Reconsideration of the generalized second law based on information geometry

T Nakamura, H H Hasegawa and D J Driebe

1 Department of Mathematical Sciences, Ibaraki University, Bunkyo, Mito, 310-8512 Japan
2 Center for Complex Quantum Systems, University of Texas, Austin, Texas, 78712 United States of America
E-mail: hiroshi.hasegawa.sci@vc.ibaraki.ac.jp

Keywords: nonequilibrium statistical mechanics, the generalized second law, information geometry

Abstract

The maximum work formulation of the second law of thermodynamics has been generalized to transitions between nonequilibrium states. The generalization involves the relative entropy between nonequilibrium states and canonical states. The relative entropy scaled by the temperature of the canonical state quantifies the work available for extraction from the nonequilibrium state. This scaled relative entropy can be interpreted as an energy-dimensional divergence in information geometry. The generalized Pythagorean theorem relating three energy-scaled divergences, which we interpret as thermodynamic distances, gives a geometrical interpretation of the generalized maximum work formulation. Under this interpretation the optimal cyclic operation to extract work from a nonequilibrium state is discussed in a simple two-level quantum system.

1. Introduction

The role of information processing in understanding the second law of thermodynamics has been evident since Maxwell [1] introduced his demon [2, 3]. Landauer [4], Bennett [5] and others [6–8], also pointed out that information processing itself may be accompanied by thermal effects, sometimes in subtle ways. These considerations have led researchers to recognize information as a resource on par with heat and work in formulating a thermodynamics of information [9–13].

We reconsider, based on information theory, the maximum work formulation of the second law generalized to transitions between nonequilibrium states [14–19]. The generalized second law is stated in terms of the initial and final relative entropies between each nonequilibrium state and corresponding canonical states, as well as the Helmholtz free energy of the canonical states. The relative entropy is known as the Kullback–Leibler (KL) divergence [20, 21] in information theory. The KL divergence is always non-negative and measures the ‘information distance’ between two probability distributions; it also serves as a bridge connecting information theory and geometry.

Information geometry [22, 23] is geometry on a Riemannian manifold. Probability distributions are the points of the Riemannian manifold. The Riemannian metric is provided by a divergence. The KL divergence furnishes the Fisher information matrix as the Riemannian metric tensor. Since the (KL) divergence is not symmetric in the two probability distributions it is not really a proper mathematical metric. Even so, Amari and Nagaoka [22, 23] demonstrated a generalized Pythagorean theorem (GPT) for a ‘right-angled triangle’ of three probability distributions and introduced projection as an important concept in information geometry.

In this paper we introduce an interpretation of the generalized second law based on information geometry. We need an energy-dimensional ‘distance’ [24] to measure extractable work from a nonequilibrium state, since ‘information distance’ measured by the KL divergence is dimensionless. This energy dimensional ‘distance,’ hereafter referred to as the ‘thermodynamic distance,’ is the KL divergence scaled by a temperature. Our temperature has an energy dimension as we set Boltzmann’s constant $k = 1$ throughout this paper.

The ‘distance’ measured through the divergence is the square of length in the sense of the Pythagorean theorem. This reference by Crooks is an important work in laying the geometrical foundation of thermodynamics.
The Riemannian metric tensor corresponding to the energy-scaled KL divergence is an energy-scaled Fisher information matrix. The geometric structure based on the energy-scaled Fisher information matrix is completely different from the geometric structure based on the original Fisher information matrix. The validity condition of the GPT based on the scaled KL divergence is an isentropic condition as will be shown in section 3. On the other hand, the validity condition of the GPT based on the bare KL divergence is an isoenergetic condition as will be shown in appendix B. This important difference has not been recognized, since many researchers mainly studied isothermal processes in which the temperature is just a constant [24]. In our approach the temperature is a variable parameter.

In the next section we start by recalling the maximum work formulation of the generalized second law. The generalized second law is universal for both classical and quantum systems. It is valid even for a finite integrable system since it can be derived using only two fundamental properties: the conservation of Gibbs–Shannon (GS) entropy and the non-negativity of the KL divergence. To clarify the role of information theory, we focus on a thermally isolated Hamiltonian system. We exclude any phenomenological assumptions such as detailed balance, course graining, or the existence of a heat reservoir that would produce a canonical ensemble.

We do not have an a priori temperature in a thermally isolated system, nevertheless; an effective temperature characterizing the nonequilibrium state can be introduced. Also, we show that the canonical distribution plays the role of a reference state for an initial nonequilibrium state, as well as a target for the final state that achieves maximum work extraction. This important role of the canonical distribution in a thermally isolated system is unconventional.

In sections 3 and 4, we reconsider the generalized second law based on information geometry. We consider the case of a cyclic operation to make clear the role of the scaled KL divergence in section 3. For a cyclic operation, the work is written only in terms of the scaled KL divergences because there is no change in the Helmholtz free energy. The work is bounded from below by the scaled KL divergence between a nonequilibrium initial state and a canonical state. The canonical state is parameterized by its temperature. The maximum work is determined by minimizing the ‘thermodynamic distance.’ The GPT for the scaled KL divergences gives orthogonality between the one-parameter line of canonical states and the isentropic surface of state space. The maximum work is given by the canonical state that intersects the surface, with the temperature determined by the isentropic condition. We extend our arguments to the case of a general non-cyclic operation in section 4. We also illustrate a concrete simple example of the GPT for the scaled KL divergences using the adiabatic expansion of the ideal gas at the end of section 4.

In section 5, we apply the geometrical interpretation of the generalized maximum work formulation to a simple two-level quantum system. The optimal cyclic operation to extract work from a nonequilibrium state is determined by minimalizing the scaled KL divergence between the final state and the final canonical state. The geometrical interpretation of the generalized maximum work formulation gives us a systematic method to figure out a protocol to realize the optimal operation.

In appendix B we show that the geometric structure based on the bare KL divergences is different from that based on the scaled KL divergences. The GPT for the bare KL divergences gives orthogonality between the one-parameter line of canonical states and the isoenergetic surface. It also gives as an immediate corollary the principle of maximum entropy.

2. The generalized second law

We start with a review of the maximum work formulation of the second law generalized to transitions between nonequilibrium states [14–17]. The generalized second law is valid for both classical and quantum systems. We consider a thermally isolated Hamiltonian system and focus on the period of the operation starting at \( t = 0 \) and finishing at \( t = T \). The state is described using the probability density, \( \rho(x,t) \) (or \( \rho_i \) in abbreviated form), at time \( t \) and phase-space point \( x \), which is either specified or is obtained by dynamical evolution from a previously specified state. The dynamics of the system is governed by its Hamiltonian, \( H(x,\kappa(t)) \) (abbreviated as \( H_i \)), that has an explicit time dependence due to the external parameters, \( \kappa(t) \), under our control. Hereafter we abbreviate a function of time \( A(t) \) as \( A_t \).

The time evolution of the state is written using the time evolution operator \( U \) as

\[
\rho_T = U^T \rho_0,
\]

where \( U^T \) is defined as

\[
U^T = \hat{T} \exp \left( -i \int_0^T L(\kappa(t)) dt \right),
\]

where \( \hat{T} \) is the time-ordered product and \( L \) is the Liouvillian for the Hamiltonian.

\[^{4}\text{The perpendicular line from the nonequilibrium initial state to the one-parameter canonical line gives the minimum 'thermodynamic distance' corresponding to the maximum work.}\]
We employ the bracket notation to describe the expectation value of an observable $A(x)$ as,

$$\langle A | \rho \rangle = \int_\Gamma A^*(x) \rho(x) \, dx,$$

where $\Gamma$ is the phase space of the Hamiltonian system and $A^*$ is the (transposed) complex conjugate of $A$. Thus the expectation value of the Hamiltonian of the system at time $t$ is given by $\langle H_t | \rho_t \rangle$. The work done on this thermally isolated system is given just by the change in the internal energy of the system, i.e.,

$$W = E_f - E_0$$

where the internal energy of the system at time $t$ is

$$E_t = \langle H_t | \rho_t \rangle.$$

The canonical distribution of the system with respect to the Hamiltonian and with a parameter $\alpha$ will play an important role:

$$p_{\text{can}}(\alpha) = \frac{e^{-\alpha H}}{Z(\alpha)}$$

Here, the partition function

$$Z(\alpha) = \langle 1 | e^{-\alpha H} \rangle$$

we are going to rewrite by using its relation to the Helmholtz free energy

$$F(\alpha) = -\alpha^{-1} \log Z(\alpha),$$

so that the canonical distribution may be expressed simply as the exponential function

$$p_{\text{can}}(\alpha) = e^{\alpha \left\{ F(\alpha) - H \right\}}.$$
Since the work is given by the change in the internal energy as in equation (4), we obtain

\[
W = \Delta F(\alpha) + \alpha^{-1} D_{KL}[\rho_T \| \rho_{can,T}(\alpha)] - \alpha^{-1} D_{KL}[\rho_0 \| \rho_{can,0}(\alpha)]
\]

(18)

where we used the conservation of the GS entropy in a thermally isolated Hamiltonian system, i.e.,

\[
S[\rho_T] = S[\rho_0] \quad \text{and} \quad D_{KL}(\alpha) \equiv F_T(\alpha) - F_0(\alpha) \quad \text{is the change in the Helmholtz free energy of the system. Since there is no heat bath and we do not have an a priori temperature, we call the parameter corresponding to an inverse temperature as } \alpha.
\]

The work equality, equation (18), is just a statement of energy accounting when we know the initial and final state of the system as well as the initial and final Hamiltonians. Let us suppose that we are just given a nonequilibrium initial state and have a Hamiltonian with external parameters under our control. Until we choose a schedule for the parameter control of the Hamiltonian the final state is not determined. Due to the nonnegativity of the KL divergence term involving the unknown final state, equation (18) provides us under the condition of a known initial state the following work inequality:

\[
W \geq \Delta F(\alpha) - \alpha^{-1} D_{KL}[\rho_0 \| \rho_{can,0}(\alpha)] \equiv W_{LB}(\alpha)
\]

(19)

where we denote the lower bound for the work as \(W_{LB}(\alpha)\).

First, we note that the inequality (19) is true for any value of \(\alpha\). Second, \(W_{LB}(\alpha)\) is a concave function of \(\alpha^{-1}\), which means that it has a global maximum as a function of \(\alpha\) as discussed in appendix A. Thus, there exists a best value for \(\alpha\); namely, the value for which \(W_{LB}(\alpha)\) is a maximum. This is so because the maximum of \(W_{LB}(\alpha)\) is a value for the work that may be achieved. A lesser value for \(W_{LB}(\alpha)\) (i.e., a more negative value when considering work extraction) while giving a true inequality would not be an achievable value for the work since this value would not satisfy the inequality of \(W \geq \max.\, W_{LB}(\alpha)\), which must be true.

To find the maximum of \(W_{LB}(\alpha)\) we make its dependence on \(\alpha\) explicit by writing the KL divergence, as in equation (16), in terms of the entropy, free energy and internal energy as

\[
W_{LB}(\alpha) = F_T(\alpha) + \alpha^{-1} S[\rho_0] - E_{ir}
\]

(20)

Now, taking the derivative with respect to \(\alpha\), and using the fact that the derivative of the free energy with respect to temperature is minus the entropy gives us

\[
\frac{\partial W_{LB}(\alpha)}{\partial \alpha} = \frac{1}{\alpha^2} S[\rho_{can,T}(\alpha)] - \frac{1}{\alpha^2} S[\rho_0].
\]

(21)

The value of \(\alpha = \tilde{\beta}\) where this equals zero we call the effective (inverse) temperature. This gives us the isentropic condition of

\[
S[\rho_{can,T}(\tilde{\beta})] = S[\rho_0].
\]

(22)

So, the effective temperature is determined as the temperature of the canonical distribution with respect to the final Hamiltonian that has the same entropy as the initial nonequilibrium state. This condition is entirely reasonable as the dynamics preserves the entropy and the final distribution being canonical means that the maximum energy permissible has been extracted. The isentropic condition is also naturally derived from the GPT in information geometry, which will be discussed in the next section.

The generalized second law for a nonequilibrium initial state in a thermally isolated system is thus,

\[
W \geq \Delta F(\tilde{\beta}) - \tilde{\beta}^{-1} D_{KL}[\rho_0 \| \rho_{can,0}(\tilde{\beta})],
\]

(23)

where the effective temperature, \(\tilde{\beta}^{-1}\), is determined by the isentropic condition, equation (22).

If the Hamiltonian is changed back to its original form, as in a cyclic process, then \(\Delta F = 0\) and the work on the system from the initial nonequilibrium state is just

\[
W \geq -\tilde{\beta}^{-1} D_{KL}[\rho_0 \| \rho_{can,0}(\tilde{\beta})].
\]

(24)

In both cases the work extracted from the system is \(-W^\ast\). The KL divergence in the right-hand-side of equation (24) measures the ‘information distance’ corresponding to the available informational resource. The effective temperature converts the available informational resource to the extractable work. Considering the Kelvin principle in the context of a thermally isolated system leads us to regard a canonical state as an equilibrium state for such a system as well

6 When the initial state is a canonical state, the effective temperature is the temperature of the canonical state from the isentropic condition. The divergence in the right-hand-side of equation (24) vanishes so that no work is extractable for any cyclic operation. This argument can be applied to an isothermal system obtained by dividing a thermally isolated total system into the system of interest and a reservoir with a canonical distribution [15]. We note that the microcanonical state is usually regarded as the equilibrium state in a thermally isolated system. As is well known though, the expectation value of any physical observable for a canonical state can be approximated by that of the corresponding microcanonical state in a many-body thermodynamic system in the sense of the law of large numbers.
3. Reconsideration based on information geometry

As shown in equation (18), the extractable work for a thermally isolated system is given as the difference between two KL divergences scaled by a temperature $\alpha^{-1}$ (and the change in the Helmholtz free energy). This work equality suggests that there is an information-geometric foundation of the generalized second law. In this section we consider a cyclic operation because this case has a clearer geometric structure than the case of a non-cyclic operation, which we will consider in the next section.

Information geometry considers the geometric structure in parameter space of families of probability distributions [22, 23]. A parameter (vector) $\theta$ specifies the distribution $\rho(\theta)$ and the notion of ‘distance’ between two distributions is provided by a divergence function $D[\rho(\theta_A) \parallel \rho(\theta_B)]$, which satisfies the following three conditions:

1. Non-negativity: $D[\rho(\theta_A) \parallel \rho(\theta_B)] \geq 0$.
2. Uniqueness: $D[\rho(\theta_A) \parallel \rho(\theta_B)] = 0$, if and only if $\theta_A = \theta_B$.
3. Positive-definiteness of the metric tensor $g_{ij}$: $D[\rho(\theta) \parallel \rho(\theta + d\theta)] = \sum g_{ij}/2 d\theta_i d\theta_j$.

The geometric structure is determined by the metric tensor appearing in the above local divergence. The KL divergence satisfies the above conditions and its metric tensor is the well-known Fisher information matrix,

$$ g_{ij} = \left\langle \frac{\partial \log \rho(\theta)}{\partial \theta_i} \frac{\partial \log \rho(\theta)}{\partial \theta_j} - \rho(\theta) \right\rangle, \quad (25) $$

Now we define the key concept, ‘thermodynamic distance.’ The ‘thermodynamic distance’ is the KL divergence between an arbitrary probability distribution and a canonical distribution scaled by the temperature of the canonical distribution, $\alpha^{-1}D_{KL}[\rho \parallel \rho_{can}(\alpha)]$. A divergence scaled by a positive parameter satisfies the above three conditions so that the scaled divergence is also a divergence. This scaled KL divergence has energy dimension and measures the extractable work in the generalized second law as a ‘thermodynamic distance.’ The metric tensor of the scaled KL divergence is the Fisher information matrix scaled by the temperature. The different metric tensor gives us completely different geometric structure as will be shown in this section.

The maximum extractable work for a cyclic process is determined by the following inequality obtained from equation (18),

$$ W \geq -\alpha^{-1}D_{KL}[\rho_0 \parallel \rho_{can,0}(\alpha)] \quad (26) $$

where we used $\Delta F(\alpha) = 0$ and the non-negativity of the KL divergence involving the final distribution. In order to find the lower bound of $W$ in equation (26) (i.e., the upper bound for the extractable work), we have to find the value of $\alpha$ that makes the right-hand-side of equation (26) maximum. In the previous section we derived the condition of the effective temperature by differentiating the right-hand-side of equation (26). Now we use the information-geometric structure to derive the same condition.

The minimum (or shortest) distance from a point to a plane is obtained by the Pythagorean theorem in elementary geometry. Similarly the minimum divergence from a probability distribution to a curved surface is obtained by the GPT in information geometry. The GPT is based on three divergences. Suppose that a point $P$ is a probability distribution and $\Sigma$ is a curved surface which does not include point $P$ in the space of probability distribution. When there exists a point $Q$ within $\Sigma$ such that the geodesic line from $P$ to $Q$ is orthogonal to $\Sigma$ as illustrated in figure 1, the GPT holds as

$$ D[P|R] = D[P|Q] + D[Q|R] \quad \text{for} \quad \forall R \in \Sigma. \quad (27) $$

From the non-negativity of the divergence, we obtain

$$ D[P|R] \geq D[P|Q] \quad \text{for} \quad \forall R \in \Sigma \quad (28) $$

which means that the divergence between $P$ and $Q$ is minimum [22, 23].

---

7 Although the divergence is not always symmetric, it is dual symmetric in information geometry [22, 23]. The local KL divergence for a canonical distribution $\rho_{can}(\beta)$ is written as $D_{KL}[\rho_{can}(\beta) \parallel \rho_{can}(\beta + \beta')] = d\beta \Delta E/\beta^2/2 = d\beta(\beta /d\beta)(\beta /d\beta)/2 = d\beta \beta /d\beta/2$ where $\Delta E^2$ is the Fisher information matrix. On the other hand, the scaled local KL divergence is also written as $D_{KL}[\rho_{can}(\beta) \parallel \rho_{can}(\beta + \beta')] = d(1/\beta \beta /d\beta)(1/\beta \beta /d\beta)/2 = d(1/\beta \beta /d\beta/2$. The dual orthogonality, $\beta \beta /d\beta(\beta /d\beta)$ is changed to $d(\beta^{-1})d\beta$ by the scale transformation. The validity conditions of the GPT are based on these dual orthogonalities as will be shown later. The dual variable $F(\Sigma)$ is related to the original $\beta (\beta^{-1})$ through the Legendre transformation based on the well-known convex function $-\beta F(\beta)$ by which the dual variable is given as $-E = d(\beta F /d\beta)(\beta = d(\beta F) /d(\beta^{-1}))$. 

---
The GPT based on three bare KL divergences has been well studied in information geometry [22, 23]. The validity condition of the GPT based on three bare KL divergences is the isoenergetic condition. It determines the minimum KL divergence from a probability distribution to a set of canonical distributions (one-parameter geodesic line of canonical distribution). The principle of maximum entropy is derived from the GPT based on three bare KL divergences. We show the details in appendix B.

Since the (information) entropy is dimensionless and the principle of maximum entropy is based on dimensionless KL divergences, we expect that the maximum work formulation is based on energy-dimensional divergences. Our energy-dimensional divergence is equal to (the minus of) the right-hand-side of equation (26) for the pair of the initial probability distribution and the initial canonical distribution. The change of scaled KL divergences is the (dissipative) work in equation (18) for a cyclic process. The minimization of this scaled KL divergence means the maximization of the work in equation (26).

The GPT based on scaled KL divergences holds as the following theorem:

**Theorem 1.** If the isentropic condition is satisfied, i.e.,

\[ S[\rho_{can,0}(\tilde{\beta})] = S[\rho_0], \tag{29} \]

then the GPT holds,

\[ \alpha^{-1}D_{KL}[\rho_0\|\rho_{can,0}(\alpha)] = \tilde{\beta}^{-1}D_{KL}[\rho_0\|\rho_{can,0}(\tilde{\beta})] + \alpha^{-1}D_{KL}[\rho_{can,0}(\tilde{\beta})\|\rho_{can,0}(\alpha)]. \tag{30} \]

**Proof.** The derivation is straightforward,

\[ \alpha^{-1}D_{KL}[\rho_0\|\rho_{can,0}(\alpha)] = \tilde{\beta}^{-1}D_{KL}[\rho_0\|\rho_{can,0}(\tilde{\beta})] - \alpha^{-1}D_{KL}[\rho_{can,0}(\tilde{\beta})\|\rho_{can,0}(\alpha)] \]

\[ = -\alpha^{-1}S[\rho_0] + \tilde{\beta}^{-1}S[\rho_0] + F_0(\tilde{\beta}) \]

\[ + \alpha^{-1}S[\rho_{can,0}(\tilde{\beta})] - \langle H_0\rho_{can,0}(\tilde{\beta}) \rangle. \tag{31} \]

Substituting the following relation into equation (31),

\[ F_0(\tilde{\beta}) - \langle H_0\rho_{can,0}(\tilde{\beta}) \rangle = -\tilde{\beta}^{-1}S[\rho_{can,0}(\tilde{\beta})], \tag{32} \]

we obtain

\[ \alpha^{-1}D_{KL}[\rho_0\|\rho_{can,0}(\alpha)] = \tilde{\beta}^{-1}D_{KL}[\rho_0\|\rho_{can,0}(\tilde{\beta})] - \alpha^{-1}D_{KL}[\rho_{can,0}(\tilde{\beta})\|\rho_{can,0}(\alpha)] \]

\[ = -\alpha^{-1}S[\rho_0] + \tilde{\beta}^{-1}S[\rho_0] + \alpha^{-1}S[\rho_{can,0}(\tilde{\beta})] - \tilde{\beta}^{-1}S[\rho_{can,0}(\tilde{\beta})] \]

\[ = (\tilde{\beta}^{-1} - \alpha^{-1})(S[\rho_0] - S[\rho_{can,0}(\tilde{\beta})]) = 0, \tag{33} \]

where we used the isentropic condition (29) in the last line.

From equation (30) and the non-negativity of the scaled KL divergence, we immediately obtain the following inequality,
Since equation (26) is valid for any $\alpha$, the right-hand-side of equation (34) provides the greatest lower bound of $W$ consistent with the isentropic condition (29); hence, its negative gives the maximum extractable work from the system.

The geometric image of the GPT based on scaled KL divergences is illustrated in figure 2. The set of canonical distributions is drawn as a line parametrized by the temperature. The (geodesic) line connecting $\rho_0$ and $\rho_{\text{can},0}(\beta)$ on the isentropic surface is orthogonal to the parametric line of canonical distributions in terms of the scaled KL divergence.

Comparing figure 2 and B1 in appendix B, we can see that the bare KL divergences and the energy-scaled KL divergences give us different geometrical structures. The validity condition of the GPT based on the bare KL divergence is an isoenergetic condition as illustrated in figure B1 in appendix B. The Riemann metric tensor corresponding to the bare KL divergence is the Fisher information matrix. The orthogonality is based on the dual relation between the inverse temperature and the internal energy in the Fisher information matrix. On the other hand, the validity condition of the GPT based on the energy-scaled KL divergence is an isentropic condition as illustrated in figure 2. The Riemann metric tensor corresponding to the energy-scaled KL divergence is an energy-scaled Fisher information matrix. The orthogonality is based on the dual relation between the temperature and the GS entropy in the energy-scaled Fisher information matrix. This important difference has not been recognized, since many researchers mainly studied isothermal processes in which the temperature is just a constant. In our approach the temperature is a variable parameter.

4. Reconsideration for a non-cyclic process

The maximum work formulation of the generalized second law for a non-cyclic operation ($\Delta F(\alpha) = 0$) may also be obtained by the GPT based on the scaled KL divergences. The work inequality is written as equation (19),

$$W \geq \Delta F(\alpha) = -\alpha^{-1}D_{KL}[^0\rho_0 \parallel ^0\rho_{\text{can},0}(\alpha)]$$

When $\beta$ is the inverse effective temperature determined by the isentropic condition (22), i.e.,

$$S[^0\rho_{\text{can},\gamma}(\beta)] = S[^0\rho_0],$$

the following inequality holds for any $\alpha$,

$$\Delta F(\beta) = \beta^{-1}D_{KL}[^0\rho_0 \parallel ^0\rho_{\text{can},0}(\beta)] \geq \Delta F(\alpha) = -\alpha^{-1}D_{KL}[^0\rho_0 \parallel ^0\rho_{\text{can},0}(\alpha)]$$

which means that the inverse effective temperature maximizes the right-hand-side of equation (19).

We can prove equation (35) as follows. First, we rewrite the difference between the two sides of equation (35) in terms of scaled KL divergences as,

$$\Delta F(\beta) = \beta^{-1}D_{KL}[^0\rho_0 \parallel ^0\rho_{\text{can},0}(\beta)] - \Delta F(\alpha) + \alpha^{-1}D_{KL}[^0\rho_0 \parallel ^0\rho_{\text{can},0}(\alpha)]$$

$$= -\beta^{-1}D_{KL}[^0\rho_0 \parallel ^0\rho_{\text{can},\gamma}(\beta)] + \alpha^{-1}D_{KL}[^0\rho_0 \parallel ^0\rho_{\text{can},\gamma}(\alpha)].$$

Then the next theorem holds:
Theorem 2. When the isentropic condition (22) is given, the GPT based on the scaled KL divergences holds as
\[
\alpha^{-1}D_{KL}[\rho_0 \| \rho_{\text{can},T}(\alpha)] = \tilde{\beta}^{-1}D_{KL}[\rho_0 \| \rho_{\text{can},T}(\tilde{\beta})] + \alpha^{-1}D_{KL}[\rho_{\text{can},T} (\tilde{\beta}) \| \rho_{\text{can},T}(\alpha)].
\] (37)

Proof. equation (37) is derived straightforwardly,
\[
\alpha^{-1}D_{KL}[\rho_0 \| \rho_{\text{can},T}(\alpha)] = \tilde{\beta}^{-1}D_{KL}[\rho_0 \| \rho_{\text{can},T}(\tilde{\beta})] - \alpha^{-1}D_{KL}[\rho_{\text{can},T}(\tilde{\beta}) \| \rho_{\text{can},T}(\alpha)]
\]
\[
= -\alpha^{-1}S[\rho_0] + \tilde{\beta}^{-1}S[\rho_0] + \alpha^{-1}S[\rho_{\text{can},T}(\tilde{\beta})] - \tilde{\beta}^{-1}S[\rho_{\text{can},T}(\tilde{\beta})]
\]
\[
= (\tilde{\beta}^{-1} - \alpha^{-1})(S[\rho_0] - S[\rho_{\text{can},T}(\tilde{\beta})]) = 0,
\] (38)
where we used the relation,
\[
F_T(\tilde{\beta}) - \langle H_T \rangle[\rho_{\text{can},T}(\tilde{\beta})] = -\tilde{\beta}^{-1}S[\rho_{\text{can},T}(\tilde{\beta})],
\] (39)
and the isentropic condition equation (22). \qed

The GPT (37) can be rearranged to yield the right-hand-side of equation (36),
\[
-\tilde{\beta}^{-1}D_{KL}[\rho_0 \| \rho_{\text{can},T}(\tilde{\beta})] + \alpha^{-1}D_{KL}[\rho_0 \| \rho_{\text{can},T}(\alpha)]
\]
\[
= \alpha^{-1}D_{KL}[\rho_{\text{can},T}(\tilde{\beta}) \| \rho_{\text{can},T}(\alpha)] \geq 0,
\] (40)
guaranteeing the non-negativity of the left-hand-side of equation (36) and the validity of equation (35).

A simple example that illustrates the GPT is a thermally isolated ideal gas which is confined to a container with a movable wall. Suppose that initially the volume of gas is $V_i$ and the temperature is $\beta_i^{-1}$. The wall is then adiabatically moved, expanding the volume to $V_f$ as work is extracted from the gas. This is a non-cyclic process.

The GPT for scaled KL divergences in this case is given by equation (37). We can describe the initial state and the final canonical state using a Maxwellian velocity distribution $\rho_{\text{Max}}$ as $\rho_0 = \rho_{\text{Max}}(\beta_i, V_i)$ and $\rho_{\text{can},T}(\alpha) = \rho_{\text{Max}}(\alpha, V_f)$, where
\[
\rho_{\text{Max}}(\alpha, V) = \frac{1}{Z(\alpha, V)} \sum_{p=1}^{3N} \exp\left(-\frac{\alpha p^2}{2m}\right)^{N/2}
\] (41)
where $N$ is the number of particles and $p$, and $m$ are momentum and mass of a particle, respectively. The indicator function $\chi_V$ serves to indicate the configuration space dependence of the distribution and $Z$ is the partition function,
\[
Z(\alpha, V) = \frac{V^N}{N!} \left(\frac{2\pi m}{\hbar^2 \alpha}\right)^{3N/2}.
\] (42)

The GPT is then written as
\[
\alpha^{-1}D_{KL}[\rho_{\text{Max}}(\beta_i, V_i) \| \rho_{\text{Max}}(\alpha, V_f)] = \tilde{\beta}^{-1}D_{KL}[\rho_{\text{Max}}(\beta_i, V_i) \| \rho_{\text{Max}}(\tilde{\beta}, V_f)]
\]
\[
+ \alpha^{-1}D_{KL}[\rho_{\text{Max}}(\tilde{\beta}, V_f) \| \rho_{\text{Max}}(\alpha, V_f)],
\] (43)
which holds for the effective temperature $\tilde{\beta}$ determined by the following condition,
\[
\tilde{\beta} = \left(\frac{V_f}{V_i}\right)^{2/3} \beta_i.
\] (44)

The proof is straightforward using equations (41) and (42) as
\[
\alpha^{-1}D_{KL}[\rho_{\text{Max}}(\beta_i, V_i) \| \rho_{\text{Max}}(\alpha, V_f)]
\]
\[
- \tilde{\beta}^{-1}D_{KL}[\rho_{\text{Max}}(\beta_i, V_i) \| \rho_{\text{Max}}(\tilde{\beta}, V_f)] - \alpha^{-1}D_{KL}[\rho_{\text{Max}}(\tilde{\beta}, V_f) \| \rho_{\text{Max}}(\alpha, V_f)]
\]
\[
= \alpha^{-1}\log\left(\frac{Z(\alpha, V_f)}{Z(\beta_i, V_i)}\right) - \frac{3}{2}N(\alpha^{-1} - \beta_i^{-1}) - \tilde{\beta}^{-1}\log\left(\frac{Z(\tilde{\beta}, V_f)}{Z(\beta_i, V_i)}\right)
\]
\[
+ \frac{3}{2}N(\tilde{\beta}^{-1} - \beta_i^{-1}) - \alpha^{-1}\log\left(\frac{Z(\alpha, V_f)}{Z(\tilde{\beta}, V_f)}\right) - \tilde{\beta}^{-1}\log\left(\frac{Z(\tilde{\beta}, V_f)}{Z(\tilde{\beta}, V_f)}\right)
\]
\[
= \left\{\log\left(\frac{V_f}{V_i}\right)^N - \log\left(\frac{\tilde{\beta}}{\beta_i}\right)^{3N/2}\right\}(\alpha^{-1} - \tilde{\beta}^{-1}),
\] (45)

Equation (43) guarantees that the final state of a Maxwellian velocity distribution with the effective temperature $\tilde{\beta}^{-1}$ gives the maximum work as discussed in this section. The condition of effective temperature...
We consider a thermally isolated two-level quantum system as a simple example. A spin-1 state to extract work that satisfies dissipationless non-quasi-static processes. Time-dependent two-level quantum systems often appear in the context of matter--nonequilibrium statistical mechanics to simple, few-degrees-of-freedom quantum systems is a current topic of much interest. The application of nonequilibrium statistical mechanics to simple, few-degrees-of-freedom quantum systems is a current topic of much interest. We apply the GPT to an optimization problem in nonequilibrium statistical mechanics. From the generalized formulation, the maximum work is extracted by an operation under which the initial state does not change and the Hamiltonian changes. The solutions for the Hamiltonian and the effective temperature. The effectively optimal state may be the ideally optimal state in a set of realizable final states. The ‘closest’ state is the experimentally optimal state to extract work that satisfies the orthogonality condition.

### 5. Application to a two-level quantum system

We apply the GPT to an optimization problem in nonequilibrium statistical mechanics. From the generalized maximum work formulation, the maximum work is extracted by an operation under which the final state is the canonical state with the final Hamiltonian and the effective temperature. The final canonical state with the effective temperature is the ideally optimal state to extract the maximum work. The ideally optimal state may be difficult to realize experimentally. It is important to figure out what is the ‘closest’ (minimum divergence) state from the ideally optimal state in a set of realizable final states. The ‘closest’ state is the experimentally optimal state to extract work that satisfies the orthogonality condition.

#### 5.1. A spin-1/2 particle operated by a magnetic field

We consider a thermally isolated two-level quantum system as a simple example. The application of nonequilibrium statistical mechanics to simple, few-degrees-of-freedom quantum systems is a current topic of much interest. Time-dependent two-level quantum systems often appear in the context of matter--nonequilibrium statistical mechanics. We consider a spin-1/2 particle embedded in a magnetic field that can be controlled. The Hamiltonian of the system is

\[
H(t) = \frac{\hbar \omega}{2} (\cos \phi_t \sigma_x + \sin \phi_t \sigma_z),
\]

where \(\sigma_i (i = x, y, z)\) are the standard Pauli matrices and we choose the direction of the magnetic field restricted to the \(x - z\) plane with \(\phi_t\) the rotation-angle of the magnetic field around the \(y\)-axis at time \(t\). The Hamiltonian can be rewritten by using rotation operators in spin space,

\[
H(t) = \frac{\hbar \omega}{2} e^{-i \gamma \phi_t/2} \sigma_z e^{i \gamma \phi_t/2}.
\]

The eigenvalues of the Hamiltonian are \(E_0 = -\hbar \omega/2\) and \(E_1 = \hbar \omega/2\), which we identify as the ground and excited state energies, respectively.

The spin-1/2 particle is externally operated from \(t = 0\) to \(t = T\). The operation is represented by the time-dependent angle in the Hamiltonian. We choose a cyclic operation, \(\phi_0 = 0\) at \(t = 0\) and \(\phi_T = 2 \pi\) at \(t = T\). The angle increases from 0 to \(2\pi\) monotonically and \(\phi_t = \pi\) at \(t = \tau\). The Hamiltonian is proportional to the Pauli matrix \(\sigma_z\) at \(t = 0\). It is proportional to \(-\sigma_z\) at \(t = \tau\). When \(t = T\) the Hamiltonian returns to its original form at \(t = 0\).

We briefly discuss how to extract the maximum work from a nonequilibrium initial state such as a pure excited state. The maximum work can be obtained by the cyclic operation including two processes: (1) The short stabilization process for \(t \in (0, \tau)\) in which the pure excited state becomes the ground state by changing \(H_0\) to \(H_1\). (2) The restoration process for \(t \in (\tau, T)\) to the original Hamiltonian, \(H_T = H_0\), without any transition to the excited state. These processes are analogous to those that have been discussed on how to extract the maximum work from a nonequilibrium initial state in a thermodynamic system. The first corresponds to the stabilization process to prevent spontaneous relaxation in a thermodynamic system. For simplicity, we consider here the sudden process, \(\tau \to 0\). The second is generally a quasi-static (isentropic) process without dissipation. After the sudden process, the initial state does not change and the Hamiltonian changes as \(H_{T+} = -H_0\). The control parameter of our operation is the rate at which the field is rotated. We will find the optimal angular frequency to extract work. The solution for finite frequencies correspond to dissipationless non-quasi-static processes.

The Schrödinger equation is

\[
i\hbar \frac{\partial}{\partial t} |\psi_t\rangle = H(t)|\psi_t\rangle = \frac{\hbar \omega}{2} e^{-i \gamma \phi_t/2} \sigma_z e^{i \gamma \phi_t/2} |\psi_t\rangle.
\]

We choose a simple linear time dependence for the rotation angle,

\[
\phi_t = \pi + \Omega t\quad \text{for} \quad 0 < t \leq T
\]

where \(\Omega = \pi/T\) is the adjustable angular frequency. Then, the Schrödinger equation can be easily solved by using the rotating amplitude \(e^{i \gamma \phi_t/2} |\psi_t\rangle\). The solution is written as,
\[ |\psi_i \rangle = u^* |\psi_0 \rangle \] (50)

where the time evolution operator \( u^t \) is given as
\[ u^t = e^{-i(\pi + \Omega t)\sigma_3/2} e^{-i\omega_0 t - i\Omega t\eta} e^{i\omega_0 t}/2, \] (51)

5.2. A vector representation of a quantum state

In this subsection, we introduce a vector representation of a quantum state. This subsection prepares for the optimization of the rotation-angle frequency discussed in the next subsection. A quantum state is represented as a point vector in parameter space in information geometry. The orthogonality condition from the GPT is written in terms of these vectors [22, 23].

A quantum state is written as a 2 \( \times \) 2 density matrix that can be expanded as a linear combination of four Hermitian matrixes: the unit matrix \( I \) and the three Pauli matrices \( \sigma_i (i = x, y, z) \). Any state except pure states is written as a positive definite density matrix,
\[ \rho = \frac{1}{2} I - \frac{1}{2} \tilde{\eta} \cdot \sigma \tanh|\eta|, \] (52)

where \( \eta \) is a three dimensional real vector, \( \tilde{\eta} = \eta/|\eta| \) is its unit vector, and \( \sigma = (\sigma_x, \sigma_y, \sigma_z) \). The condition of total probability is obvious, \( \text{Tr}[\rho] = 1 \). The positive definiteness is guaranteed from the range, \( |\tanh|\eta|| < 1 \), since the eigenvalues of \( \rho \) are \((1 \pm \tanh|\eta|)/2 \). A density matrix corresponding to a pure state is positive semi-definite. We consider the pure state in the limit of \( |\tanh|\eta|| \to 1 \) for \( |\eta| \to \infty \). The pure ground state of \( H_0 \) is \( \rho = (I - \sigma_z)/2 \), the initial energy \( E_0 = \text{Tr}[\rho_0 H_0] \) is positive, so that work is extractable. We chose the \( x \) component \( \tilde{\eta}_0 = 0 \) for simplicity. We may be able to adjust \( \tilde{\eta}_0 \) by an additional operation around the \( y \)-axis.

From the solution of the Schrödinger equation, the time evolution of the density matrix is given,
\[ \rho_t = u^t \rho_0 u^{-t} = \frac{1}{2} I - \frac{1}{2} \tilde{\eta}_t \cdot \sigma \tanh|\eta_t|, \] (57)

where
\[ |\eta_t| = \sqrt{\frac{1}{2} \text{Tr}(u^t(\eta_0 \cdot \sigma)u^{-t})^2} = |\eta_0|. \] (59)

The final state at time \( t = T \) is written as
\[ \rho_T = \frac{1}{2} I - \frac{1}{2} \tilde{\eta}_T \cdot \sigma \tanh|\eta_T|. \] (60)

We used the orthogonality of Pauli matrices, \( \text{Tr}[\sigma_i \sigma_j] = 2\delta_{i,j} \). The norm \( |\eta| \) is constant in time under the unitary time evolution,
\[ \rho_{\text{can},t} = \frac{e^{-\tilde{\beta}t \mathcal{H}}}{Z_{\text{can}}} = \frac{e^{-\eta_{\text{can}}/\sigma}}{Z_{\text{can}}} \]  
(61)

where \( \eta_{\text{can}} = \text{Tr}[\sigma \tilde{\beta} \mathcal{H}_j]/2 \) and

\[ Z_{\text{can}} = \text{Tr}[e^{-\tilde{\beta}t \mathcal{H}}] = 2 \cosh \left( \frac{\tilde{\beta} \hbar \omega}{2} \right) \]  
(62)

The partition function \( Z_{\text{can}} \) is constant in time, since the trace is preserved under the rotation operation. The initial canonical state before the sudden process is the same as the final canonical state because of the cyclic operation. The initial/final state is written in E-representation as

\[ \rho_{\text{can},0/T} = \frac{e^{-\eta_{\text{can},0}/\sigma}}{Z_{\text{can}}} \]  
(63)

where \( 0/T \) means 0 or \( T \) and \( \eta_{\text{can},0/T} = \tilde{\beta} \hbar \omega/2 \) and \( \eta_{\text{can},0/T} = (0, 0, 1) \). Similarly, the initial/final canonical state is written in M-representation as

\[ \rho_{\text{can},0/T} = \frac{1}{2} I - \frac{1}{2} \eta_{\text{can},0/T} \cdot \sigma \tanh \left( \frac{\tilde{\beta} \hbar \omega}{2} \right) \]  
(64)

We note that the final canonical state does not depend on the controllable time period \( T \), since \( \phi_{\text{T}} = 2\pi \) for any \( T \).

We briefly discuss the isentropic condition to determine the effective temperature, \( S[\rho_{\text{can},T}(\tilde{\beta})] = S[\rho_0] \).

The entropy of the initial state is calculated by using both E- and M-representations,

\[ S[\rho_0] = -\text{Tr}[\rho_0 \log \rho_0] \]  
(65)

\[ = -\text{Tr}[\left( \frac{1}{2} I - \frac{1}{2} \eta_0 \cdot \sigma \tanh(\eta_0 I - \eta_0 \cdot \sigma) \right) (-\log(2 \cosh|\eta_0| I - \eta_0 \cdot \sigma))] \]  
(66)

\[ = \log(2 \cosh|\eta_0|) - |\eta_0| \tanh|\eta_0|. \]  
(67)

Similarly, the entropy of the final canonical state is

\[ S[\rho_{\text{can},T}(\tilde{\beta})] = \log(2 \cosh|\eta_{\text{can},T}|) - |\eta_{\text{can},T}| \tanh|\eta_{\text{can},T}| \]  
(68)

where \( |\eta_{\text{can},T}| = \tilde{\beta} \hbar \omega/2 \). The isentropic condition is simply rewritten as \( |\eta_{\text{can},T}| = |\eta_0| \). The effective temperature is determined from the initial condition,

\[ \frac{\tilde{\beta} \hbar \omega}{2} = |\eta_0|. \]  
(69)

**5.3. Optimization of the rotation rate**

We consider the quantum mechanical version of the work identity equation \( (18) \) for the effective temperature \( \tilde{\beta} \),

\[ W = \tilde{\beta}^{-1} D_{KL}[\rho_T \parallel \rho_{\text{can},T}(\tilde{\beta})] - \tilde{\beta}^{-1} D_{KL}[\rho_0 \parallel \rho_{\text{can},0}(\tilde{\beta})] \]  
(70)

where we used \( \Delta F(\tilde{\beta}) = 0 \) for a cyclic operation. The quantum relative entropy is defined as the following quantum-mechanical analog of the KL divergence,

\[ D_{KL}[\rho_A \parallel \rho_B] = \langle \log \rho_A | \rho_A \rangle - \langle \log \rho_B | \rho_A \rangle \]  
(71)

where the bra-ket notation is defined as

\[ \langle A|B \rangle = \text{Tr}[AB]. \]  
(72)

We used the same notation \( D_{KL}[\rho_A \parallel \rho_B] \) for the quantum relative entropy, since equation \( (71) \) is the same as equation \( (13) \).

The extractable work is the negative of the work done on the system,

\[ -W = W_{\text{max}} - \tilde{\beta}^{-1} D_{KL}[\rho_T \parallel \rho_{\text{can},T}(\tilde{\beta})] \]  
(73)

where the maximum extractable work is defined as

\[ W_{\text{max}} = \tilde{\beta}^{-1} D_{KL}[\rho_0 \parallel \rho_{\text{can},0}(\tilde{\beta})] \geq 0. \]  
(74)

Now we try to find the optimal operation with the angular frequency \( \Omega^* = \pi/T^* \) for which the final state satisfies the minimum scaled KL divergence,
\[ T^* = \arg \min_T \beta^{-1} D_{KL}[\rho_T \| \rho_{\text{can}, \beta}(\beta)]. \] (75)

where we replaced the final canonical state with the initial canonical state to make clear the absence of \( T \) dependence of the final canonical state, since \( H_T = H_0 \) for any \( T \).

First we consider the validity condition of the GPT \( \Delta = 0^8 \) where

\[
\Delta = \beta^{-1} D_{KL}[\rho_T \| \rho_{\text{can}, \beta}(\beta)] - \beta^{-1} D_{KL}[\rho_T \| \rho_{T^*}] \\
- \beta^{-1} D_{KL}[\rho_{T^*} \| \rho_{\text{can}, \beta}(\beta)].
\] (76)

The KL divergence may be divided into the sum of the negative entropy and the cross entropy,

\[ D_{KL}[\rho_A \| \rho_B] = -S[\rho_A] + S_C[\rho_A, \rho_B], \] (77)

where the cross entropy is

\[ S_C[\rho_A, \rho_B] = -\text{Tr}[\rho_A \log \rho_B]. \] (78)

After substituting the above sum into \( \Delta \),

\[
\Delta = \beta^{-1} S_C[\rho_T, \rho_{\text{can}, \beta}(\beta)] - \beta^{-1} S_C[\rho_T, \rho_{T^*}] \\
+ \beta^{-1} S_C[\rho_{T^*}, \rho_{T^*}] - \beta^{-1} S_C[\rho_{T^*}, \rho_{\text{can}, \beta}(\beta)].
\] (79)

where we used \( S[\rho_{T^*}] = S_C[\rho_{T^*}, \rho_{T^*}]. \)

Similar to the entropy, the cross entropy between state A and state B is calculated by using both E- and M-representations,

\[ S[\rho_A, \rho_B] = -\text{Tr}[\rho_A \log \rho_B] \] (80)

\[
= \log(2 \cosh(\frac{\beta h\omega}{2})) - \frac{\beta h\omega}{2} \tanh(\frac{\beta h\omega}{2}) \hat{\eta}_A \cdot \hat{\eta}_B
\] (81)

where we chose \( |\eta_A| = |\eta_B| = \beta h\omega/2. \) After substituting the above expression of the cross entropy, we obtain

\[ \Delta = \frac{h\omega}{2} \tanh(\frac{\beta h\omega}{2})(\hat{\eta}_{T^*} - \hat{\eta}_T) \cdot (\hat{\eta}_{\text{can},0} - \hat{\eta}_{T^*}). \] (82)

Quantum states are represented as vectors such as \( \eta_T \). The validity condition of the GPT is written as the orthogonality condition \( (\hat{\eta}_{T^*} - \hat{\eta}_T) \cdot (\hat{\eta}_{\text{can},0} - \hat{\eta}_{T^*}) = 0 \) in the vector space. In information geometry, a state is represented as a vector in a space of parameters. The ‘distance’ (square of length in the sense of the Pythagorean theorem) between two states is measured through the divergence. In nonequilibrium statistical mechanics, the energy-dimensional ‘distance’ between two states is measured through the scaled KL divergence.

Suppose \( \hat{\eta}_{T^*} = \hat{\eta}_{\text{can},0} \), then the GPT is valid globally. The vector representation of the ideally optimal state to extract the maximum work is \( \hat{\eta}_{\text{can},0} \). However, if \( \hat{\eta}_{\beta,0} \neq 0 \), then \( \hat{\eta}_T \) cannot be the ideally optimal state \( \hat{\eta}_{\text{can},0} \) for any \( T \) in general, since the rotation is limited around the \( y \)-axis. Although any realizable final states cannot be the ideally optimal state, we can figure out what is the ‘closest’ (minimum divergence) state from the ideally optimal state in a set of realizable final states. The ‘closest’ state is the experimentally optimal state to extract work and satisfies the local orthogonal condition (See figure 3).

The validity condition of the local GPT becomes the following local orthogonal condition between the tangent vector at \( \rho_{T^*} \) and the geodesic line connecting \( \rho_{\text{can}, \beta}(\beta) \) and \( \rho_{T^*} \),

\[
\lim_{T \to T^*} \frac{\Delta}{T - T^*} \propto \frac{d\hat{\eta}_{T^*}}{dT} \cdot (\hat{\eta}_{\text{can},0} - \hat{\eta}_{T^*}) = \frac{d\hat{\eta}_{T^*}}{dT} \cdot \hat{\eta}_{\text{can},0} = \frac{d\hat{\eta}_{T^*}}{dT} \cdot \hat{\eta}_{z,T^*} = 0
\] (83)

where we used \( d(\hat{\eta}_T \cdot \hat{\eta}_T)/dT = d1/dT = 0 \) and \( \hat{\eta}_{\text{can},0} = (0, 0, 1) \). \( \hat{\eta}_{T,z} \) is written as

\[ \hat{\eta}_{T,z} = \frac{1}{2} \text{Tr}[\sigma^z (\sigma^z \hat{\eta}_0 \cdot \sigma^z \hat{\eta}_0^* - \sigma^z \hat{\eta}_0 \cdot \sigma^z \hat{\eta}_0^*)] \] (84)

where

\[ u^T = e^{-i\gamma} e^{-i\omega T \sigma_z - i\pi \gamma}/\sqrt{e^{i\pi \gamma}/2}. \]

The calculation of \( d\hat{\eta}_{T,z}/dT \) is straightforward but tedious. We show only the final result,

\[ \hat{\eta}_{T,z} = [1 - \cos(\sqrt{\omega^2 T^2 + \pi^2})] \frac{\pi}{\omega^2 T^2 + \pi^2} (-\omega T \hat{\eta}_{\beta,0} + \pi \hat{\eta}_{\text{can},z}) - \hat{\eta}_{0,z} \] (86)

\(^8\) The temperature has already been fixed as the effective temperature so the scale factor is common for three divergences in equation (76). Thus, the validity condition of the GPT of the scaled KL divergences is the same as that of the bare KL divergences in this section.
and

\[
\frac{d\hat{\eta}_{T,z}}{dT} = \frac{2\pi\omega}{\omega^2 T^2 + \frac{\pi^2}{2}} \sin \left( \frac{\sqrt{\omega^2 T^2 + \frac{\pi^2}{2}}}{2} \right) \cos \left( \frac{\sqrt{\omega^2 T^2 + \frac{\pi^2}{2}}}{2} \right) \frac{\omega T}{\sqrt{\omega^2 T^2 + \frac{\pi^2}{2}}}(-\omega T\hat{\eta}_{0,y} + \pi\hat{\eta}_{0,z})
\]

\[
+ \sin \left( \frac{\sqrt{\omega^2 T^2 + \frac{\pi^2}{2}}}{2} \right) \left( -\frac{2\omega T}{\omega^2 T^2 + \frac{\pi^2}{2}} (-\omega T\hat{\eta}_{0,y} + \pi\hat{\eta}_{0,z}) - \hat{\eta}_{0,y} \right) \right) \right) \right) \right)
\]

\[ (87) \]

There are clearly two ways of solving \(d\hat{\eta}_{T,z}/dT = 0\) that yield two sets of solutions for \(T^*\). One way is when \(\sin(\sqrt{\omega^2 T^2 + \frac{\pi^2}{2}}/2) = 0\) so that \(\omega T^* = \sqrt{4n^2 - 1}\pi (n = 1, 2, ...). For these \(T^*\), the scaled KL divergence \(\beta^{-1}D_{KL}[\rho_{T^*}||\rho_{can,0}(\beta)]\) takes the constant value since \(\hat{\eta}_{T,z} = -\hat{\eta}_{0,z}\) and the divergence only depends on \(\hat{\eta}_{T,z}\), i.e.,

\[
\beta^{-1}D_{KL}[\rho_{T^*}||\rho_{can,0}(\beta)] = \frac{\hbar\omega}{2} \tanh \left( \frac{\beta \hbar\omega}{2} \right) (1 - \hat{\eta}_{T,z})
\]

\[
= \frac{\hbar\omega}{2} \tanh \left( \frac{\beta \hbar\omega}{2} \right) (1 + \hat{\eta}_{0,z})
\]

\[ (88) \]

This value also corresponds to the quasi-static process of \(T \to \infty (\Omega \to 0)\) as illustrated in figures 4 and 5.

Another set of solutions is obtained when the terms inside the curly brackets in equation (87) sum to zero. (In this case a simple expression for \(T^*\) is not available.) The scaled KL divergences \(\beta^{-1}D_{KL}[\rho_{T^*}||\rho_{can,0}(\beta)]\) at these \(T^*\) become local maximum or local minimum according to the initial conditions as illustrated in figures 4 and 5. For \(\varphi = 0\) in \(\hat{\eta}_{0}\) (see figure 4), the divergences of the second series take local maximum. However, for \(\varphi = \pi/3\) (see figure 5), the divergence takes the globally minimum, 0.024792... at \(\Omega^* = \pi/T^* = 0.358859...\) that is one of the solutions of the second series. This interesting result tells us that we can extract more work than the quasi-static process by choosing the optimal parameter \(\Omega^*\) in the case of \(\hat{\eta}_{0,y} \neq 0\). The local GPT or the generalized projection theorem gives us candidates of the realizable optimal state that is ‘closest’ to the ideal optimal state.

6. Concluding remarks

An alternative information-geometric foundation of the generalized second law has been presented. We introduced ‘thermodynamic distance’ as an energy-dimensional KL divergence scaled by temperature. The change of ‘thermodynamic distance’ gives us the dissipative work. The GPT based on three ‘thermodynamic distances’ gives us a geometric interpretation of the maximum work formulation of the generalized second law.

9 For \(\varphi = 0\) the state is the ground state after the sudden change at \(t = 0\). The former set of operations including the quasi-static operation keeps the ground state. The maximum work is realized for the former set of operations. For \(\varphi = \pi/3\), the state has higher energy than the ground state after the sudden change at \(t = 0\). The former set of operations including the quasi-static operation keep this higher energy. On the other hand some quick operations in the latter set induce transitions to make the energy of the state lower. As a result, we can extract more work than the quasi-static process by the optimal operation in the latter set.
From the nonequilibrium initial state, the canonical state with the effective temperature is the closest ‘point’ in ‘thermodynamic distance.’ Since our new concept of ‘thermodynamic distance’ has an important role in nonequilibrium statistical mechanics, it can also be called as thermodynamic divergence.

One important recognition in this paper is that the geometric structure based on the KL divergence scaled by a variable temperature is different from that based on the bare KL divergence. This geometric difference does not play any role for a constant temperature in an isothermal process. For a variable temperature, the validity condition of the GPT based on the energy-scaled KL divergences is an isentropic condition. On the other hand, the validity condition of the GPT based on the bare KL divergences is an isoenergetic condition.

The geometrical interpretation of the generalized maximum work formulation gives us a systematic method to figure out a protocol to realize the optimal operation. In this paper we applied the geometrical interpretation of the generalized maximum work formulation to a simple two-level quantum system. The optimal cyclic operation to extract work from a nonequilibrium state was determined by minimizing the scaled KL divergence between the final state and the final canonical state.

In this paper we discussed only adiabatic processes in thermally isolated Hamiltonian systems in order to highlight the information-geometrical foundation of the thermodynamics. Although we excluded any phenomenological assumptions such as the existence of a heat reservoir that would produce a canonical ensemble, we would need high and low temperature heat reservoirs to consider a heat engine. Since the information geometry is flexibly applicable for any parametrized probability distributions, the scaled KL divergence would play important roles in the heat engine. One may consider scaling of the KL divergence by a quantity with power dimension, such as temperature divided by time. Then a GPT based on this new scaled KL divergence would give us important knowledge for optimizing the power. We will discuss this problem in the near future.
Acknowledgments

TN deeply appreciates A and R Nakamura for their great helps. HHH would like to thank N Hasegawa, L Reichl and T Washio for their supports. TN and HHH are grateful to S Amari for his suggestions and encouragement. DJD appreciates the suggestions of E Karpov, V Basios and J Lutsko. This work was performed under the Cooperative Research Program of ‘Network Joint Research Center for Materials and Devices.’ This work was also supported by Hitachi Automotive Systems, Ltd.

Appendix A. Concavity of $\mathcal{W}_{LB}$ with respect to the temperature

To determine the effective inverse temperature we use the fact that $\mathcal{W}_{LB}(\alpha)$ has one global maximum as a function of $\alpha$. This is most easily seen from the concavity of $\mathcal{W}_{LB}$ as a function of the temperature, $\alpha^{-1}$. To show this we use the facts (known from thermodynamics and also holding for a canonical distribution) that the derivative of the Helmholtz free energy with respect to temperature is minus the entropy and the derivative of the entropy with respect to temperature is positive.

We have,

$$\frac{\partial}{\partial \alpha^{-1}} \mathcal{W}_{LB}(\alpha) = -S[\rho_{\text{can},T}(\alpha)] + S[\rho_0].$$  \hspace{1cm} (A.1)

Then,

$$\frac{\partial^2}{(\partial \alpha^{-1})^2} \mathcal{W}_{LB}(\alpha) = - \frac{\partial}{\partial \alpha^{-1}} S[\rho_{\text{can},T}(\alpha)].$$  \hspace{1cm} (A.2)

The derivative of the entropy with respect to temperature is proportional to the heat capacity or equivalently to the variance of the energy, which is positive. Thus we have

$$\frac{\partial^2}{(\partial \alpha^{-1})^2} \mathcal{W}_{LB}(\alpha) < 0,$$  \hspace{1cm} (A.3)

which means that $\mathcal{W}_{LB}(\alpha)$ is a concave function of $\alpha^{-1}$.

Suppose $\frac{\partial}{\partial \alpha^{-1}} \mathcal{W}_{LB}(\alpha) = 0$ at $\alpha^{-1} = \beta^{-1}$. The concavity of $\mathcal{W}_{LB}(\alpha)$ tells us that $\frac{\partial}{\partial \alpha^{-1}} \mathcal{W}_{LB}(\alpha)$ is a strictly monotonic function of $\alpha^{-1}$ so that $\frac{\partial}{\partial \alpha^{-1}} \mathcal{W}_{LB}(\alpha)$ cannot be zero for any other value of $\alpha^{-1}$ than $\beta^{-1}$. Since $\alpha^{-1}$ and $\alpha$ are in one-to-one correspondence, $\mathcal{W}_{LB}(\alpha)$ has then only one global maximum as a function of $\alpha$; the one at $\alpha = \beta$.

Appendix B. The generalized Pythagorean theorem based on the KL divergence

The importance of the GPT based on the scaled KL divergence is shown in sections 3 and 4. Here, we will describe the GPT based on the bare (non-scaled) KL divergence for comparison. The geometric structure based on the bare KL divergence is different from the geometric structure based on the scaled KL divergence. We also derive the well-known principle of maximum entropy. While its proof usually just uses the non-negativity of the KL divergence \[21, 25\], it can also be derived from the GPT \[23\]. In this Appendix, for brevity we suppress the subscripts representing time.

Using bare KL divergences gives us the following theorem;

Theorem 3. Under the isoenergetic condition, i.e.,

$$\langle H|\rho\rangle = \langle H|\rho_{\text{can}}(\beta)\rangle,$$  \hspace{1cm} (B.1)

the GPT based on the KL divergence holds,

$$D_{KL}[\rho]\parallel \rho_{\text{can}}(\alpha) = D_{KL}[\rho]\parallel \rho_{\text{can}}(\beta) + D_{KL} [\rho_{\text{can}}(\beta)\parallel \rho_{\text{can}}(\alpha)].$$  \hspace{1cm} (B.2)

Proof. This theorem is easy to confirm as,

$$D_{KL}[\rho]\parallel \rho_{\text{can}}(\alpha) = D_{KL}[\rho]\parallel \rho_{\text{can}}(\beta) - D_{KL} [\rho_{\text{can}}(\beta)\parallel \rho_{\text{can}}(\alpha)]$$

$$\equiv -S[\rho] - \alpha (F(\alpha) - \langle H|\rho\rangle) - \langle H|\rho_{\text{can}}(\beta)\rangle - \beta (F(\beta) - \langle H|\rho_{\text{can}}(\beta)\rangle)$$

$$\equiv \alpha \langle H|\rho\rangle - \beta \langle H|\rho_{\text{can}}(\beta)\rangle - \alpha (\langle H|\rho_{\text{can}}(\beta)\rangle + \beta \langle H|\rho_{\text{can}}(\beta)\rangle)$$

$$\equiv (\beta - \alpha) (\langle H|\rho_{\text{can}}(\beta)\rangle - \langle H|\rho\rangle) = 0,$$  \hspace{1cm} (B.3)

where we used the isoenergetic condition of (B.1) in the last line.

\[\square\]
The geometric image of the GPT based on KL divergences is illustrated in figure B1. The (geodesic) line connecting \( \rho \) and \( \rho_{\text{can}} (\beta) \) on the isoenergetic surface is orthogonal to the parametric line of canonical distributions in terms of the KL divergence.

By substituting the following expressions of the KL divergences into the GPT, equation (B.2),

\[
\begin{align*}
D_{\text{KL}} [\rho || \rho_{\text{can}} (\alpha)] &= - S [\rho] - \alpha (F (\alpha) - \langle H \rangle) \\
D_{\text{KL}} [\rho_{\text{can}} (\beta) || \rho_{\text{can}} (\alpha)] &= - S [\rho_{\text{can}} (\beta)] - \alpha (F (\alpha) - \langle H \rangle) \rho_{\text{can}} (\beta)
\end{align*}
\]

equation (B.2) is rewritten as

\[
\begin{align*}
-S [\rho] &= - \alpha (F (\alpha) - \langle H \rangle) \\
D_{\text{KL}} [\rho || \rho_{\text{can}} (\beta)] &= - S [\rho_{\text{can}} (\beta)] - \alpha (F (\alpha) - \langle H \rangle) \rho_{\text{can}} (\beta).
\end{align*}
\]

From the isoenergetic condition equation (B.1),

\[
-S [\rho] &= D_{\text{KL}} [\rho || \rho_{\text{can}} (\beta)] - S [\rho_{\text{can}} (\beta)].
\]

Using the non-negativity of the KL divergence in the right-hand-side, we can obtain

\[
S [\rho] \leq S [\rho_{\text{can}} (\beta)].
\]

This inequality holds for any probability distribution \( \rho \) satisfying the isoenergetic condition (B.1). The canonical distribution has the maximum entropy among isoenergetic distributions. This property is the well-known principle of maximum entropy.

ORCID iDs

T Nakamura @ https://orcid.org/0000-0001-8639-7341

References

[1] Maxwell J C 1888 Theory of Heat (Longmans: London and New York) (Reprinted by Dover Publications, New York, 2001)
[2] Leff H S and Rex A F 2003 Maxwell’s Demon 2: Entropy, Classical and Quantum Information, Computing (Institute of Physics Publishing: Bristol and Philadelphia)
[3] Maruyama K, Nori F and Vedral V 2009 Rev. Mod. Phys. 81 1–23
[4] Landauer R 1961 IBM J. Res. Dev. 5 183–91
[5] Bennett C H 1987 Sci. Am. 257 108–16
[6] Sagawa T and Ueda M 2008 Phys. Rev. Lett. 100 080403
[7] Sagawa T and Ueda M 2009 Phys. Rev. Lett. 102 250602
[8] Bérut A, Arakelyan A, Petrosyan A, Ciliberto S, Dillenschneider R and Lutz E 2012 Nature 483 187–9
[9] Parrondo J, Horowitz J and Sagawa T 2015 Nat. Phys. 11 131–9 and references therein
[10] Deffner S and Jarzynski C 2013 Phys. Rev. X 3 041003
[11] Park J, Kim K-H, Sagawa T and Kim S 2013 Phys. Rev. Lett. 111 230402
[12] Bera M, Riera A, Lewenstein M and Winter A 2017 Nat. Commun. 8 2180
[13] Ren L-H and Fan H 2017 Phys. Rev. A 96 042304
[14] Hasagawa H H, Ishikawa J, Takara K and Driebe D 2010 Phys. Lett. A 374 1001–4
[15] Takara K, Hasagawa H H and Driebe D 2010 Phys. Lett. A 375 88–92
[16] Ishikawa J, Takara K, Hasagawa H H and Driebe D 2014 Entropy 16 3471–81
[17] Hasagawa H H, Nakamura T and Driebe D 2017 Chaos 27 104606
[18] Esposito M, Lindenberg K and Van den Broeck C 2010 New J. Phys. 12 013013
[19] Esposito M and Van den Broeck C 2011 Europhys. Lett. 95 40004
[20] Nielsen M and Chuang I 2000 Quantum Computation and Quantum Information (Cambridge: Cambridge University Press)
[21] Cover T and Thomas J 2006 Elements of Information Theory (Hoboken: Wiley-Interscience)
[22] Amari S and Nagaoka H 2000 Methods of Information Geometry (Translations of Mathematical Monographs vol 191) (Providence: American Mathematical Society/Oxford: Oxford University Press) Translated from the 1993 Japanese original by Daishi Harada

[23] Amari S 2016 Information Geometry and Its Applications (Japan: Springer)

[24] Crooks G E 2007 Phys. Rev. Lett. 99 100602

[25] Pauli W 2000 Pauli Lectures on Physics. vol. 4: Statistical Mechanics (Cambridge: MIT Press) 1973 and reprinted by Dover in 2000

[26] Ribeiro W, Landi G and Semiao F 2016 Am. J. Phys. 84 948–57

[27] Batalhao T et al 2014 Phys. Rev. Lett. 113 140601

[28] Messiah A 2014 Quantum mechanics: Two volumes bound as one (New York: Dover)