Phase space gradient of dissipated work and information: A role of relative Fisher information

TAKUYA YAMANO
Department of Mathematics and Physics, Faculty of Science, Kanagawa University, 2946, 6-233 Tsuchiya, Hiratsuka, Kanagawa 259-1293, Japan

We show that an information theoretic distance measured by the relative Fisher information between canonical equilibrium phase densities corresponding to forward and backward processes is intimately related to the gradient of the dissipated work in phase space. We present a universal constraint on it via the logarithmic Sobolev inequality. Furthermore, we point out that a possible expression of the lower bound indicates a deep connection in terms of the relative entropy and the Fisher information of the canonical distributions.

I. INTRODUCTION

Phase space gradient of thermodynamic quantities in a nonequilibrium process and those limitations by information theoretic ones should bring fundamental insights into the understandings of the system. Among others, dissipation and work are central quantities in thermodynamic operation between equilibrium states. The connecting path between the two distinct equilibrium states in phase space has a diversified range depending on the steps of the procedure and it determines the amount of the mechanical work needed to perform the process. It is well recognized that as long as the process is not quasistatic, the free energy at the end of the process becomes less than that of the initial state plus the work invested from outside [1]. In other words, there is dissipative loss of work into the surroundings of the system – the second law of thermodynamics.

The dissipated work is thus an indicator of the excess of the injected work into the system, and it is defined in terms of the difference in the equilibrium free energy $\Delta F$ as $W_{\text{diss}}(\Gamma, \lambda) = W(\Gamma, \lambda) - \Delta F$, where $\lambda$ is a protocol parameter. $W(\Gamma, \lambda)$ denotes the average work done on the system by the external operator as perturbation and it is a function of the position in phase space. The dissipated work also represents the total change in entropy of the system as a result of the operated transition. The free energy difference between terminal states is also called the reversible work. Instead of taking one instance in the definition, we consider a statistical ensemble of realizations of the process as the consequence of infinitely many repetitions of this process. Therefore, we refer to the dissipated work in the sense of mean $\langle W_{\text{diss}}(\Gamma, \lambda) \rangle$ throughout this paper, but for simplicity we omit the angular brackets in the following.

Any displacement in phase space from a specific equilibrium state to another equilibrium phase point induced by a dynamics has a counterpart process by reversing time. The to-and-fro movement makes us to expect finding a universal relation in the dissipated work in an averaged way. One such intriguing example was discovered as a relation [2, 3]

$$\frac{\langle W_{\text{diss}} \rangle}{kT} = D_{KL}(P_F \parallel P_B), \quad (1)$$

under a sequence of procedure where the system is initially at canonical equilibrium with temperature $T$, and then it is detached from the heat reservoir to let the system evolve according to the Liouville equation and lastly equilibrate again to a new canonical state determined by the protocol parameter by attaching the same bath at the end of the process. The equality is replaced by inequality in case the system is kept contacting with the bath [3]. $P_F$ and $P_B$ are phase densities of the forward and the time-reversed processes at a particular time, respectively. Whether or not the prompt thermalization for each instance at any attained phase space point is realized remains a matter of careful thought, however, it is a standard device for consideration commonly used in the literature. The relation gives a meaning of a degree of time-reversal asymmetry measured by the Kullback-Leibler (KL) relative entropy $D_{KL}(P_F \parallel P_B)$ from $P_F$ and $P_B$ [4, 5]. The left-hand side of Eq.(1) is a physical content in units of thermal energy with the Boltzmann constant $k$ and, on the other hand, the right-hand side represents purely information theoretic quantity.

*Email: yamano@amy.hi-ho.ne.jp
In this paper, based on the above consideration, we see how the gradient of the dissipated work in phase space is inextricably linked to an information theoretic distance between phase density functions of forward and backward processes. Specifically, we show that a novel constraint on the averaged square of the gradient of the dissipated work \(|\nabla W_{\text{diss}}|^2\) can be obtained via the relative Fisher information and the logarithmic Sobolev inequality. We further present a possible inequality expression indicating a deep connection in terms of two kinds of information quantities. These studies are in conformity with a modern approach that bases the construction of statistical thermodynamics upon the concept of information (e.g., [1]).

In Sec.II, we first present a general relation corresponding to Eq.(1) when we employ an alternative to the KL relative entropy in order to recognize a distance notion in more general context in terms of information measure. The main ingredient of our consideration is introduced in Sec.III. We then show a constraint of the obtained relation with the inequality in Sec.IV. We give a concluding summary in Sec. V.

## II. A GENERAL RELATION FOR DISSIPATED WORK

Let \(M^n\) be the \(n\)-dimensional configuration space of the present system. The two participating distributions \(\mathcal{P}_F\) and \(\mathcal{P}_B\) are compared at equal-time when calculating the general relative entropy, which is specified by a convex function \(\chi(\mathcal{P}_F/\mathcal{P}_B)\) with \(\chi(1) = 0\),

\[
D_G(\mathcal{P}_F||\mathcal{P}_B) := \int \mathcal{P}_F \chi \left( \frac{\mathcal{P}_F}{\mathcal{P}_B} \right) d\Gamma. \tag{2}
\]

This form of the general relative entropy is first introduced in [3, 4] and it is now well recognized as the Csiszár-Morimoto divergence (less known in physics, though). It was employed in the proof of the H-theorem for Markov processes [5, 6]. In general, we do not require the asymmetry in the arguments but assume positivity. Let \(T^* M^n\) be the cotangent bundle to \(M^n\), i.e., phase space. The system whose Hamiltonian is parameter dependent such as \(H_\lambda(\Gamma_\lambda)\) starts its evolution from a phase space point \(\Gamma_0 \in T^* M^n\) at time \(t_0\) and reaches another point \(\Gamma_1 \in T^* M^n\) at \(t_1\). During this interval of time, the protocol parameter changes from \(\lambda(t_0) = \lambda_0\) to \(\lambda(t_1) = \lambda_1\). Initially, the system is put in the canonical equilibrium form \(\mathcal{P}_F(\Gamma_0) = e^{-\beta H_{\lambda_0}(\Gamma_0)}/Z_0\) with \(\beta = (kT)^{-1}\). Similarly, prior to the backward process the system is supposed to have a density function \(\mathcal{P}_B(\Gamma_1) = e^{-\beta H_{\lambda_1}(\Gamma_1)}/Z_1\) by preparing the canonical form at the start of the return. The reverse protocol that goes from \(\lambda_0\) to \(\lambda_1\) corresponds to \(\Gamma_1\) and \(\Gamma_0\) in the phase space. The free energy difference between the initial and the final states is given by \(\Delta F = -\beta^{-1}(\ln Z_1 - \ln Z_0)\) with the partition functions \(Z_0\) and \(Z_1\) for each state. We note that the protocol path in \(T^* M^n\) joining \(\Gamma_0\) to \(\Gamma_1\) that represents the operation does not necessarily proceed along the geodesic associated with it. The geodesic on the surface of the phase space manifold is a significant locus, however, the protocol given by an external agent or by an arbitrary schedule does not always go through the shortest path. Moreover, even if an operation for one realization takes the geodesic, other realizations do not trace exactly the same route at every time. They fluctuate around the geodesic path. That is why the work performed on the system is statistically distributed representing the ensemble of realizations and it is averaged over the ensemble.

Let \(C\) be a curve (path) on a surface of \(M^n\) connecting \(\Gamma_0\) and \(\Gamma_1\). Then, the amount of work the system receives on completion of the process depends not only on the location of the two terminal points \(\Gamma_0\) and \(\Gamma_1\), but also on the route \(C\). It should be precisely denoted as \(W_C(\Gamma_0,\Gamma_1)\) and equals to the difference in the Hamiltonian \(H_{\lambda_1}(\Gamma_1) - H_{\lambda_0}(\Gamma_0) = \int_{\lambda_0}^{\lambda_1} d\lambda \partial H_{\lambda}(\Gamma_\lambda)/\partial \lambda\) (the first law of thermodynamics). For simplicity, we omit both the subscript and arguments, and we denote \(W\) in the following.

We note that for the well-defined distance \(D_G\), the relative density \(\mathcal{P}_F/\mathcal{P}_B\) must have a finite value on a support (that is, \(\mathcal{P}_F\) is absolutely continuous with respect to \(\mathcal{P}_B\)), otherwise it is defined as infinity. Then, the relative phase density becomes

\[
\frac{\mathcal{P}_F}{\mathcal{P}_B} = \frac{Z_1}{Z_0} e^{-\beta(H_{\lambda_0}(\Gamma_0) - H_{\lambda_1}(\Gamma_1))} = e^{\beta(W - \Delta F)}. \tag{3}
\]

Inserting this into Eq.(2), we obtain the general reformulation of Eq.(1) as the following form

\[
D_G(\mathcal{P}_F||\mathcal{P}_B) = \left\langle \chi(e^{\beta(W - \Delta F)}) \right\rangle_{\mathcal{P}_F}, \tag{4}
\]
where \( \langle \cdot \rangle_{PF} \) denotes the average with respect to the distribution \( P_F \). This is the most general expression relating the dissipated work and the distance measure in this setting. We remark that the choice \( \chi(x) = \ln x \) in Eq. (2) with \( x = P_F/P_B \) defines the KL distance and recovers exactly the result Eq. (1). It is also worth mentioning another example. If we choose \( \chi(x) = 1/\sqrt{x} \), we have the overlap distance between \( P_F \) and \( P_B \), i.e.,

\[
D_O(P_F\parallel P_B) := \int \sqrt{P_F/P_B} d\Gamma = \langle \sqrt{P_F/P_B} \rangle_{PF}.
\]

If two phase densities are identical, it gives unity. However, \( P_F \) and \( P_B \) can never be identical \((P_F \neq P_B)\) in our consideration. Substituting Eq. (3) stemming from the canonical forms into the definition of the overlap distance, we have

\[
D_O(P_F\parallel P_B) = \langle e^{\frac{\beta}{2}(W-\Delta F)} \rangle_{PF} = e^{\frac{\beta}{2}\Delta F} \langle e^{-\frac{\beta}{2}W} \rangle_{PF},
\]

\( \text{J.E.} = 0. \) (5)

The second line follows from the fact that the \( \Delta F \) is not a statistically distributed quantity. The last line is due to the nonequilibrium work relation called the Jarzynski equality (J.E.) \( \langle e^{-\beta W} \rangle = e^{-\beta\Delta F} \) [9, 10], where the angular brackets denotes an average over a statistical ensemble of realizations of a single process through phase space. It is known that for general circumstances including the Hamilton’s evolution, this average is equivalent to taking an average using the initial equilibrium distribution designated by \( \lambda_0 \) [11]. Furthermore, the J.E. can be potentially useful in various equilibrium statistical physics, in which the estimation of the free energy differences is difficult to obtain. In these cases, such an equilibrium property is obtained from observations (experiments) of work distributions performed on the system. Note that distinct distributions in phase space do not share the common domain, so that Eq. (5) means that the vanishing overlap between the two equilibrium states is consistent with the fact that the J.E. holds. On the other hand, the J.E. is irrelevant to the derivation of Eq. (1), i.e. the case of the KL distance. We also remark that when the system realizes \( W = \Delta F \) (reversible process), the overlap \( D_o \) vanishes, meaning that it can be a measure of how much the nonequilibrium behavior deviates from the reversibility.

### III. RELATIVE FISHER INFORMATION AND DISSIPATED WORK

In Section II, we recognized that Eq. (1) is just one realization of the information theoretic expression of the averaged dissipated work among other possibilities. The choice of \( \chi \) is basically at our disposal and there is no legitimate criterion in terms of information theory. For operational purposes such as statistical inferences and estimation of measurement values, relative entropy with parameters can be considered advantageous over the one without them. Such a generalized distance measure with keeping the original properties of the Kullback-Leibler has been proposed [12] (see also e.g., [13]). But we do not delve into such possibilities including the overlap distance in this paper. Instead, to advance the understanding in line with previous works, we hereafter set the gauge of the distance measure as KL in our consideration. As we pronounced in Introduction, we are concerned with the quantity \( \nabla_T W_{\text{diss}} \), where \( \nabla_T \) indicates the gradient in phase space, because the dissipated work depends on the location in phase space and its gradient reflects a local structure (geometry) of the dissipation occurred as an excess of work performed on the system.

The relative Fisher information from \( f \) to \( g \) is defined as (e.g. [14], p.278)

\[
D_{RFI}(f\parallel g) := \int f \left| \nabla_T \left( \ln \left| \frac{f}{g} \right| \right) \right|^2 d\Gamma.
\]

(6)

It is non-negative and achieves zero iff \( f = g \). Further, it is asymmetric, i.e., \( D_{RFI}(f\parallel g) \neq D_{RFI}(g\parallel f) \), meaning that it is directed as is the case of the KL distance. This form can be defined independently of whether the distributions possess estimation parameters or not. We can also employ it when the two distributions are parametrized by the same family (say, \( \theta \)). We can say that this metric reflects a local comparison of the two distributions in that we take the derivative of them, whereas the KL distance returns a coarse-grained quantity by gathering each contribution of the displacement (or difference) between distributions. The same remark mentioned before Eq. (3) also applies to this relative distance: the relative phase density must have finite value on the domain of phase space. To the best of our knowledge, there is no study to give a physical connection to that measure. Now, we show one of the applications below. That is, substituting Eq. (3) into the above definition, we have a relation (see Appendix A)

\[
D_{RFI}(P_F\parallel P_B) = \beta^2 \left( \left| \nabla_T W_{\text{diss}}(\Gamma) \right| \right)^2_{PF}.
\]

(7)
This identity articulates that an average of the square of the gradient of the dissipated work generated by the transient process, taken over a forward equilibrium state, can be equated with the distance between the forward and backward distributions in the phase space measured by the relative Fisher information.

Our setting of formulation represented by an operational parameter is in a position of Hamilton equations (different equations on symplectic manifolds). Then, in order to consider nonequilibrium processes in phase space perspective, it is principally necessary to deal with the trajectory on a manifold (either on symplectic or on Riemannian). It is known that paracompact $C^\infty$-manifolds always have their Riemannian metrics. Since the usual manifold is paracompact, the Riemannian metric is introduced if we regard the phase space as a manifold. Then, the metric expression of the above identity is of concern. To this end, we remark that a gradient vector of a differentiable function $f$ on the Riemannian manifold associated with the covector $g^{ij}$ is defined by the derivative of $f$ along $a$. Therefore, the gradient vector $\nabla f$ satisfies $df(a) = \nabla f$. Then, substituting $(\nabla f)^i$ into this, we find that the square of the modulus of the gradient $f$ can be written as $|\nabla f|^2 = df(\nabla f) = g^{ij}(\partial x_i f)g^{ij}(\partial x_j f)$. Since we now take $f = W_{\text{diss}}(\Gamma)$ in the present consideration, therefore, the right-hand side of Eq. (7) can be expressed as

$$\beta^2 \left\langle \sum_{ij} \frac{\partial W_{\text{diss}}(\Gamma)}{\partial x^i} g^{ij} \frac{\partial W_{\text{diss}}(\Gamma)}{\partial x^j} \right\rangle_{\mathcal{P}_F}.$$  

We can obtain a more direct physical meaning for the quantity of the Dirichlet form $(|\nabla_W W_{\text{diss}}(\Gamma)|^2)_{\mathcal{P}_F}$ appearing in Eq. (7) via the gradient flow interpretation as we will see shortly. To this end, we recall again what $\nabla f$ expresses. It means that the work done on the system when it starts from $\Gamma$ and reaches $\Gamma_1$ by changing the control parameter $\lambda \in [\lambda_0, \lambda_1]$ along a path $C$ on the surface of the phase space manifold. Since $\lambda$ designates the instance of the evolution from any starting point $\Gamma_0$, we rewrite it as $W(\lambda)(\Gamma)$ by denoting $\Gamma_0$ as $\Gamma$ on a specific $C$. Accordingly, the dissipated work up to the intermediate value of $\lambda$ can be denoted as $W_{\text{diss}}(\Gamma_\lambda)$, where the phase point $\Gamma_\lambda \in T^* M^n$ should be read as $\hat{T}_\lambda \Gamma$ with an evolution operator $\hat{T}_\lambda$ that represents the protocol. Recall now that the gradient flow associated with the velocity vector field $-\nabla(\delta G/\delta \rho) := v(\Gamma_\lambda)$ is defined (Appendix B) as

$$\frac{\partial \rho}{\partial t} = \text{div} \left( \rho \nabla G \right),$$  

where $G(\rho)$ is an energy functional. In our case, we can take it as

$$G(\rho) = \int W_{\text{diss}}(\Gamma_\lambda) \rho(\Gamma) d\Gamma.$$  

With abuse of notation above, we have used $\rho(\Gamma)$ as the probability density function in phase space at each time instant $t$. This is the phase space average of the dissipated work in our consideration and the first variation of it is $\delta G/\delta \rho = W_{\text{diss}}(\Gamma_\lambda)$.

Therefore, the average kinetic energy $K_\lambda$ accompanied to the dissipation up to the protocol $\lambda$ is found to be just the squared mean of the gradient of the dissipated work evaluated by the initial equilibrium state $\mathcal{P}_F$ that is prepared for the start of the process:

$$K_\lambda = \int |v(\Gamma_\lambda)|^2 \mathcal{P}_F(\Gamma) d\Gamma = \left\langle |\nabla_W W_{\text{diss}}(\Gamma)|^2 \right\rangle_{\mathcal{P}_F}.$$  

In the whole process, the total kinetic energy becomes $\int_{\lambda_0}^{\lambda_1} K_\lambda d\lambda$.

IV. A LOWER BOUND FOR AVERAGE DISSIPATED-WORK GRADIENT

In this section, we show that our primarily focused quantity can be bounded by the KL distance via the logarithmic Sobolev inequality (LSI) [13]. In this sense, the identity that connects the relative Fisher information and the quantity $(|\nabla_W W_{\text{diss}}(\Gamma)|^2)_{\mathcal{P}_F}$ derived in the previous section can be regarded as an intermediate result. The LSI has wide range of applications and it takes several mathematically equivalent forms [14]. In this paper, we employ the form (Lemma 6.1 in [15])

$$\int |f|^2 \ln |f| d\mu \leq c \int |\nabla f|^2 d\mu + \|f\|^2_2 \ln \|f\|^2_2,$$  

where $c$ is a constant.
which holds for all functions \( f \), whose gradient \( \nabla f \) and \( f \) itself are square integrable in the domain. \( \| f \|_2 \) denotes \( (\int f^2 d\mu)^{1/2} \) and the constant \( c > 0 \) independent of \( f \). The physical interpretation of this constant is not obvious in general circumstances and would depend on the physical model. However, when an equilibrium state satisfies the LSI and the probability density function has the corresponding gradient flow, the constant has a clear meaning of how fast the system equilibrates with it. The corresponding relaxation rate takes the exponential form and it appears as a pre-factor to the KL distance between the initial and the equilibrium density functions. This feature is a direct consequence of the result provided in \(^{16}\) (see also e.g., \(^{14}\), p.288 and Appendix C). If we substitute \( d\mu = \mathcal{P}_B(\Gamma) d\Gamma \) for the probability measure and choosing \( f = \sqrt{P_F/P_B} \) then multiplying both sides by 2, we find

\[
\int \mathcal{P}_F \ln \frac{P_F}{P_B} d\Gamma \leq 2c \int |\nabla \sqrt{\frac{P_F}{P_B}}|^2 \mathcal{P}_B d\Gamma = \frac{c}{2} \int \mathcal{P}_F |\nabla \left( \ln \frac{P_F}{P_B} \right)|^2 d\Gamma. 
\]

In what follows, we set \( c = 1 \) without impeding our consideration. Combining Eq.4 and Eq.13, we find that the mean of the gradient dissipated work is lower bounded by the information theoretic distance between forward and backward phase densities:

\[
\left\langle (\nabla_W \text{diss}(\Gamma))^2 \right\rangle_{\mathcal{P}_F} \geq 2(kT)^2 D_{KL}(\mathcal{P}_F||\mathcal{P}_B). 
\]

We further pursue a reformulation of thus obtained constraint from an information point of view, that is, in terms of Fisher information in statistics.

We begin with recalling the followings. For a family of probability distributions \( \rho_\theta(x) \) parametrized by \( \theta \), the Fisher information in estimation theory is defined as \( I(\theta) := \left\langle \left( \nabla_\theta \rho/\rho \right)^2 \right\rangle_{\rho_\theta(x)} \). In the same way, it is defined also as \( I(\rho(x)) := \left\langle \left( \nabla_\rho \rho/\rho \right)^2 \right\rangle_{\rho(x)} \) for a differentiable distribution \( \rho(x) \), and it measures how much two neighboring density functions are statistically distinguishable. The former form is invariant against any shift \( \theta \), which signifies the diagonal entries of the Fisher information matrix, thereby implying i.i.d. data. Indeed, when we choose as \( \rho_\theta(x) = \rho(x-\theta) \), we easily find that due to \( \nabla_x \rho(x-\theta) = -\nabla_\theta \rho(x-\theta) \), the Fisher information does not depend on \( \theta \) and \( I(\theta) = I(\rho(x)) \) follows.

There have been comprehensive efforts to understand various physical laws in terms of this information \(^{17}\). It also plays a crucial role to upper bound the entropy production (e.g. \(^{19}\) and references therein). Since the present canonical equilibrium distribution specified by a protocol parameter satisfies \( \nabla \mathcal{P}(\Gamma) = -\beta H(\Gamma) \mathcal{P}(\Gamma) \), it becomes

\[
I(\mathcal{P}) = \left\langle \left( \frac{\nabla \mathcal{P}}{\mathcal{P}} \right)^2 \right\rangle_{\mathcal{P}} = \int \frac{\nabla \mathcal{P}}{\mathcal{P}} (-\beta \nabla H(\Gamma)) \mathcal{P} d\Gamma.
\]

Integrating by parts, the right-hand side reduces further to

\[
-\beta \int \nabla H(\Gamma) \nabla \mathcal{P}(\Gamma) d\Gamma = \beta \left\langle \Delta H(\Gamma) \right\rangle_{\mathcal{P}},
\]

where we have assumed that the phase density vanishes at the boundary \( \partial S \) of the system, so that \( [\mathcal{P} \nabla H(\Gamma)]_{\partial S} = 0 \), and \( \Delta \) denotes the Laplacian in phase space. Therefore, we obtain a striking relation

\[
I(\mathcal{P}) = \beta \left\langle \Delta H(\Gamma) \right\rangle_{\mathcal{P}}, \tag{16}
\]

which tells that the inverse temperature of the system can be intimately linked with the information quantity (Upon revision of this manuscript, the author became aware that this relation has derived also in Ref. \(^{21}\)). In general coordinates, we note that \( \Delta H(\Gamma) \) can be expressed as \( \sqrt{g}^{-1} \partial_i (\sqrt{g} g^{ij} \partial_j H(\Gamma)) \), where \( g \) is the determinant of the metric \( g = det(g_{ij}) \). A way to interpret this relation is that the temperature of the system can be defined by the Fisher information of the system and by the averaged second order differential (curvature) of energy \(^{20}\). As we shall use below, this relation is critical for the present study. Since the system contacts with the same heat bath (with common \( \beta \)) both at the start and the end of the process as described in Sec. II (or we could also restate it as follows; the heat reservoir is so large compared with the system, so that temperature of the reservoir is not disturbed by the heat discarded by the system), we have readily a relation from Eq.16

\[
\frac{I(\mathcal{P}_F)}{\left\langle \Delta H(\Gamma) \right\rangle_{\mathcal{P}_F}} = \frac{I(\mathcal{P}_B)}{\left\langle \Delta H(\Gamma) \right\rangle_{\mathcal{P}_B}} = \beta. \tag{17}
\]
An immediate but profound implication of this consequence is that the ratio of the Fisher information associated with probability distributions of the forward and backward processes is equivalent to the ratio of the averaged curvature of the Hamiltonian. This is true if the system takes the canonical form in distribution. The former ratio has the origin of the information quantity and the latter has the physical one. Next, substituting Eq. (17) into Eq. (13), we readily have an inequality

$$\frac{\langle |\nabla f W_{\text{diss}}(\Gamma)\rangle_{P_F}^2}{\langle \Delta H(\Gamma) \rangle_{P_\gamma}^2} \geq 2 \frac{D_{KL}(P_F \| P_B)}{\langle I(\Gamma) \rangle_{P_\gamma}^2}.$$  \hspace{1cm} (18)

where the symbol $\gamma$ denotes either $F$ or $B$ representing the forward and the backward equilibrium states. The lower bound on the ratio relevant to the averaged physical quantities (the left-hand side) is nicely bounded from below in terms of the ratio of information-associated quantities only. A further interesting observation can be derived for this relation Eq. (18) from the well-known Cramer-Rao inequality in statistical estimation theory. For simplicity’s sake, we consider it for one-dimensional case. The Cramer-Rao inequality tells a tradeoff relation between Fisher information and the variance of the distribution $\sigma_X^2$, i.e., $I(P(X)) \geq 1/\sigma_X^2$, where $X$ is a random variable. Then, we find that the lower bound in Eq. (18) is upper bounded by $2D_{KL}(P_F \| P_B)\sigma_X^4$. Now that $P(X)$ is of the canonical form in our setting, the equality can be achieved when the Hamiltonian $H(X)$ is of quadratic form.

V. SUMMARY

To deepen the understandings of a profound information theoretic relation between work and dissipation in nonequilibrium systems, we have derived a universal relation that connect the gradient of the dissipated work and the relative Fisher information within a framework of the setup repeatedly employed in the previous studies. Considering the gradient of the dissipated work at each point in phase space enables us to get geometric information that cannot be obtained from KL entropy only. The relative Fisher information plays the role. By way of this, we have established the information based lower bound for the quantity relevant to the dissipation in phase space. The instantaneous equilibration was a premise to assure the well-posedness of the free energy difference between the two canonical equilibrium states associated to the forward and backward processes. However, the notion of the nonequilibrium free energy change $\Delta F_{\text{neq}}$ has recently considered to refine the dissipation occurring in far from thermodynamic equilibrium, where the dissipated work is defined by average work minus $\Delta F_{\text{neq}}$ instead of the equilibrium free energy change $\Delta F$. An extension of the present result to such a case surely needed if one is to understand and to gain deeper insights into biological systems.

Acknowledgments

The author wishes to thank Hiroaki Yoshida for a valuable discussion on the relative Fisher information at the Ochanomizu University in August 2012.

Appendix A: Relative Fisher information $D_{RFI}$

The concept of the relative Fisher information measured from $f$ to $g$, $D_{RFI}(f\|g)$ is less acknowledged in application by the general physics community compared with the relative entropy (Kullback-Leibler divergence). We briefly present here a physical origin of this form that is unrecognized in the literature (see also [23] and references therein for a more general discussion). We consider the time change of the KL divergence between two distribution functions $f = f(x,t)$ and $g = g(x,t)$ that obey a heat equation ($\partial f = \mathcal{D} \nabla^2 f$ etc.), where $\mathcal{D}$ is a diffusion constant:

$$\frac{d}{dt} D_{KL}(f\|g) = \frac{d}{dt} \int f \ln \frac{f}{g} \, dx = \int f \ln \frac{f}{g} \, dx + \int g \left( \frac{\dot{f}}{g} - \frac{f}{g^2} \dot{g} \right) \, dx, \hspace{1cm} (A1)$$

where $\dot{f} = \partial f/\partial t$ etc. The first term is calculated by integration by parts as

$$\mathcal{D} \int (\nabla^2 f) \ln \frac{f}{g} \, dx = -\mathcal{D} \int (\nabla f) \nabla \left( \ln \frac{f}{g} \right) \, dx, \hspace{1cm} (A2)$$
where we have used the boundary conditions that $f$ and $\nabla f$ vanish when $|x| \to 0$. Similarly, under the conditions that $f$, $g$ and $\nabla g$ vanish when $|x| \to 0$, the second term becomes

\[ \mathcal{D} \int f \left( \frac{\nabla g}{g} \right) \nabla \left( \ln \frac{f}{g} \right) dx. \]  

(A3)

Therefore, combining Eq.(A2) and Eq.(A3), we have

\[ \frac{d}{dt} D_{KL}(f\parallel g) = -\mathcal{D} \int f \left( \frac{\nabla f - \nabla g}{f} \right) \nabla \left( \ln \frac{f}{g} \right) dx \]

(A4)

Except for the diffusion coefficient $\mathcal{D}$, the right-hand side provides $D_{RFI}$, which is called the de Bruijn-type identity [23]. The information involves how fast the KL distance changes under the heat equation. In a more general context [23], we have

\[ \frac{d}{dt} D_{KL}(f\parallel g) = \left\langle \left( \frac{j_f - j_g}{f} \right) \nabla \left( \ln \frac{f}{g} \right) \right\rangle_f, \]  

(A5)

where $j_f$ and $j_g$ are the flows associated, respectively, with $f$ and $g$ in the continuity equation $\partial_t \rho = -\nabla \cdot j$, and $\langle \cdot \rangle_f$ denotes the averaging with respect to $f$.

1. $D_{RFI}$ between two canonical equilibrium distributions

The relative Fisher information between forward and backward phase densities is calculated as

\[ D_{RFI}(P_F\parallel P_B) = \int P_F \left| \nabla_{\Gamma} \left( \ln \frac{P_F}{P_B} \right) \right|^2 d\Gamma \]  

(by definition Eq.(6))

\[ = \int P_F \left| \nabla_{\Gamma} \left( \ln e^{\beta(W-\Delta F)} \right) \right|^2 d\Gamma \]  

(by relation Eq.(3))

(A6)

\[ = \beta^2 \int P_F \left| \nabla_{\Gamma}(W-\Delta F) \right|^2 d\Gamma. \]

Finally, by the definition of the dissipated work, we have the relation Eq.(7).

Appendix B: Gradient flow

An evolution equation of the form

\[ \frac{\partial \rho}{\partial t} = \text{grad}_W G(\rho) \]  

(B1)

for some functionals $G(\rho)$, is called a gradient flow, where $\text{grad}_W G(\rho)$ is the Wasserstein gradient on the Wasserstein space (e.g., [14]). A wide class of partial differential equations can be understood in light of this approach. The Wasserstein gradient here $\text{div}(\rho v(\Gamma_{\lambda}))$ for the potential functional $G$ causes the time change in the density function. In other words, the vector field makes the backward density function $P_B$ different from the forward one $P_F$, during which the work dissipation is completed. The variation of $G$ with respect to $\rho$ is the gradient in $L^2$, and corresponds to the dissipated work $W_{diss}(\Gamma_{\lambda})$ in the present consideration. To be more convinced, it is appropriate to consider a linear Fokker-Planck equation whose free energy functional $G(\rho)$ can be given by

\[ G(\rho) = \int u(x) \rho(x) dx + T \int \rho(x) \ln \rho(x) dx, \]  

(B2)

with $u(x)$ and $T$ denoting the internal energy and temperature, respectively. Then, the gradient of the functional $\delta G/\delta \rho = u(x) + \log \rho(x) + 1$ leads to the gradient flow $\partial \rho/\partial t = \text{div}(\nabla \rho + \rho \nabla u(x))$. In terms of the continuity equation, it is equivalent to that the system has a flow $j = -\nabla \rho + \rho \nabla u(x)$. That is, it consists of the Fick’s law plus the gradient of the internal energy. Substituting this into Eq.(A5), we immediately recover the de Bruijn-type identity Eq.(A4).
Appendix C: Approach to equilibrium controlled by a constant in the logarithmic Sobolev inequality

The constant \( c \) in the LSI (Eq. [12]): \[ \int |f|^2 \ln |f| d\mu - (\int |f|^2 d\mu) \ln(\int |f|^2 d\mu) \leq c \int |\nabla f|^2 d\mu \] determines the rate at which the system approaches to an equilibrium state when measured with the KL distance. Consider the KL distance between a density function \( f_t \), \( t \geq 0 \) and an equilibrium one \( f_\infty \). Its time derivative is then

\[
\frac{d}{dt} D_{KL}(f_t \| f_\infty) = \int f_t \ln \frac{f_t}{f_\infty} d\mu.
\]

(C1)

Since we take \( G = D_{KL}(f_t \| f_\infty) \) in the definition (Eq. [9]) and then \( \delta G/\delta f_t = 1 + \log f_t/f_\infty \), the gradient flow associated with it is

\[
\frac{\partial f_t}{\partial t} = \nabla \left( f_t \ln \frac{f_t}{f_\infty} \right).
\]

(C2)

Substituting this into the above and performing the integration by parts under the vanishing boundary condition, we have

\[
\frac{d}{dt} D_{KL}(f_t \| f_\infty) = -\int f_t |\nabla \ln \frac{f_t}{f_\infty}|^2 d\mu = -D_{RFI}(f_t \| f_\infty).
\]

(C3)

This is the de Bruijn-type identity under a gradient flow with the equilibrium state \( f_\infty \) as a reference density function. On the other hand, choosing the probability measure as \( d\mu = f_\infty d\mu \) and putting \( f = \sqrt{f_t/f_\infty} \) in the LSI, we have

\[
\frac{1}{2} \int f_t \ln \left( \frac{f_t}{f_\infty} \right) f_\infty d\mu \leq c \int |\nabla \sqrt{f_t/f_\infty}|^2 f_\infty d\mu
\]

\[
= \frac{c}{4} \int f_t |\nabla \ln \frac{f_t}{f_\infty}|^2 f_\infty d\mu.
\]

The right-hand side equals just to \( c/4D_{RFI}(P_F \| P_B) \). Therefore, from Eq. (C3) we have an inequality

\[
\frac{d}{dt} D_{KL}(f_t \| f_\infty) \leq -\frac{c}{2} D_{KL}(f_t \| f_\infty).
\]

(C4)

Recall that the Gronwall’s inequality states that if two continuous functions \( w(t) \) and \( \gamma(t) \) defined on an interval \( I = [a, \infty) \) satisfies a differential inequality \( w'(t) \leq \gamma(t) w(t) \), then \( w(t) \) is bounded as \( w(t) \leq w(a) \exp(\int_a^t \gamma(s) ds) \) for \( \forall t \in I \). Applying this to Eq. (C4) by setting \( a = 0 \), we readily have

\[
D_{KL}(f_t \| f_\infty) \leq D_{KL}(f_0 \| f_\infty) \exp(-\int_0^t \frac{c}{2} ds) = e^{-\frac{c}{2} t} D_{KL}(f_0 \| f_\infty),
\]

(C5)

where \( f_0 \) is the initial density function. This indicates explicitly that the distance between the initial and equilibrium density functions converges by the factor \( e^{-2t/c} \). The smaller the value of \( c \), the faster it converges.

[1] L. D. Landau and E. M. Lifshitz, *Statistical Physics Part 1, Course of Theoretical Physics* Vol. 5, §20, (Pergamon Press Ltd., 1980).
[2] C. Jarzynski, Phys. Rev. E 73 046105 (2006).
[3] R. Kawai, J. M. R. Parrondo and C. Van den Broeck, Phys. Rev. Lett. 98 080602 (2007).
[4] S. Kullback and R. A. Leibler, Ann. Math. Stat. 22 79 (1951).
[5] S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959).
[6] Arieh Ben-Naim, *A Farewell to Entropy: statistical thermodynamics based on information*, World Scientific, (2008).
[7] I. Csiszár, Publ. Math. Inst. Hungar. Acad. Sci. 8 85 (1963).
[8] T. Morimoto, J. Phys. Soc. Jpn. 18 328 (1963).
[9] C. Jarzynski, Phys. Rev. Lett. 78 2690 (1997).
[10] G. E. Crooks, J. Stat. Phys. 90 1481 (1998).
[11] C. Jarzynski, Eur. Phys. J. B 64 331 (2008).
[12] T. Yamano, J. Math. Phys. 50 043302 (2009); Phys. Lett. A 374 3116 (2010).
Special Issue: Distance in information and statistical mechanics, Entropy, edited by T. Yamano (Basel, Switzerland, 2011), see http://www.mdpi.com/journal/entropy/special_issues/distance-info-stat-physics.

C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics Vol.58 (American Mathematical Society, 2000).

L. Gross, Amer. J. Math. 97 1061 (1975).

A. R. Barron, Ann. Probab. 14 336 (1986).

B. R. Frieden, *Science from Fisher Information - A Unification* (Cambridge University Press, 2004).

T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley-Interscience, 2006).

T. Yamano, J. Math. Phys. 53 043301 (2012).

It is worth conferring that when we regard $P$ as a function of $\beta$, then the Fisher information provides the fluctuation of energy for the canonical distribution: $I(P(\beta)) \equiv \int P(\beta)(d\ln P(\beta)/d\beta)^2d\beta = \langle H^2(\Gamma) \rangle_p - \langle H(\Gamma) \rangle^2_p$.

K. R. Narayanan and A. R. Srinivasa, Phys. Rev. E 85 031151 (2012).

D. A. Sivak and G. E. Crooks, Phys. Rev. Lett. 108 150601 (2012).

T. Yamano, Eur. J. Phys. B 86 363 (2013).