Ergod. Th. & Dynam. Sys. (2009), **29**, 273–279
doi:10.1017/S0143385708000163
© *2008 Cambridge University Press*
Printed in the United Kingdom

Abstract fluctuation theorem

MACIEJ P. WOJTKOWSKI
Department of Mathematics and Informatics, University of Warmia and Mazury, ul. Zolnierska 14, 10-561 Olsztyn, Poland
(e-mail: maciejw@math.arizona.edu)

(*Received* 29 *September* 2007 *and accepted in revised form* 10 *January* 2008)

Abstract. We formulate an abstract fluctuation theorem which sheds light on mathematical relations between the fluctuation theorems of Bochkov and Kuzovlev [Contribution to the general theory of thermal fluctuations in nonlinear systems. *Sov. Phys.–JETP* **45** (1977), 125] and Jarzynski [Hamiltonian derivation of a detailed fluctuation theorem. *J. Stat. Phys.* **98** (2001), 77–102] on the one hand, and those of Evans and Searles [Equilibrium microstates which generate second law violating steady states. *Phys. Rev.* E **50** (1994), 1645–1648] and Gallavotti and Cohen [Dynamical ensembles in stationary states. *J. Stat. Phys.* **80** (1995), 931–970] on the other.

1. *Algebraic preliminaries*

Let $\Phi : M \to M$ and $\Psi : M \to M$ be two dynamical systems in a broad sense; later on we will specify them to be measurable maps, measure preserving or not, or diffeomorphisms. We start with a general framework, assuming only that Φ and Ψ are one-to-one maps of the phase space M.

Let $R : M \to M$ be an involution, i.e. $R^2 = R \circ R = \mathrm{Id}$. We say that Φ is an *R-inverse* of Ψ if $R\Psi R\Phi = \mathrm{Id}$, i.e. if the following diagram commutes

$$\begin{array}{ccc} M & \xrightarrow{\Phi} & M \\ {\scriptstyle R}\downarrow & & \downarrow{\scriptstyle R} \\ M & \xrightarrow{\Psi^{-1}} & M \end{array}$$

Clearly if Φ is an R-inverse of Ψ, then Ψ is an R-inverse of Φ. If Φ is an R-inverse of itself, then Φ is called *R-reversible*, or simply reversible.

PROPOSITION 1. *Ψ is an R-inverse of Φ if and only if for $S = R \circ \Phi$ and the involution $U : M \times M \to M \times M$, $U(x_1, x_2) = (Rx_2, Rx_1)$, the following diagram commutes*

$$\begin{array}{ccc} M & \xrightarrow{S} & M \\ {\scriptstyle \widetilde{\Phi}}\downarrow & & \downarrow{\scriptstyle \widetilde{\Psi}} \\ M \times M & \xrightarrow{U} & M \times M \end{array}$$

where $\widetilde{\Phi}(x) = (x, \Phi x)$ and $\widetilde{\Psi}(x) = (x, \Psi x)$.

https://doi.org/10.1017/S0143385708000163 Published online by Cambridge University Press

Proof. $U\widetilde{\Phi}x = U(x, \Phi x) = (R\Phi x, Rx) = (Sx, \Psi R\Phi x) = \widetilde{\Psi}Sx$. □

We can immediately extend Proposition 1 to the space of trajectories of length $k+1$, $k \geq 1$. Let $M^{k+1} = M \times \cdots \times M$, $k \geq 1$, and define $\widetilde{\Phi} : M \to M^{k+1}$ and $\widetilde{\Psi} : M \to M^{k+1}$ by

$$\widetilde{\Phi}(x) = (x, \Phi x, \ldots, \Phi^k x), \quad \widetilde{\Psi}(x) = (x, \Psi x, \ldots, \Psi^k x). \tag{1}$$

Then, again, we have that the involution $U = U_k : M^{k+1} \to M^{k+1}$,

$$U(x_0, \ldots, x_k) = (Rx_k, Rx_{k-1}, \ldots, Rx_0),$$

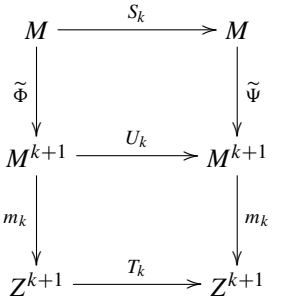
takes the trajectory of Φ starting at x into the trajectory of Ψ starting at Sx, where $S = S_k = R \circ \Phi^k$. In other words, we have again that $U_k \circ \widetilde{\Phi} = \widetilde{\Psi} \circ S_k$.

Let us now consider a set Z and a map $m : M \to Z$, which we think of as a measurement on the phase space M. In particular, Z can be much 'smaller' than M. We assume further that the involution R can be factored onto Z, i.e. there is an involution $r : Z \to Z$ such that $rm = mR$. We then obtain that the involution $T = T_k : Z^{k+1} \to Z^{k+1}$, $T(z_0, z_1, \ldots, z_k) = (rz_k, rz_{k-1}, \ldots, rz_0)$ takes the 'time series' of Φ starting at x into the 'time series' of Ψ starting at $S_k x$. More precisely, we have the following.

PROPOSITION 2. *Let $m = m_k : M^{k+1} \to Z^{k+1}$ be the map*

$$m_k(x_0, x_1, \ldots, x_k) = (mx_0, mx_1, \ldots, mx_k).$$

Then we have the commuting diagram



Proof. The commutativity of the upper part of the diagram is the generalized Proposition 1, thus we only need to establish the commutativity of the lower part, i.e. $m_k U_k = T_k m_k$. This can be seen as follows:

$$\begin{aligned} m_k U_k(x_0, x_1, \ldots, x_k) &= m_k(Rx_k, \ldots, Rx_1, Rx_0) = (mRx_k, \ldots, mRx_1, mRx_0) \\ &= (rmx_k, rmx_{k-1}, \ldots, rmx_0) = T(mx_0, rmx_1, \ldots, mx_k). \end{aligned}$$ □

In the special case of a reversible dynamical system, i.e. when $\Psi = \Phi$, we get that $S = S_k = R \circ \Phi^k$, $k \geq 1$, is an involution; moreover, although the dynamics itself cannot be factored in general to the space of measurements Z (or Z^k), the involution S has a natural factor T on Z^k, or more precisely on the space of time series (that is, the space of measurements on finite trajectories of Φ, since this is exactly the range of $m_k \circ \widetilde{\Phi}$). Let us stress that the only requirement is that the involution R factors to Z.

https://doi.org/10.1017/S0143385708000163 Published online by Cambridge University Press

Finally, let us examine how this formalism works in the case of time-dependent systems; consider two sequences of maps of M, $\Phi_1, \ldots, \Phi_k$ and $\Psi_1, \ldots, \Psi_k$, $k \geq 1$. We assume that each Ψ_i is an R-inverse of Φ_i, for $i = 1, \ldots, k$. Let us consider the time-dependent dynamical systems $\Phi^j = \Phi_j \circ \cdots \circ \Phi_2 \circ \Phi_1$ and $\Psi^j = \Psi_{k-j+1} \circ \Psi_{k-j+2} \circ \cdots \Psi_k$, $j = 1, \ldots, k$.

Note the reversal of order in the definition of Ψ^j; it gives the impression of lacking physical meaning. However, in some examples it can be interpreted as the arranged 'time reversal of controls'.

We can formulate an obvious time-dependent generalization of Proposition 2: if $\widetilde{\Phi}$ and $\widetilde{\Psi}$ are again defined by (1), with the modified meaning of Φ^j, Ψ^j ($j = 1, \ldots, k$), then the commuting diagram of Proposition 2 holds.

2. *Factors of measures*

We keep all the assumptions of §1 with the addition that $M = (M, \mathcal{B}, \mu)$ is a measurable space, i.e. $\mathcal{B}$ is a σ-algebra of subsets of M and μ is a σ-finite measure. We assume that Φ, Ψ and R are measurable maps, that the involution R preserves the measure μ, and that $\Phi_* \mu = f\mu$ for some positive function f on M.

PROPOSITION 3. *We have* $\Psi_* \mu = g\mu$, *where* $g = (f \circ \Phi \circ R)^{-1}$. *It follows that* $g \circ \Psi = (f \circ R)^{-1}$.

Proof. We have $\Phi_*^{-1} \mu = (f \circ \Phi)^{-1} \mu$ and $\Psi = R \circ \Phi^{-1} \circ R$. It follows that $\Psi_* \mu = (f \circ \Phi \circ R)^{-1} \mu$. □

Clearly, if $\Phi_*^k \mu = f^k \mu$, then $\Psi_*^k \mu = g^k \mu$ where $g^k = (f^k \circ \Phi^k \circ R)^{-1}$, for a fixed value of $k \geq 1$. (Note that k in f^k and g^k is a superscript, not a power.) Moreover, this formula applies also to the case of time-dependent systems described in §1. More precisely, if $(\Phi_i)_* \mu = f_i \mu$, $i = 1, \ldots, k$, then $\Phi_*^k \mu = f^k \mu$ with

$$f^k = \prod_{i=1}^{k} f_i \circ \Phi_{i+1}^{-1} \circ \cdots \circ \Phi_k^{-1}.$$

Let us consider the augmented maps

$$\widehat{\Phi} : M \to Z^{k+1} \times \mathbb{R}, \quad \widehat{\Psi} : M \to Z^{k+1} \times \mathbb{R},$$

defined by

$$\widehat{\Phi} x = (m_k \widetilde{\Phi} x, \ln f^k \circ \Phi^k) = \left(mx, m\Phi x, \ldots, m\Phi^k x, \ln \prod_{i=1}^{k} f_i(\Phi^i x) \right)$$

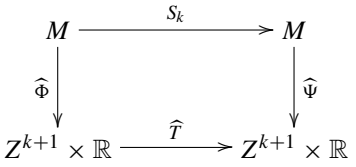
and the respective formula for $\widehat{\Psi}$. We also introduce the involution

$$\widehat{T} : Z^{k+1} \times \mathbb{R} \to Z^{k+1} \times \mathbb{R},$$
$$\widehat{T}(z_0, z_1, \ldots, z_k, a) = (rz_k, \ldots, rz_1, rz_0, -a).$$

Then we have the following measure counterpart of Proposition 2.

https://doi.org/10.1017/S0143385708000163 Published online by Cambridge University Press

THEOREM 1. *The diagram*



commutes, and

$$\widehat{T}_* \widehat{\Phi}_* \mu = e^{-a} \widehat{\Psi}_* \mu.$$

In the case of a probability measure μ, the measure $\widehat{\Phi}_* \mu$ represents the statistics of the time series of Φ augmented by the value of $\ln f^k \circ \Phi^k$. The measure $\widehat{\Psi}_* \mu$ has a similar meaning and the two probability distributions are related by the involutive symmetry $\widehat{T}$; in other words, the image of $\widehat{\Phi}_* \mu$ under $\widehat{T}$ has density e^{-a} with respect to $\widehat{\Psi}_* \mu$.

The last claim in Theorem 1 can be also formulated as

$$\widehat{T}_* (e^{-a/2} \widehat{\Phi}_* \mu) = e^{-a/2} \widehat{\Psi}_* \mu.$$

Hence, in the case of R-reversible Φ (i.e. $\Psi = \Phi$), we get that the measure $e^{-(a/2)} \widehat{\Phi}_* \mu$ is invariant under the involution $\widehat{T}$.

Proof. In view of Proposition 2, to establish commutativity of the diagram we need only check that

$$g^k \circ \Psi^k \circ S_k = (f^k \circ \Phi^k)^{-1},$$

which follows immediately from Proposition 3.

Once the commutativity is established, we obtain

$$\widehat{T}_* \widehat{\Phi}_* \mu = \widehat{\Psi}_* (R_* \Phi^k_* \mu) = \widehat{\Psi}_* (f^k \circ R \mu).$$

We conclude the proof by observing that, by Proposition 3, the function $f^k \circ R$ is equal to $(g^k \circ \Psi^k)^{-1}$ and hence factors under $\widehat{\Psi}$ to the function on $Z^{k+1} \times \mathbb{R}$ which equals e^{-a}, where a is the last coordinate in $Z^{k+1} \times \mathbb{R}$. □

Let us assume further that Z comes with a reference measure ν which is invariant under the involution r. We introduce the measure λ on $Z^{k+1} \times \mathbb{R}$, equal to the product of ν^{k+1} on Z^{k+1} and the Lebesgue measure on $\mathbb{R}$. Clearly, the measure λ is invariant under the action of the involution $\widehat{T}$. Under these assumptions, we deduce immediately the following corollary of Theorem 1.

THEOREM 2. *If the measures $\widehat{\Phi}_* \mu$ and $\widehat{\Psi}_* \mu$ are absolutely continuous with respect to the measure λ on $Z^{k+1} \times \mathbb{R}$, with densities equal to $p_+(z_0, z_1, \ldots, z_k, a)$ and $p_-(z_0, z_1, \ldots, z_k, a)$, then*

$$p_+(z_0, z_1, \ldots, z_k, a) = e^a p_-(r z_k, \ldots, r z_1, r z_0, -a).$$

https://doi.org/10.1017/S0143385708000163 Published online by Cambridge University Press

3. *The Bochkov–Kuzovlev and Jarzynski scenarios*

Theorem 1 provides a general scheme, special cases of which appeared earlier in different setups. For simplicity, in this section we will consider only the case $k = 1$. We assume that the phase space M is a symplectic manifold with the symplectic form ω, and that the involution R is anti-symplectic, i.e. $R^*\omega = -\omega$. We consider a time-dependent Hamiltonian $H(x, t)$ on M. We further assume that the Hamiltonian H is at every moment of time invariant under the involution R, i.e. $H(Rx, t) = H(x, t)$. Let us fix a time interval $[0, \tau]$ and let Φ be the 'after time τ' symplectic map of M defined by the Hamiltonian dynamics of H. Consider the Hamiltonian $G(x, t) = H(x, \tau - t)$, and take Ψ to be the 'after time τ' symplectic map of M defined by the Hamiltonian dynamics of G. In the scenario of Jarzynski [**J**], the Hamiltonian G governs the process with the '*time reversed protocols*'.

PROPOSITION 4. *Ψ is an R-inverse of Φ.*

Proof. By the definition of Ψ, we get that Ψ^{-1} is the 'after time τ' map of the Hamiltonian $-H(x, t)$. At the same time, since the involution R is anti-symplectic, the map $R\Phi R$ is also the 'after time τ' map of the Hamiltonian $-H(x, t)$. □

In the Bochkov–Kuzovlev [**BK**] scenario, the Hamiltonian $H(x, t)$ is of the form $H(x, t) = H_0(x) + H_1(x, t)$, with $H_1(x, t)$ representing a time-dependent action on the system. Let χ be the Liouville measure on M and let $\mu = (1/z)e^{-\beta H_0}\chi$ be the Gibbs probability measure, where z is the normalizing factor. We consider $Z = M$ with m being the identity map, i.e. all variables are observed. We obtain $\widehat{\Phi}x = (x, \Phi x, \beta H_0(\Phi x) - \beta H_0(x))$ and $\widehat{\Psi}x = (x, \Psi x, \beta H_0(\Psi x) - \beta H_0(x))$. Let us denote by P the probability defined by the Gibbs measure μ. Theorem 1 can be now reformulated as the following.

COROLLARY 1. *For any subsets $A_0, A_1 \subset M$ and any real a and $\epsilon > 0$,*

$$e^{\beta(a-\epsilon)} \leq \frac{P(x \in A_0, \Phi x \in A_1, |H_0(\Phi x) - H_0(x) - a| \leq \epsilon)}{P(x \in RA_1, \Psi x \in RA_0, |H_0(\Psi x) - H_0(x) + a| \leq \epsilon)} \leq e^{\beta(a+\epsilon)}.$$

In particular, by taking $A_0 = A_1 = M$, we obtain from Corollary 1 the following.

COROLLARY 2. *For any real a and $\epsilon > 0$,*

$$e^{\beta(a-\epsilon)} \leq \frac{P(|H_0(\Phi x) - H_0(x) - a| \leq \epsilon)}{P(|H_0(\Psi x) - H_0(x) + a| \leq \epsilon)} \leq e^{\beta(a+\epsilon)}.$$

COROLLARY 3. *If $p_+ = p_+(a)$ and $p_- = p_-(a)$ are the densities with respect to the Lebesgue measure on $\mathbb{R}$ of the distributions of values of $H_0(\Phi x) - H_0(x)$ and $H_0(\Psi x) - H_0(x)$, respectively, then we have*

$$p_+(a) = e^{\beta a} p_-(-a).$$

In the Jarzynski scenario, there is a Hamiltonian system on a symplectic manifold Z interacting with 'heat baths', represented by another Hamiltonian system on a symplectic manifold Y. The joint Hamiltonian system lives on $M = Y \times Z$, and we take $m : M \to Z$ to be the projection on the second component. The Liouville measure χ on M is invariant under Φ, Ψ and R; however, it is in general infinite. We introduce a probabilistic density

https://doi.org/10.1017/S0143385708000163 Published online by Cambridge University Press

$\rho = \rho(y)$ on Y, a '*prepared state of the heat baths*', and apply Theorems 1 and 2 to the measure $\mu = \rho\chi$. We have

$$\Phi_*\mu = f\mu = \frac{\rho \circ \Phi^{-1}}{\rho}\mu.$$

In the Jarzynski scenario, $\ln f \circ \Phi = \ln \rho - \ln \rho \circ \Phi$ quantifies the entropy production. The measure $\mu = \rho\chi$ on M is not probabilistic, but it disintegrates into a family of probabilistic measures after fixing the Z component. Consequently, the measure $\widehat{\Phi}_*\mu$ on $Z \times Z \times \mathbb{R}$ disintegrates into a family of probabilistic measures upon fixing the first coordinate z_0. This gives us the 'conditional' probabilities $P_+(\cdot|z_0)$. Similarly, we obtain the family of 'conditional' probabilities $P_-(\cdot|z_0)$ by considering the disintegration of the measure $\widehat{\Psi}_*\mu$. We now obtain the Jarzynski detailed fluctuation theorem as a consequence of Theorem 2.

COROLLARY 4. *Denoting by $p_+(z_1, a|z_0)$ and $p_-(z_1, a|z_0)$ the densities of the probability measures $P_+(\cdot|z_0)$ and $P_-(\cdot|z_0)$, respectively, we have*

$$p_+(z_1, a|z_0) = e^a\, p_-(rz_0, -a|rz_1).$$

4. *The Evans–Searles and Gallavotti–Cohen scenarios*

Let Φ be an R-reversible diffeomorphism of a Riemannian manifold M, and let μ be a probabilistic volume element invariant under the involution R. Let $J(x) = |\det \mathcal{D}_x\Phi|$ be the Jacobian of the diffeomorphism Φ and, more generally, let $J_k(x) = |\det \mathcal{D}_x\Phi^k|$ be the Jacobian of Φ^k. We have $\Phi_*\mu = f\mu$ with $f(x) = (J(\Phi^{-1}x))^{-1}$.

Hence the quantity $\ln \prod_{i=1}^{k} f(\Phi^i x) = -\ln J_k(x)$.

Taking $\Psi = \Phi$, $Z = M$ and $m = \mathrm{Id}$, we get $\widehat{\Phi} = (x, \Phi x, \ldots, \Phi^k x, -\ln J_k(x))$, and Theorem 1 then gives us the Evans–Searles fluctuation theorem [**ES**].

COROLLARY 5. *For any subsets $A_0, A_1, \ldots, A_k \subset M$, any real a, and any positive ϵ,*

$$e^{-a-\epsilon} \le \frac{P\{x \in A_0, \ldots, \Phi^k(x) \in A_k, |\ln J_k(x) - a| \le \epsilon\}}{P\{x \in RA_k, \ldots, \Phi^k(x) \in RA_0, |\ln J_k(x) + a| \le \epsilon\}} \le e^{-a+\epsilon},$$

where P denotes the probability defined by the probabilistic volume element μ.

In particular, taking $A_0 = A_1 = \cdots = A_k = M$ in Corollary 5, we obtain the following.

COROLLARY 6.

$$e^{-a-\epsilon} \le \frac{P\{|\ln J_k(x) - a| \le \epsilon\}}{P\{|\ln J_k(x) + a| \le \epsilon\}} \le e^{-a+\epsilon}.$$

COROLLARY 7. *Let $p_k = p_k(a)$ denote the density with respect to the Lebesgue measure on $\mathbb{R}$ of the distribution of values of $\ln J_k$. Then we have*

$$p_k(a) = e^{-a}\, p_k(-a). \tag{2}$$

The property in Corollary 7 is stronger than the one in the Gallavotti–Cohen fluctuation theorem [**GC**]; however, it holds for the probability P defined by the reference Lebesgue measure μ on M, and not for the asymptotic state of the system.

https://doi.org/10.1017/S0143385708000163 Published online by Cambridge University Press

We can get a little closer to the Gallavotti–Cohen formulation by observing that the distribution of values of

$$\sum_{i=-k}^{k} \ln J(\Phi^i x) = \ln J_{2k+1}(\Phi^{-k} x)$$

with respect to the probability measure $\Phi_*^k \mu$ coincides with the distribution of values of $\ln J_{2k+1}(x)$ with respect to the measure μ; moreover, the density of this distribution, being equal to $p_{2k+1}(a)$, has the symmetry (2). In general, however, it is not clear how to take the limit $k \to \infty$.

Acknowledgements. I am grateful to Lai-Sang Young who spent many hours discussing fluctuation theorems with me, starting from my visit to the Courant Institute in March 2006. I would also like to thank Jan Wehr and Bill Faris, who were the first to draw my attention to the subject. While working on this paper in the Spring of 2007, I enjoyed the hospitality of MSRI in Berkeley.

This work was partially supported by the EC FP6 Marie Curie ToK programme SPADE2, MTKD-CT-2004-014508 and the Polish MNiSW SPB-M.

References

[BK] G. N. Bochkov and Yu. E. Kuzovlev. Contribution to the general theory of thermal fluctuations in nonlinear systems. *Zh. Eksp. Teor. Fiz.* **72** (1977), 238–247 (Engl. Transl. *Sov. Phys.–JETP* **45** (1977), 125).

[ES] D. J. Evans and D. J. Searles. Equilibrium microstates which generate second law violating steady states. *Phys. Rev.* E **50** (1994), 1645–1648.

[J] C. Jarzynski. Hamiltonian derivation of a detailed fluctuation theorem. *J. Stat. Phys.* **98** (2001), 77–102.

[GC] G. Gallavotti and E. G. D. Cohen. Dynamical ensembles in stationary states. *J. Stat. Phys.* **80** (1995), 931–970.

https://doi.org/10.1017/S0143385708000163 Published online by Cambridge University Press