid when the macroscopic part is treated as a fixed protocol [5]), but is not useful as it is since the right-hand side cannot be controlled in general.

4 Jarzynski-type equality and the second law

We treat the general initial distribution (2.2). For technical convenience we first assume that $\rho_0(X)$ is an arbitrary distribution which satisfies[4] $\rho_0(X) > 0$ for any $X \in \Gamma_{\mathrm{macro}}$.

Let $\tilde{\rho}(X', x') > 0$ be an arbitrary function[5] of $X' \in \Gamma_{\mathrm{macro}}$ and $x' \in \Gamma_{\mathrm{micro}}$ which satisfies

$$\int dX'\, \tilde{\rho}(X', x') = 1, \tag{4.1}$$

[4]We have made the requirement that $\rho_0(X) > 0$ in order to avoid dividing by zero. This requirement can be relaxed after deriving the results. See the end of the section.

[5]We have made the requirement that $\tilde{\rho}(X', x') > 0$ in order to safely consider $\log \tilde{\rho}(X', x')$. This requirement can be relaxed.

for any x'. It should be noted that $\tilde{\rho}(X', x')$ is not a distribution on $\Gamma_{\text{macro}} \times \Gamma_{\text{micro}}$.

Then we have Jarzynski-type equality

$$\boxed{\left\langle \frac{\tilde{\rho}(\widehat{X}_T(X,x), \hat{x}_T(X,x))}{\rho_0(X)} e^{\beta W(X,x)} \right\rangle_{\rho_0} = 1,} \tag{4.2}$$

where $W(X,x)$ is the energy increase of the macroscopic part defined in (2.9).

Derivation is straightforward. From the definitions (2.7) and (2.9), the left-hand side of (4.2) becomes

$$\begin{aligned}
&\int dX\, dx\, \frac{\tilde{\rho}(\widehat{X}_T(X,x), \hat{x}_T(X,x))}{\rho_0(X)} e^{\beta\{h(x)-h(\hat{x}_T(X,x))\}} \rho_0(X) \frac{e^{-\beta h(x)}}{Z} \\
&= \int dX'\, dx'\, \tilde{\rho}(X', x') \frac{e^{-\beta h(x')}}{Z} = \int dx' \frac{e^{-\beta h(x')}}{Z} = 1,
\end{aligned} \tag{4.3}$$

where we have again made the change of variables (3.4), and used (4.1).

The equality (4.2) itself is too abstract and not enlightening. We shall now derive from (4.2) a useful second law. Let us rewrite (4.2) as

$$\left\langle \exp[\beta W(X,x) + \log \tilde{\rho}(\widehat{X}_T(X,x), \hat{x}_T(X,x)) - \log \rho_0(X)] \right\rangle_{\rho_0} = 1. \tag{4.4}$$

Then the Jensen inequality

$$\langle e^{f(X,x)} \rangle_{\rho_0} \geq \exp[\langle f(X,x) \rangle_{\rho_0}], \tag{4.5}$$

which is valid for any $f(X,x)$ (and indeed for any average), implies that

$$\langle W(X,x) \rangle_{\rho_0} \leq kT\{-\langle \log \tilde{\rho}(\widehat{X}_T(X,x), \hat{x}_T(X,x)) \rangle_{\rho_0} + \langle \log \rho_0(X) \rangle_{\rho_0}\}. \tag{4.6}$$

From the initial distribution (2.2), we have

$$\langle \log \rho_0(X) \rangle_{\rho_0} = \int dX\, \rho_0(X) \log \rho_0(X) = -\frac{1}{k} S[\rho_0(\cdot)], \tag{4.7}$$

where we have defined the Schannon entropy of an arbitrary distribution $\rho(\cdot)$ on Γ_{macro} as

$$S[\rho(\cdot)] := -k \int dX\, \rho(X) \log \rho(X). \tag{4.8}$$

As for the arbitrary function $\tilde{\rho}(X', x')$, we set

$$\tilde{\rho}(X', x') = \rho_T(X') := \int DX\, dx\, \delta(\widehat{X}_T(X,x) - X')\, \rho_0(X) \frac{e^{-\beta h(x)}}{Z}, \tag{4.9}$$

which is the distribution of the macroscopic variable at $t = T$. Then we have

$$\begin{aligned}
\langle \log \tilde{\rho}(\widehat{X}_T(X,x), \hat{x}_T(X,x)) \rangle_{\rho_0} &= \int dX\, dx\, \log \rho_T(\widehat{X}_T(X,x))\, \rho_0(X)\, \frac{e^{-\beta h(x)}}{Z} \\
&= \int dX'\, dX\, dx\, \log \rho_T(X')\, \delta(\widehat{X}_T(X,x) - X')\, \rho_0(X)\, \frac{e^{-\beta h(x)}}{Z} \\
&= \int dX'\, \rho_T(X')\, \log \rho_T(X') = -\frac{1}{k}\, S[\rho_T(\cdot)]. \qquad (4.10)
\end{aligned}$$

Then the inequality (4.6) becomes

$$\boxed{\langle W(X,x) \rangle_{\rho_0} \leq T\{\, S[\rho_T(\cdot)] - S[\rho_0(\cdot)]\,\}. \qquad (4.11)}$$

At this stage, we can relax the requirement that $\rho_0(X) > 0$ with $\rho_0(X) \geq 0$ by taking a suitable limit[6]. The inequality (4.11) was derived in [9] using the relative entropy method.

The inequality (4.11) states that *the energy gain of the macroscopic part cannot exceed the entropy difference of the macroscopic part multiplied by* T. This statement can be regarded as a version of the second law of thermodynamics. In particular if one has

$$S[\rho_T(\cdot)] - S[\rho_0(\cdot)] = O(k), \qquad (4.12)$$

then (4.11) becomes

$$\langle W(X,x) \rangle_{\rho_0} \leq O(kT), \qquad (4.13)$$

which is the desired second law in the case of collision. It states that one may never see a "super-elastic" collision except for those with the energy gain of order kT. This is consistent with the observation made at the end of section 3.

In [9], we have proved that (4.12) is indeed the case for certain initial distribution $\rho_0(X)$. It should be noted that for the initial distribution $\rho_0(X) = \delta(X - X_0)$, one has $S[\rho_0(\cdot)] = -\infty$ and (4.11) becomes a trivial inequality $\langle W(X,x) \rangle_{X_0} \leq \infty$.

5 The case with the delta function initial condition

Fix $X_0 \in \Gamma_{\text{macro}}$. We shall treat the case when the macroscopic part initially has a fixed (mechanical) state described by the distribution $\rho_0(X) = \delta(X - X_0)$. We derive relations due to Sasa [13], which somehow resembles Jarzynski equality and the second law.

[6]Note that $\tilde{\rho}(X', x')$ is here set to $\rho_T(X')$ and may or may not satisfy $\tilde{\rho}(X', x') > 0$.

Let us first consider the average of $e^{\beta W(X,x)}$, where $W(X,x)$ is the energy gain of the macroscopic part defined in (2.9). By using the formula (2.8) for the average, we have

$$\begin{aligned}\langle e^{\beta W(X,x)} \rangle_{X_0} &= \int dX\, dx\, \delta(X - X_0)\, \frac{e^{-\beta h(x'(x,X))}}{Z} \\ &= \int dX'\, dx'\, \delta(\widehat{X}_{-T}(X',x') - X_0)\, \frac{e^{-\beta h(x')}}{Z},\end{aligned} \tag{5.1}$$

where we have again made the change of variables (3.4), and used the Liouville theorem. Now we fix x', and first integrate over X'. By defining

$$J(x') := \int dX'\, \delta(\widehat{X}_{-T}(X',x') - X_0), \tag{5.2}$$

(5.1) becomes a Jarzynski-type relation

$$\langle e^{\beta W(X,x)} \rangle_{X_0} = \int dx'\, J(x')\, \frac{e^{-\beta h(x')}}{Z} =: \langle J(x') \rangle_{\text{micro}}. \tag{5.3}$$

Here the right-hand side is the average over the canonical distribution of the microscopic part. This, with the Jensen inequality (4.5), readily gives a "second law"

$$\langle W(X,x) \rangle_{X_0} \leq \frac{1}{\beta} \log\langle J(x') \rangle_{\text{micro}}. \tag{5.4}$$

Whether the relations (5.3) and (5.4) are useful or not depends on how well we can control the quantity $J(x')$ and its average $\langle J(x') \rangle_{\text{micro}}$.

To see this, let us define

$$\mathcal{S}_{x'} := \{X' \in \Gamma_{\text{macro}} \,|\, \widehat{X}_{-T}(X',x') = X_0\}. \tag{5.5}$$

For simplicity we assume that, for any fixed x', the set $\mathcal{S}_{x'}$ consists of a finite number of points. Then (5.2) becomes

$$J(x') = \sum_{X' \in \mathcal{S}_{x'}} \left| \frac{\partial \widehat{X}_{-T}(X',x')}{\partial X'} \right|^{-1}. \tag{5.6}$$

Thus $J(x')$ is in general a sum of Jacobians, and may be vanishing if the set $\mathcal{S}_{x'}$ is empty.

One naively expects that $\mathcal{S}_{x'}$ consists of a single point for any $x' \in \Gamma_{\text{micro}}$. If this is the case, $J(x')$ is the Jacobian itself, and one may further regard $J(x')$ as a function of x via the relation $\tilde{J}(\hat{x}_{-T}(X',x')) := J(x')$ for $X' \in \mathcal{S}_{x'}$. Then we have another Jarzynski equality

$$\langle \tilde{J}(x)^{-1}\, e^{\beta W(X,x)} \rangle_{X_0} = 1, \tag{5.7}$$

and the corresponding second law.

Unfortunately it is too naive to expect that $\mathcal{S}_{x'}$ always consists of a single point. We shall study in Appendix A a very simple example exhibiting a "super-elastic impact." The model has a constant Jacobian, and hence the Jarzynski-type relation (5.1) (with the expectation that $\mathcal{S}_{x'}$ consists of a single point) predicts that the quantity $\langle e^{\beta W(X,x)} \rangle_{X_0}$ is constant and equal to the Jacobian. Our explicit calculation shows this is not the case.

A A simple examples of "super-elastic impact"

Let us describe a very simple example which exhibits a "super-elastic impact", and in which one can evaluate the Jacobian in (5.6) explicitly. We will see that the naive expectation about the set (5.5) does not hold.

A.1 Setting

Consider a system of two particles, which we call 1 and 2, with masses m and λm, respectively. We assume that the particles move in one-dimensional space. The two particles interact via an infinite wall potential

$$U(x_2 - x_1) = \begin{cases} \infty & \text{if } x_1 > x_2 \text{ or } x_2 - x_1 > R; \\ 0 & \text{otherwise,} \end{cases} \tag{A.1}$$

where x_1 and x_2 are the positions of 1 and 2, respectively, and $R > 0$ is a constant. The motion of the two particles are restricted so that 1 is always at the left of 2, and that the distance between the two does not exceed R. Otherwise the two particles move freely. We denote by v_1 and v_2 the velocities of the particle 1 and 2, respectively.

When the center of mass is fixed, the two particles may be in expanding motion with $v_1 = -\lambda v$ and $v_2 = v$ or in the shrinking motion with $v_1 = \lambda v$ and $v_2 = -v$, where $v > 0$ characterizes the relative motion.

When the center of mass has a velocity V, the velocities of the two particles are $v_1 = V - \lambda v$ and $v_2 = V + v$ in the expanding motion, and $v_1 = V + \lambda v$ and $v_2 = V - v$ in the shrinking motion. We assume that $V > 0$. Thus the center of mass moves to the right while the two particles perform expanding and shrinking motions in turn. We here treat an anomalous situation where the center of mass has a very slow velocity compared with the internal velocity, i.e., $0 < V \ll v$.

There is a hard wall at a fixed position, and the particle is reflected elastically when hitting the wall. Obviously only the particle 2 can hit and be reflected by the wall. We further assume that particle 2 is reflected by the wall only once[7].

A.2 Jacobian

One can easily calculate the Jacobian in (5.6) which is expected to characterize the impact. We denote by x_1, x_2, v_1, v_2 the positions and the velocities of the particles at $t = 0$ which is slightly before the impact, and by $x_1', x_2', v_1' = v_1, v_2' = -v_1$ those at $t = T$ which is slightly

[7]One can show that, if $V \ll v$, the particle never comes back to the position of the wall after the first impact. We can also make an artificial assumption that the wall is removed immediately after the first impact so that the second never takes place. This makes the total Hamiltonian of the model time-dependent, but all the results in the main body of the present note remain valid since both the energy conservation law and the Liouville theorem hold. For simplicity we shall employ the second interpretation.

after the impact[8]. According to (5.6), we should calculate the Jacobian by fixing the internal coordinates after the collision. This means that we should make small modifications to the final values as

$$x_1' \to x_1' + \Delta x, \quad x_2' \to x_2' + \Delta x, \quad v_1' \to v_1' + \Delta v, \quad v_2' \to v_2' + \Delta v. \tag{A.2}$$

An inspection shows that the corresponding changes in the initial coordinates are

$$x_1 \to x_1 + \Delta x - T\,\Delta v, \quad x_2 \to x_2 - \Delta x + T\,\Delta v, \quad v_1 \to v_1 + \Delta v, \quad v_2 \to v_2 - \Delta v. \tag{A.3}$$

Then the changes in the position $x_{\rm cm} = (x_1 + \lambda\, x_2)/(1+\lambda)$ and the velocity $v_{\rm cm} = (v_1 + \lambda\, v_2)/(1+\lambda)$ are found to be

$$x_{\rm cm} \to x_{\rm cm} + \frac{1-\lambda}{1+\lambda}(\Delta x - T\,\Delta v), \quad v_{\rm cm} \to v_{\rm cm} + \frac{1-\lambda}{1+\lambda}\Delta v, \tag{A.4}$$

which means

$$\frac{\partial x_{\rm cm}}{\partial x'_{\rm cm}} = \frac{1-\lambda}{1+\lambda}, \quad \frac{\partial x_{\rm cm}}{\partial v'_{\rm cm}} = -T\frac{1-\lambda}{1+\lambda}, \quad \frac{\partial v_{\rm cm}}{\partial x'_{\rm cm}} = 0, \quad \frac{\partial v_{\rm cm}}{\partial v'_{\rm cm}} = \frac{1-\lambda}{1+\lambda}. \tag{A.5}$$

From this we can calculate the desired Jacobian as

$$\left|\frac{\partial \widehat{X}_{-T}(X',x')}{\partial X'}\right|_{x' \text{ is fixed}} = \begin{vmatrix} \dfrac{\partial x_{\rm cm}}{\partial x'_{\rm cm}} & \dfrac{\partial x_{\rm cm}}{\partial v'_{\rm cm}} \\ \dfrac{\partial v_{\rm cm}}{\partial x'_{\rm cm}} & \dfrac{\partial v_{\rm cm}}{\partial v'_{\rm cm}} \end{vmatrix} = \left(\frac{1-\lambda}{1+\lambda}\right)^2. \tag{A.6}$$

Note that the Jacobian is a constant. This is of course a very special feature of the present situation in which only a single elastic collision takes place.

A.3 Expectation values

Knowing that the Jacobian is constant, the Jarzynski-type equality (5.3) and a naive expectation (that $\mathcal{S}_{x'}$ consists of a single point) leads us to the conjecture

$$\langle\, e^{\beta\, W(X,x)}\,\rangle_{X_0} = \left(\frac{1+\lambda}{1-\lambda}\right)^2. \tag{A.7}$$

We will see that this relation fails drastically (and indeed in a neat manner) in the present problem.

Before the impact, the two particle system is in the motion characterized by the velocity V of the center of mass and the velocity v of the internal motion. The total kinetic energy (either in the expanding or shrinking motion) is

$$E_{\rm tot} = \frac{m}{2}(V \mp \lambda v)^2 + \frac{\lambda m}{2}(V \pm v)^2 = \frac{(1+\lambda)m}{2}V^2 + \frac{m}{2}\lambda(1+\lambda)v^2. \tag{A.8}$$

[8]The Jacobian does not depend on the choice of the initial and the final moments, provided that they are before and after the impact.

In the right-most hand, the first term represents the kinetic energy of the center of mass and the second term represents that for the internal motion.

When $V < v$, the impact between the wall and the particle 2 takes place only when the system is in the expanding motion. We assume this is always the case. Then the velocities of the two particle immediately before the impact is $v_1 = V - \lambda v$ and $v_2 = V + v$. After the impact, the velocity of the particle 2 changes to $v_2' = -(V + v)$. Thus the velocity of the center of mass after the impact is

$$V' = \frac{v_1 + \lambda v_2'}{1+\lambda} = \frac{-2\lambda v + (1-\lambda)V}{1+\lambda} \simeq \frac{-2\lambda v}{1+\lambda}, \tag{A.9}$$

where the final expression is valid when $V \ll v$, i.e., when the center of mass has a very slow velocity. The system clearly exhibits a "super-elastic impact" since $|V'| \gg V$. The energy gain of the macroscopic part is given by

$$W = \frac{(1+\lambda)m}{2}(V')^2 - \frac{(1+\lambda)m}{2}V^2 \simeq \frac{m}{2}\frac{4\lambda^2}{1+\lambda}v^2, \tag{A.10}$$

where the final expression is again for $V \ll v$.

Let us assume that the internal velocity v is distributed according to the canonical distribution

$$p_{\mathrm{can}} \propto \exp[-\beta \frac{m}{2}\lambda(1+\lambda)v^2]. \tag{A.11}$$

Then the average of the energy gain W is

$$\langle W \rangle \simeq \frac{2\lambda}{(1+\lambda)^2}kT, \tag{A.12}$$

when $V \ll \sqrt{kT/m}$. The naive second law is violated here, but only by the order of kT. This is consistent[9] with the rigorous second law (4.13). In the same limit, we also get

$$\langle e^{\beta W} \rangle \simeq \left|\frac{1+\lambda}{1-\lambda}\right|, \tag{A.13}$$

which somehow resembles but does not coincide with the conjectured (A.7). This example clearly shows that the naive guess that the set $\mathcal{S}_{x'}$ consists of a single element for each x' does not hold[10].

I wish to thank Hisao Hayakawa and Shin-ichi Sasa for indispensable discussions and various useful suggestions.

[9]But $\langle W \rangle \sim kT$ may not be the universal behavior for impacts with extremely slow center of mass velocity. In a very similar problem of two particles coupled by a harmonic potential, one has $\langle W \rangle \sim \sqrt{mV^2kT}$ [14].

[10]The fact that the final internal velocity is given by $v' \simeq v(\lambda - 1)/(1+\lambda)$ may be used to explain the difference between (A.7) and (A.13).

References

[1] A. Lenard, *Thermodynamic proof of the Gibbs formula for elementary quantum systems*, J. Stat. Phys. **19**, 575–586 (1978).

[2] W. Pusz and S. L. Woronoicz, *Passive states and KMS states for general quantum systems*, Comm. Math. Phys. **58**, 273–290 (1978).

[3] W. Thirring, *Quantum Mechanics of Large Systems* (Springer, 1983).

[4] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[5] C. Jarzynski, J. Stat. Phys. **96**, 415 (1999).

[6] H. Tasaki, Statistical mechanical derivation of the second law of thermodynamics (unpublished note), `cond-mat/0009206`.

[7] H. Tasaki, The second law of Thermodynamics as a theorem in quantum mechanics (unpublished note), `cond-mat/0011321`.

[8] H. Tasaki, The coefficient of restitution does not exceed unity (submitted to J. Stat. Phys.), `cond-mat/0408481`

[9] C. Maes and H. Tasaki, *Second law of thermodynamics for macroscopic mechanics coupled to thermodynamic degrees of freedom*, `cond-mat/0511419`.

[10] C. Jarzynski, *Hamiltonian Derivation of a Detailed Fluctuation Theorem*, J. Stat. Phys. **98**, 77–102 (2000).

[11] C. Maes and K. Netocny, *Time-reversal and entropy*, J. Stat. Phys. **110**, 269–310 (2003).

[12] K. Kuninaka and H. Hayakawa, private communication.

[13] S. Sasa, unpublished note (2000).

[14] H. Tasaki, unpublished.