From Sosuke's abstract for WOST IV, 2023, concerning this very article (PRR version):
Though this geometric decomposition is mathematically equivalent to the decomposition introduced by Maes and Netocny for the overdamped Langevin system, we reformulated this decomposition under a somewhat different concept based on the Benamou-Brenier formula in optimal transport theory [2, 5].

[1] M. Nakazato and S. Ito, Phys. Rev. Res. 3, 043093, (2021).
[2] A. Dechant, S-I Sasa and S. Ito, Phys. Rev. Res. 4, L012034 (2022). Phys. Rev. E 106, 024125 (2022).
[3] K. Yoshimura, A. Kolchinsky, A. Dechant and S. Ito, Phys. Rev. Res. 4, L012034 (2023).
[4] A. Kolchinsky, A. Dechant, K. Yoshimura and S. Ito. arXiv:2206.14599 (2022).
[5] S. Ito, Information geometry, 1-42 (2023).

Geometrical aspects of entropy production in stochastic thermodynamics based on Wasserstein distance

Muka Nakazato[1] and Sosuke Ito[1, 2, 3]

[1] *Department of Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0031, Japan*
[2] *Universal Biology Institute, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[3] *JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan*

(Dated: March 5, 2021)

We study a relationship between optimal transport theory and stochastic thermodynamics for the Fokker-Planck equation. We show that the entropy production is bounded by the action measured by the path length of the L^2-Wasserstein distance, which is a measure of optimal transport. By using its geometrical interpretation of the entropy production, we obtain a lower bound on the entropy production, which is a trade-off relation between the transition time and the entropy production during this transition time. This trade-off relation can be regarded as a variant of thermodynamic speed limits. We discuss stochastic thermodynamics for the subsystem and derive a lower bound on the partial entropy production by the L^2-Wasserstein distance, which is a generalization of the second law of information thermodynamics. We also discuss a stochastic heat engine and show a geometrical constraint by the L^2-Wasserstein distance. Because the L^2-Waserstein distance is a measure of the optimal transport, our formalism leads to the optimal protocol to minimize the entropy production.

So a type of SLT.

I. INTRODUCTION

The concept of the difference between two probability distributions has been attracted by many researchers in information theory and statistical physics. For example, the Kullback-Leibler divergence has been used as a measure of the difference between two probability distributions [1], and it is useful in equilibrium statistical physics [2] and nonequilibrium stochastic thermodynamics [3, 6]. For example, the Kullback-Leibler divergence between two probabilities of forward and backward processes gives the entropy production [4], which is a measure of irreversibly in stochastic thermodynamics. In information geometry, the Kullback-Leibler divergence gives a differential geometry of the probability simplex. This differential geometry is naturally introduced from the Taylor expansion of the Kullback-Leibler divergence [7, 8]. Because the Kullback-Leibler divergence is strongly related to the entropy production in stochastic thermodynamics, information geometry has been recently discussed in stochastic thermodynamics [9–20] as a generalization of differential geometry in equilibrium thermodynamics and statistical physics [21–27].

In the field of optimal transport theory [28, 29], another measure of the difference between two probability distributions has been attracted. The L^2-Wasserstein distance is a well-known measure of optimal transport, which quantifies a difference between two probability distributions and introduces a differential geometry. In optimal transport theory, a relationship between L^2-Wasserstein distance and thermodynamic relaxation has been discussed, especially for the Fokker-Planck equation. For example, R. Jordan, D. Kinderlehrer, and F. Otto showed that the time evolution of the Fokker-Planck equation minimizes the sum of the free energy and the L^2-Wasserstein distance [30]. A trend to thermodynamic equilibrium for the Fokker-Planck equation has also been discussed using the L^2-Wasserstein distance [31]. Remarkably, the terminology of the entropy production is also used in optimal transport theory [28], and a connection between the entropy production and the L^2-Wasserstein distance has been discussed [32–34]. Moreover, a relationship between the L^2-Wasserstein distance and information geometry has been attracted recently [39, 40].

In the last decade, optimal transport theory has been used in stochastic thermodynamics to find a heat minimization protocol [35]. E. Aurell *et al.* have derived the lower bound on the entropy production [36] using the Banamou-Brenier formula [37] in optimal transport theory. A. Dechant and Y. Sakurai have also recently pointed out this lower bound on the entropy production in the context of thermodynamic trade-off relations [38]. This connection is strongly related to the recent studies of thermodynamic trade-off relations such as the thermodynamic uncertainty relations [41–62], the thermodynamic speed limits [9, 11, 12, 16, 18–20, 63, 64], and the universal bound on the efficiency [51, 65–68] because these trade-off relations come from a geometric feature of stochastic thermodynamics. For example, some of these trade-off relation can be derived from a mathematical feature of the Fisher information, which is a metric of information geometry [9, 11, 51, 52, 61, 68]. Based on this connection between optimal transport theory and stochastic thermodynamics, the efficiency of the stochastic heat engine has been discussed [69]. A similar con-

Trajectory-level work can be derived from the case of countable state spaces simply by replacing the Hamiltonian in that case with the function V. Similarly, trajectory level EF is given by a line integral of the product of (the time-derivative of the probability along) that trajectory and V_t(x), rather than by a sum over state transitions in a trajectory as in the countable state space case. Trajectory level EP is then given by combining that with the trajectory-level entropy, via Eq. 7.

nection between optimal transport theory and stochastic thermodynamics exists even for the Markov jump process, and a generalization of these trade-off relations has been derived without the L^2-Wasserstein distance [70].

This paper shows a connection between optimal transport theory and stochastic thermodynamics for the Fokker-Planck equation more deeply based on a connection between the entropy production rate and the L^2-Wasserstein distance [33]. We show that the entropy production is bounded by the time integral of the square of the velocity, namely the action in differential geometry, measured by the space of the L^2-Wasserstein distance. Furthermore, the entropy production can be proportional to the action with some assumptions where the force is given by the potential. Using this geometrical expression of the entropy production, we obtained a lower bound on the entropy production as a generalization of the thermodynamic speed limit, which is tighter than the previous result [36, 38]. Remarkably, the derivation of the thermodynamic speed limit is same as the original derivation of the thermodynamic speed limit in stochastic thermodynamics of information geometry [9]. Moreover, we discuss stochastic thermodynamics of the subsystem [71–74] and stochastic heat engine [75] by using the L^2-Wasserstein distance. We obtain a tighter bound on the partial entropy production as a generalization of the second law of information thermodynamics [10, 60, 71–74, 76–87], and a geometrical constraint of the heat engine's efficiency. We illustrate our results by using the example of the harmonic potential where the entropy production is proportional to the action, and analytically derive the optimization protocol [88] to minimize the entropy production based on a geometrical interpretation of the entropy production.

This all assumes a single bath and LDB, so no non-cons. forces. OTOH, the same eq.s 1,2 would allow non-cons. forces if we added some non-cons. f(x) to del V in Eq. 2

Also, this is all in the overdamped limit. (N.b., del V is a velocity, not an acceleration.)

II. FOKKER-PLANCK EQUATION AND STOCHASTIC THERMODYNAMICS

In this paper, we consider the probability distribution $p_t(\mathbf{x})$ of a particle in a Euclid d-dimensional position $\mathbf{x} \in X (= \mathbb{R}^d)$ at time t. The time evolution of $p_t(\mathbf{x})$ is described by the following Fokker-Planck equation for a particle driven by potential $V_t(x)$ with mobility μ attached to a heat bath at temperature T,

-mu del V_t(x) is u in VDB&E's FP thermo. paper, and -mu T is their D.

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = -\nabla \cdot (\boldsymbol{\nu}_t(\mathbf{x}) p_t(\mathbf{x})), \tag{1}$$

$$\boldsymbol{\nu}_t(\mathbf{x}) := -\mu \nabla [V_t(\mathbf{x}) + T \ln p_t(\mathbf{x})], \tag{2}$$

Intuitively, the second term on the RHS of Eq. 2 reflects the effects of diffusion and how it varies in space, "pushing against" the drift term, which being a potential's gradient is just the applied force.

where ∇ is the del oparator, and $\boldsymbol{\nu}_t(\mathbf{x})$ is a quantity called the mean local velocity. We here set ~~the~~ the Boltzmann constant to unity $k_{\mathrm{B}} = 1$. As a continuity equation, the mean local velocity $\boldsymbol{\nu}_t(\mathbf{x})$ is regarded as the velocity field. In stochastic thermodynamics [3], the internal energy U, the extracted work dW, the heat received from the heat bath dQ, and the entropy of the system S_{sys} at time t are defined as follows,

Note how intuitive eq.'s 3-6 are.

$$U := \int d\mathbf{x}\, V_t(\mathbf{x}) p_t(\mathbf{x}), \tag{3}$$

$$S_{\mathrm{sys}} := -\int d\mathbf{x}\, p_t(\mathbf{x}) \ln p_t(\mathbf{x}), \tag{4}$$

$$\frac{dW}{dt} := \int d\mathbf{x}\, \frac{\partial V_t(\mathbf{x})}{\partial t} p_t(\mathbf{x}), \tag{5}$$

$$\frac{dQ}{dt} := \int d\mathbf{x}\, V_t(\mathbf{x}) \frac{\partial p_t(\mathbf{x})}{\partial t}. \tag{6}$$

By definition, the heat dQ satisfies the first law of thermodynamics $dU/dt = dW/dt + dQ/dt$. From these definitions (3)-(6), the entropy production rate at time t

So (1/T) dQ/dt is the EF

$$\sigma_t := \frac{dS_{\mathrm{sys}}}{dt} - \frac{1}{T}\frac{dQ}{dt} \tag{7}$$

is calculated as

Combine Eq. 1, 6 to derive dQ/dt = \int dx F(x) dx/dt, where F is given by del V, and dx / dt is given by the Langevin formulation of Eq. 1.

$$\sigma_t = \frac{1}{\mu T} \int d\mathbf{x}\, [-\mu V_t(\mathbf{x}) - \mu T \ln p_t(\mathbf{x})] \frac{\partial p_t(\mathbf{x})}{\partial t} \tag{8}$$

$$= \frac{1}{\mu T} \int d\mathbf{x}\, \|\boldsymbol{\nu}_t(\mathbf{x})\|^2 p_t(\mathbf{x}), \tag{9}$$

where we used Eq. (1) and the normalization of the probability $(d/dt)[\int d\mathbf{x} p_t(\mathbf{x})] = 0$, and assumed that $p_t(\mathbf{x})$ vanishes at infinity. The symbol $\|\boldsymbol{\nu}_t\|^2 := \boldsymbol{\nu}_t \cdot \boldsymbol{\nu}_t$ indicates the square of L^2 norm. Thus, the entropy production rate σ_t is given by the expected value of the square of the mean local velocity divided by the factor μT. The entropy production from time $t = 0$ to time $t = \tau$ is defined as the time integral of the entropy production rate,

$$\Sigma := \int_0^\tau dt \sigma_t. \tag{10}$$

III. L^2-WASSERSTEIN DISTANCE

Next, we discuss the geometric measure of optimal transport called the L^2-Wasserstein distance [29]. We consider the distance $c(\mathbf{x}, \mathbf{y})$ on the space X as a cost function of transporting a single particle at the point $\mathbf{x} \in X$ to the point $\mathbf{y} \in X$. We first introduce the Monge-Kantrovich distance [89] as an indicator of how far apart the two probability distributions $p(\mathbf{x}), q(\mathbf{y})$ are on the manifold of the probability simplex. The Monge-Kantrovich distance for $c(\mathbf{x}, \mathbf{y})$ between $p(\mathbf{x})$ and $q(\mathbf{y})$ is defined as

$$C(p, q) := \min_{\Pi \in \mathcal{P}(p,q)} \int d\mathbf{x} d\mathbf{y}\; c(\mathbf{x}, \mathbf{y}) \Pi(\mathbf{x}, \mathbf{y}), \tag{11}$$

where the lower bound is taken over the entire set $\mathcal{P}(p, q)$ of joint probability distributions $\Pi(\mathbf{x}, \mathbf{y})$ on $X \times X$,

$$\mathcal{P}(p, q) := \{\Pi | p(\mathbf{x}) = \int d\mathbf{y} \Pi(\mathbf{x}, \mathbf{y}),\; q(\mathbf{y}) = \int d\mathbf{x} \Pi(\mathbf{x}, \mathbf{y}), \Pi(\mathbf{x}, \mathbf{y}) \geq 0\}, \tag{12}$$

In Eq. 15 of VDB&E's FP thermo. paper, they argue that LDB in this context is simply the condition that the function V(x) is the energy of the system in state x, and that Einstein's relation relates D, mu, and T.

I'm pretty sure that since the term multiplying the second term on the RHS of Eq. 2 does not vary spatially, they are assuming a constant diffusion term. See xu.simon levin....and esposito.vandenbroeck.fokker.planck...

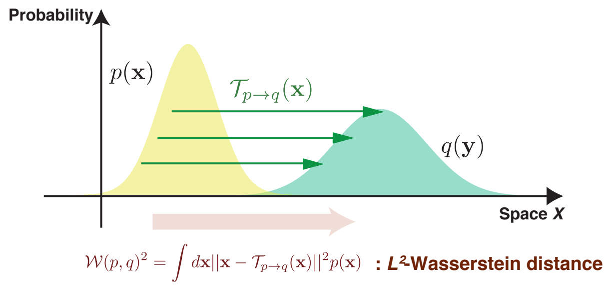

FIG. 1. Schematic of the L^2-Wasserstein distance. We here consider optimal transport from the probability distribution $p(\mathbf{x})$ to the probability distribution $p(\mathbf{y})$. The length of the green arrow shows the optimal transportation distance $||\mathbf{x}-\mathcal{T}_{p\to q}(\mathbf{x})||$, and the square of the L^2-Wasserstein distance is given by the expected value of the square of its optimal transportation distance.

where marginal distributions of $\Pi(\mathbf{x},\mathbf{y})$ in the set $\mathcal{P}(p,q)$ are given by $p(\mathbf{x})$ and $q(\mathbf{y})$. Therefore, the Monge-Kantrovich distance gives a minimum value of the expected value of the distance $c(\mathbf{x},\mathbf{y})$ for the joint distribution $\Pi(\mathbf{x},\mathbf{y})$. We call the value of Π that minimizes the expected value of the distance as the optimal transport plan Π^*, which is defined as

The min being over all such Pi(x, y) whose marginals are p and q.

$$\Pi^*(\mathbf{x},\mathbf{y}) := \mathrm{argmin}_{\Pi\in\mathcal{P}(p,q)} \int d\mathbf{x}d\mathbf{y}\; c(\mathbf{x},\mathbf{y})\Pi(\mathbf{x},\mathbf{y}). \tag{13}$$

The L^2-Wasserstein distance $\mathcal{W}(p,q)$, which plays an important role in this paper, is introduced by the square root of the Monge-Kantrovich distance for $c(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|^2$. Explicitly, the L^2-Wasserstein distance $\mathcal{W}(p,q)$ between p and q is defined as

$$\mathcal{W}(p,q)^2 := \min_{\Pi\in\mathcal{P}(p,q)} \int d\mathbf{x}d\mathbf{y}\; \|\mathbf{x}-\mathbf{y}\|^2\Pi(\mathbf{x},\mathbf{y}). \tag{14}$$

The L^2-Wasserstein distance is well defined [29] if two probability distributions p and q satisfy

$$\int d\mathbf{x}p(\mathbf{x})\|\mathbf{x}\|^2 < \infty, \int d\mathbf{y}q(\mathbf{y})\|\mathbf{y}\|^2 < \infty. \tag{15}$$

We assume this condition Eq. (15) in this paper.

Furthermore, it is known that there exists a map $\mathcal{T}_{p\to q}(\mathbf{x})$ such that $\Pi^*(\mathbf{x},\mathbf{y}) = p(\mathbf{x})\delta(\mathbf{y}-\mathcal{T}_{p\to q}(\mathbf{x}))$ for the L^2-Wasserstein distance $c(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|^2$ on the space $\mathbb{R}^d$, where $\delta(\mathbf{x})$ is the delta function [29]. This map $\mathcal{T}_{p\to q}$ is called the optimal transport map from p to q. Using the fact that the marginal distributions of $\Pi^*(\mathbf{x},\mathbf{y})$ are $p(\mathbf{x})$ and $q(\mathbf{y})$, we can obtain

$$\int d\mathbf{y} f(\mathbf{y})q(\mathbf{y}) = \int d\mathbf{x} \int d\mathbf{y} f(\mathbf{y})\Pi^*(\mathbf{x},\mathbf{y}) = \int d\mathbf{x} f(\mathcal{T}_{p\to q}(\mathbf{x}))p(\mathbf{x}) \tag{16}$$

for any differential and measurable function $f(\mathbf{x})$. If we consider the change of variables $\mathbf{y} = \mathcal{T}_{p\to q}(\mathbf{x})$ and $d\mathbf{y} = d\mathbf{x}|\det(\nabla\mathcal{T}_{p\to q}(\mathbf{x}))|$, we obtain the Jacobian equation [29]

$$p(\mathbf{x}) = q(\mathcal{T}_{p\to q}(\mathbf{x}))|\det(\nabla\mathcal{T}_{p\to q}(\mathbf{x}))|, \tag{17}$$

where $|\det(\nabla\mathcal{T}_{p\to q}(\mathbf{x}))|$ denotes the determinant of the Jacobian matrix $\nabla\mathcal{T}_{p\to q}$ at $\mathbf{x}$. By using the optimal transport map, the L^2-Wasserstein distance is calculated as

$$\mathcal{W}(p,q)^2 = \int d\mathbf{x}\; \|\mathbf{x}-\mathcal{T}_{p\to q}(\mathbf{x})\|^2 p(\mathbf{x}). \tag{18}$$

Thus, the L^2-Wasserstein distance can be regarded as the expected value of the optimal transportation distance $||\mathbf{x}-\mathcal{T}_{p\to q}(\mathbf{x})||$ (see Fig. 1).

We briefly introduce the Benamou-Brenier formula [37], which is related to a relation between the entropy production and the L^2-Wasserstein distance in this paper. If dynamics of the probability $q_t(\mathbf{x})$ are driven by the continuity equation with the velocity field $\mathbf{v}_t(\mathbf{x})$,

$$\frac{\partial q_t(\mathbf{x})}{\partial t} = -\nabla\cdot(\mathbf{v}_t(\mathbf{x})q_t(\mathbf{x})), \tag{19}$$

the L^2-Wasserstein distance gives the lower bound on the expected value of the square of the velocity field,

$$\mathcal{W}(q_0,q_\tau)^2 \leq \tau\int_0^\tau dt\int d\mathbf{x}||\mathbf{v}_t(\mathbf{x})||^2 q_t(\mathbf{x}). \tag{20}$$

where we consider the time integral from time $t=0$ to $t=\tau$. Because the velocity field of the Fokker-Planck equation is the mean local velocity, we obtain a relation between the entropy production rate and the L^2-Wasserstein distance as discussed in the next section.

IV. RELATION BETWEEN WASSERSTEIN DISTANCE AND ENTROPY PRODUCTION RATE

In this section, we discuss a relation between the L^2-Wasserstein distance and the entropy production rate. We set that dynamics of the probability distribution $p_t(\mathbf{x})$ are described by the Fokker-Planck equation (1). We define the path length on the probability simplex measured by the L^2-Wasserstein distance from time $t=0$ to time $t=\tau$ as

$$\mathcal{L}_\tau := \lim_{\Delta t\downarrow 0}\sum_{k=0}^{\lceil\tau/\Delta t\rceil}\mathcal{W}(p_{k\Delta t},p_{(k+1)\Delta t}), \tag{21}$$

So q(y) = \int dx Pi*(x, y) = \int dx p(x) \delta[y - T_{p->q}(x)]
Intuitively, we are taking the mass p(x) at each x and moving it to T_{p->q}(x), to get q(y)

See van vu.saito.Topological speed limit.arxiv where the positive integer $\lceil \tau/\Delta t \rceil$ is given by the ceiling function $\lceil \cdots \rceil$. The entropy production rate is bounded by

$$\sigma_t \geq \frac{1}{\mu T} \left(\frac{d\mathcal{L}_t}{dt} \right)^2, \tag{22}$$

which is consistent with the previous result in Ref. [33]. This equation gives a relation between the L^2-Wasserstein distance and the entropy production rate for the Fokker-Planck equation. In terms of the L^2-Wasserstein distance, the quantity $(d\mathcal{L}_t/dt)^2$ is given by

Everything from here to Eq. 43 is deriving Eq. 22.

$$\left(\frac{d\mathcal{L}_t}{dt} \right)^2 = \frac{\mathcal{W}(p_{t+\Delta t}, p_t)^2}{\Delta t^2} + \mathcal{O}(\Delta t) \tag{23}$$

$$= \lim_{\Delta t \downarrow 0} \frac{\mathcal{W}(p_{t+\Delta t}, p_t)^2}{\mu T \Delta t^2}, \tag{24}$$

for the short time Δt. Thus, this inequality can be regarded as the Benamou-Brenier formula [37] for the short time $\tau = \Delta t$,

$$\mathcal{W}(p_{t+\Delta t}, p_t)^2 \leq \Delta t \int_0^{\Delta t} dt \int d\mathbf{x} \, \|\boldsymbol{\nu}_t(\mathbf{x})\|^2 p_t(\mathbf{x}) + O(\Delta t^3). \tag{25}$$

We next discuss the situation that the equality in Eq. (43) holds. We introduce a non-negative term σ_t^{rot} defined as

$$\sigma_t^{\mathrm{rot}} = \sigma_t - \frac{1}{\mu T} \left(\frac{d\mathcal{L}_t}{dt} \right)^2 \geq 0, \tag{26}$$

and consider the situation $\sigma_t^{\mathrm{rot}} = 0$. Here, we consider the Taylor expansion of the optimal transport map $\mathcal{T}_{p_t \to p_{t+\Delta t}}(\mathbf{x})$ up to the order Δt,

$$\mathcal{T}_{p_t \to p_{t+\Delta t}}(\mathbf{x}) = \mathbf{x} + \boldsymbol{a}_1(\mathbf{x}) \Delta t + \mathcal{O}(\Delta t^2), \tag{27}$$

where $\boldsymbol{a}_1(\mathbf{x})$ is the first order of the Taylor coefficient. From Eq. (18), we obtain an expression of $(d\mathcal{L}_t/dt)^2$,

$$\left(\frac{d\mathcal{L}_t}{dt} \right)^2 = \int d\mathbf{x} \, \|\boldsymbol{a}_1(\mathbf{x})\|^2 p_t(\mathbf{x}) + \mathcal{O}(\Delta t). \tag{28}$$

Thus, if the mean local velocity gives an optimal transport map, that is $\boldsymbol{\nu}_t(\mathbf{x}) = \boldsymbol{a}_1(\mathbf{x})$, the equality holds and $\sigma_t^{\mathrm{rot}} = 0$.

Next, we consider the difference between $\boldsymbol{a}_1(\mathbf{x})$ and $\boldsymbol{\nu}_t(\mathbf{x})$. By substituting $(p_t, p_{t+\Delta t})$ into (p, q), the Jacobian equation in Eq. (17) is given by

$$p_t(\mathbf{x}) = p_{t+\Delta t}(\mathcal{T}_{p_t \to p_{t+\Delta t}}(\mathbf{x})) |\det(\nabla \mathcal{T}_{p_t \to p_{t+\Delta t}}(\mathbf{x}))|. \tag{29}$$

We calculate the Taylor expansions of the determinant up to the order Δt as follows

$$|\det(\nabla \mathcal{T}_{p_t \to p_{t+\Delta t}}(\mathbf{x}))| = 1 + \nabla \cdot \boldsymbol{a}_1(\mathbf{x}) \Delta t + \mathcal{O}(\Delta t^2). \tag{30}$$

From the Fokker-Planck equation (1), we also obtain

$$p_{t+\Delta t}(\mathbf{x}) = p_t(\mathbf{x}) - \nabla \cdot (\boldsymbol{\nu}_t(\mathbf{x}) p_t(\mathbf{x})) \Delta t + \mathcal{O}(\Delta t^2), \tag{31}$$

which is the discretized version of the Fokker-Planck equation for the short time Δt. By inserting Eqs. (27), (30) and (31) into Eq. (29), we obtain

$$0 = \nabla \cdot [(\boldsymbol{a}_1(\mathbf{x}) - \boldsymbol{\nu}_t(\mathbf{x})) p_t(\mathbf{x})] \Delta t + \mathcal{O}(\Delta t^2).$$

By considering the first-order terms of Δt, we obtain

$$\nabla \cdot [(\boldsymbol{a}_1(\mathbf{x}) - \boldsymbol{\nu}_t(\mathbf{x})) p_t(\mathbf{x})] = 0. \tag{32}$$

Because of Helmholtz's decomposition, this equation implies the existence of a vector potential $\boldsymbol{A}_t(\mathbf{x})$ such as

$$\boldsymbol{a}_1(\mathbf{x}) p_t(\mathbf{x}) = \boldsymbol{\nu}_t(\mathbf{x}) p_t(\mathbf{x}) + \nabla \times \boldsymbol{A}_t(\mathbf{x}). \tag{33}$$

Thus, this vector potential $\boldsymbol{A}_t$ quantifies a difference between optimal transport plan and the time evolution of the Fokker-Planck equation from time t to time $t + \Delta t$.

To find the expression of σ_t^{rot}, we use the formula for the time derivative of the L^2-Wasserstein distance [29]. The following formula

$$\frac{d}{ds} \left(\frac{\mathcal{W}(p, p_{t+s})^2}{2} \right) \Bigg|_{s=0} = - \int d\mathbf{x} (\mathbf{x} - \mathcal{T}_t(\mathbf{x})) \cdot \boldsymbol{\nu}_t(\mathcal{T}_t(\mathbf{x})) p(\mathbf{x}) \tag{34}$$

holds for any probability distribution $p(\mathbf{x})$, where we used the notation $\mathcal{T}_t = \mathcal{T}_{p \to p_t}$. The proof of this formula (34) is shown in Appendix A. By applying the Taylor expansion Eq. (30) to the formula Eq. (34) for $(p, p_{t+s}) = (p_t, p_{t+\Delta t+s})$, we obtain the following equation,

$$\begin{aligned}
&\frac{d}{ds} \left(\frac{\mathcal{W}(p_t, p_{t+\Delta t+s})^2}{2} \right) \Bigg|_{s=0} \\
&= - \Delta t \int d\mathbf{x} [\boldsymbol{a}_1(\mathbf{x}) \cdot \boldsymbol{\nu}_t(\mathcal{T}_{p_t \to p_{t+\Delta t}}(\mathbf{x}))] p_t(\mathbf{x}) \\
&= \Delta t \int d\mathbf{x} [\boldsymbol{a}_1(\mathcal{T}_{p_t \to p_{t+\Delta t}}(\mathbf{x})) \cdot \boldsymbol{\nu}_t(\mathcal{T}_{p_t \to p_{t+\Delta t}}(\mathbf{x}))] p_t(\mathbf{x}) + \mathcal{O}(\Delta t^2) \\
&= \Delta t \int d\mathbf{y} [\boldsymbol{a}_1(\mathbf{y}) \cdot \boldsymbol{\nu}_t(\mathbf{y})] p_{t+\Delta t}(\mathbf{y}) + \Delta t \mu T \sigma_t^{\mathrm{rot}} + \mathcal{O}(\Delta t^2) \\
&= \Delta t \int d\mathbf{x} [\boldsymbol{a}_1(\mathbf{x}) \cdot \boldsymbol{\nu}_t(\mathbf{x})] p_t(\mathbf{x}) + \mathcal{O}(\Delta t^2),
\end{aligned} \tag{35}$$

where we used Eq. (16). From the definition of the path length Eq. (21), we obtain

$$\mathcal{W}(p_{t+s}, p_t) = \frac{d\mathcal{L}_t}{dt} s + \mathcal{O}(s^2), \tag{36}$$

for small s. Therefore, we also obtain

$$\begin{aligned}
&\frac{d}{ds} \left(\frac{\mathcal{W}(p_t, p_{t+\Delta t+s})^2}{2} \right) \Bigg|_{s=0} \\
&= \lim_{s \downarrow 0} \frac{\mathcal{W}(p_{t+\Delta t+s}, p_t) - \mathcal{W}(p_{t+\Delta t}, p_t)}{s} \mathcal{W}(p_{t+\Delta t}, p_t) \\
&= \frac{d\mathcal{L}_{t+\Delta t}}{dt} \frac{d\mathcal{L}_t}{dt} \Delta t + \mathcal{O}(\Delta t^2) \\
&= \left(\frac{d\mathcal{L}_t}{dt} \right)^2 \Delta t + \mathcal{O}(\Delta t^2).
\end{aligned} \tag{37}$$

By comparing Eq. (37) with Eq. (35), we obtain another expression of $(d\mathcal{L}_t/dt)^2$

$$\left(\frac{d\mathcal{L}_t}{dt}\right)^2 = \int d\mathbf{x}[\boldsymbol{a}_1(\mathbf{x}) \cdot \boldsymbol{\nu}_t(\mathbf{x})]p_t(\mathbf{x}) + \mathcal{O}(\Delta t^2). \quad (38)$$

We finally obtain an expressions of σ_t^{rot} from Eq. (38),

$$\sigma_t^{\mathrm{rot}} = -\frac{1}{\mu T}\int d\mathbf{x}[\nabla \times \boldsymbol{A}_t(\mathbf{x})] \cdot \boldsymbol{\nu}_t(\mathbf{x}) \quad (39)$$

We also obtain another expression by comparing Eq. (28) with Eq. (38),

$$\sigma_t^{\mathrm{rot}} = -\frac{1}{\mu T}\int d\mathbf{x}[\nabla \times \boldsymbol{A}_t(\mathbf{x})] \cdot \boldsymbol{\nu}_t(\mathbf{x}) \quad (40)$$

$$= \frac{1}{\mu T}\int d\mathbf{x}\frac{||\nabla \times \boldsymbol{A}_t(\mathbf{x})||^2}{p_t(\mathbf{x})} \geq 0. \quad (41)$$

Thus, σ_t^{rot} is non-negative, and zero if $||\nabla \times \boldsymbol{A}_t(\mathbf{x})|| = 0$. Because σ_t^{rot} is proportional to the mean value of the square of $||\nabla \times \boldsymbol{A}_t(\mathbf{x})||/p_t(\mathbf{x})$, this value σ_t^{rot} quantifies the amount of the rotation in a difference between the optimal transport and the mean local velocity, and it might not be zero if the system is in nonequilibrium steady state. The non-potential force is needed to achieve nonequilibrium steady state, and the steady flow and the steady force should be cyclic because of the Schnakenberg network theory [90]. When the mean local velocity $\boldsymbol{\nu}_t(\mathbf{x})$ is given by the potential $\boldsymbol{\nu}_t(\mathbf{x}) = -\nabla\Phi_t$ with $\Phi_t = \mu(V_t(\boldsymbol{x}) + T\ln p_t(\boldsymbol{x}))$ as we assumed in Eq. (2), the quantity σ_t^{rot} is given by

$$\sigma_t^{\mathrm{rot}} = \frac{1}{\mu T}\int d\mathbf{x}[\nabla \times \boldsymbol{A}_t(\mathbf{x})] \cdot \nabla\Phi_t$$

$$= \frac{1}{\mu T}\int d\mathbf{x}\nabla \cdot [\Phi_t \nabla \times \boldsymbol{A}_t(\mathbf{x})]$$

$$= \frac{1}{\mu T}\int d\boldsymbol{S} \cdot [\Phi_t \nabla \times \boldsymbol{A}_t(\mathbf{x})] \quad (42)$$

where $\int d\boldsymbol{S}$ denotes the surface integral. If the quantity $|\Phi_t \nabla \times \boldsymbol{A}_t(\mathbf{x})|$ vanishes at infinity, the quantity σ_t^{rot} becomes zero. The quantity $|\nabla \times \boldsymbol{A}_t(\mathbf{x})|$ is bounded by $|\nabla \times \boldsymbol{A}_t(\mathbf{x})| \leq 2\max(|\boldsymbol{a}_1(\mathbf{x})|, |\boldsymbol{\nu}_t(\mathbf{x})|)p_t(\mathbf{x})$ where $|\cdots|$ denotes the absolute value. Thus, if we assume that the probability $p_t(\mathbf{x})$ converges to zero and the product $2\max(|\boldsymbol{a}_1(\mathbf{x})|, |\boldsymbol{\nu}_t(\mathbf{x})|)p_t(\mathbf{x})|\Phi_t|$ also vanishes at infinity, we obtain $\sigma_t^{\mathrm{rot}} = 0$ and

$$\sigma_t = \frac{1}{\mu T}\left(\frac{d\mathcal{L}_t}{dt}\right)^2. \quad (43)$$

This condition might be well achieved for a realistic physical situation, because the probability $p_t(\mathbf{x})$ is physically localized and the probability $p_t(\mathbf{x})$ decays much faster than the factor $1/[2\max(|\boldsymbol{a}_1(\mathbf{x})|, |\boldsymbol{\nu}_t(\mathbf{x})|)|\Phi_t|]$. For example, in the section IX, we obtain this equality (43) for the Brownian particle trapped in the harmonic potential.

The term σ_t^{rot} might only be important for the case that non-potential force exists in the Fokker-Planck equation, where the mean local velocity $\boldsymbol{\nu}_t(\mathbf{x})$ cannot be written by the potential $-\nabla\Phi_t$. Thus, the quantity σ_t^{rot} might play an important role in the steady state thermodynamics [91].

V. LOWER BOUND ON ENTROPY PRODUCTION

We here discuss a lower bound on the entropy production $\Sigma := \int dt\sigma_t$ based on Eq. (43). By using Eq. (43), the entropy production from time $t = 0$ to $t = \tau$ is bounded by

$$\Sigma = \int_0^\tau dt\sigma_t$$

$$\geq \frac{1}{\mu T}\int_0^\tau dt\left(\frac{d\mathcal{L}_t}{dt}\right)^2. \quad (44)$$

In differential geometry, the quantity $\mathcal{C} = (1/2)\int_0^\tau dt\,(d\mathcal{L}_t/dt)^2$ called as the action, and Eq. (43) implies that the entropy production for the Fokker-Planck equation is bounded by the action measured by the path length of the Wasserstein L^2 distance,

$$\Sigma \geq \frac{2\mathcal{C}}{\mu T}. \quad (45)$$

If $\sigma^{\mathrm{rot}} = 0$, the entropy production is proportional to the action measured by the path length of the Wasserstein L^2 distance $\Sigma = 2\mathcal{C}/(\mu T)$. Here, we consider the following Cauchy-Schwarz inequality

$$2\tau\mathcal{C} = \left(\int_0^\tau dt\right)\left(\int_0^\tau dt\left(\frac{d\mathcal{L}_t}{dt}\right)^2\right)$$

$$\geq \left(\int_0^\tau dt\frac{d\mathcal{L}}{dt}\right)^2$$

$$= \mathcal{L}_\tau^2, \quad (46)$$

which gives a lower bound on the action. In information geometry, this inequality has been considered [24] as a trade-off relation between time τ and the action $\mathcal{C}$. By considering $(d\mathcal{L}_t/dt)^2$ as the Fisher information of time, several variants of thermodynamic speed limits can be derived from this inequality for the Markov jump process [9], the Fokker-Planck equation [11] and the rate equation [18] in stochastic thermodynamics of information geometry. In the same way, we obtain a lower bound on the entropy production by considering the action measured by the L^2-Wasserstein distance (see also Fig. 2),

Recall T is temp., mu is constant
in FP equation, and tau is time.

$$\Sigma \geq \frac{\mathcal{L}_\tau^2}{\mu T\tau}. \quad (47)$$

Because this inequality implies a trade-off relation between time and the entropy production, this result can

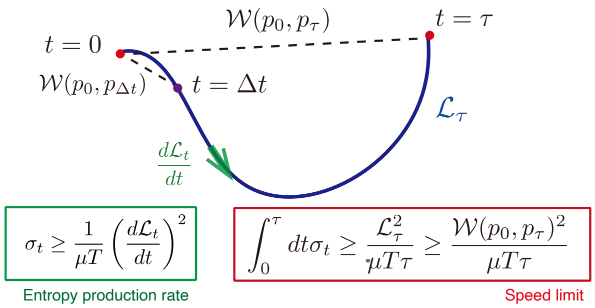

FIG. 2. Schematic of the entropy production and the L^2-Wasserstein distance. The lower bound on the entropy production is obtained from geometry of the L^2-Wasserstein distance. The entropy production $\Sigma = \int_0^\tau dt\sigma_t$ is bounded by the length measured by the L^2-Wasserstein distance $\mathcal{L}_\tau$ as a tighter bound, and the L^2-Wasserstein distance itself $\mathcal{W}(p_0, p_\tau)$ as a lower bound. These inequalities are generalizations of thermodynamic speed limits.

also be regarded as a generalization of thermodynamic speed limits. Since we use the Cauchy-Schwarz inequality, the equality can be achieved when the probability distribution moves with a constant velocity on the L^2-Wasserstein distance space, that is, when it satisfies the following equation

$$\frac{d\mathcal{L}_t}{dt} = \frac{\mathcal{L}_\tau}{\tau}, \tag{48}$$

for any $0 \leq t \leq \tau$.

Using the fact that the L^2-Wasserstein distance satisfies the triangle inequality for probabilities p, q and r,

$$W(p,r) \leq W(p,q) + W(q,r), \tag{49}$$

we obtain the following inequality,

$$\mathcal{L}_\tau \geq \mathcal{W}(p_0, p_\tau). \tag{50}$$

from the definition of $\mathcal{L}_\tau$. Using Eq. (47) and the above inequality, we can obtain the previously known inequality in Refs. [36, 38],

$$\Sigma \geq \frac{\mathcal{W}(p_0, p_\tau)^2}{\mu T \tau}, \tag{51}$$

which is equivalent to the Benamou-Brenier formula [37] because the entropy production rate is given by the expected value of the square of the velocity field $\boldsymbol{\nu}_t(\mathbf{x})$. Considering the above derivation, the condition for the equality to hold is when the probability distribution changes at a constant speed on a straight line as measured by the L^2-Wasserstein distance,

$$\mathcal{L}_\tau = \mathcal{W}(p_\tau, p_0), \tag{52}$$

$$\frac{d\mathcal{L}_t}{dt} = \frac{\mathcal{W}(p_\tau, p_0)}{\tau}. \tag{53}$$

In this case, the entropy production is minimized with constraints p_0 and p_τ. Moreover, when the initial distribution p_0, the final distribution p_τ, and the time interval τ are specified, the protocol to achieve this equality can be numerically obtained by the algorithm of the fluid mechanics [37]. In other words, by using this algorithm, we can construct an efficient heat engine for small systems with the minimum entropy production.

Similarly, we obtain another lower bound by applying the Cauchy-Schwartz inequality Eq. (46) and the triangle inequality Eq. (49). Let us consider the time interval $t_i = \tau(i/N)$. Because the entropy production is given by

$$\Sigma \geq \sum_{i=0}^{N-1} \frac{1}{\mu T} \int_{t_i}^{t_{i+1}} dt \left(\frac{d\mathcal{L}_t}{dt}\right)^2, \tag{54}$$

another lower bound on the entropy production can be obtained in a similar way as follows

$$\Sigma \geq \sum_{i=0}^{N-1} \hat{\Sigma}(t_i; t_{i+1}), \tag{55}$$

where $\hat{\Sigma}(t;s)$ is the lower bound on the entropy production by the Benamou-Brenier formula from time t to time s,

$$\hat{\Sigma}(t;s) = \frac{\mathcal{W}(p_t, p_s)^2}{\mu T(s-t)}. \tag{56}$$

Moreover, in the case $\sigma_t^{\mathrm{rot}} = 0$, we obtain

$$\Sigma = \lim_{N\to\infty} \sum_{i=0}^{N-1} \hat{\Sigma}(t_i; t_{i+1}), \tag{57}$$

because the change from p_{t_i} to $p_{t_{i+1}}$ is at a constant rate on a straight line as measured by the L^2-Wasserstein distance in the limit $t_{i+1} - t_i = \tau/N \to 0$. Remarkably, a calculation of $\hat{\Sigma}(t_i; t_{i+1})$ does not require information of the joint probability distribution at time t_i and t_{i+1}, while the experimental estimation of the entropy production based on the fluctuation theorem needs information of the joint probability distribution [92]. It is relatively difficult to estimate the joint probability in an experiment with a small number of samples, compared to two probabilities. This fact might be useful to estimate the entropy production in an experiment by using Eq. (57). This estimation of the entropy production by using Eq. (57) is similar to the estimation of the entropy production based on thermodynamic trade-off relations such as thermodynamic uncertainty relations [55–58]. Because the algorithm of the fluid mechanics [37] provides a proper estimation of the mean local velocity numerically, this estimation of the entropy production by using Eq. (57) might be better than the estimation of the entropy production based on thermodynamic uncertainty relations [55–58] for a Brownian particle, where its dynamics are given by the Fokker-Planck equation with the potential force.

VI. STOCHASTIC THERMODYNAMICS OF SUBSYSTEM

In this section, we discuss a relationship between the L^2-Wasserstein distance of the subsystem and thermodynamics. We start with two-dimensional systems X and Y. Stochastic dynamics of two positions $x \in X(=\mathbb{R})$ and $y \in Y(=\mathbb{R})$ are driven by the following Fokker-Planck equation

$$\begin{aligned}\frac{\partial p_t(x,y)}{\partial t} =& -\frac{\partial}{\partial x}(\nu_t^X(x,y)p_t(x,y)) - \frac{\partial}{\partial y}(\nu_t^Y(x,y)p_t(x,y)),\\ \nu_t^X(x,y) :=& -\mu\frac{\partial}{\partial x}[V_t(x,y) + T\ln p_t(x,y)],\\ \nu_t^Y(x,y) :=& -\mu\frac{\partial}{\partial y}[V_t(x,y) + T\ln p_t(x,y)]. \end{aligned} \tag{58}$$

We first consider the situation that the position y is the hidden degree of freedom and we can only observe the position x. Thus, we can only measure the marginal distribution of X defined as

$$p_t^X(x) = \int dy\ p_t(x,y). \tag{59}$$

The time evolution of the marginal distribution is given by

$$\frac{\partial p_t^X(x)}{\partial t} = -\frac{\partial}{\partial x}\left(\bar{\nu}_t^X(x)p_t^X(x)\right), \tag{60}$$

$$\begin{aligned}\bar{\nu}_t^X(x) =& \frac{\int dy\ \nu_t^X(x,y)p_t(x,y)}{p_t^X(x)}\\ =& -\mu\frac{\partial}{\partial x}\left[\int dy p_t^{Y|X}(y|x)[V_t(x,y)+T\ln p_t(x,y)]\right],\end{aligned} \tag{61}$$

where $\bar{\nu}_t^X(x)$ is the marginal mean local velocity of X, $p_t^{Y|X}(y|x) := p_t(x,y)/p_t^X(x)$ is the conditional probability of Y under the condition of X, and we assumed that $p_t(x,y)$ vanishes at infinity. If we want to measure the entropy production rate for this system, we only obtain the apparent entropy production rate of X,

$$\bar{\sigma}_t^X = \frac{1}{\mu T}\int dx[\bar{\nu}_t^X(x)]^2 p_t^X(x), \tag{62}$$

which is different from the partial entropy production rate of X,

See Eq. 9. This can sort of be viewed as the trajectory instantaneous EP, marginalized to one subsystem, and then averaged over the states of that subsystem.

$$\sigma_t^X = \frac{1}{\mu T}\int dx \int dy[\nu_t^X(x,y)]^2 p_t(x,y). \tag{63}$$

From the Cauchy-Schwarz inequality, we obtain the inequality

$$\begin{aligned}&\sigma_t^X - \bar{\sigma}_t^X\\ =&\frac{1}{\mu T}\int dx \frac{\left(\int dy[\nu_t^X(x,y)]^2 p_t(x,y)\right)\left(\int dy p_t(x,y)\right)}{p_t^X(x)}\\ &-\frac{1}{\mu T}\int dx\frac{\left(\int dy \nu_t^X(x,y)p_t(x,y)\right)^2}{p_t^X(x)}\\ \geq& 0.\end{aligned} \tag{64}$$

Thus, the apparent entropy production rate $\bar{\sigma}_t^X$ is always smaller than the partial entropy production rate σ_t^X. The apparent entropy production rate is equivalent to the partial entropy production when $\nu_t^X(x,y) = \bar{\nu}_t^X(x)$. This condition implies that the potential force $-\partial V_t(x,y)/\partial x$ does not depend on y, and the systems X and Y are statistically independent $p_t(x,y) = p_t^X(x)p_t^Y(y)$ with $p_t^Y(y) := \int dx\ p_t(x,y)$.

If we define the path length of X from time $t=0$ to time $t=\tau$ as

$$\mathcal{L}_\tau^X := \lim_{\Delta t \downarrow 0} \sum_{k=0}^{\lceil \tau/\Delta t \rceil} \mathcal{W}(p_{k\Delta t}^X, p_{(k+1)\Delta t}^X), \tag{65}$$

our result for the path length of X gives the apparent entropy production rate of X,

$$\bar{\sigma}_t^X \geq \frac{1}{\mu T}\left(\frac{d\mathcal{L}_t^X}{dt}\right)^2 = \lim_{\Delta t\to 0}\frac{\mathcal{W}(p_{t+\Delta t}^X, p_t^X)^2}{\mu T \Delta t^2}. \tag{66}$$

Because $\bar{\nu}_t^X(x)$ is given by the potential $\bar{\nu}_t^X(x) = -(\partial/\partial x)\Phi_t^X$ with

$$\Phi_t^X = \mu\left[\int dy p_t^{Y|X}(y|x)[V_t(x,y)+T\ln p_t(x,y)]\right], \tag{67}$$

we might obtain the equality

$$\bar{\sigma}_t^X = \frac{1}{\mu T}\left(\frac{d\mathcal{L}_t^X}{dt}\right)^2. \tag{68}$$

under the assumption that $p^X(x)$ vanishes at infinity and the distribution is localized. We also obtain a lower bound on the apparent entropy production rate of X as follows,

$$\begin{aligned}\bar{\Sigma}^X :=& \int_0^\tau dt \bar{\sigma}_t^X && (69)\\ \geq& \frac{(\mathcal{L}_\tau^X)^2}{\tau\mu T} && (70)\\ \geq& \frac{\mathcal{W}(p_0^X, p_\tau^X)^2}{\tau\mu T}. && (71)\end{aligned}$$

Now, we discuss a relationship between two subsystems X and Y. We introduce the marginal mean local velocity, the apparent entropy production rate and the partial entropy production rate of Y as follows

$$\frac{\partial p_t^Y(x)}{\partial t} = -\frac{\partial}{\partial y}\left(\bar{\nu}_t^Y(x) p_t^Y(x)\right), \tag{72}$$

$$\bar{\nu}_t^Y(x) = \frac{\int dx\ \nu_t^Y(x,y) p_t(x,y)}{p_t^Y(y)}, \tag{73}$$

$$\bar{\sigma}_t^Y = \frac{1}{\mu T}\int dy [\bar{\nu}_t^Y(x)]^2 p_t^Y(y), \tag{74}$$

$$\sigma_t^Y = \frac{1}{\mu T}\int dx \int dy [\nu_t^Y(x,y)]^2 p_t(x,y). \tag{75}$$

The entropy production rate is given by the sum of the partial entropy production rates,

$$\sigma_t = \sigma_t^X + \sigma_t^Y. \tag{76}$$

Because $\sigma_t^X \geq \bar{\sigma}_t^X$ and $\sigma_t^Y \geq \bar{\sigma}_t^Y$, the inequality

$$\sigma_t - \bar{\sigma}_t^X - \bar{\sigma}_t^Y \geq 0, \tag{77}$$

is satisfied.

VII. INFORMATION THERMODYNAMICS

We here discuss information thermodynamics, which explains a paradox of the Maxwell's demon [80]. In information thermodynamics, we consider a relation between the partial entropy production and information flow for the 2D Fokker-Planck equation (58) or the 2D Langevin equations [74, 76, 81]. The partial entropy production rates of X and Y for Eq. (58) are calculated as

$$\sigma_t^X = \sigma_{\mathrm{bath};t}^X + \sigma_{\mathrm{sys};t}^X - \dot{\mathcal{I}}^X, \tag{78}$$

$$\sigma_t^Y = \sigma_{\mathrm{bath};t}^Y + \sigma_{\mathrm{sys};t}^Y - \dot{\mathcal{I}}^Y, \tag{79}$$

$$\sigma_{\mathrm{bath};t}^X = \frac{1}{T}\int dx \int dy \left[-\frac{\partial V_t(x,y)}{\partial x}\right] \nu_t^X(x,y) p_t(x,y), \tag{80}$$

$$\sigma_{\mathrm{bath};t}^Y = \frac{1}{T}\int dx \int dy \left[-\frac{\partial V_t(x,y)}{\partial y}\right] \nu_t^Y(x,y) p_t(x,y), \tag{81}$$

$$\sigma_{\mathrm{sys};t}^X = \int dx \int dy \left[-\frac{\partial \ln p_t^X(x)}{\partial x}\right] \nu_t^X(x,y) p_t(x,y), \tag{82}$$

$$\sigma_{\mathrm{sys};t}^Y = \int dx \int dy \left[-\frac{\partial \ln p_t^Y(y)}{\partial y}\right] \nu_t^Y(x,y) p_t(x,y), \tag{83}$$

$$\dot{\mathcal{I}}^X = \int dx \int dy \left[\frac{\partial}{\partial x}\left(\ln \frac{p_t(x,y)}{p_t^X(x) p_t^Y(y)}\right)\right] \nu_t^X(x,y) p_t(x,y), \tag{84}$$

$$\dot{\mathcal{I}}^Y = \int dx \int dy \left[\frac{\partial}{\partial y}\left(\ln \frac{p_t(x,y)}{p_t^X(x) p_t^Y(y)}\right)\right] \nu_t^Y(x,y) p_t(x,y), \tag{85}$$

where $\sigma_{\mathrm{bath};t}^X$ ($\sigma_{\mathrm{bath};t}^Y$) is the entropy change of the system X (Y), $\sigma_{\mathrm{bath};t}^X$ ($\sigma_{\mathrm{bath};t}^Y$) is the entropy change of the heat bath attached to the system X (Y), and $\dot{\mathcal{I}}^X$ ($\dot{\mathcal{I}}^Y$) is information flow from X to Y (Y to X).

We explain the decomposition of the partial entropy production rates Eqs. (78) and (79). The entropy changes of the system X and Y are given by the differential entropy change,

$$\sigma_{\mathrm{sys};t}^X = \int dx \frac{\partial p_t^X(x)}{\partial t}[-\ln p_t^X(x)] = \frac{d}{dt} S_{\mathrm{sys}}^X, \tag{86}$$

$$S_{\mathrm{sys}}^X = \int dx [-p_t^X(x) \ln p_t^X(x))], \tag{87}$$

$$\sigma_{\mathrm{sys};t}^Y = \frac{d}{dt} S_{\mathrm{sys}}^X, \tag{88}$$

$$S_{\mathrm{sys}}^Y = \int dy [-p_t^Y(y) \ln p_t^Y(y)], \tag{89}$$

where we used the partial integral and the normalization of the probability $(d/dt)\int dx p_t^X(x) = 0$. The sum of the entropy changes of the heat bath gives the total entropy changes of the heat bathes

$$\sigma_{\mathrm{bath};t}^X + \sigma_{\mathrm{bath};t}^Y = \frac{1}{T}\int dx \int dy \frac{\partial p_t(x,y)}{\partial t}[-V_t(x,y)] = -\frac{1}{T}\frac{dQ}{dt}, \tag{90}$$

where we used the partial integral. The sum of information flows gives the change of the mutual information I between X and Y ,

$$\dot{\mathcal{I}}^X + \dot{\mathcal{I}}^Y = \int dx \int dy \frac{\partial p_t(x,y)}{\partial t}\left(\ln \frac{p_t(x,y)}{p_t^X(x) p_t^Y(y)}\right) = \frac{dI}{dt}, \tag{91}$$

$$I = \int dx \int dy p_t(x,y) \ln \frac{p_t(x,y)}{p_t^X(x) p_t^Y(y)}, \tag{92}$$

where we used the partial integral, the marginalization $\int dy p_t(x,y) = p_t^X(x)$ and $\int dx p_t(x,y) = p_t^Y(y)$, and the normalization of the probability $(d/dt)\int dx p_t^X(x) = 0$, $(d/dt)\int dx p_t^Y(x) = 0$, and $(d/dt)\int dx dy p_t(x,y) = 0$. Additionally, we obtain

$$\sigma_{\mathrm{sys};t}^X + \sigma_{\mathrm{sys};t}^Y - \dot{\mathcal{I}}^X - \dot{\mathcal{I}}^Y = \frac{dS_{\mathrm{sys}}}{dt}, \tag{93}$$

thus the sum of the partial entropy production rates gives the total entropy production rate.

The non-negativity of the partial entropy production rates gives the second laws of information thermodynamics for the subsystem [71–74, 76, 81],

$$\sigma_{\mathrm{bath};t}^X + \sigma_{\mathrm{sys};t}^X \geq \dot{\mathcal{I}}^X, \tag{94}$$

$$\sigma_{\mathrm{bath};t}^Y + \sigma_{\mathrm{sys};t}^Y \geq \dot{\mathcal{I}}^Y, \tag{95}$$

which implies that the entropy changes of the system and heat bath are bounded by information flow in the presence of the subsystem. These inequalities explains a conversion between information and thermodynamic quantities in the context of the Maxwell's demon. The sum of two inequalities

$$\sigma^X_{\mathrm{bath};t} + \sigma^X_{\mathrm{sys};t} - \dot{\mathcal{I}}^X + \sigma^Y_{\mathrm{bath};t} + \sigma^Y_{\mathrm{sys};t} - \dot{\mathcal{I}}^Y \quad \geq 0, \quad (96)$$

gives the second law of thermodynamics for the total system

$$\sigma_t \geq 0. \quad (97)$$

Based on the results Eqs. (64) and (66), we obtain tighter inequalities compared to the second law of information thermodynamics as follows

$$\sigma^X_{\mathrm{bath};t} + \sigma^X_{\mathrm{sys};t} \geq \dot{\mathcal{I}}^X + \lim_{\Delta t \to 0} \frac{\mathcal{W}(p^X_{t+\Delta t}, p^X_t)^2}{\mu T \Delta t^2} \geq \dot{\mathcal{I}}^X, \quad (98)$$

$$\sigma^Y_{\mathrm{bath};t} + \sigma^Y_{\mathrm{sys};t} \geq \dot{\mathcal{I}}^Y + \lim_{\Delta t \to 0} \frac{\mathcal{W}(p^Y_{t+\Delta t}, p^Y_t)^2}{\mu T \Delta t^2} \geq \dot{\mathcal{I}}^Y. \quad (99)$$

Thus, the entropy changes of the system and heat bath are tightly bounded by both information flow and the L^2-Wasserstein distance.

We now consider the case $\sigma^{\mathrm{rot}}_t = 0$. Because the sum of the partial entropy production rates gives the total entropy production rate, the sum of two tighter inequalities gives non-negativity of a measure $I^{\mathcal{W}}$,

$$\lim_{\Delta t \to 0} \frac{I^{\mathcal{W}}}{\Delta t^2} \geq 0, \quad (100)$$

$$I^{\mathcal{W}} = \mathcal{W}(p_{t+\Delta t}, p_t)^2 - \mathcal{W}(p^X_{t+\Delta t}, p^X_t)^2 - \mathcal{W}(p^Y_{t+\Delta t}, p^Y_t)^2. \quad (101)$$

The equality holds when

$$\nu^X_t(x,y) = \bar{\nu}^X_t(x), \nu^Y_t(x,y) = \bar{\nu}^Y_t(x). \quad (102)$$

The measure $I^{\mathcal{W}}$ quantifies both the statistical independence and the independence of the potential, while the mutual information I only quantifies the statistical independence. Thus, $I^{\mathcal{W}}$ could be an interesting measure of the independence between two systems when stochastic dynamics of two systems are driven by the Fokker-Planck equation, and its non-negativity is decomposed by tighter inequalities of information thermodynamics Eqs. (98) and (99).

VIII. STOCHASTIC HEAT ENGINE

Let us consider a stochastic heat engine [75] driven by the potential V_t that is not quasi-static. The cycle of a stochastic engine consists of the following four steps (see also Fig. 3).

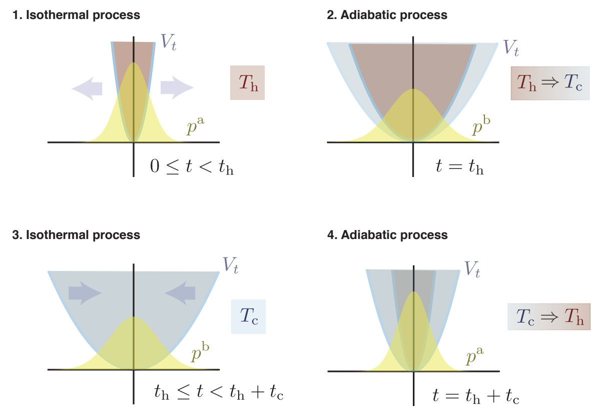

FIG. 3. An example of a stochastic heat engine. Because the initial state at time $t = 0$ and the final state at time $t = t_{\mathrm{h}} + t_{\mathrm{c}}$ are same, the four steps gives the cycle of a stochastic heat engine. The work $-W_{\mathrm{h}}$ is extracted during time $0 \leq t < t_{\mathrm{h}}$, and the work W_{c} is done during time $t_{\mathrm{h}} \leq t < t_{\mathrm{h}} + t_{\mathrm{c}}$. The total amount of the work through one cycle $-W = -W_{\mathrm{h}} + W_{\mathrm{c}} > 0$ is extracted.

1. An isothermal process of varying the potential $V_t(\mathbf{x})$ during time $0 \leq t < t_{\mathrm{h}}$ at temperature T_{h}. During this step, the probability distribution changes from p^{a} to p^{b}, and the entropy change of the system is given by $\Delta S := \int d\mathbf{x} p^{\mathrm{a}}(\mathbf{x}) \ln p^{\mathrm{a}}(\mathbf{x}) - \int d\mathbf{x} p^{\mathrm{b}}(\mathbf{x}) \ln p^{\mathrm{b}}(\mathbf{x})$. In this step, the work is extracted $-W_{\mathrm{h}} := \int_0^{t_{\mathrm{h}}} dt (dW/dt) > 0$ for the external system.

2. The temperature is changed from T_{h} to $T_{\mathrm{c}} (< T_{\mathrm{h}})$ instantaneously at time $t = t_{\mathrm{h}}$. During this time, the distribution p^{b} does not change. Therefore, the entropy of the system also did not change, and this step can be interpreted as an adiabatic process.

3. An isothermal process that returns the potential $V_{t_{\mathrm{h}}}(\mathbf{x})$ to $V_0(\mathbf{x}) = V_{t_{\mathrm{h}}+t_{\mathrm{c}}}(\mathbf{x})$ during time $t_{\mathrm{h}} \leq t < t_{\mathrm{h}} + t_{\mathrm{c}}$ at temperature T_{c}. During this step, the probability distribution changes from p^{b} to p^{a}, and the entropy change of the system is $-\Delta S$. In this step, the system is assumed to be given work $W_{\mathrm{c}} := \int_{t_{\mathrm{h}}}^{t_{\mathrm{h}}+t_{\mathrm{c}}} dt (dW/dt) > 0$ by the external system.

4. The temperature is changed from T_{c} to T_{h} instantaneously at time $t = t_{\mathrm{h}} + t_{\mathrm{c}}$. During this time, the distribution does not change. Therefore, the entropy of the system also did not change, and this step can be interpreted as an adiabatic process.

If we consider the harmonic potential and the initial distribution p^{a} is Gaussian, thermodynamic quantities such as the entropy change and the work are calculated, and we can find an optimal protocol to minimize the entropy production can be obtained analytically [75]. As we shown in the section IX, $\sigma_t^{\rm rot}=0$ and the entropy production rate is proportional to the action.

Here we consider a general case that the potential is not necessarily harmonic and the probability distribution at time t is not necessarily Gaussian. When the time $t_{\rm h}$ and $t_{\rm c}$ are long enough and the potential $V_t(\mathbf{x})$ is a harmonic potential, the efficiency of the heat engine becomes the Carnot efficiency asymptotically, and the heat engine can be considered as a stochastic extension of the Carnot cycle. The extracted work of the heat engine through the one cycle is

$$-W := W_{\rm h} - W_{\rm c} = (T_{\rm h}-T_{\rm c})\Delta S - T_{\rm h}\Sigma_{\rm h} - T_{\rm c}\Sigma_{\rm c}, \tag{103}$$

where $\Sigma_{\rm h} := \int_0^{t_{\rm h}} dt \sigma_t$ is entropy production in the isothermal step 1 at temperature $T_{\rm h}$ and $\Sigma_{\rm c} := \int_{t_{\rm h}}^{t_{\rm h}+t_{\rm c}} dt\sigma_t$ is entropy production in the isothermal step 3 at temperature T_c. If we assumed that the extracted work is positive $-W>0$, the condition $\Delta S \geq 0$ should be needed because of the second law of thermodynamics $\Sigma_{\rm h}\geq 0$ and $\Sigma_{\rm c}\geq 0$.

By using Eq. (51), we can obtain the following inequality for the extracted work $-W$,

$$-W \leq (T_{\rm h}-T_{\rm c})\Delta S - \frac{\mathcal{W}(p^{\rm a},p^{\rm b})^2}{\mu t_r}, \tag{104}$$

$$\frac{1}{t_{\rm r}} := \frac{1}{t_{\rm h}} + \frac{1}{t_{\rm c}}, \tag{105}$$

where $t_{\rm r}$ is called the reduced time. When we impose the positive extracted work in the whole cycle, i.e., $-W>0$, we obtain the following inequality for the reduced time $t_{\rm r}$ from Eq.(103),

$$\frac{1}{t_{\rm r}} \leq \frac{\mu(T_{\rm h}-T_{\rm c})\Delta S}{\mathcal{W}(p^{\rm a},p^{\rm b})^2}. \tag{106}$$

This inequality implies that the reduced time in the engine is generally bounded by the entropy change and the the L^2-Wasserstein distance $\mathcal{W}(p^{\rm a},p^{\rm b})$, which are given by the initial distribution p_a and the final distribution p_b.

Because the efficiency of the heat engine η is defined as

$$\eta = \frac{-W}{T_{\rm h}\Delta S - T_{\rm h}\Sigma_{\rm h}}. \tag{107}$$

Because the second law of thermodynamics $\Sigma_{\rm h}+\Sigma_{\rm c}\geq 0$ holds, we obtain the fact that the efficiency is generally bounded by the Carnot efficiency η_C [93],

$$\eta \leq \frac{T_{\rm h}-T_{\rm c}}{T_{\rm h}} := \eta_{\rm C}. \tag{108}$$

When we considered the situation that the entropy production is minimized as follows

$$T_{\rm c}\Sigma_{\rm c} = \frac{\mathcal{W}(p^{\rm a},p^{\rm b})^2}{\mu t_{\rm c}}, \tag{109}$$

$$T_{\rm h}\Sigma_{\rm h} = \frac{\mathcal{W}(p^{\rm a},p^{\rm b})^2}{\mu t_{\rm h}}, \tag{110}$$

the efficiency η is given by

$$\eta = \frac{T_{\rm h}-T_{\rm c} - \frac{\mathcal{W}(p^{\rm a},p^{\rm b})^2}{\mu\Delta S t_{\rm r}}}{T_{\rm h} - \frac{\mathcal{W}(p^{\rm a},p^{\rm b})^2}{\mu\Delta S t_h}}, \tag{111}$$

and reaches to the Carnot efficiency $\eta_{\rm C}$ in the limit $t_{\rm h}\to\infty$ and $t_{\rm c}\to\infty$. This fact is also discussed in Ref. [69]. In the limit $t_{\rm h}\to\infty$ and $t_{\rm c}\to\infty$. the square of the L^2-Wasserstein distance plays the same role as the irreversible "action" $A_{\rm irr}$ in Ref. [75].

When $\sigma_t^{\rm rot}=0$, we obtain a geometric interpretation of the efficiency from Eq. (43),

$$\eta = \frac{T_{\rm h}-T_{\rm c} - \frac{1}{\mu\Delta S}\int_0^{t_{\rm h}+t_{\rm c}} dt\left(\frac{d\mathcal{L}_t}{dt}\right)^2}{T_{\rm h} - \frac{1}{\mu\Delta S}\int_0^{t_{\rm h}} dt\left(\frac{d\mathcal{L}_t}{dt}\right)^2}. \tag{112}$$

In this case, we obtain a lower bound on the efficiency

$$\eta_{\rm C} - \frac{2\mathcal{C}}{\mu\Delta S T_{\rm h}} \leq \eta \leq \eta_{\rm C}, \tag{113}$$

where $\mathcal{C}=(1/2)\int_0^{t_{\rm h}+t_{\rm c}} dt(d\mathcal{L}_t/dt)^2$ is the action measured by the L^2-Wasserstein distance. The efficiency η can reach to the Carnot efficiency η_C when the ratio between the action and the Shannon entropy change $\mathcal{C}/\Delta S$ converges to zero. In general $\sigma_t^{\rm rot}\neq 0$ and this lower bound by the action is generally violated, especially for the case of the non-potential force. Thus, the quantity $\sigma_t^{\rm rot}$ might play an important role in a stochastic heat engine with a non-potential force.

IX. EXAMPLE: BROWINAN PARTICLE IN HARMONIC POTENTIAL

We here show the case of a Browinan particle in harmonic potential as an example of stochastic thermodyanmics based on L^2-Wassserstein distance. In this case, we can show $\sigma^{\rm rot}=0$, and obtain the protocol of minimizing the entropy production analytically. In terms of the Langevin equation, the time evolution of the position $x(t)$ at time t is given by

$$\frac{dx(t)}{dt} = -\mu\frac{\partial V_t(x)}{\partial x} + \sqrt{2\mu T}\xi(t), \tag{114}$$

with the harmonic potential

$$V_t(x) = \frac{1}{2}k_t(x-a_t)^2, \tag{115}$$

where $\xi(t)$ is the Gaussian noise with the mean $\langle\xi(t)\rangle=0$ and the variance $\langle\xi(t)\xi(t')\rangle=\delta(t-t')$. This Langevin equation corresponds to the Fokker-Planck equation [94],

$$\frac{\partial p_t(x)}{\partial t} = -\frac{\partial}{\partial x}(\nu_t(x)p_t(x)), \tag{116}$$

$$\nu_t(x) := -\mu\frac{\partial}{\partial x}[V_t(x) + T\ln p_t(x)]. \tag{117}$$

We now assume that the probability distribution at the initial time is Gaussian. For the harmonic potential, the probability distribution at time t is Gaussian if the probability distribution at the initial time is Gaussian,

$$p_t(x) = \frac{1}{\sqrt{2\pi \mathrm{Var}[x]_t}} \exp\left(-\frac{(x-\mathrm{E}[x]_t)^2}{2\mathrm{Var}[x]_t}\right), \tag{118}$$

$$\mathrm{E}[x]_t = \int dx x p_t(x), \tag{119}$$

$$\mathrm{Var}[x]_t = \int dx x^2 p_t(x) - (\mathrm{E}[x]_t)^2. \tag{120}$$

For this Fokker-Planck equation, the time evolution of $\mathrm{E}[x]_t$ and $\mathrm{Var}[x]_t$ are given by

$$\frac{d}{dt}\mathrm{E}[x]_t = \mu k_t (a_t - \mathrm{E}[x]_t), \tag{121}$$

$$\frac{d}{dt}\mathrm{Var}[x]_t = -2\mu \left(k_t \mathrm{Var}[x]_t - T\right). \tag{122}$$

Therefore, the mean local velocity $\nu_t(x)$ is analytically calculated as

$$\nu_t(x) = -\mu k_t (\mathrm{E}[x]_t - a_t) + \left(\frac{\mu T}{\mathrm{Var}[x]_t} - \mu k_t\right)(x - \mathrm{E}[x]_t), \tag{123}$$

and the entropy production rate is also calculated as

$$\begin{aligned} \sigma_t &= \frac{1}{\mu T}\int dx |\nu_t(x)|^2 p_t(x) \qquad (124)\\ &= \frac{\mu}{T}\left\{\left(k_t - \frac{T}{\mathrm{Var}[x]_t}\right)^2 \mathrm{Var}[x]_t + k_t^2 (\mathrm{E}[x]_t - a_t)^2\right\}. \qquad (125) \end{aligned}$$

The Wasserstein distance can be concretely calculated for the Gaussian distribution [95, 96]. For two probability distributions

$$p^{\mathrm{a}}(x) = \frac{1}{\sqrt{2\pi \mathrm{Var}[x]^{\mathrm{a}}}} \exp\left(-\frac{(x-\mathrm{E}[x]^{\mathrm{a}})^2}{2\mathrm{Var}[x]^{\mathrm{a}}}\right) \tag{126}$$

and

$$p^{\mathrm{b}}(x) = \frac{1}{\sqrt{2\pi \mathrm{Var}[x]^{\mathrm{b}}}} \exp\left(-\frac{(x-\mathrm{E}[x]^{\mathrm{b}})^2}{2\mathrm{Var}[x]^{\mathrm{b}}}\right), \tag{127}$$

the L^2-Wasserstein distance can be written as follows

$$\mathcal{W}(p^{\mathrm{a}}, p^{\mathrm{b}})^2 = (\mathrm{E}[x]^{\mathrm{a}} - \mathrm{E}[x]^{\mathrm{b}})^2 + \left(\sqrt{\mathrm{Var}[x]^{\mathrm{a}}} - \sqrt{\mathrm{Var}[x]^{\mathrm{b}}}\right)^2. \tag{128}$$

This L^2-Wasserstein distance is also known as the Fréchet distance [97]. Thus, we can confirm $\sigma^{\mathrm{rot}} = 0$ as follows

$$\begin{aligned} \left(\frac{d\mathcal{L}_t}{dt}\right)^2 &= \lim_{\Delta t \downarrow 0} \frac{\mathcal{W}(p_t, p_{t+\Delta t})^2}{\Delta t^2} \\ &= \left(\frac{d\mathrm{E}[x]_t}{dt}\right)^2 + \left(\frac{d\sqrt{\mathrm{Var}[x]_t}}{dt}\right)^2 \\ &= \mu^2 \left\{\frac{(k_t \mathrm{Var}[x]_t - T)^2}{\mathrm{Var}[x]_t} + k_t^2 (\mathrm{E}[x]_t - a_t)^2\right\} \\ &= \mu T \sigma_t. \end{aligned} \tag{129}$$

We also can see that the entropy production Σ is minimized if Eq. (53) holds. The minimum value of the entropy production Σ for fixed p_0 and p_τ is calculated as

$$\Sigma = \frac{\int_0^\tau dt \left[\left(\frac{d\mathrm{E}[x]_t}{dt}\right)^2 + \left(\frac{d\sqrt{\mathrm{Var}[x]_t}}{dt}\right)^2\right]}{\mu T} \tag{130}$$

$$\geq \frac{\left(\int_{t=0}^{t=\tau} d\mathrm{E}[x]_t\right)^2 + \left(\int_{t=0}^{t=\tau} d\sqrt{\mathrm{Var}[x]_t}\right)^2}{\mu T \tau} \tag{131}$$

$$= \frac{(\mathrm{E}[x]_\tau - \mathrm{E}[x]_0)^2 + \left(\sqrt{\mathrm{Var}[x]_\tau} - \sqrt{\mathrm{Var}[x]_0}\right)^2}{\mu T \tau}, \tag{132}$$

where we used the Cauchy-Schwarz inequality $\tau \int_0^\tau dt (d\theta_t/dt)^2 \geq (\int_0^\tau dt (d\theta_t/dt))^2$ with $\theta_t = \mathrm{E[x]_t}$ and $\theta_t = \sqrt{\mathrm{Var}[x]_t}$. The minimum value can be achieved if $d\theta_t/dt$ is constant. This condition of the minimum value can be rewritten as

$$\mathrm{E}[x]_t = \left(1 - \frac{t}{\tau}\right)\mathrm{E}[x]_0 + \frac{t}{\tau}\mathrm{E}[x]_\tau \tag{133}$$

$$\sqrt{\mathrm{Var}[x]_t} = \left(1 - \frac{t}{\tau}\right)\sqrt{\mathrm{Var}[x]_0} + \frac{t}{\tau}\sqrt{\mathrm{Var}[x]_\tau}, \tag{134}$$

or equivalently,

$$\frac{d\mathrm{E}[x]_t}{dt} = \frac{\mathrm{E}[x]_\tau - \mathrm{E}[x]_0}{\tau}, \tag{135}$$

$$\frac{d\sqrt{\mathrm{Var}[x]_t}}{dt} = \frac{\sqrt{\mathrm{Var}[x]_\tau} - \sqrt{\mathrm{Var}[x]_0}}{\tau}. \tag{136}$$

Under this condition, $\mathcal{W}(p_0, p_\tau)/\tau$ is calculated as

$$\begin{aligned} &\frac{\mathcal{W}(p_0, p_\tau)}{\tau} \\ &= \frac{1}{\tau}\sqrt{(\mathrm{E}[x]_\tau - \mathrm{E}[x]_0)^2 + \left(\sqrt{\mathrm{Var}[x]_\tau} - \sqrt{\mathrm{Var}[x]_0}\right)^2} \\ &= \sqrt{\left(\frac{d\mathrm{E}[x]_t}{dt}\right)^2 + \left(\frac{d\sqrt{\mathrm{Var}[x]_t}}{dt}\right)^2} \\ &= \frac{d\mathcal{L}_t}{dt}, \end{aligned} \tag{137}$$

which is the condition that the probability distribution changes at a constant rate on a straight line as measured by the L^2-Wasserstein distance Eq. (53). By comparing Eqs. (135) and (136) with (121) and (122), the optimal protocol that minimizes the entropy production is given by

$$\mu k_t(a_t - \mathrm{E}[x]_t) = \frac{\mathrm{E}[x]_\tau - \mathrm{E}[x]_0}{\tau}, \tag{138}$$

$$-\mu\left(k_t \mathrm{Var}[x]_t - T\right) = \sqrt{\mathrm{Var}[x]_t}\frac{\sqrt{\mathrm{Var}[x]_\tau} - \sqrt{\mathrm{Var}[x]_0}}{\tau}. \tag{139}$$

In terms of the parameters of the harmonic potential $V_t(x)$, the optimal protocol that minimizes the entropy production is given by

$$k_t = T - \frac{\sqrt{\mathrm{Var}[x]_\tau} - \sqrt{\mathrm{Var}[x]_0}}{\mu\tau\sqrt{\mathrm{Var}[x]_t}}, \tag{140}$$

$$a_t = \mathrm{E}[x]_t + \frac{\mathrm{E}[x]_\tau - \mathrm{E}[x]_0}{k_t \mu\tau}. \tag{141}$$

Thus, we obtain k_t and a_t which realizes such an optimal protocol in practice

$$k_t = T - \frac{\sqrt{\mathrm{Var}[x]_\tau} - \sqrt{\mathrm{Var}[x]_0}}{\mu\left[\tau\sqrt{\mathrm{Var}[x]_0} + t(\sqrt{\mathrm{Var}[x]_\tau} - \sqrt{\mathrm{Var}[x]_0})\right]}, \tag{142}$$

$$a_t = \left(1 - \frac{t}{\tau}\right)\mathrm{E}[x]_0 + \frac{t}{\tau}\mathrm{E}[x]_\tau + \frac{\mathrm{E}[x]_\tau - \mathrm{E}[x]_0}{k_t\mu\tau}. \tag{143}$$

If we assume that k_t is always non-negative, the following inequality

$$\tau \geq \frac{1 - t\mu T}{\mu T}\frac{\sqrt{\mathrm{Var}[x]_\tau} - \sqrt{\mathrm{Var}[x]_0}}{\sqrt{\mathrm{Var}[x]_0}} \tag{144}$$

$$\geq \frac{1}{\mu T}\frac{\sqrt{\mathrm{Var}[x]_\tau} - \sqrt{\mathrm{Var}[x]_0}}{\sqrt{\mathrm{Var}[x]_0}}. \tag{145}$$

must hold for this optimal protocol. The results show that when the variance gets smaller, i.e., $\mathrm{Var}[x]_\tau < \mathrm{Var}[x]_0$, we can use this optimal protocol for all $\tau > 0$, but when the variance gets larger, i.e., $\mathrm{Var}[x]_\tau \geq \mathrm{Var}[x]_0$, there is a limit to the time τ for the process to achieve this optimal protocol.

X. DISCUSSION

We show a geometrical feature of stochastic thermodynamics for the Fokker-Planck equation based on the L^2-Wasserstein distance. As shown in this paper, the L^2-Wasserstein distance is strongly related to the entropy production in stochastic thermodynamics for the Fokker-Planck equation. Thus, based on L^2-Wasserstein distance, we can consider a differential geometry of stochastic thermodynamics for the Fokker-Planck equation, which is closely related to the entropy production.

It might be interesting to consider a relation between the L^2-Wasserstein distance and the Fisher information matrix because the Fisher information matrix gives a metric in information geometry, which is a possible choice of differential geometry of stochastic thermodynamics. For example, the entropy production is also given by the projection in information geometry. Thus, there might be a deep connection between information geometry and optimal transport by the L^2-Wasserstein distance. For example, the HWI inequality, the logarithmic Sobolev inequalities, and the Talagrand inequalities are considered as a trade-off relation among the L^2-Wasserstein distance, the relative Fisher information, and the Shannon entropy [29, 33]. As shown in Ref. [11], we have analogy between the entropy production rate and the Fisher information of time for the Fokker-Planck equation. This analogy is also pointed out in Ref. [98]. Moreover, the entropy production is also obtained from the projection in information geometry [10]. Thus, we might unify two directions of researches of information geometry and the L^2-Wasserstein distance for the Fokker-Planck equation based on the entropy production. The unification of information geometry and geometry of the L^2-Wasserstein distance has been recently discussed [39, 40], and our results might provide a new direction in this topic.

If we consider thermodynamics based on information geometry, we can consider not only stochastic thermodynamics for the Fokker-Planck equation [11] but also stochastic thermodynamics for the Markov jump process [9] and chemical thermodynamics for the rate equation [18]. Thus, it might be interesting to seek a correspondence of the L^2-Wasserstein distance for the Markov jump process and the rate equation. Indeed, Y. Hasegawa and T. Van Vu derived a generalization of thermodynamic speed limits for the Markov jump process [70], then a thermodynamic correspondence of the L^2-Wasserstein distance for the Markov jump process might be the distance discussed in Ref. [70]. In nonequilibrium steady state, the quantity σ_t^{rot} might play an important role, and its physical meaning is interesting. Under the existence of the non-potential force, the entropy production rate is generally decomposed into two non-negative parts, the Wasserstein part $(d\mathcal{L}_t/dt)^2/(\mu T)$ and the rotational part σ_t^{rot}. This fact is very similar to the case of the steady state thermodynamics [91], where the entropy production is decomposed into the excess entropy production and the house keeping heat.

ACKNOWLEDGMENTS

We thank Shun-ichi Amari, Masafumi Oizumi, Kohei Yoshimura, Andreas Dechant, Shin-ichi Sasa, Tan Van Vu and Hideo Higuchi for fruitful discussions. We also thank Andreas Dechant and Shin-ichi Sasa for critical comments to the previous vesrion of this manuscript.

Sosuke Ito is supported by JSPS KAKENHI Grant No. 19H05796, JST Presto Grant No. JPMJPR18M2 and UTEC-UTokyo FSI Research Grant Program.

Appendix A: Proof of the formula Eq. (34)

To obtain the formula Eq. (34), we introduce the map $\mathcal{M}_{t\to s}$ for the trajectory of the particle according to the Fokker-Planck equation from time t to time s. The map $\mathcal{M}_{t\to t+s}$ is given by the following differential equations for $s \geq 0$,

$$\frac{d}{ds}\mathcal{M}_{t\to t+s}(\mathbf{x}) = \boldsymbol{\nu}_{t+s}(\mathcal{M}_{t\to t+s}(\mathbf{x})), \tag{A1}$$

with the initial condition $\mathcal{M}_{t\to t}(\mathbf{x}) := \mathbf{x}$. The map $\mathcal{M}_{t\to t-s}$ for $s \geq 0$ is also given by

$$\frac{d}{dt}\mathcal{M}_{t\to t-s}(\mathbf{x}) = -\boldsymbol{\nu}_{t-s}(\mathcal{M}_{t\to t-s}(\mathbf{x})). \tag{A2}$$

with the initial condition $\mathcal{M}_{t\to t}(\mathbf{x}) := \mathbf{x}$. These differential equations correspond to the Lagrangian descriptions of the Fokker-Planck equation as a continuity equation. Because the composite map $\mathcal{M}_{t\to t+s} \circ \mathcal{T}_t(\mathbf{x}) = \mathcal{M}_{t\to t+s}(\mathcal{T}_t(\mathbf{x}))$ is a non-optimal transport plan from p to p_{t+s}, we obtain the inequality

$$\begin{aligned}\mathcal{W}(p, p_{t+s})^2 &= \int d\mathbf{x}\, \|\mathbf{x} - \mathcal{T}_{t+s}(\mathbf{x})\|^2 p(\mathbf{x}) \\ &\leq \int d\mathbf{x} \|\mathbf{x} - \mathcal{M}_{t\to t+s}(\mathcal{T}_t(\mathbf{x}))\|^2 p(\mathbf{x}).\end{aligned} \tag{A3}$$

By using Eqs. (A1) and (A3), we obtain

$$\begin{aligned}&\frac{d}{ds}\left(\frac{\mathcal{W}(p, p_{t+s})^2}{2}\right)\bigg|_{s=0} \\ &= \lim_{s\downarrow 0}\frac{1}{s}\left(\frac{\mathcal{W}(p, p_{t+s})^2}{2} - \frac{\mathcal{W}(p, p_t)^2}{2}\right) \\ &\leq \int d\mathbf{x} p(\mathbf{x})\left[\lim_{s\downarrow 0}\frac{\|\mathbf{x} - \mathcal{M}_{t\to t+s}(\mathcal{T}_t(\mathbf{x}))\|^2 - \|\mathbf{x} - \mathcal{T}_t(\mathbf{x})\|^2}{2s}\right] \\ &= -\int d\mathbf{x}(\mathbf{x} - \mathcal{T}_t(\mathbf{x}))\cdot \boldsymbol{\nu}_t(\mathcal{T}_t(\mathbf{x}))p(\mathbf{x}).\end{aligned} \tag{A4}$$

Similarly, we obtain

$$\begin{aligned}&\frac{d}{ds}\left(\frac{\mathcal{W}(p, p_{t+s})^2}{2}\right)\bigg|_{s=0} \\ &= \lim_{s\downarrow 0}\frac{1}{s}\left(\frac{\mathcal{W}(p, p_{t+s})^2}{2} - \frac{\mathcal{W}(p, p_t)^2}{2}\right) \\ &\geq \int d\mathbf{x} p(\mathbf{x})\left[\lim_{s\downarrow 0}\frac{\|\mathbf{x} - \mathcal{T}_{t+s}(\mathbf{x})\|^2 - \|\mathbf{x} - \mathcal{M}_{t+s\to t}(\mathcal{T}_{t+s}(\mathbf{x}))\|^2}{2s}\right] \\ &= -\int d\mathbf{x}(\mathbf{x} - \mathcal{T}_t(\mathbf{x}))\cdot \boldsymbol{\nu}_t(\mathcal{T}_t(\mathbf{x}))p(\mathbf{x}),\end{aligned} \tag{A5}$$

because the composite map $\mathcal{M}_{t+s\to t} \circ \mathcal{T}_{t+s}$ is a non-optimal transport plan from p to p_t. From Eqs. (A4) and (A5), we finally obtain the formula Eq. (34).

[1] T. M. Cover and J. A. Thomas, *Elements of information theory* (John Wiley & Sons , 2012).
[2] Jaynes, E. T. *Information theory and statistical mechanics.* Physical Review, **106**, 620 (1957).
[3] K. Sekimoto, *Stochastic energetics* (Springer, 2010).
[4] R. Kawai, J. M. R. Parrondo and C. Van den Broeck, *Dissipation: The phase-space perspective*, Phys. Rev. Lett. **98**, 080602 (2007).
[5] U. Seifert, *Stochastic thermodynamics, fluctuation theorems and molecular machines*, Rep. Prog. Phys. **75**, 126001 (2012).
[6] U. Seifert, *Stochastic thermodynamics, fluctuation theorems and molecular machines*, Rep. Prog. Phys. **75**, 126001 (2012).
[7] C. R. Rao, *Information and the accuracy attainable in the estimation of statistical parameters*, Bullet. Calcutta Math. Soc. **37**, 81 (1945).
[8] S.-i. Amari, H. Nagaoka, *Methods of Information Geometry*, Vol. 191 (Oxford University Press, New York, 2000).
[9] S. Ito, *Stochastic thermodynamic interpretation of information geometry*, Phys. Rev. Lett. **121**, 030605 (2018).
[10] S. Ito, M. Oizumi, and S. I. Amari, *Unified framework for the entropy production and the stochastic interaction based on information geometry.* Physical Review Research, **2**, 033048 (2020).
[11] S. Ito and A. Dechant, *Stochastic time evolution, information geometry, and the Cramér-Rao bound.* Physical Review X, **10**, 021056 (2020).
[12] K. Ashida and K. Oka, K. *Stochastic thermodynamic limit on E. coli adaptation by information geometric approach.* Biochemical and biophysical research communications, **508**, 690-694 (2019).
[13] T. Nakamura, H. H. Hasegawa, and D. J. Driebe, *Reconsideration of the generalized second law based on information geometry.* Journal of Physics Communications, **3**, 015015 (2019).
[14] S. Ito, *Thermodynamics of information geometry as a generalization of the Glansdorff-Prigogine criterion for stability*, arXiv:1908.09446 (2019).
[15] N. Shiraishi and K. Saito, *Information-theoretical bound of the irreversibility in thermal relaxation processes.* Physical review letters, **123**, 110603 (2019).

[16] D. Gupta and D. M. Busiello, *Tighter thermodynamic bound on the speed limit in systems with unidirectional transitions.* Physical Review E, **102**, 062121 (2020).

[17] T. Parr, L. Da Costa, and K. Friston, K. *Markov blankets, information geometry and stochastic thermodynamics.* Philosophical Transactions of the Royal Society A, **378**, 20190159 (2020).

[18] K. Yoshimura, and S. Ito, *Information geometric inequalities of chemical thermodynamics*, Physical Review Research, **3**, 013175 (2021).

[19] S. B. Nicholson, L. P. Garcia-Pintos, A. del Campo, and J. R. Green, *Time–information uncertainty relations in thermodynamics.* Nature Physics, **16**, 1211-1215 (2020)

[20] K. Ashida, K. Aoki, and S. Ito, *Experimental evaluation of thermodynamic cost and speed limit in living cells via information geometry.* bioRxiv. (2020). doi: https://doi.org/10.1101/2020.11.29.403097

[21] F. Weinhold, *Metric geometry of equilibrium thermodynamics*, J. Chem. Phys. **63**, 2479 (1975).

[22] G. Ruppeiner, *Riemannian geometry in thermodynamic fluctuation theory*, Rev. Mod. Phys. **67**, 605 (1995).

[23] P. Salamon and R. S. Berry, *Thermodynamic Length and Dissipated Availability*, Phys. Rev. Lett. **51**, 1127 (1983).

[24] G. E. Crooks, *Measuring thermodynamic length*, Phys. Rev. Lett. **99**, 100602 (2007).

[25] D. A. Sivak and G. E. Crooks, *Thermodynamic Metrics and Optimal Paths*, Phys. Rev. Lett. **108**, 190602 (2012).

[26] G. M. Rotskoff and G. E. Crooks, *Optimal control in nonequilibrium systems: Dynamic Riemannian geometry of the Ising model*, Phys. Rev. E, **92**, 060102(R) (2015).

[27] M. Aguilera, S. A. Moosavi, and H. Shimazaki, *A unifying framework for mean-field theories of asymmetric kinetic Ising systems.* Nature Communications, **12**, 1-12 (2021).

[28] C. Villani, *Topics in optimal transportation.* American Mathematical Soc. (2003).

[29] C. Villani, *Optimal transport: old and new.* (Springer, 2008).

[30] R. Jordan, D. Kinderlehrer, and F. Otto, *The Variational Formulation of the Fokker–Planck Equation*,SIAM J. MATH. ANAL, **29**, 1, 1 (1998).

[31] P.A. Markowich, and C. Villani.*On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis.* Mat. Contemp, **19**, 1 (2000).

[32] C. Villani, *Entropy production and convergence to equilibrium.* In Entropy methods for the Boltzmann equation. Springer, Berlin, Heidelberg (2008).

[33] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality.* Journal of Functional Analysis, **173**, 361-400 (2000).

[34] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter, *On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations.* Comm. Partial Differential Equations **26**, 43 (2001).

[35] E. Aurell, C. Mejía-Monasterio, and P. Muratore-Ginanneschi, *Optimal protocols and optimal transport in stochastic thermodynamics.* Physical review letters, **106**, 250601 (2011).

[36] E. Aurell, K. Gawedzki, C. Mejía-Monasterio, R. Mohayaee, and P. Muratore-Ginanneschi, *Refined second law of thermodynamics for fast random processes.* Journal of statistical physics, **147**, 487 (2012).

[37] J. D. Benamou and Y. Brenier. *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem.* Numerische Mathematik 84.3: 375-393 (2000).

[38] A. Dechant., and Y. Sakurai, *Thermodynamic interpretation of Wasserstein distance.* arXiv preprint arXiv:1912.08405.(2019).

[39] S. I. Amari, R. Karakida, and M. Oizumi, *Information geometry connecting Wasserstein distance and Kullback–Leibler divergence via the entropy-relaxed transportation problem.* Information Geometry, **1**, 13 (2018).

[40] S. I. Amari, R. Karakida, M. Oizumi, and M. Cuturi, *Information geometry for regularized optimal transport and barycenters of patterns.* Neural computation, **31**, 827 (2019).

[41] J. M. Horowitz and T. R. Gingrich, *Thermodynamic uncertainty relations constrain non-equilibrium fluctuations*, Nat. Phys., **16**, 15 (2019).

[42] A. C. Barato and U. Seifert, *Thermodynamic uncertainty relation for biomolecular processes*, Phys. Rev. Lett. **114**, 158101 (2015).

[43] P. Pietzonka, A. C. Barato and U. Seifert, *Universal bounds on current fluctuations*, Phys. Rev. E, **93**, 052145 (2016).

[44] T. R. Gingrich, J. M. Horowitz, N. Perunov and J. L. England, *Dissipation bounds all steady-state current fluctuations*, Phys. Rev. Lett. **116**, 120601 (2016).

[45] M. Polettini, A. Lazarescu and M. Esposito, *Tightening the uncertainty principle for stochastic currents*, Phys. Rev. E, **94**, 052104 (2016).

[46] C. Maes, *Frenetic bounds on the entropy production*, Phys. Rev. Lett. **119**, 160601 (2017).

[47] J. M. Horowitz and T. R. Gingrich, *roof of the finite-time thermodynamic uncertainty relation for steady-state currents*, Phys. Rev. E, **96**, 020103(R) (2017).

[48] K. Proesmans and C. Van den Broeck, *Discrete-time thermodynamic uncertainty relation*, Europhys. Lett., **119**, 20001 (2017).

[49] P. Pietzonka, F. Ritort, and U. Seifert, *Finite-time generalization of the thermodynamic uncertainty relation.* Physical Review E, **96**, 012101 (2017).

[50] A. Dechant and S. I. Sasa, *Current fluctuations and transport efficiency for general Langevin systems*, J. Stat. Mech. **2018**, 063209 (2018).

[51] A. Dechant, *Multidimensional thermodynamic uncertainty relations*, J. Phys. A, **52**, 035001 (2018).

[52] Y. Hasegawa, and T. Van Vu, *Uncertainty relations in stochastic processes: An information inequality approach*, Phys. Rev. E, **99**, 062126 (2019).

[53] A. M. Timpanaro, G. Guarnieri, J. Goold, and G. T. Landi, *Thermodynamic uncertainty relations from exchange fluctuation theorems.* Physical review letters, **123**, 090604 (2019).

[54] Y. Hasegawa and T. Van Vu *Fluctuation theorem uncertainty relation.* Physical review letters, **123**, 110602 (2019).

[55] J. Li, J. M. Horowitz, T. R. Gingrich, and N. Fakhri, *Quantifying dissipation using fluctuating currents.* Nature communications, **10**, 1 (2019).

[56] S. K. Manikandan, D. Gupta, D and S. Krishnamurthy, *Inferring entropy production from short experiments.* Physical review letters, 124(12), 120603 (2020).

[57] T. Van Vu and Y. Hasegawa, *Entropy production estimation with optimal current.* Physical Review E, **101**, 042138 (2020).

[58] S. Otsubo, S. Ito, A. Dechant, T. Sagawa, *Estimating entropy production by machine learning of short-time fluctuating currents*, to appear in Phys. Rev. E (2020).

[59] A. Dechant, and S. I. Sasa, *Fluctuation–response inequality out of equilibrium.*Proceedings of the National Academy of Sciences, **117**, 6430-6436 (2020).

[60] D. H. Wolpert, Uncertainty relations and fluctuation theorems for Bayes nets. Physical Review Letters, **125**, 200602 (2020).

[61] K. Liu, Z. Gong, and M. Ueda, *Thermodynamic uncertainty relation for arbitrary initial states.* Physical Review Letters, **125**, 140602 (2020).

[62] J. Korbel, and D. H. Wolpert, *Stochastic thermodynamics and fluctuation theorems for non-linear systems.* arXiv preprint arXiv:2011.06866 (2020).

[63] N. Shiraishi, K. Funo, and K. Saito, *Speed limit for classical stochastic processes.* Phys. Rev. Lett. **121**, 070601 (2018).

[64] G. Falasco and M. Esposito, *Dissipation-time uncertainty relation.* Phys. Rev. Lett. **125**, 120604 (2020).

[65] K. Proesmans, B. Cleuren, and C. Van den Broeck, Power-Efficiency-Dissipation Relations in Linear Thermo- dynamics, Phys. Rev. Lett. **116**, 220601 (2016).

[66] N. Shiraishi, K. Saito, and H. Tasaki, *Universal Trade-Off Relation between Power and Efficiency for Heat Engines*, Phys. Rev. Lett. **117**, 190601 (2016).

[67] P. Pietzonka and U. Seifert, *Universal Trade-Off between Power, Efficiency, and Constancy in Steady-State Heat Engines Patrick*, Phys. Rev. Lett. **120**, 190602 (2018).

[68] K. Brandner, and K. Saito, *Thermodynamic geometry of microscopic heat engines.* Physical review letters, **124**, 040602 (2020).

[69] R. Fu, A. Taghvaei, Y. Chen and T. T. Georgiou, *Maximal power output of a stochastic thermodynamic engine.* Automatica, **123**, 109366 (2021).

[70] T. Van Vu and Y. Hasegawa, *Geometrical bounds of the irreversibility in Markovian systems.* Physical Review Letters, **126**, 010601 (2021).

[71] S. Ito and T. Sagawa, *Information thermodynamics on causal networks*, Phys. Rev. Lett. **111**, 180603 (2013).

[72] J. M. Horowitz and M. Esposito, *Thermodynamics with continuous information flow*, Phys. Rev. X **4**, 031015 (2014).

[73] D. Hartich, A. C. Barato, and U. Seifert, *Stochastic thermodynamics of bipartite systems: transfer entropy inequalities and a Maxwell's demon interpretation*, J. Stat. Mech. P02016 (2014).

[74] J. M. Horowitz and H. Sandberg, *Second-law-like inequalities with information and their interpretations.* New Journal of Physics, **16**, 125007 (2014).

[75] T. Schmiedl and U. Seifert, *Efficiency at maximum power: An analytically solvable model for stochastic heat engines*, J. Chem. Phys., **126**, 044101 (2007).

[76] A. E. Allahverdyan, D. Janzing and G. Mahler, *Thermodynamic efficiency of information and heat flow*, J. Stat. Mech. P09011 (2009).

[77] T. Sagawa and M. Ueda, *Generalized Jarzynski Equality under Nonequilibrium Feedback Control*, Phys. Rev. Lett. **104**, 090602 (2010).

[78] S. Still, D. A. Sivak, A. J. Bell and G. E. Crooks, *Thermodynamics of prediction*, Phys. Rev. Lett. **109**, 120604 (2012).

[79] T. Sagawa and M. Ueda, *Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics*, Phys. Rev. Lett. **109**, 180602 (2012).

[80] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, *Thermodynamics of information*, Nat. Phys. **11**, 131 (2015).

[81] S. Ito and T. Sagawa, *Maxwell's demon in biochemical signal transduction with feedback loop*, Nat. Commun. **6**, 7498 (2015).

[82] N. Shiraishi and T. Sagawa, *Fluctuation theorem for partially masked nonequilibrium dynamics*, Phys. Rev. E **91**, 3 (2015).

[83] N. Shiraishi, S. Ito, K. Kawaguchi, and T. Sagawa, *Role of measurement-feedback separation in autonomous Maxwell's demons.* New Journal of Physics, **17**, 045012 (2015).

[84] R. E. Spinney, J. T. Lizier, and M. Prokopenko, *Transfer entropy in physical systems and the arrow of time.* Physical Review E, **94**, 022135 (2016).

[85] Ito, S. *Backward transfer entropy: Informational measure for detecting hidden Markov models and its interpretations in thermodynamics, gambling and causality.* Scientific reports, **6**, 36831 (2016).

[86] G. E. Crooks and S. Still, *Marginal and conditional second laws of thermodynamics.* EPL (Europhysics Letters), **125**, 40005 (2019).

[87] A. Auconi, A. Giansanti, and E. Klipp, *Information Thermodynamics for Time Series of Signal-Response Models.* Entropy, **21**, 177 (2019).

[88] T. Schmiedl and U. Seifert, *Optimal finite-time processes in stochastic thermodynamics.* Physical review letters, **98**, 108301 (2007).

[89] L. V. Kantorovich. *On the translocation of masses.* Journal of mathematical sciences, **133**, 1381 (2006).

[90] J. Schnakenberg, *Network theory of microscopic and macroscopic behavior of master equation systems.* Reviews of Modern physics, **48**, 571 (1976).

[91] T. Hatano, and S. I. Sasa, *Steady-state thermodynamics of Langevin systems.* Physical review letters, **86**, 3463 (2001).

[92] S. Ciliberto, *Experiments in stochastic thermodynamics: Short history and perspectives.* Physical Review X, **7**, 021051 (2017).

[93] C. Van den Broeck, *Stochastic thermodynamics: A brief introduction.* Physics of Complex Colloids, **184**, 155-193 (2013).

[94] H. Risken, *The Fokker-Planck Equation* (Springer Berlin, 1986).

[95] C. R. Givens, and R. M. Shortt. *A class of Wasserstein metrics for probability distributions.* The Michigan Mathematical Journal, **31**, 231-240(1984).

[96] A. Takatsu, *Wasserstein geometry of Gaussian measures.* Osaka Journal of Mathematics, **48**, 1005 (2011).

[97] D. C. Dowson, and B. V. Landau, *The Fréchet distance between multivariate normal distributions.* Journal of multivariate analysis, **12**, 450-455 (1982).

[98] W. Li and J. Zhao, *Wasserstein information matrix.* arXiv preprint arXiv:1910.11248 (2019).