RESEARCH PAPER

Geometric thermodynamics for the Fokker–Planck equation: stochastic thermodynamic links between information geometry and optimal transport

Sosuke Ito[1]

Abstract

We propose a geometric theory of non-equilibrium thermodynamics, namely geometric thermodynamics, using our recent developments of differential-geometric aspects of entropy production rate in non-equilibrium thermodynamics. By revisiting our recent results on geometrical aspects of entropy production rate in stochastic thermodynamics for the Fokker–Planck equation, we introduce a geometric framework of non-equilibrium thermodynamics in terms of information geometry and optimal transport theory. We show that the proposed geometric framework is useful for obtaining several non-equilibrium thermodynamic relations, such as thermodynamic trade-off relations between the thermodynamic cost and the fluctuation of the observable, optimal protocols for the minimum thermodynamic cost and the decomposition of the entropy production rate for the non-equilibrium system. We clarify several stochastic-thermodynamic links between information geometry and optimal transport theory via the excess entropy production rate based on a relation between the gradient flow expression and information geometry in the space of probability densities and a relation between the velocity field in optimal transport and information geometry in the space of path probability densities.

Communicated by Shun-ichi Amari.

✉ Sosuke Ito
sosuke.ito@ubi.s.u-tokyo.ac.jp

1 Universal Biology Institute, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

1 Introduction

A geometric interpretation of thermodynamics originates from the geometric picture of the thermodynamic potential proposed by W. Gibbs in equilibrium thermodynamics and chemical thermodynamics [1]. In non-equilibrium thermodynamics, second-order thermodynamic fluctuations around the equilibrium or steady state have been studied [2–6]. A differential geometry for equilibrium thermodynamics has been proposed by Weinhold [7] and Ruppeiner [8] by considering the fluctuation around the equilibrium state, and the length called thermodynamic length in Weinhold geometry has been proposed to quantify the dissipated availability [9]. Because this geometry for equilibrium thermodynamics is based on the second-order fluctuation of entropy, its generalization [10, 11] has been regarded as information geometry [12, 13], which is the differential geometry for the Fisher metric [14].

In recent years, differential geometry for non-equilibrium thermodynamics, especially for stochastic thermodynamics [15, 16] and non-equilibrium chemical thermodynamics [17, 18], has been used to investigate mathematical properties of entropy production for non-equilibrium transitions and fluctuations around non-equilibrium steady states [11, 19–32]. Because stochastic thermodynamics is based on stochastic processes [33] such as the Fokker–Planck equation [34], differential geometry for non-equilibrium thermodynamics is related to information geometry [12, 13] and optimal transport theory [35, 36].

In this paper, we summarize our recent development of differential geometry for non-equilibrium thermodynamics [22, 24, 25, 29, 31, 37–45] and propose several relations between these studies by focusing on non-equilibrium dynamics of the Fokker–Planck equation. Because entropy production for the Fokker–Planck equation can be discussed from the viewpoint of both information geometry and optimal transport theory, these relations provide links between information geometry and optimal transport theory. Our proposed geometrical framework for non-equilibrium thermodynamics, namely *geometric thermodynamics*, offers a new perspective on links between information geometry and optimal transport theory [46–49] and the unification of non-equilibrium thermodynamic geometry.

2 Fokker–Planck equation and stochastic thermodynamics

2.1 Setup

We consider the time evolution of a probability density described by the (over-damped) Fokker–Planck equation. Let $t \in \mathbb{R}$ and $x \in \mathbb{R}^d$ ($d \in \mathbb{N}$) be time and the d-dimensional position, respectively. The probability density of x at time t will be denoted by $P_t(x)$, which satisfies $P_t(x) \geq 0$ and $\int dx\, P_t(x) = 1$. The Fokker–Planck equation is given by the following continuity equation,

$$\partial_t P_t(x) = -\nabla \cdot (v_t(x) P_t(x)),$$
$$v_t(x) = \mu(F_t(x) - T \nabla \ln P_t(x)). \tag{1}$$

Here, $\boldsymbol{v}_t(\boldsymbol{x}) \in \mathbb{R}^d$ and $\boldsymbol{F}_t(\boldsymbol{x}) \in \mathbb{R}^d$ are the vector functions at position $\boldsymbol{x}$, $\mu \in \mathbb{R}_{>0}$ and $T \in \mathbb{R}_{>0}$ are positive constants, and $\nabla\cdot$ and ∇ are the divergence and the gradient operators, respectively. Physically, the Fokker–Planck equation is used to describe the time evolution of the probability density of an over-damped Brownian particle. For Brownian motion, μ, T, and $\boldsymbol{F}_t(\boldsymbol{x})$ physically represent the mobility of the Brownian particle, the temperature of the medium scaled by the Boltzmann constant, and the force on the Brownian particle, respectively [33]. The force field $\boldsymbol{v}_t(\boldsymbol{x})$ is called the mean local velocity because it quantifies the ensemble average of the Brownian particle's velocity in $\boldsymbol{x}$ at time t [16].

This Fokker–Planck equation corresponds to the over-damped Langevin equation, which describes the position of the Brownian particle $\boldsymbol{X}(t) \in \mathbb{R}^d$ at time t, that is

$$\dot{\boldsymbol{X}}(t) = \mu \boldsymbol{F}_t(\boldsymbol{X}(t)) + \sqrt{2\mu T}\boldsymbol{\xi}(t). \tag{2}$$

Here, $\dot{\boldsymbol{X}}(t)$ is the time derivative of position $\boldsymbol{X}(t)$, and $(\boldsymbol{\xi}(t))_i = \xi_i(t)$ ($i \in \{1, 2, \ldots, d\}$) is the white Gaussian noise that satisfies $\langle \xi_i(t)\xi_j(t') \rangle = \delta_{ij}\delta(t - t')$ and $\langle \xi_i(t) \rangle = 0$, where $\langle \cdot \rangle$, δ_{ij}, and $\delta(t - t')$ stand for the ensemble average, the Kronecker delta, and the delta function, respectively ($j \in \{1, 2, \ldots, d\}$). The product of $\sqrt{2\mu T}$ and $\boldsymbol{\xi}(t)$ is given by the Ito integral. Mathematically, this correspondence between the Fokker–Planck equation and the over-damped Langevin equation indicates that these two descriptions provide the same transition probability density from position $\boldsymbol{X}(\tau) = \boldsymbol{x}_\tau$ to position $\boldsymbol{X}(\tau + dt) = \boldsymbol{x}_{\tau+dt}$ during the positive infinitesimal time interval $dt > 0$. The transition probability density from $\boldsymbol{x}_\tau$ to $\boldsymbol{x}_{\tau+dt}$ is given by the Onsager–Machlup theory [33],

$$\mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) = \frac{1}{(4\pi\mu T dt)^{\frac{d}{2}}} \exp\left[-\frac{\|\boldsymbol{x}_{\tau+dt} - \boldsymbol{x}_\tau - \mu\boldsymbol{F}_\tau(\boldsymbol{x}_\tau)dt\|^2}{4\mu T dt} \right], \tag{3}$$

where $\|\cdot\|$ stands for the L^2 norm, and the transition probability density satisfies $\int d\boldsymbol{x}_{\tau+dt}\mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) = 1$ and $\mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) \geq 0$. Let be $\mathcal{X}_\tau$ be the random variable corresponding to state $\boldsymbol{x}_\tau$. The joint probability of $\mathcal{X}_{\tau+dt}$ and $\mathcal{X}_\tau$ being in $\boldsymbol{x}_{\tau+dt}$ and $\boldsymbol{x}_\tau$ is defined as

$$\mathbb{P}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) = \mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau)P_\tau(\boldsymbol{x}_\tau), \tag{4}$$

which satisfies $\int d\boldsymbol{x}_{\tau+dt}d\boldsymbol{x}_\tau\mathbb{P}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) = 1$ and $\mathbb{P}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \geq 0$. This joint probability $\mathbb{P}$ is called the forward path probability density because it is the probability of the forward path from time $t = \tau$ to time $t = \tau + dt$.

2.2 Entropy production rate

We introduce stochastic thermodynamics [15, 16], that is a framework for non-equilibrium thermodynamics described by a stochastic process such as the Fokker–Planck equation. In stochastic thermodynamics, the entropy production rate is

introduced as a measure of thermodynamic dissipation [34]. The entropy production rate is defined as follows.

Definition 1 For the Fokker–Planck equation (1), *the entropy production rate at time τ is defined as*

$$\sigma_\tau = \frac{1}{\mu T} \int d\boldsymbol{x} \|\boldsymbol{v}_\tau(\boldsymbol{x})\|^2 P_\tau(\boldsymbol{x}). \tag{5}$$

Remark 1 This entropy production rate is definitely non-negative, and its non-negativity $\sigma_\tau \geq 0$ is known as the second law of thermodynamics [16].

Remark 2 The entropy production rate is regarded as the sum of the entropy changes in the heat bath and the system [16]. If we assume that $P_\tau(\boldsymbol{x})$ decays sufficiently rapidly at infinity, the entropy production rate can be rewritten as

$$
\begin{aligned}
\sigma_\tau &= \frac{1}{\mu T} \int d\boldsymbol{x} (\boldsymbol{v}_\tau(\boldsymbol{x}) \cdot \boldsymbol{v}_\tau(\boldsymbol{x})) P_\tau(\boldsymbol{x}) \\
&= \frac{\int d\boldsymbol{x} (\boldsymbol{v}_\tau(\boldsymbol{x}) \cdot \boldsymbol{F}_\tau(\boldsymbol{x})) P_\tau(\boldsymbol{x})}{T} - \int d\boldsymbol{x} (\boldsymbol{v}_\tau(\boldsymbol{x}) P_\tau(\boldsymbol{x})) \cdot \nabla \ln P_\tau(\boldsymbol{x}) \\
&= \frac{\int d\boldsymbol{x} (\boldsymbol{v}_\tau(\boldsymbol{x}) \cdot \boldsymbol{F}_\tau(\boldsymbol{x})) P_\tau(\boldsymbol{x})}{T} + \int d\boldsymbol{x} \nabla \cdot (\boldsymbol{v}_\tau(\boldsymbol{x}) P_\tau(\boldsymbol{x})) \ln P_\tau(\boldsymbol{x}) \\
&= \frac{\int d\boldsymbol{x} (\boldsymbol{v}_\tau(\boldsymbol{x}) \cdot \boldsymbol{F}_\tau(\boldsymbol{x})) P_\tau(\boldsymbol{x})}{T} + \partial_\tau \left[-\int d\boldsymbol{x} P_\tau(\boldsymbol{x}) \ln P_\tau(\boldsymbol{x}) \right],
\end{aligned} \tag{6}
$$

where we used Eq. (1), $\int d\boldsymbol{x} P_\tau(\boldsymbol{x})(\partial_\tau \ln P_\tau(\boldsymbol{x})) = \partial_\tau \int d\boldsymbol{x} P_\tau(\boldsymbol{x}) = 0$, and $\int d\boldsymbol{x} \nabla \cdot (\boldsymbol{v}_\tau(\boldsymbol{x}) P_\tau(\boldsymbol{x}) \ln P_\tau(\boldsymbol{x})) = 0$ because of the assumption that $P_\tau(\boldsymbol{x})$ decays sufficiently rapidly at infinity. The term

$$\dot{S}_\tau = \partial_\tau \left[-\int d\boldsymbol{x} P_\tau(\boldsymbol{x}) \ln P_\tau(\boldsymbol{x}) \right], \tag{7}$$

is the time derivative of the differential entropy [50], which is regarded as the entropy change of the system. The term

$$-\dot{Q}_\tau = \int d\boldsymbol{x} (\boldsymbol{v}_\tau(\boldsymbol{x}) \cdot \boldsymbol{F}_\tau(\boldsymbol{x})) P_\tau(\boldsymbol{x}), \tag{8}$$

is the heat dissipation rate and $-\dot{Q}_\tau/T$ is regarded as the entropy change of the heat bath. Thus, the entropy production rate is given by the sum of the entropy changes in the heat bath and the system,

$$\sigma_\tau = \dot{S}_\tau - \frac{\dot{Q}_\tau}{T}. \tag{9}$$

Its non-negativity $\sigma_\tau \geq 0$ provides the Clausius inequality for the Fokker–Planck equation,

$$\dot{S}_\tau \geq \frac{\dot{Q}_\tau}{T}, \tag{10}$$

which is an expression of the second law of thermodynamics.

2.3 Kullback–Leibler divergence and entropy production rate

We introduce an expression of the entropy production rate in terms of the Kullback–Leibler divergence, which was discussed in the context of the fluctuation theorem [16, 51–53]. We assume that the parity of state $\boldsymbol{x}_\tau$ is even; in other words, the sign of $\boldsymbol{x}_\tau$ does not change under the time reversal transformation. Let $\mathbb{P}^\dagger(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau)$ be backward path probability density defined as $\mathbb{P}^\dagger(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) = \mathbb{T}_\tau(\boldsymbol{x}_\tau \mid \boldsymbol{x}_{\tau+dt}) P_{\tau+dt}(\boldsymbol{x}_{\tau+dt})$. Now, we consider the Kullback–Leibler divergence between $\mathbb{P}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau)$ and $\mathbb{P}^\dagger(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau)$ defined as

$$D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^\dagger) = \int d\boldsymbol{x}_\tau d\boldsymbol{x}_{\tau+dt} \mathbb{P}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \ln \frac{\mathbb{P}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau)}{\mathbb{P}^\dagger(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau)}. \tag{11}$$

The entropy production rate σ_τ is given by this Kullback-Leibler divergence as follows.

Lemma 1 *The entropy production rate σ_τ is given by*

$$\sigma_\tau = \lim_{dt \to 0} \frac{D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^\dagger)}{dt}. \tag{12}$$

Proof We rewrite the Kullback–Leibler divergence $D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^\dagger)$ as

$$
\begin{aligned}
D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^\dagger) &= \int d\boldsymbol{x}_\tau d\boldsymbol{x}_{\tau+dt} \mathbb{P}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \ln \frac{\mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) P_\tau(\boldsymbol{x}_\tau)}{\mathbb{T}_\tau(\boldsymbol{x}_\tau \mid \boldsymbol{x}_{\tau+dt}) P_\tau(\boldsymbol{x}_{\tau+dt})} \\
&\quad + \int d\boldsymbol{x}_{\tau+dt} P_{\tau+dt}(\boldsymbol{x}_{\tau+dt}) \ln \frac{P_\tau(\boldsymbol{x}_{\tau+dt})}{P_{\tau+dt}(\boldsymbol{x}_{\tau+dt})} \\
&= \int d\boldsymbol{x}_\tau d\boldsymbol{x}_{\tau+dt} \mathbb{P}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \ln \frac{\mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) P_\tau(\boldsymbol{x}_\tau)}{\mathbb{T}_\tau(\boldsymbol{x}_\tau \mid \boldsymbol{x}_{\tau+dt}) P_\tau(\boldsymbol{x}_{\tau+dt})} \\
&\quad - dt \int d\boldsymbol{x}_{\tau+dt} \partial_{\tau'} P_{\tau'}(\boldsymbol{x}_{\tau+dt}) \mid_{\tau'=\tau+dt} + O(dt^2) \\
&= \int d\boldsymbol{x}_\tau d\boldsymbol{x}_{\tau+dt} \mathbb{P}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \ln \frac{\mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) P_\tau(\boldsymbol{x}_\tau)}{\mathbb{T}_\tau(\boldsymbol{x}_\tau \mid \boldsymbol{x}_{\tau+dt}) P_\tau(\boldsymbol{x}_{\tau+dt})} + O(dt^2),
\end{aligned} \tag{13}
$$

where $O(dt^2)$ means the term that satisfies $\lim_{dt \to 0} O(dt^2)/dt = 0$ and we used $\int d\boldsymbol{x}_\tau \mathbb{P}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) = P_{\tau+dt}(\boldsymbol{x}_{\tau+dt})$ and $\int d\boldsymbol{x}_{\tau+dt} \partial_{\tau'} P_{\tau'}(\boldsymbol{x}_{\tau+dt}) \mid_{\tau'=\tau+dt} = 0$. From

Eq. (3), we obtain

$$
\ln \frac{\mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau)P_\tau(\boldsymbol{x}_\tau)}{\mathbb{T}_\tau(\boldsymbol{x}_\tau \mid \boldsymbol{x}_{\tau+dt})P_\tau(\boldsymbol{x}_{\tau+dt})}
$$
$$
=(\boldsymbol{x}_{\tau+dt} - \boldsymbol{x}_\tau) \circ \left[\frac{\boldsymbol{v}_\tau(\boldsymbol{x}_\tau)}{\mu T} + O(dt) \right] + O(\|\boldsymbol{x}_{\tau+dt} - \boldsymbol{x}_\tau\|^2), \qquad (14)
$$

where $\circ$ stands for the Stratonovich integral defined as $(\boldsymbol{x}_{\tau+dt} - \boldsymbol{x}_\tau) \circ \boldsymbol{v}_\tau(\boldsymbol{x}_\tau) = (\boldsymbol{x}_{\tau+dt} - \boldsymbol{x}_\tau) \cdot \boldsymbol{v}_\tau([\boldsymbol{x}_\tau + \boldsymbol{x}_{\tau+dt}]/2)$, $O(dt)$ means the term that satisfies $\lim_{dt \to 0} O(dt) = 0$ and $O(\|\boldsymbol{x}_{\tau+dt} - \boldsymbol{x}_\tau\|^2)$ is the higher order term which satisfies $\int d\boldsymbol{x}_{\tau+dt} \mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) O(\|\boldsymbol{x}_{\tau+dt} - \boldsymbol{x}_\tau\|^2) = O(dt)$. Because the Gaussian integral provides

$$
\frac{1}{dt} \int d\boldsymbol{x}_\tau P_\tau(\boldsymbol{x}_\tau) \int d\boldsymbol{x}_{\tau+dt} \mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) \ln \frac{\mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau)P_\tau(\boldsymbol{x}_\tau)}{\mathbb{T}_\tau(\boldsymbol{x}_\tau \mid \boldsymbol{x}_{\tau+dt})P_\tau(\boldsymbol{x}_{\tau+dt})}
$$
$$
= \int d\boldsymbol{x}_\tau P_\tau(\boldsymbol{x}_\tau) \int d\boldsymbol{x}_{\tau+dt} \mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) \frac{\boldsymbol{x}_{\tau+dt} - \boldsymbol{x}_\tau}{dt} \circ \left[\frac{\boldsymbol{v}_\tau(\boldsymbol{x}_\tau)}{\mu T} + O(dt) \right] + O(dt)
$$
$$
= \frac{1}{\mu T} \int d\boldsymbol{x}_\tau P_\tau(\boldsymbol{x}_\tau) \|\boldsymbol{v}_\tau(\boldsymbol{x}_\tau)\|^2 + O(dt), \qquad (15)
$$

we obtain

$$
\lim_{dt \to 0} \frac{D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^\dagger)}{dt} = \frac{1}{\mu T} \int d\boldsymbol{x}_\tau \|\boldsymbol{v}_\tau(\boldsymbol{x}_\tau)\|^2 P_\tau(\boldsymbol{x}_\tau) + \lim_{dt \to 0} O(dt) = \sigma_\tau, \qquad (16)
$$

from Eqs. (13) and (15).

Remark 3 The Kullback–Leibler divergence is always non-negative $D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^\dagger) \geq 0$ and zero if and only if $\mathbb{P} = \mathbb{P}^\dagger$. Thus, $\sigma_\tau = 0$ if and only if $\mathbb{P} = \mathbb{P}^\dagger$. Physically, $\mathbb{P} = \mathbb{P}^\dagger$ means the reversibility of stochastic dynamics, and $\sigma_\tau = 0$ means that the system is in equilibrium.

Remark 4 The time integral of the entropy production rate $\Sigma(\tau'; \tau) = \int_\tau^{\tau'} dt \sigma_t$ is called the entropy production from time $t = \tau$ to time $t = \tau'$. The Lemma 1 implies that the Kullback–Leibler divergence $D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^\dagger)$ is equivalent to the entropy production from time $t = \tau$ to $t = \tau + dt$ up to $O(dt^2)$,

$$
D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^\dagger) = \Sigma(\tau + dt; \tau) + O(dt^2), \qquad (17)
$$

where $O(dt^2)$ means the term $\lim_{dt \to 0} O(dt^2)/dt = 0$.

Remark 5 Because the Fokker–Planck equation describes the Markov process, each increments are independent and the results in this paper can be generalized for the entire path from $t = 0$ to $t = \tau$. Thus, the results for the entropy production rate σ_τ in this paper can be generalized for the entropy production $\Sigma(\tau; 0)$ based on the expression,

$$
\Sigma(\tau; 0) = \lim_{dt \to 0 \mid dt N = \tau} D_{\mathrm{KL}}(\hat{\mathbb{P}}\|\hat{\mathbb{P}}^\dagger), \qquad (18)
$$

where $\hat{\mathbb{P}}$ and $\hat{\mathbb{P}}^\dagger$ is the forward and backward path probability densities for the entire path $\Gamma = (x_{Ndt}, \ldots, x_{2dt}, x_{dt}, x_0)$ with $Ndt = \tau$, defined as $\hat{\mathbb{P}}(\Gamma) = \prod_{i=1}^{N} \mathbb{T}_{(i-1)dt}(x_{idt} \mid x_{(i-1)dt}) P_0(x_0)$ and $\hat{\mathbb{P}}^\dagger(\Gamma) = \prod_{i=1}^{N} \mathbb{T}_{(i-1)dt}(x_{(i-1)dt} \mid x_{idt}) P_\tau(x_\tau)$ respectively. The Kullback–Leibler divergence is defined as $D_{\mathrm{KL}}(\hat{\mathbb{P}} \| \hat{\mathbb{P}}^\dagger) = \int d\Gamma \hat{\mathbb{P}}(\Gamma) \ln[\hat{\mathbb{P}}(\Gamma)/\hat{\mathbb{P}}^\dagger(\Gamma)]$.

This link between the entropy production rate and the Kullback–Leibler divergence in Lemma 1 leads to an information-geometric interpretation of the entropy production rate.

3 Information geometry and entropy production

3.1 Projection theorem and entropy production

We discuss an information-geometric interpretation of the entropy production based on the projection theorem [13], which is obtained for a general Markov jump process in Ref. [25]. The entropy production can be understood in terms of the information-geometric projection onto the backward manifold defined as follows.

Definition 2 Let $\mathbb{Q}(x_{\tau+dt}, x_\tau)$ be a probability density that satisfies $\mathbb{Q}(x_{\tau+dt}, x_\tau) \geq 0$ and $\int dx_{\tau+dt} dx_\tau \mathbb{Q}(x_{\tau+dt}, x_\tau) = 1$. The set of the probability density

$$\mathcal{M}_{\mathrm{B}}(\mathbb{P}) = \left\{ \mathbb{Q} \middle| \mathbb{Q}(x_{\tau+dt}, x_\tau) = \mathbb{T}_\tau(x_\tau \mid x_{\tau+dt}) \int dx_\tau \mathbb{Q}(x_{\tau+dt}, x_\tau) \right\}, \quad (19)$$

is called *the backward manifold*.

Remark 6 $\mathcal{M}_{\mathrm{B}}(\mathbb{P})$ depends on $\mathbb{P}$ because $\mathcal{M}_{\mathrm{B}}(\mathbb{P})$ depends on $\mathbb{T}_\tau$. $\mathbb{T}_\tau$ is given by the function of $\mathbb{P}$ such that $\mathbb{T}_\tau(x_{\tau+dt} \mid x_\tau) = \mathbb{P}(x_{\tau+dt}, x_\tau)/[\int dx_{\tau+dt} \mathbb{P}(x_{\tau+dt}, x_\tau)]$ and $\mathbb{T}_\tau(x_\tau \mid x_{\tau+dt})$ is given by the change of variables.

Remark 7 $\mathbb{P}^\dagger \in \mathcal{M}_{\mathrm{B}}(\mathbb{P})$ because $\mathbb{P}^\dagger(x_{\tau+dt}, x_\tau) = \mathbb{T}_\tau(x_\tau \mid x_{\tau+dt}) P_{\tau+dt}(x_{\tau+dt})$ and $P_{\tau+dt}(x_{\tau+dt}) = \int dx_\tau \mathbb{P}^\dagger(x_{\tau+dt}, x_\tau)$.

This backward path probability density $\mathbb{P}^\dagger$ is given by the information-geometric projection from $\mathbb{P}$ onto $\mathcal{M}_{\mathrm{B}}(\mathbb{P})$. This information-geometric projection is formulated based on the following generalized Pythagorean theorem.

Lemma 2 *For any $\mathbb{Q} \in \mathcal{M}_{\mathrm{B}}(\mathbb{P})$, the generalized Pythagorean theorem*

$$D_{\mathrm{KL}}(\mathbb{P} \| \mathbb{Q}) = D_{\mathrm{KL}}(\mathbb{P} \| \mathbb{P}^\dagger) + D_{\mathrm{KL}}(\mathbb{P}^\dagger \| \mathbb{Q}), \quad (20)$$

holds.

Proof $\mathbb{Q} \in \mathcal{M}_{\mathrm{B}}(\mathbb{P})$ is given by $\mathbb{Q}(x_{\tau+dt}, x_\tau) = \mathbb{T}_\tau(x_\tau \mid x_{\tau+dt}) Q_{\tau+dt}(x_{\tau+dt})$ where $Q_{\tau+dt}(x_{\tau+dt}) = \int dx_\tau \mathbb{Q}(x_{\tau+dt}, x_\tau)$. Thus,

$$D_{KL}(\mathbb{P}\|\mathbb{Q}) = \int d\boldsymbol{x}_\tau d\boldsymbol{x}_{\tau+dt} \mathbb{P}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \ln \frac{\mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) P_\tau(\boldsymbol{x}_\tau)}{\mathbb{T}_\tau(\boldsymbol{x}_\tau \mid \boldsymbol{x}_{\tau+dt}) Q_{\tau+dt}(\boldsymbol{x}_{\tau+dt})}$$

$$= \int d\boldsymbol{x}_\tau d\boldsymbol{x}_{\tau+dt} \mathbb{P}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \ln \frac{\mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) P_\tau(\boldsymbol{x}_\tau)}{\mathbb{T}_\tau(\boldsymbol{x}_\tau \mid \boldsymbol{x}_{\tau+dt}) P_{\tau+dt}(\boldsymbol{x}_{\tau+dt})}$$

$$+ \int d\boldsymbol{x}_{\tau+dt} P_{\tau+dt}(\boldsymbol{x}_{\tau+dt}) \ln \frac{P_{\tau+dt}(\boldsymbol{x}_{\tau+dt})}{Q_{\tau+dt}(\boldsymbol{x}_{\tau+dt})}$$

$$= D_{KL}(\mathbb{P}\|\mathbb{P}^\dagger) + D_{KL}(\mathbb{P}^\dagger\|\mathbb{Q}), \tag{21}$$

where we used $\int d\boldsymbol{x}_{\tau+dt} \mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) = 1$, $\int d\boldsymbol{x}_\tau \mathbb{T}_\tau(\boldsymbol{x}_\tau \mid \boldsymbol{x}_{\tau+dt}) = 1$, and $P_{\tau+dt}(\boldsymbol{x}_{\tau+dt})/Q_{\tau+dt}(\boldsymbol{x}_{\tau+dt}) = \mathbb{P}^\dagger(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau)/\mathbb{Q}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau)$.   □

Remark 8 The generalized Pythagorean theorem Eq. (20) can be rewritten as

$$\int d\boldsymbol{x}_\tau d\boldsymbol{x}_{\tau+dt} (\mathbb{P}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) - \mathbb{P}^\dagger(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau)) \ln \frac{\mathbb{P}^\dagger(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau)}{\mathbb{Q}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau)} = 0. \tag{22}$$

Information-geometrically, Eq. (22) implies that the *m*-geodesic between two points $\mathbb{P}$ and $\mathbb{P}^\dagger$ is orthogonal to the *e*-geodesic between two points $\mathbb{P}^\dagger$ and $\mathbb{Q}$ [12]. The definition of the *m*-geodesic between two points $\mathbb{P}$ and $\mathbb{P}^\dagger$ is given by $(1-\theta)\mathbb{P} + \theta\mathbb{P}^\dagger$, where θ is an affine parameter $\theta \in [0, 1]$. The definition of the *e*-geodesic between two points $\mathbb{P}^\dagger$ and $\mathbb{Q}$ is also given by $(1-\theta)\ln\mathbb{P}^\dagger + \theta\ln\mathbb{Q}$.

The orthogonality based on the generalized Pythagorean theorem leads to the projection theorem, which provides a minimization problem of the Kullback–Leibler divergence. Thus, Lemma 2 implies that the entropy production rate can be obtained from a minimization problem of the Kullback–Leibler divergence.

Theorem 3 *The entropy production rate σ_τ is given by the minimization problem,*

$$\sigma_\tau = \lim_{dt \to 0} \inf_{\mathbb{Q} \in \mathcal{M}_B(\mathbb{P})} \frac{D_{KL}(\mathbb{P}\|\mathbb{Q})}{dt}. \tag{23}$$

The entropy production $\Sigma(\tau + dt; \tau)$ is also given by the minimization problem,

$$\Sigma(\tau + dt; \tau) = \inf_{\mathbb{Q} \in \mathcal{M}_B(\mathbb{P})} D_{KL}(\mathbb{P}\|\mathbb{Q}) + O(dt^2). \tag{24}$$

Proof From Lemma 2, Eq. (20) implies

$$D_{KL}(\mathbb{P}\|\mathbb{P}^\dagger) = \inf_{\mathbb{Q} \in \mathcal{M}_B(\mathbb{P})} D_{KL}(\mathbb{P}\|\mathbb{Q}), \tag{25}$$

because $D_{KL}(\mathbb{P}^\dagger\|\mathbb{Q}) \geq 0$ and $\mathbb{P}^\dagger \in \mathcal{M}_B(\mathbb{P})$. By combining Eqs. (12) and (17) with Eq. (25), we obtain Eqs. (23) and (24), respectively.   □

Thus, Theorem 3 implies that the entropy production $\Sigma(\tau + dt; \tau)$ can be obtained from the information-geometric projection onto the backward manifold $\mathcal{M}_B(\mathbb{P})$. This

result is helpful in estimating the entropy production $\Sigma(\tau + dt; \tau)$ numerically by calculating the optimization problem to minimize the Kullback–Leibler divergence $D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{Q})$.

3.2 Interpolated dynamics and Fisher information

We can consider not only the m-geodesic between $\mathbb{P}$ and $\mathbb{P}^\dagger$ in Lemma 2 but also the e-geodesic between $\mathbb{P}$ and $\mathbb{P}^\dagger$. This geodesic can be discussed in terms of the interpolated dynamics, which has been substantially introduced in Refs. [40, 54]. By considering this interpolation, we obtain an expression of the entropy production rate by the Fisher metric, which provides a trade-off relation between the entropy production rate and the fluctuation of any observable. We start with the definition of interpolated dynamics as follows.

Definition 3 Dynamics described by the following continuity equation are called *interpolated dynamics for two force fields* $\boldsymbol{v}_t(\boldsymbol{x}) = \mu(\boldsymbol{F}_t(\boldsymbol{x}) - T\nabla \ln P_t(\boldsymbol{x}))$ and $\boldsymbol{v}'_t(\boldsymbol{x}) \in \mathbb{R}^d$,

$$\partial_t P_t(\boldsymbol{x}) = -\nabla \cdot (\boldsymbol{v}_t^\theta(\boldsymbol{x}) P_t(\boldsymbol{x})), \tag{26}$$

$$\boldsymbol{v}_t^\theta(\boldsymbol{x}) = (1 - \theta)\boldsymbol{v}_t(\boldsymbol{x}) + \theta \boldsymbol{v}'_t(\boldsymbol{x}), \tag{27}$$

where $\theta \in [0, 1]$ is an interpolation parameter.

Remark 9 The corresponding over-damped Langevin equation of interpolated dynamics is given by

$$\dot{X}(t) = \mu \boldsymbol{F}_t(X(t)) + \theta[\boldsymbol{v}'_t(X(t)) - \boldsymbol{v}_t(X(t))] + \sqrt{2\mu T}\boldsymbol{\xi}(t). \tag{28}$$

Definition 4 *The path probability density of interpolated dynamics for two force fields* $\boldsymbol{v}_\tau(\boldsymbol{x}_\tau) = \mu(\boldsymbol{F}_\tau(\boldsymbol{x}_\tau) - T\nabla \ln P_\tau(\boldsymbol{x}_\tau))$ and $\boldsymbol{v}'_\tau(\boldsymbol{x}_\tau)$ is defined as

$$\mathbb{P}^\theta_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) = \mathbb{T}^\theta_{\tau; \boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) P_\tau(\boldsymbol{x}_\tau), \tag{29}$$

$$\mathbb{T}^\theta_{\tau; \boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) = \frac{\exp\left[-\frac{\|\boldsymbol{x}_{\tau+dt} - \boldsymbol{x}_\tau - \mu \boldsymbol{F}_\tau(\boldsymbol{x}_\tau)dt - \theta[\boldsymbol{v}'_\tau(\boldsymbol{x}_\tau) - \boldsymbol{v}_\tau(\boldsymbol{x}_\tau)]dt]\|^2}{4\mu T dt}\right]}{(4\pi\mu T dt)^{\frac{d}{2}}}. \tag{30}$$

Remark 10 The parameter θ quantifies a difference between interpolated dynamics and original Fokker–Planck dynamics (1) because $\theta = 0$ provides the transition rate $\mathbb{T}^0_{\tau; \boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) = \mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau)$ and the forward path probability density $\mathbb{P}^0_{\boldsymbol{v}'_\tau} = \mathbb{P}$ of original Fokker–Planck dynamics $\partial_t P_t(\boldsymbol{x}) = -\nabla \cdot (\boldsymbol{v}_t(\boldsymbol{x}) P_t(\boldsymbol{x}))$.

Remark 11 Because the path probability density is given by

$$\ln \mathbb{P}^\theta_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) = \theta\left[\frac{[\boldsymbol{x}_{\tau+dt} - \boldsymbol{x}_\tau - \mu \boldsymbol{F}_\tau(\boldsymbol{x}_\tau)dt] \cdot [\boldsymbol{v}'_\tau(\boldsymbol{x}_\tau) - \boldsymbol{v}_\tau(\boldsymbol{x}_\tau)]}{2\mu T}\right]$$
$$+ \ln \mathbb{P}^0_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) + O(dt), \tag{31}$$

the parameter θ can be regarded as a theta coordinate system for the exponential family in information geometry [13]. By neglecting $O(dt)$, $\ln \mathbb{P}^{\theta}_{\boldsymbol{v}'_{\tau}}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_{\tau})$ can be rewritten as

$$\ln \mathbb{P}^{\theta}_{\boldsymbol{v}'_{\tau}}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_{\tau}) = (1 - \theta) \ln \mathbb{P}^{0}_{\boldsymbol{v}'_{\tau}}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_{\tau}) + \theta \ln \mathbb{P}^{1}_{\boldsymbol{v}'_{\tau}}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_{\tau}), \quad (32)$$

which implies that $\mathbb{P}^{\theta}_{\boldsymbol{v}'_{\tau}}$ gives the e-geodesic between two points $\mathbb{P}^{0}_{\boldsymbol{v}'_{\tau}} = \mathbb{P}$ and $\mathbb{P}^{1}_{\boldsymbol{v}'_{\tau}}$.

We next consider the backward path probability density $\mathbb{P}^{\dagger}$ in terms of interpolated dynamics. The backward path probability density $\mathbb{P}^{\dagger}$ can be regarded as $\mathbb{P}^{1}_{-\boldsymbol{v}_{\tau}}$ as follows.

Lemma 4 *The backward path probability density $\mathbb{P}^{\dagger}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_{\tau})$ is given by*

$$\ln \mathbb{P}^{\dagger}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_{\tau}) = \ln \mathbb{P}^{1}_{-\boldsymbol{v}_{\tau}}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_{\tau}) + O(dt). \quad (33)$$

Proof The backward path probability density $\mathbb{P}^{\dagger}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_{\tau})$ is calculated as

$$
\begin{aligned}
&\ln \mathbb{P}^{\dagger}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_{\tau}) \\
&= \ln \mathbb{P}^{1}_{-\boldsymbol{v}_{\tau}}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_{\tau}) + \ln \frac{P_{\tau+dt}(\boldsymbol{x}_{\tau+dt})}{P_{\tau}(\boldsymbol{x}_{\tau})} - \frac{\|\boldsymbol{x}_{\tau} - \boldsymbol{x}_{\tau+dt} - \mu \boldsymbol{F}_{\tau}(\boldsymbol{x}_{\tau+dt})dt\|^2}{4\mu T dt} \\
&\quad + \frac{\|\boldsymbol{x}_{\tau+dt} - \boldsymbol{x}_{\tau} + \mu \boldsymbol{F}_{\tau}(\boldsymbol{x}_{\tau})dt - 2\mu T \nabla \ln P_{\tau}(\boldsymbol{x}_{\tau})dt\|^2}{4\mu T dt} \\
&= \ln \mathbb{P}^{1}_{-\boldsymbol{v}_{\tau}}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_{\tau}) + O(dt),
\end{aligned}
\quad (34)
$$

where we used $\ln[P_{\tau+dt}(\boldsymbol{x}_{\tau+dt})/P_{\tau}(\boldsymbol{x}_{\tau})] = (\boldsymbol{x}_{\tau+dt} - \boldsymbol{x}_{\tau}) \cdot \nabla \ln P_{\tau}(\boldsymbol{x}_{\tau}) + O(dt)$. $\square$

Lemma 4 implies that the path probability density of interpolated dynamics $\mathbb{P}^{\theta}_{-\boldsymbol{v}_{\tau}}$ gives the e-geodesic between $\mathbb{P}$ and $\mathbb{P}^{\dagger}$.

We discuss an information-geometric interpretation of the entropy production based on $\mathbb{P}^{\theta}_{-\boldsymbol{v}_{\tau}}$. To discuss it, we introduce the following lemma proposed in Ref. [55].

Lemma 5 *Let $\theta \in [0, 1]$ and $\theta' \in [0, 1]$ be two interpolation parameters. The Kullback–Leibler divergence between $\mathbb{P}^{\theta}_{\boldsymbol{v}'_{\tau}}$ and $\mathbb{P}^{\theta'}_{\boldsymbol{v}'_{\tau}}$ is given by*

$$D_{\mathrm{KL}}(\mathbb{P}^{\theta}_{\boldsymbol{v}'_{\tau}} \| \mathbb{P}^{\theta'}_{\boldsymbol{v}'_{\tau}}) = \frac{(\theta - \theta')^2 dt}{4\mu T} \int d\boldsymbol{x}_{\tau} \|\boldsymbol{v}'_{\tau}(\boldsymbol{x}_{\tau}) - \boldsymbol{v}_{\tau}(\boldsymbol{x}_{\tau})\|^2 P_{\tau}(\boldsymbol{x}_{\tau}). \quad (35)$$

Proof The quantity $\ln[\mathbb{T}^{\theta}_{\tau;\boldsymbol{v}'_{\tau}}(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_{\tau})/\mathbb{T}^{\theta'}_{\tau;\boldsymbol{v}'_{\tau}}(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_{\tau})]$ is calculated as

$$
\begin{aligned}
\ln \frac{\mathbb{T}^{\theta}_{\tau;\boldsymbol{v}'_{\tau}}(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_{\tau})}{\mathbb{T}^{\theta'}_{\tau;\boldsymbol{v}'_{\tau}}(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_{\tau})} &= \frac{(\theta - \theta')(\boldsymbol{x}_{\tau+dt} - \boldsymbol{x}_{\tau} - \mu \boldsymbol{F}_{\tau}(\boldsymbol{x}_{\tau})dt) \cdot (\boldsymbol{v}'_{\tau}(\boldsymbol{x}_{\tau}) - \boldsymbol{v}_{\tau}(\boldsymbol{x}_{\tau}))}{2\mu T} \\
&\quad + \frac{(\theta'^2 - \theta^2)\|\boldsymbol{v}'_{\tau}(\boldsymbol{x}_{\tau}) - \boldsymbol{v}_{\tau}(\boldsymbol{x}_{\tau})\|^2 dt}{4\mu T}.
\end{aligned}
\quad (36)
$$

Thus, the Kullback–Leibler divergence is calculated as

$$
\begin{aligned}
D_{\mathrm{KL}}(\mathbb{P}^{\theta}_{\mathbf{v}'_{\tau}} \| \mathbb{P}^{\theta'}_{\mathbf{v}'_{\tau}}) &= \int d\mathbf{x}_{\tau} P_{\tau}(\mathbf{x}_{\tau}) \int d\mathbf{x}_{\tau+dt} \mathbb{T}^{\theta}_{\tau;\mathbf{v}'_{\tau}}(\mathbf{x}_{\tau+dt} \mid \mathbf{x}_{\tau}) \ln \frac{\mathbb{T}^{\theta}_{\tau;\mathbf{v}'_{\tau}}(\mathbf{x}_{\tau+dt} \mid \mathbf{x}_{\tau})}{\mathbb{T}^{\theta'}_{\tau;\mathbf{v}'_{\tau}}(\mathbf{x}_{\tau+dt} \mid \mathbf{x}_{\tau})} \\
&= \int d\mathbf{x}_{\tau} P_{\tau}(\mathbf{x}_{\tau}) \left[\frac{(\theta'^2 - 2\theta\theta' + \theta^2)\|\mathbf{v}'_{\tau}(\mathbf{x}_{\tau}) - \mathbf{v}_{\tau}(\mathbf{x}_{\tau})\|^2 dt}{4\mu T} \right] \\
&= \frac{(\theta - \theta')^2 dt}{4\mu T} \int d\mathbf{x}_{\tau} \|\mathbf{v}'_{\tau}(\mathbf{x}_{\tau}) - \mathbf{v}_{\tau}(\mathbf{x}_{\tau})\|^2 P_{\tau}(\mathbf{x}_{\tau}).
\end{aligned}
\tag{37}
$$

$\square$

Remark 12 Let $g_{\theta(\mathbf{v}'_{\tau})\theta(\mathbf{v}'_{\tau})}(\mathbb{P})$ be the Fisher information of the interpolation parameter θ for $\mathbb{P}^{\theta}_{\mathbf{v}'_{\tau}}$ defined as

$$
g_{\theta(\mathbf{v}'_{\tau})\theta(\mathbf{v}'_{\tau})}(\mathbb{P}) = \lim_{\Delta\theta \to 0} \frac{2D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^{\Delta\theta}_{\mathbf{v}'_{\tau}})}{(\Delta\theta)^2}.
\tag{38}
$$

Information-geometrically, this Fisher information $g_{\theta(\mathbf{v}'_{\tau})\theta(\mathbf{v}'_{\tau})}(\mathbb{P})$ can be regarded as a particular Riemannian metric called the Fisher metric [13] at point $\mathbb{P}$. If we consider $\mathbb{P}^{\theta}_{\mathbf{v}'_{\tau}} = \mathbb{P}^0_{\mathbf{v}'_{\tau}} = \mathbb{P}$ and $\mathbb{P}^{\theta'}_{\mathbf{v}'_{\tau}} = \mathbb{P}^{\Delta\theta}_{\mathbf{v}'_{\tau}}$ in Lemma 5, the Fisher metric is given by

$$
g_{\theta(\mathbf{v}'_{\tau})\theta(\mathbf{v}'_{\tau})}(\mathbb{P}) = \frac{dt}{2\mu T} \int d\mathbf{x}_{\tau} \|\mathbf{v}'_{\tau}(\mathbf{x}_{\tau}) - \mathbf{v}_{\tau}(\mathbf{x}_{\tau})\|^2 P_{\tau}(\mathbf{x}_{\tau}).
\tag{39}
$$

Based on Lemma 5, we obtain an information-geometric interpretation of the entropy production, which is substantially obtained in Refs. [40, 54].

Theorem 6 *Let $\theta \in [0, 1]$ and $\theta' \in [0, 1]$ be any interpolation parameters. The entropy production rate σ_{τ} for original Fokker–Planck dynamics* (1) *is given by*

$$
\sigma_{\tau} = \lim_{dt \to 0} \frac{D_{\mathrm{KL}}(\mathbb{P}^{\theta}_{-\mathbf{v}_{\tau}} \| \mathbb{P}^{\theta'}_{-\mathbf{v}_{\tau}})}{(\theta - \theta')^2 dt}.
\tag{40}
$$

The entropy production $\Sigma(\tau + dt; \tau)$ for original Fokker–Planck dynamics (1) *is also given by*

$$
\Sigma(\tau + dt; \tau) = \frac{D_{\mathrm{KL}}(\mathbb{P}^{\theta}_{-\mathbf{v}_{\tau}} \| \mathbb{P}^{\theta'}_{-\mathbf{v}_{\tau}})}{(\theta - \theta')^2} + O(dt^2).
\tag{41}
$$

In terms of the Fisher information, the entropy production $\Sigma(\tau + dt; \tau)$ for original Fokker–Planck dynamics (1) *is given by half of the Fisher metric,*

$$
\Sigma(\tau + dt; \tau) = \sigma_{\tau} dt + O(dt^2) = \frac{1}{2} g_{\theta(-\mathbf{v}_{\tau})\theta(-\mathbf{v}_{\tau})}(\mathbb{P}) + O(dt^2).
\tag{42}
$$

Proof From Lemma 5, the Kullback–Leibler divergence $D_{\mathrm{KL}}(\mathbb{P}^{\theta}_{-\boldsymbol{\nu}_{\tau}}\|\mathbb{P}^{\theta'}_{-\boldsymbol{\nu}_{\tau}})$ is calculated as

$$
\begin{aligned}
D_{\mathrm{KL}}(\mathbb{P}^{\theta}_{-\boldsymbol{\nu}_{\tau}}\|\mathbb{P}^{\theta'}_{-\boldsymbol{\nu}_{\tau}}) &= \frac{(\theta-\theta')^2 dt}{\mu T} \int d\boldsymbol{x}_{\tau} \|\boldsymbol{\nu}_{\tau}(\boldsymbol{x}_{\tau})\|^2 P_{\tau}(\boldsymbol{x}_{\tau}) \\
&= (\theta-\theta')^2 dt\sigma_{\tau} \\
&= (\theta-\theta')^2 \Sigma(\tau+dt;\tau) + O(dt^2).
\end{aligned}
\tag{43}
$$

Thus, Eqs. (40) and (41) holds. If we consider $\mathbb{P}^{\theta}_{-\boldsymbol{\nu}_{\tau}} = \mathbb{P}^{0}_{-\boldsymbol{\nu}_{\tau}} = \mathbb{P}$ and $\mathbb{P}^{\theta'}_{-\boldsymbol{\nu}_{\tau}} = \mathbb{P}^{\Delta\theta}_{-\boldsymbol{\nu}_{\tau}}$ for Eq. (43) and use Eq. (38), we obtain Eq. (42). □

Remark 13 The entropy production can be regarded as half of the Fisher metric, and thus the entropy production can also be a particular Riemannian metric of differential geometry. Based on the Fisher metric, we can introduce the square of the line element ds^2_{path} defined as $ds^2_{\mathrm{path}} = g_{\theta(-\boldsymbol{\nu}_{\tau})\theta(-\boldsymbol{\nu}_{\tau})}(\mathbb{P})d\theta^2 = 2\Sigma(\tau+dt;\tau)d\theta^2 + O(dt^2)$, where θ is the interpolation parameter for $\mathbb{P}^{\theta}_{-\boldsymbol{\nu}_{\tau}}$ and this line element is introduced on the corresponding e-geodesic.

3.3 Thermodynamic uncertainty relations

The link between the entropy production rate and the Fisher metric leads to thermodynamic trade-off relations between the entropy production rate and the fluctuation of any observable. A particular case of thermodynamic trade-off relations was proposed as the thermodynamic uncertainty relations [56, 57]. In Refs. [24, 40, 43, 58–60], several links between the Cramér–Rao bound and generalizations of the thermodynamic uncertainty relations have been discussed. Here, we newly propose a generalization of the thermodynamic uncertainty relations based on the fact that the entropy production is regarded as half of the Fisher information in Theorem 6. To obtain the proposed thermodynamic uncertainty relation, we start with the Cramér–Rao bound.

Lemma 7 *Let $R(\boldsymbol{x}_{\tau+dt},\boldsymbol{x}_{\tau}) \in \mathbb{R}$ be any function of states $\boldsymbol{x}_{\tau+dt} \in \mathbb{R}^d$ and $\boldsymbol{x}_{\tau} \in \mathbb{R}^d$. The Fisher metric $g_{\theta(\boldsymbol{\nu}'_{\tau})\theta(\boldsymbol{\nu}'_{\tau})}(\mathbb{P})$ is bounded by the Cramér–Rao bound as follows,*

$$
g_{\theta(\boldsymbol{\nu}'_{\tau})\theta(\boldsymbol{\nu}'_{\tau})}(\mathbb{P}) \geq \frac{\left(\partial_{\theta}\mathbb{E}_{\mathbb{P}^{\theta}_{\boldsymbol{\nu}'_{\tau}}}[R]\right)^2\Big|_{\theta=0}}{\mathbb{E}_{\mathbb{P}^0_{\boldsymbol{\nu}'_{\tau}}}\left[(\Delta_{\mathbb{P}^0_{\boldsymbol{\nu}'_{\tau}}}R)^2\right]},
\tag{44}
$$

where $|_{\theta=0}$ stands for substitution $\theta=0$, $\mathbb{E}_{\mathbb{P}^{\theta}_{\boldsymbol{\nu}'_{\tau}}}[R]$ is the expected value defined as

$$
\mathbb{E}_{\mathbb{P}^{\theta}_{\boldsymbol{\nu}'_{\tau}}}[R] = \int d\boldsymbol{x}_{\tau+dt}d\boldsymbol{x}_{\tau}\mathbb{P}^{\theta}_{\boldsymbol{\nu}'_{\tau}}(\boldsymbol{x}_{\tau+dt},\boldsymbol{x}_{\tau})R(\boldsymbol{x}_{\tau+dt},\boldsymbol{x}_{\tau}),
\tag{45}
$$

and the deviation $\Delta_{\mathbb{P}^{\theta}_{\boldsymbol{\nu}'_{\tau}}}R(\boldsymbol{x}_{\tau+dt},\boldsymbol{x}_{\tau})$ is defined as $\Delta_{\mathbb{P}^{\theta}_{\boldsymbol{\nu}'_{\tau}}}R(\boldsymbol{x}_{\tau+dt},\boldsymbol{x}_{\tau}) = R(\boldsymbol{x}_{\tau+dt},\boldsymbol{x}_{\tau}) - \mathbb{E}_{\mathbb{P}^{\theta}_{\boldsymbol{\nu}'_{\tau}}}[R]$.

Proof The Fisher metric $g_{\theta(\boldsymbol{v}'_\tau)\theta(\boldsymbol{v}'_\tau)}(\mathbb{P})$ is calculated as

$$g_{\theta(\boldsymbol{v}'_\tau)\theta(\boldsymbol{v}'_\tau)}(\mathbb{P}^0_{\boldsymbol{v}'_\tau}) = \lim_{\theta \to 0} \frac{2 \int d\boldsymbol{x}_{\tau+dt} d\boldsymbol{x}_\tau \mathbb{P}^0_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \ln \frac{\mathbb{P}^0_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau)}{\mathbb{P}^\theta_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau)}}{\theta^2}$$

$$= \int d\boldsymbol{x}_{\tau+dt} d\boldsymbol{x}_\tau \mathbb{P}^0_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \left(\partial_\theta \ln \mathbb{P}^\theta_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \right)^2 \Big|_{\theta=0} . \tag{46}$$

From the Cauchy–Schwartz inequality, we obtain the Cramér–Rao bound,

$$(\partial_\theta \mathbb{E}_{\mathbb{P}^\theta_{\boldsymbol{v}'_\tau}}[R])^2 \Big|_{\theta=0}$$

$$= \left[\int d\boldsymbol{x}_{\tau+dt} d\boldsymbol{x}_\tau \left(\sqrt{\mathbb{P}^0_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau)} \right)^2 \Delta_{\mathbb{P}^0_{\boldsymbol{v}'_\tau}} R(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \left(\partial_\theta \ln \mathbb{P}^\theta_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \right) \Big|_{\theta=0} \right]^2$$

$$\leq \left[\int d\boldsymbol{x}_{\tau+dt} d\boldsymbol{x}_\tau \mathbb{P}^0_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \left(\Delta_{\mathbb{P}^0_{\boldsymbol{v}'_\tau}} R(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \right)^2 \right]$$

$$\times \left[\int d\boldsymbol{x}_{\tau+dt} d\boldsymbol{x}_\tau \mathbb{P}^0_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \left(\partial_\theta \ln \mathbb{P}^\theta_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \right)^2 \Big|_{\theta=0} \right]$$

$$= \mathbb{E}_{\mathbb{P}^0_{\boldsymbol{v}'_\tau}} \left[(\Delta_{\mathbb{P}^0_{\boldsymbol{v}'_\tau}} R)^2 \right] g_{\theta(\boldsymbol{v}'_\tau)\theta(\boldsymbol{v}'_\tau)}(\mathbb{P}^0_{\boldsymbol{v}'_\tau}), \tag{47}$$

where we used $\int d\boldsymbol{x}_{\tau+dt} d\boldsymbol{x}_\tau \partial_\theta \mathbb{P}^\theta_{\boldsymbol{v}'_\tau}(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) = 0$. $\qquad\square$

By plugging Eq. (42) into the Cramér–Rao bound, the proposed generalized thermodynamic uncertainty relation, which is a trade-off relation between the entropy production rate σ_τ and the fluctuation of any observable $\mathrm{Var}[R] = \mathbb{E}_\mathbb{P}[(\Delta_\mathbb{P} R)^2]$, can be obtained as follows.

Proposition 8 *Let $R(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \in \mathbb{R}$ be any function of states $\boldsymbol{x}_{\tau+dt} \in \mathbb{R}^d$ and $\boldsymbol{x}_\tau \in \mathbb{R}^d$ such that $\mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) R(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau) \to 0$ at infinity of $\boldsymbol{x}_{\tau+dt}$. The entropy production rate σ_τ is bounded by the generalized thermodynamic uncertainty relation,*

$$\sigma_\tau dt \geq \frac{2(\mathcal{J}[\tilde{\boldsymbol{R}}])^2}{\mathrm{Var}[R]/dt} dt + O(dt^2), \tag{48}$$

where $\mathrm{Var}[R]$ is the variance defined as $\mathrm{Var}[R] = \mathbb{E}_\mathbb{P}[(\Delta_\mathbb{P} R)^2]$ and $\mathcal{J}[\tilde{\boldsymbol{R}}]$ is the generalized current defined as

$$\mathcal{J}[\tilde{\boldsymbol{R}}] = \int d\boldsymbol{x}_\tau \boldsymbol{v}(\boldsymbol{x}_\tau) \cdot \tilde{\boldsymbol{R}}(\boldsymbol{x}_\tau) P_\tau(\boldsymbol{x}_\tau), \tag{49}$$

$$\tilde{\boldsymbol{R}}(\boldsymbol{x}_\tau) = \int d\boldsymbol{x}_{\tau+dt} \mathbb{T}_\tau(\boldsymbol{x}_{\tau+dt} \mid \boldsymbol{x}_\tau) \nabla_{\boldsymbol{x}_{\tau+dt}} R(\boldsymbol{x}_{\tau+dt}, \boldsymbol{x}_\tau). \tag{50}$$

Here, $\nabla_{\boldsymbol{x}}$ is a gradient operator for $\boldsymbol{x} \in \mathbb{R}^d$.

Proof By plugging $\mathbb{P}^{\theta}_{\boldsymbol{\nu}'_{\tau}} = \mathbb{P}^{\theta}_{-\boldsymbol{\nu}_{\tau}}$ into Lemma. 7, we obtain

$$g_{\theta(-\boldsymbol{\nu}_{\tau})\theta(-\boldsymbol{\nu}_{\tau})}(\mathbb{P}) \geq \frac{\left(\partial_{\theta}\mathbb{E}_{\mathbb{P}^{\theta}_{-\boldsymbol{\nu}_{\tau}}}[R]\right)^2\Big|_{\theta=0}}{\mathrm{Var}[R]}, \tag{51}$$

where we used $\mathbb{P}^0_{-\boldsymbol{\nu}_{\tau}} = \mathbb{P}$. From Eq. (42), we obtain $g_{\theta(-\boldsymbol{\nu}_{\tau})\theta(-\boldsymbol{\nu}_{\tau})}(\mathbb{P}) = 2\sigma_{\tau}dt + O(dt^2)$. $\partial_{\theta}\mathbb{E}_{\mathbb{P}^{\theta}_{-\boldsymbol{\nu}_{\tau}}}[R]\Big|_{\theta=0}$ is calculated as

$$\begin{aligned}
&\partial_{\theta}\mathbb{E}_{\mathbb{P}^{\theta}_{-\boldsymbol{\nu}_{\tau}}}[R]\Big|_{\theta=0} \\
&= \int d\boldsymbol{x}_{\tau+dt}d\boldsymbol{x}_{\tau}R(\boldsymbol{x}_{\tau+dt},\boldsymbol{x}_{\tau})P_{\tau}(\boldsymbol{x}_{\tau})\partial_{\theta}\mathbb{T}^{\theta}_{\tau;-\boldsymbol{\nu}_{\tau}}(\boldsymbol{x}_{\tau+dt}\mid\boldsymbol{x}_{\tau})|_{\theta=0} \\
&= \frac{-\int d\boldsymbol{x}_{\tau}P_{\tau}(\boldsymbol{x}_{\tau})\boldsymbol{\nu}(\boldsymbol{x}_{\tau})\cdot\left[\int d\boldsymbol{x}_{\tau+dt}R(\boldsymbol{x}_{\tau+dt},\boldsymbol{x}_{\tau})(\boldsymbol{x}_{\tau+dt}-\boldsymbol{x}_{\tau}-\mu\boldsymbol{F}_{\tau}(\boldsymbol{x}_{\tau})dt)\mathbb{T}_{\tau}(\boldsymbol{x}_{\tau+dt}\mid\boldsymbol{x}_{\tau})\right]}{\mu T} \\
&= 2dt\int d\boldsymbol{x}_{\tau}P_{\tau}(\boldsymbol{x}_{\tau})\boldsymbol{\nu}(\boldsymbol{x}_{\tau})\cdot\left[\int d\boldsymbol{x}_{\tau+dt}R(\boldsymbol{x}_{\tau+dt},\boldsymbol{x}_{\tau})\nabla_{\boldsymbol{x}_{\tau+dt}}\mathbb{T}_{\tau}(\boldsymbol{x}_{\tau+dt}\mid\boldsymbol{x}_{\tau})\right] \\
&= -2dt\int d\boldsymbol{x}_{\tau}P_{\tau}(\boldsymbol{x}_{\tau})\boldsymbol{\nu}(\boldsymbol{x}_{\tau})\cdot\left[\int d\boldsymbol{x}_{\tau+dt}\mathbb{T}_{\tau}(\boldsymbol{x}_{\tau+dt}\mid\boldsymbol{x}_{\tau})\nabla_{\boldsymbol{x}_{\tau+dt}}R(\boldsymbol{x}_{\tau+dt},\boldsymbol{x}_{\tau})\right] \\
&= -2\mathcal{J}[\tilde{\boldsymbol{R}}]dt, \tag{52}
\end{aligned}$$

where we used $\int d\boldsymbol{x}_{\tau+dt}\nabla_{\boldsymbol{x}_{\tau+dt}}[\mathbb{T}_{\tau}(\boldsymbol{x}_{\tau+dt}\mid\boldsymbol{x}_{\tau})R(\boldsymbol{x}_{\tau+dt},\boldsymbol{x}_{\tau})] = 0$ because of the assumption $\mathbb{T}_{\tau}(\boldsymbol{x}_{\tau+dt}\mid\boldsymbol{x}_{\tau})R(\boldsymbol{x}_{\tau+dt},\boldsymbol{x}_{\tau}) \to 0$ at infinity of $\boldsymbol{x}_{\tau+dt}$. From $\left(\partial_{\theta}\mathbb{E}_{\mathbb{P}^{\theta}_{-\boldsymbol{\nu}_{\tau}}}[R]\right)^2\Big|_{\theta=0} = 4(\mathcal{J}[\tilde{\boldsymbol{R}}])^2(dt)^2$ and $g_{\theta(-\boldsymbol{\nu}_{\tau})\theta(-\boldsymbol{\nu}_{\tau})}(\mathbb{P}) = 2\sigma_{\tau}dt + O(dt^2)$, Eq. (51) can be rewritten as Eq. (48). □

Remark 14 In Ref. [58], a special case of the proposed thermodynamic uncertainty relation (48) was discussed for

$$R(\boldsymbol{x}_{\tau+dt},\boldsymbol{x}_{\tau}) = (\boldsymbol{x}_{\tau+dt}-\boldsymbol{x}_{\tau})\circ\boldsymbol{w}(\boldsymbol{x}_{\tau}) = (\boldsymbol{x}_{\tau+dt}-\boldsymbol{x}_{\tau})\cdot\boldsymbol{w}\left(\frac{\boldsymbol{x}_{\tau+dt}+\boldsymbol{x}_{\tau}}{2}\right), \tag{53}$$

where $\boldsymbol{w}(\boldsymbol{x}_{\tau}) \in \mathbb{R}^d$ is any function of $\boldsymbol{x}_{\tau}$. In this case, we obtain $\tilde{\boldsymbol{R}}(\boldsymbol{x}_{\tau}) = \boldsymbol{w}(\boldsymbol{x}_{\tau}) + O(dt)$ and $\mathrm{Var}[R]/dt = 2\mu T\int d\boldsymbol{x}_{\tau}\|\boldsymbol{w}(\boldsymbol{x}_{\tau})\|^2 P_{\tau}(\boldsymbol{x}_{\tau}) + O(dt)$. Thus, the generalized thermodynamic uncertainty relation Eq. (48) can be rewritten as

$$\sigma_{\tau}dt \geq \frac{(\mathcal{J}[\boldsymbol{w}])^2}{\mu T\int d\boldsymbol{x}_{\tau}\|\boldsymbol{w}(\boldsymbol{x}_{\tau})\|^2 P_{\tau}(\boldsymbol{x}_{\tau})}dt + O(dt^2). \tag{54}$$

This result can also be easily obtained from the Cauchy-Schwartz inequality as follows,

$$(\mathcal{J}[\boldsymbol{w}])^2 = \left(\int d\boldsymbol{x}_\tau \, \boldsymbol{v}(\boldsymbol{x}_\tau) \cdot \boldsymbol{w}(\boldsymbol{x}_\tau) P_\tau(\boldsymbol{x}_\tau)\right)^2$$

$$\leq \left(\int d\boldsymbol{x}_\tau \, \|\boldsymbol{w}(\boldsymbol{x}_\tau)\|^2 P_\tau(\boldsymbol{x}_\tau)\right) \left(\int d\boldsymbol{x}_\tau \, \|\boldsymbol{v}(\boldsymbol{x}_\tau)\|^2 P_\tau(\boldsymbol{x}_\tau)\right)$$

$$= \mu T \sigma_\tau \int d\boldsymbol{x}_\tau \, \|\boldsymbol{w}(\boldsymbol{x}_\tau)\|^2 P_\tau(\boldsymbol{x}_\tau). \tag{55}$$

4 Optimal transport theory and entropy production

4.1 L^2-Wasserstein distance and minimum entropy production

We next discuss a relation between the L^2-Wasserstein distance in optimal transport theory [36] and the entropy production. We start with the Benamou–Brenier formula [61], which gives the definition of the L^2-Wasserstein distance.

Definition 5 Let $P(\boldsymbol{x})$ and $Q(\boldsymbol{x})$ be probability densities at the position $\boldsymbol{x} \in \mathbb{R}^d$ that satisfy $\int d\boldsymbol{x} \, P(\boldsymbol{x}) = \int d\boldsymbol{x} \, Q(\boldsymbol{x}) = 1$, $P(\boldsymbol{x}) \geq 0$ and $Q(\boldsymbol{x}) \geq 0$, where the second-order moments $\int d\boldsymbol{x} \|\boldsymbol{x}\|^2 P(\boldsymbol{x})$ and $\int d\boldsymbol{x} \|\boldsymbol{x}\|^2 Q(\boldsymbol{x})$ are finite. Let $\Delta\tau \geq 0$ be a non-negative time interval. *The L^2-Wasserstein distance between probability densities P and Q is* defined as

$$\mathcal{W}_2(P, Q) = \sqrt{\inf \Delta\tau \int_\tau^{\tau+\Delta\tau} dt \int d\boldsymbol{x} \|\boldsymbol{v}_t(\boldsymbol{x})\|^2 \mathcal{P}_t(\boldsymbol{x})}, \tag{56}$$

where the infimum is taken among all paths $(\boldsymbol{v}_t(\boldsymbol{x}), \mathcal{P}_t(\boldsymbol{x}))_{\tau \leq t \leq \tau+\Delta\tau}$ satisfying the continuity equation

$$\partial_t \mathcal{P}_t(\boldsymbol{x}) = -\nabla \cdot (\boldsymbol{v}_t(\boldsymbol{x}) \mathcal{P}_t(\boldsymbol{x})), \tag{57}$$

with the boundary conditions

$$\mathcal{P}_\tau(\boldsymbol{x}) = P(\boldsymbol{x}), \tag{58}$$

$$\mathcal{P}_{\tau+\Delta\tau}(\boldsymbol{x}) = Q(\boldsymbol{x}). \tag{59}$$

Remark 15 This definition of the L^2-Wasserstein distance is consistent with the definitions used in the Monge–Kantorovich problem [35]. The definition in the Monge–Kantrovich problem is as follows. Let $\Pi(\boldsymbol{x}, \boldsymbol{x}')$ be the joint probability density at positions $\boldsymbol{x} \in \mathbb{R}^d$ and $\boldsymbol{x}' \in \mathbb{R}^d$ that satisfies $\Pi(\boldsymbol{x}, \boldsymbol{x}') \geq 0$, $\int d\boldsymbol{x}' \Pi(\boldsymbol{x}, \boldsymbol{x}') = P(\boldsymbol{x})$, and $\int d\boldsymbol{x} \Pi(\boldsymbol{x}, \boldsymbol{x}') = Q(\boldsymbol{x}')$. The L^2-Wasserstein distance is also defined as

$$\mathcal{W}_2(P, Q) = \sqrt{\inf_{\Pi(\boldsymbol{x}, \boldsymbol{x}')} \int d\boldsymbol{x} d\boldsymbol{x}' \|\boldsymbol{x} - \boldsymbol{x}'\|^2 \Pi(\boldsymbol{x}, \boldsymbol{x}')}. \tag{60}$$

Remark 16 The L^2-Wasserstein distance is a distance [35], since it is symmetric $\mathcal{W}_2(P, Q) = \mathcal{W}_2(Q, P)$, non-negative $\mathcal{W}_2(P, Q) \geq 0$, and zero $\mathcal{W}_2(P, Q) = 0$ if and only if $P = Q$. The triangle inequality $\mathcal{W}_2(P, Q) \leq \mathcal{W}_2(P, P') + \mathcal{W}_2(P', Q)$ for any probability density $P'(\boldsymbol{x})$ with a finite second-order moment is satisfied.

Based on the Benamou–Breiner formula, we can consider the minimum entropy production for the Fokker–Planck equation in terms of the L^2-Wasserstein distance. This link between the L^2-Wasserstein distance and the entropy production was initially pointed out in the field of the optimal transport theory (for example, in Refs. [36, 62]). After that, it was also discussed in the context of stochastic thermodynamics [19, 63]. This link has been recently revisited in terms of thermodynamic trade-off relations such as the thermodynamic speed limit [31, 64] and the thermodynamic uncertainty relation [40]. The decomposition of the entropy production rate based on the optimal transport theory has also been proposed in Refs. [31, 39].

For a stochastic process evolving according to the Fokker–Planck equation, the entropy production $\Sigma(\tau + \Delta\tau; \tau)$ for a fixed initial probability density P_τ and a fixed final probability density $P_{\tau+\Delta\tau}$ is bounded by the L^2-Wasserstein distance as follows. This result is regarded as a thermodynamic speed limit [31, 64], which is a trade-off relation between the finite time interval $\Delta\tau$ and the entropy production $\Sigma(\tau + \Delta\tau; \tau)$.

Lemma 9 *The entropy production $\Sigma(\tau + \Delta\tau; \tau)$ for fixed probability densities P_τ and $P_{\tau+\Delta\tau}$ is bounded by*

$$\Sigma(\tau + \Delta\tau; \tau) \geq \frac{[\mathcal{W}_2(P_\tau, P_{\tau+\Delta\tau})]^2}{\mu T \Delta\tau}. \tag{61}$$

The entropy production rate σ_τ is also bounded by

$$\sigma_\tau \geq \frac{1}{\mu T} \lim_{\Delta t \to 0} \frac{[\mathcal{W}_2(P_\tau, P_{\tau+\Delta t})]^2}{(\Delta t)^2}. \tag{62}$$

Proof From the definition of the L^2-Wasserstein distance, we obtain Eq. (61)

$$
\begin{aligned}
[\mathcal{W}_2(P_\tau, P_{\tau+\Delta t})]^2 &= \inf \Delta\tau \int_\tau^{\tau+\Delta\tau} dt \int d\boldsymbol{x} \|\boldsymbol{v}_t(\boldsymbol{x})\|^2 P_t(\boldsymbol{x}) \\
&\leq \Delta\tau \int_\tau^{\tau+\Delta\tau} dt \int d\boldsymbol{x} \|\boldsymbol{v}_t(\boldsymbol{x})\|^2 P_t(\boldsymbol{x}) \\
&= \mu T \Sigma(\tau + \Delta\tau; \tau),
\end{aligned}
\tag{63}
$$

because the time evolution of $P_t(\boldsymbol{x})$ with boundary conditions $P_\tau(\boldsymbol{x})$ and $P_{\tau+\Delta t}(\boldsymbol{x})$ is described by the Fokker–Planck equation $\partial_t P_t(\boldsymbol{x}) = -\nabla \cdot (\boldsymbol{v}_t(\boldsymbol{x}) P_t(\boldsymbol{x}))$, which is a kind of continuity equation. By using $\Delta\tau = \Delta t$, we obtain $[\mathcal{W}_2(P_\tau, P_{\tau+\Delta t})]^2 \leq (\Delta t)^2(\sigma_\tau + O(\Delta t))$, and thus Eq. (62) holds. □

By considering the geometry of the L^2-Wasserstein distance and introducing the L^2-Wasserstein path length, we can obtain another lower bound on the entropy production,

which is the tighter than Eq. (61). This bound is proposed as a tighter version of the thermodynamic speed limit in Ref. [31]. The L^2-Wasserstein path length is defined as follows.

Definition 6 Let $t \in \mathbb{R}$ and $s \in \mathbb{R}$ indicate time. For a fixed trajectory of probability density $(P_t)_{\tau \leq t \leq \tau + \Delta \tau}$, the Wasserstein path length from time $t = \tau$ to time $t = \tau + \Delta \tau$ is defined as

$$\mathcal{L}(\tau + \Delta \tau; \tau) = \mathcal{L}_{\tau + \Delta \tau} - \mathcal{L}_\tau = \int_\tau^{\tau + \Delta \tau} ds \left[\lim_{\Delta t \to 0} \frac{\mathcal{W}_2(P_s, P_{s + \Delta t})}{\Delta t} \right], \quad (64)$$

$$\mathcal{L}_t = \int_{-\infty}^t ds \left[\lim_{\Delta t \to 0} \frac{\mathcal{W}_2(P_s, P_{s + \Delta t})}{\Delta t} \right]. \quad (65)$$

Remark 17 We can obtain $\mathcal{L}(\tau + \Delta \tau; \tau) \geq \mathcal{W}_2(P_\tau, P_{\tau + \Delta \tau})$ by using the triangle inequality of the L^2-Wasserstein distance. Thus, the L^2-Wasserstein distance $\mathcal{W}_2(P_\tau, P_{\tau + \Delta \tau})$ can be regarded as the minimum Wasserstein path length, that is the geodesic between two points P_τ and $P_{\tau + \Delta \tau}$.

Remark 18 In terms of the Wasserstein path length, Eq. (62) can be rewritten as

$$\sigma_\tau \geq \frac{1}{\mu T} \left. (\partial_s \mathcal{L}_s)^2 \right|_{s = \tau}. \quad (66)$$

Theorem 10 *The entropy production $\Sigma(\tau + \Delta \tau; \tau)$ is bounded by*

$$\Sigma(\tau + \Delta \tau; \tau) \geq \frac{[\mathcal{L}(\tau + \Delta \tau; \tau)]^2}{\mu T \Delta \tau}. \quad (67)$$

This lower bound on the entropy production is tighter than the bound Eq. (61) as follows,

$$\Sigma(\tau + \Delta \tau; \tau) \geq \frac{[\mathcal{L}(\tau + \Delta \tau; \tau)]^2}{\mu T \Delta \tau} \geq \frac{[\mathcal{W}_2(P_\tau, P_{\tau + \Delta \tau})]^2}{\mu T \Delta \tau}. \quad (68)$$

Proof From the Cauchy–Schwartz inequality, we obtain

$$\int_\tau^{\tau + \Delta \tau} dt \left. (\partial_s \mathcal{L}_s)^2 \right|_{s = t} \int_\tau^{\tau + \Delta \tau} dt \geq \left[\int_\tau^{\tau + \Delta \tau} dt (\partial_s \mathcal{L}_s) \Big|_{s = t} \right]^2$$
$$= [\mathcal{L}(\tau + \Delta \tau; \tau)]^2, \quad (69)$$

where we used $\partial_s \mathcal{L}_s \geq 0$. From Eq. (66), we obtain Eq. (67) as follows,

$$\Sigma(\tau + \Delta \tau; \tau) = \int_\tau^{\tau + \Delta \tau} dt \sigma_t$$

$$\geq \frac{1}{\mu T} \int_\tau^{\tau+\Delta\tau} dt \, (\partial_s \mathcal{L}_s)^2 \bigg|_{s=t}$$

$$\geq \frac{[\mathcal{L}(\tau + \Delta\tau; \tau)]^2}{\mu T \Delta\tau}. \tag{70}$$

From the inequality $\mathcal{L}(\tau + \Delta\tau; \tau) \geq \mathcal{W}_2(P_\tau, P_{\tau+\Delta\tau})$ and Eq. (67), we obtain Eq. (68).
□

4.2 Geometric decomposition of entropy production rate

Based on Eq. (66), we can obtain a decomposition of the entropy production into two non-negative parts, namely the housekeeping entropy production rate and the excess entropy production rate. This decomposition has been substaintially obtained in Ref. [31], and discussed in Refs. [39, 40] from the viewpoint of the thermodynamic uncertainty relation.

Here, we define the housekeeping entropy production rate and the excess entropy production rate based on Eq. (66).

Definition 7 *The excess entropy production rate* is defined as

$$\sigma_\tau^{\text{ex}} = \frac{1}{\mu T} (\partial_s \mathcal{L}_s)^2 \bigg|_{s=\tau} = \frac{1}{\mu T} \lim_{\Delta t \to 0} \frac{[\mathcal{W}_2(P_\tau, P_{\tau+\Delta t})]^2}{(\Delta t)^2}, \tag{71}$$

and *the housekeeping entropy production rate* is defined as

$$\sigma_\tau^{\text{hk}} = \sigma_\tau - \frac{1}{\mu T} (\partial_s \mathcal{L}_s)^2 \bigg|_{s=\tau}. \tag{72}$$

Remark 19 The excess entropy production rate is definitely non-negative $\sigma_\tau^{\text{ex}} \geq 0$. The housekeeping entropy production is non-negative $\sigma_\tau^{\text{hk}} \geq 0$ because of Eq. (66). The entropy production rate is decomposed as $\sigma_\tau = \sigma_\tau^{\text{ex}} + \sigma_\tau^{\text{hk}}$.

Remark 20 The excess entropy production rate becomes zero $\sigma_\tau^{\text{ex}} = 0$ if and only if the system is in the steady-state $\partial_t P_t(\boldsymbol{x}) = 0$, or equivalently $P_\tau = P_{\tau+dt}$ for an infinitesimal time interval dt. The decomposition of the entropy production rate such that the excess entropy production rate becomes zero in the steady state is not unique, and another example of the decomposition of the entropy production rate has been obtained in the study of the steady-state thermodynamics [65]. Our definitions of the excess entropy production rate and the housekeeping entropy production rate are generally different from the excess entropy production rate and the housekeeping entropy production rate proposed in Ref. [65]. We discussed this difference in Ref. [40].

Remark 21 The thermodynamic speed limit can be tightened by using the excess entropy production rate as follows.

$$\Sigma(\tau + \Delta\tau; \tau) \geq \int_{\tau}^{\tau+\Delta\tau} dt\, \sigma_t^{\text{ex}} \geq \frac{[\mathcal{L}(\tau + \Delta\tau; \tau)]^2}{\mu T \Delta t} \geq \frac{[\mathcal{W}_2(P_\tau, P_{\tau+\Delta\tau})]^2}{\mu T \Delta\tau}. \quad (73)$$

The contribution of the housekeeping entropy production rate does not affect in the thermodynamic speed limit and the lower bound becomes tighter if $\sigma_t^{\text{hk}} = 0$. The lower bound $\Sigma(\tau + \Delta\tau; \tau) = [\mathcal{W}_2(P_\tau, P_{\tau+\Delta\tau})]^2/(\mu T \Delta\tau)$ is achieved when

$$(\partial_s \mathcal{L}_s)|_{s=t} = \frac{\mathcal{W}_2(P_\tau, P_{\tau+\Delta\tau})}{\Delta\tau} = \text{const.} \quad (74)$$

and $\sigma_t^{\text{hk}} = 0$ for $\tau \leq t \leq \tau + \Delta\tau$. This implies that the geodesic in the space of the L^2-Wasserstein distance is related to the optimal protocol that minimizes the entropy production in a finite time. The condition $\sigma_t^{\text{hk}} = 0$ is related to the condition of the potential force as discussed below. Thus, if we want to minimize the entropy production in a finite time, the probability density P_t should be changed along the geodesic in the space of the L^2-Wasserstein distance by the potential force [31].

To discuss a physical interpretation of this decomposition, we focus on another expression of the optimal protocol proposed in Ref. [61].

Lemma 11 *The L^2-Wasserstein distance is given by*

$$\mathcal{W}_2(P_\tau, P_{\tau+\Delta\tau}) = \sqrt{\Delta\tau \int_{\tau}^{\tau+\Delta\tau} dt \int dx \|v_t^*(x)\|^2 P_t(x)}. \quad (75)$$

Here, $v_t^*(x) \in \mathbb{R}^d$ is a vector field, namely an optimal mean local velocity, that satisfies

$$\partial_t P_t(x) = -\nabla \cdot (v_t^*(x) P_t(x)), \quad (76)$$

$$v_t^*(x) = \nabla\phi_t(x), \quad (77)$$

$$\partial_t \phi_t(x) = -\frac{1}{2}\|\nabla\phi_t(x)\|^2, \quad (78)$$

with a potential $\phi_t(x) \in \mathbb{R}$ and a time evolution of $P_t(x)$ that connects $P_\tau(x)$ and $P_{\tau+\Delta\tau}(x)$.

Proof Using the method of Lagrange multipliers, the optimization problem in Eq. (56) with the constraint $\partial_t P_t(x) = -\nabla \cdot (v_t^*(x) P_t(x))$ can be solved by the calculus of variations for $(P_t)_{\tau < t < \tau+\Delta\tau}$ and $(v_t^*)_{\tau \leq t < \tau+\Delta\tau}$,

$$\partial_{P_t(x)} \mathbb{L}(\{P_t\}, \{v_t^*\}, \{\phi_t\}) = \partial_{(v_t^*(x))_i} \mathbb{L}(\{P_t\}, \{v_t^*\}, \{\phi_t\}) = 0, \quad (79)$$

with the Lagrangian

$$\mathbb{L}(\{P_t\}, \{v_t^*\}, \{\phi_t\})$$

$$
= \int_\tau^{\tau+\Delta\tau} dt \int d\boldsymbol{x} \left[\frac{1}{2} \|\boldsymbol{v}_t^*(\boldsymbol{x})\|^2 P_t(\boldsymbol{x}) + \phi_t(\boldsymbol{x})[\partial_t P_t(\boldsymbol{x}) + \nabla \cdot (\boldsymbol{v}_t^*(\boldsymbol{x}) P_t(\boldsymbol{x}))] \right]
$$

$$
= \int_\tau^{\tau+\Delta\tau} dt \int d\boldsymbol{x} \left[\frac{1}{2} \|\boldsymbol{v}_t^*(\boldsymbol{x})\|^2 P_t(\boldsymbol{x}) - P_t(\boldsymbol{x})\partial_t \phi_t(\boldsymbol{x}) - \nabla\phi_t(\boldsymbol{x}) \cdot \boldsymbol{v}_t^*(\boldsymbol{x}) P_t(\boldsymbol{x}) \right]
$$

$$
+ \int_\tau^{\tau+\Delta\tau} dt \int d\boldsymbol{x} \left[\partial_t[\phi_t(\boldsymbol{x}) P_t(\boldsymbol{x})] + \nabla \cdot (\phi_t(\boldsymbol{x})\boldsymbol{v}_t^*(\boldsymbol{x}) P_t(\boldsymbol{x})) \right], \tag{80}
$$

where $\partial_{P_t(\boldsymbol{x})}$ and $\partial_{(\boldsymbol{v}_t^*(\boldsymbol{x}))_i}$ stand for functional derivatives, $(\boldsymbol{v}_t^*(\boldsymbol{x}))_i$ stands for i-th component of $\boldsymbol{v}_t^*(\boldsymbol{x})$, and $\phi_t(\boldsymbol{x})$ is the Lagrange multiplier. The variations $\partial_{(\boldsymbol{v}_t^*(\boldsymbol{x}))_i}\mathbb{L}(\{P_t\}, \{\boldsymbol{v}_t^*\}, \{\phi_t\}) = 0$ and $\partial_{P_t(\boldsymbol{x})}\mathbb{L}(\{P_t\}, \{\boldsymbol{v}_t^*\}, \{\phi_t\}) = 0$ in Eq. (79) are calculated as $\boldsymbol{v}_t^*(\boldsymbol{x}) = \nabla\phi_t(\boldsymbol{x})$ and $\partial_t \phi_t(\boldsymbol{x}) = \|\boldsymbol{v}_t^*(\boldsymbol{x})\|^2/2 - (\nabla\phi_t(\boldsymbol{x}) \cdot \boldsymbol{v}_t^*(\boldsymbol{x})) = -\|\nabla\phi_t(\boldsymbol{x})\|^2/2$, respectively.   □

From this optimal protocol, the excess entropy production rate and the housekeeping entropy production rate can be regarded as a potential contribution and a non-potential contribution to the entropy production rate, respectively. This fact is given by the following theorem.

Theorem 12 *Let $\phi_\tau(\boldsymbol{x}) \in \mathbb{R}$ be the potential that satisfies*

$$
\partial_\tau P_\tau(\boldsymbol{x}) = -\nabla \cdot (\boldsymbol{v}_\tau(\boldsymbol{x}) P_\tau(\boldsymbol{x})) = -\nabla \cdot (\boldsymbol{v}_\tau^*(\boldsymbol{x}) P_\tau(\boldsymbol{x})), \tag{81}
$$

where $\boldsymbol{v}_\tau^*(\boldsymbol{x})$ is the optimal mean local velocity defined as $\boldsymbol{v}_\tau^*(\boldsymbol{x}) = \nabla\phi_\tau(\boldsymbol{x})$, and $\partial_\tau P_\tau(\boldsymbol{x}) = -\nabla \cdot (\boldsymbol{v}_\tau(\boldsymbol{x}) P_\tau(\boldsymbol{x}))$ is the Fokker–Planck equation with the mean local velocity $\boldsymbol{v}_\tau(\boldsymbol{x}) = \mu(\boldsymbol{F}_\tau(\boldsymbol{x}) - T\nabla \ln P_\tau(\boldsymbol{x}))$. We assume that $P_\tau(\boldsymbol{x})$ decays sufficiently rapidly at infinity. The excess entropy production rate is given by

$$
\sigma_\tau^{\mathrm{ex}} = \frac{1}{\mu T} \int d\boldsymbol{x} \|\boldsymbol{v}_\tau^*(\boldsymbol{x})\|^2 P_\tau(\boldsymbol{x}), \tag{82}
$$

and the housekeeping entropy production rate is given by

$$
\sigma_\tau^{\mathrm{hk}} = \frac{1}{\mu T} \int d\boldsymbol{x} \|\boldsymbol{v}_\tau(\boldsymbol{x}) - \boldsymbol{v}_\tau^*(\boldsymbol{x})\|^2 P_\tau(\boldsymbol{x}). \tag{83}
$$

Proof From Lemma 11, the excess entropy production rate is calculated as

$$
\sigma_\tau^{\mathrm{ex}} = \frac{1}{\mu T} \lim_{\Delta t \to 0} \frac{[\mathcal{W}_2(P_\tau, P_{\tau+\Delta t})]^2}{(\Delta t)^2}
$$

$$
= \frac{1}{\mu T} \int d\boldsymbol{x} \|\boldsymbol{v}_\tau^*(\boldsymbol{x})\|^2 P_\tau(\boldsymbol{x}), \tag{84}
$$

where $\boldsymbol{v}_\tau^*(\boldsymbol{x})$ is the optimal mean local velocity that satisfies $\boldsymbol{v}_\tau^*(\boldsymbol{x}) = \nabla\phi_\tau(\boldsymbol{x})$ and $\partial_\tau P_\tau(\boldsymbol{x}) = -\nabla \cdot (\boldsymbol{v}_\tau^*(\boldsymbol{x}) P_\tau(\boldsymbol{x}))$. Here, the condition of Eq. (78) is not needed in the definition of the excess entropy production rate because Eq. (84) does not include the time evolution of $\phi_\tau(\boldsymbol{x})$. By combining the Fokker–Planck equation $\partial_\tau P_\tau(\boldsymbol{x}) =$

$-\nabla \cdot (\boldsymbol{v}_\tau(\boldsymbol{x}) P_\tau(\boldsymbol{x}))$ with the condition $\partial_\tau P_\tau(\boldsymbol{x}) = -\nabla \cdot (\boldsymbol{v}_\tau^*(\boldsymbol{x}) P_\tau(\boldsymbol{x}))$, we obtain Eq. (81). We can calculate the quantity $\int d\boldsymbol{x} \boldsymbol{v}_\tau^*(\boldsymbol{x}) \cdot (\boldsymbol{v}_\tau(\boldsymbol{x}) - \boldsymbol{v}_\tau^*(\boldsymbol{x})) P_\tau(\boldsymbol{x})$ as follows.

$$
\begin{aligned}
\int d\boldsymbol{x} \boldsymbol{v}_\tau^*(\boldsymbol{x}) \cdot (\boldsymbol{v}_\tau(\boldsymbol{x}) - \boldsymbol{v}_\tau^*(\boldsymbol{x})) P_\tau(\boldsymbol{x}) &= \int d\boldsymbol{x} \nabla \phi_\tau(\boldsymbol{x}) \cdot (\boldsymbol{v}_\tau(\boldsymbol{x}) - \boldsymbol{v}_\tau^*(\boldsymbol{x})) P_\tau(\boldsymbol{x}) \\
&= -\int d\boldsymbol{x} \phi_\tau(\boldsymbol{x}) \nabla \cdot [(\boldsymbol{v}_\tau(\boldsymbol{x}) - \boldsymbol{v}_\tau^*(\boldsymbol{x})) P_\tau(\boldsymbol{x})] \\
&= 0,
\end{aligned}
\tag{85}
$$

where we used $\nabla \cdot ((\boldsymbol{v}_\tau(\boldsymbol{x}) - \boldsymbol{v}_\tau^*(\boldsymbol{x})) P_\tau(\boldsymbol{x})) = 0$ in Eq. (81), and $\int d\boldsymbol{x} \nabla \cdot [\phi_\tau(\boldsymbol{x})(\boldsymbol{v}_\tau(\boldsymbol{x}) - \boldsymbol{v}_\tau^*(\boldsymbol{x})) P_\tau(\boldsymbol{x})] = 0$ because of the assumption that $P_\tau(\boldsymbol{x})$ decays sufficiently rapidly at infinity. Thus, the housekeeping entropy production rate is calculated as

$$
\begin{aligned}
\sigma_\tau^{\mathrm{hk}} &= \frac{1}{\mu T} \int d\boldsymbol{x} \|\boldsymbol{v}_\tau(\boldsymbol{x})\|^2 P_\tau(\boldsymbol{x}) - \frac{1}{\mu T} \int d\boldsymbol{x} \|\boldsymbol{v}_\tau^*(\boldsymbol{x})\|^2 P_\tau(\boldsymbol{x}) \\
&\quad - \frac{2}{\mu T} \int d\boldsymbol{x} \boldsymbol{v}_\tau^*(\boldsymbol{x}) \cdot (\boldsymbol{v}_\tau(\boldsymbol{x}) - \boldsymbol{v}_\tau^*(\boldsymbol{x})) P_\tau(\boldsymbol{x}) \\
&= \frac{1}{\mu T} \int d\boldsymbol{x} \|\boldsymbol{v}_\tau(\boldsymbol{x}) - \boldsymbol{v}_\tau^*(\boldsymbol{x})\|^2 P_\tau(\boldsymbol{x}).
\end{aligned}
\tag{86}
$$

$\square$

Remark 22 If we introduce the inner product $\langle \boldsymbol{a}, \boldsymbol{b} \rangle_{P_\tau/(\mu T)} = \int d\boldsymbol{x} [\boldsymbol{a}(\boldsymbol{x}) \cdot \boldsymbol{b}(\boldsymbol{x})] P_\tau(\boldsymbol{x})/(\mu T)$ for $\boldsymbol{a}(\boldsymbol{x}) \in \mathbb{R}^d$ and $\boldsymbol{b}(\boldsymbol{x}) \in \mathbb{R}^d$, the entropy production rate, the excess entropy production rate, and the housekeeping entropy production rate are given by $\sigma_\tau = \langle \boldsymbol{v}_\tau, \boldsymbol{v}_\tau \rangle_{P_\tau/(\mu T)}$, $\sigma_\tau^{\mathrm{ex}} = \langle \boldsymbol{v}_\tau^*, \boldsymbol{v}_\tau^* \rangle_{P_\tau/(\mu T)}$, and $\sigma_\tau^{\mathrm{hk}} = \langle \boldsymbol{v}_\tau - \boldsymbol{v}_\tau^*, \boldsymbol{v}_\tau - \boldsymbol{v}_\tau^* \rangle_{P_\tau/(\mu T)}$, respectively. Thus, the decomposition $\sigma_\tau = \sigma_\tau^{\mathrm{ex}} + \sigma_\tau^{\mathrm{hk}}$ can be regarded as the Pythagorean theorem

$$
\langle \boldsymbol{v}_\tau, \boldsymbol{v}_\tau \rangle_{P_\tau/(\mu T)} = \langle \boldsymbol{v}_\tau^*, \boldsymbol{v}_\tau^* \rangle_{P_\tau/(\mu T)} + \langle \boldsymbol{v}_\tau - \boldsymbol{v}_\tau^*, \boldsymbol{v}_\tau - \boldsymbol{v}_\tau^* \rangle_{P_\tau/(\mu T)},
\tag{87}
$$

where $\boldsymbol{v}_\tau(\boldsymbol{x})$ is orthogonal to $\boldsymbol{v}_\tau(\boldsymbol{x}) - \boldsymbol{v}_\tau^*(\boldsymbol{x})$ because of $\langle \boldsymbol{v}_\tau^*, \boldsymbol{v}_\tau - \boldsymbol{v}_\tau^* \rangle_{P_\tau/(\mu T)} = 0$. Because the orthogonality is based on $\boldsymbol{v}_\tau^*(\boldsymbol{x}) = \nabla \phi_\tau(\boldsymbol{x})$ and $\nabla \cdot ((\boldsymbol{v}_\tau(\boldsymbol{x}) - \boldsymbol{v}_\tau^*(\boldsymbol{x})) P_\tau(\boldsymbol{x})) = 0$, this decomposition is related to the Helmholtz–Hodge decomposition, and the mean local velocity is given by $\boldsymbol{v}_t(\boldsymbol{x}) = \nabla \phi_\tau(\boldsymbol{x}) + \boldsymbol{v}_\tau(\boldsymbol{x})$ for $\boldsymbol{v}_\tau(\boldsymbol{x}) \in \mathbb{R}^d$ that satisfies $\nabla \cdot (\boldsymbol{v}_\tau(\boldsymbol{x}) P_\tau(\boldsymbol{x})) = 0$. The decomposition of the entropy production rate based on $\nabla \cdot (\boldsymbol{v}_\tau(\boldsymbol{x}) P_\tau(\boldsymbol{x})) = 0$ was discussed in Ref. [66] without considering optimal transport theory.

Remark 23 Let us consider the case that a force $\boldsymbol{F}_\tau(\boldsymbol{x})$ is a potential force $\boldsymbol{F}_\tau(\boldsymbol{x}) = -\nabla U_\tau(\boldsymbol{x})$ where $U(\boldsymbol{x}) \in \mathbb{R}$ is a potential. In this case, the local mean velocity is given by $\boldsymbol{v}_\tau(\boldsymbol{x}) = \nabla(-\mu U_\tau(\boldsymbol{x}) - \mu T \ln P_\tau(\boldsymbol{x}))$ and $\phi_\tau(\boldsymbol{x})$ can be $\phi_\tau(\boldsymbol{x}) = -\mu U_\tau(\boldsymbol{x}) - \mu T \ln P_\tau(\boldsymbol{x})$. Thus, we obtain $\boldsymbol{v}_\tau(\boldsymbol{x}) = \boldsymbol{v}_\tau^*(\boldsymbol{x})$, $\sigma_\tau^{\mathrm{ex}} = \sigma_\tau$ and $\sigma_\tau^{\mathrm{hk}} = 0$ for a potential force. This fact implies that the excess entropy production rate and the housekeeping entropy production rate quantify contributions of a potential force and a non-potential force to the entropy production rate, respectively.

Based on the expression in Theorem 12, we also obtain a thermodynamic uncertainty relation for the excess entropy production rate [39, 40], which was substaintially obtained in Refs. [24, 67] for the entropy production rate.

Theorem 13 *Let $r(\boldsymbol{x}) \in \mathbb{R}$ be any time-independent function of $\boldsymbol{x} \in \mathbb{R}^d$. We assume that $P_\tau(\boldsymbol{x})$ decays sufficiently rapidly at infinity. The entropy production rate is bounded by*

$$\sigma_\tau \geq \sigma_\tau^{\mathrm{ex}} \geq \frac{[\partial_\tau \mathbb{E}_{P_\tau}(r)]^2}{\mu T \int d\boldsymbol{x} \|\nabla r(\boldsymbol{x})\|^2 P_\tau(\boldsymbol{x})}, \tag{88}$$

where $\mathbb{E}_{P_\tau}(r)$ is the expected value defined as

$$\mathbb{E}_{P_\tau}(r) = \int d\boldsymbol{x} \, P_\tau(\boldsymbol{x}) r(\boldsymbol{x}). \tag{89}$$

Proof The quantity $\partial_\tau \mathbb{E}_{P_\tau}(r)$ is calculated as

$$\begin{aligned}
\partial_\tau \mathbb{E}_{P_\tau}(r) &= \int d\boldsymbol{x} \, \partial_\tau P_\tau(\boldsymbol{x}) r(\boldsymbol{x}) \\
&= -\int d\boldsymbol{x} \, \nabla \cdot (\boldsymbol{v}_\tau^*(\boldsymbol{x}) P_\tau(\boldsymbol{x})) r(\boldsymbol{x}) \\
&= \int d\boldsymbol{x} (\boldsymbol{v}_\tau^*(\boldsymbol{x}) P_\tau(\boldsymbol{x})) \cdot \nabla r(\boldsymbol{x}),
\end{aligned} \tag{90}$$

where we used $\int d\boldsymbol{x} \nabla \cdot (\boldsymbol{v}_\tau^*(\boldsymbol{x}) P_\tau(\boldsymbol{x}) r(\boldsymbol{x})) = 0$ because of the assumption that $P_\tau(\boldsymbol{x})$ decays sufficiently rapidly at infinity. From the Cauchy–Schwartz inequality, we obtain

$$\begin{aligned}
[\partial_\tau \mathbb{E}_{P_\tau}(r)]^2 &= \left[\int d\boldsymbol{x} (\boldsymbol{v}_\tau^*(\boldsymbol{x}) P_\tau(\boldsymbol{x})) \cdot \nabla r(\boldsymbol{x}) \right]^2 \\
&\leq \left[\int d\boldsymbol{x} \|\boldsymbol{v}_\tau^*(\boldsymbol{x})\|^2 P_\tau(\boldsymbol{x}) \right] \left[\int d\boldsymbol{x} \|\nabla r(\boldsymbol{x})\|^2 P_\tau(\boldsymbol{x}) \right] \\
&= \mu T \sigma_\tau^{\mathrm{ex}} \left[\int d\boldsymbol{x} \|\nabla r(\boldsymbol{x})\|^2 P_\tau(\boldsymbol{x}) \right].
\end{aligned} \tag{91}$$

By combining Eq. (91) with $\sigma_\tau \geq \sigma_\tau^{\mathrm{ex}}$, we obtain Eq. (88).   □

Remark 24 The weaker inequality $\sigma_\tau \geq [\partial_\tau \mathbb{E}_{P_\tau}(r)]^2 / [\mu T \int d\boldsymbol{x} \|\nabla r(\boldsymbol{x})\|^2 P_\tau(\boldsymbol{x})]$ can be regarded as the thermodynamic uncertainty relation Eq. (54) for $\boldsymbol{w}(\boldsymbol{x}) = \nabla r(\boldsymbol{x})$ where the generalized current is calculated as $\mathcal{J}[\nabla r(\boldsymbol{x})] = \partial_\tau \mathbb{E}_{P_\tau}(r)$. Thus, this result is also regarded as a consequence of the Cramér–Rao bound for the path probability density. In the context of optimal transport theory, a mathematically equivalent inequality, namely the Wasserstein–Cramér–Rao bound, was also proposed in Ref. [68].

5 Thermodynamic links between information geometry and optimal transport

5.1 Gradient flow and information geometry in space of probability densities

In terms of the excess entropy production rate, we can obtain a thermodynamic link between information geometry in the space of probability densities and optimal transport theory. To discuss this thermodynamic link, we start with the definition of the pseudo energy $U_t^*(x)$ and the pseudo canonical distribution $P_t^{\mathrm{pcan}}(x)$ proposed in Ref. [40].

Definition 8 Let $\phi_t(x) \in \mathbb{R}^d$ be a potential which provides the optimal mean local velocity $v_t^*(x) = \nabla \phi_t(x) \in \mathbb{R}^d$ for an infinitesimal time such that $\lim_{\Delta t \to 0}[\mathcal{W}_2(P_t, P_{t+\Delta t})]^2/(\Delta t)^2 = \int dx \|v_t^*(x)\|^2 P_t(x)$. The pseudo energy $U_t^*(x) \in \mathbb{R}$ is defined as

$$\phi_t(x) = -\mu U_t^*(x) - \mu T \ln P_t(x), \tag{92}$$

and the pseudo canonical distribution $P_t^{\mathrm{pcan}}(x)$ is defined as

$$P_t^{\mathrm{pcan}}(x) = \frac{\exp\left[-\frac{U_t^*(x)}{T}\right]}{\int dx \exp\left[-\frac{U_t^*(x)}{T}\right]}, \tag{93}$$

which is a probability density that satisfies $P_t^{\mathrm{pcan}}(x) \geq 0$ and $\int dx\, P_t^{\mathrm{pcan}}(x) = 1$.

Remark 25 The pseudo energy can be defined for a non-potential force $F_t(x)$ that satisfies $\nabla \cdot ((F_t(x) + \nabla U_t^*(x))P_t(x)) = 0$. Thus, the pseudo energy $U_t^*(x)$ is not generally unique. If a force $F_t(x)$ is given by a potential force $F_t(x) = -\nabla U_t(x)$, a potential energy can trivially be a pseudo energy $U_t^*(x) = U_t(x)$.

The time evolution of $P_t(x)$ is given by a gradient flow expression. The concept of the gradient flow is originated from the Jordan–Kinderlehrer–Otto scheme [62]. We rewrite the gradient flow expression in Ref. [69] by using a functional derivative of the Kullback–Leibler divergence for general Markov jump processes [43]. The following proposition is a special case of a gradient flow expression [43] for the Fokker–Planck equation.

Proposition 14 *The time evolution of $P_t(x)$ under the Fokker–Planck equation is described by a gradient flow expression,*

$$\partial_t P_t(x) = \mathsf{D}[\partial_{P_t(x)} D_{\mathrm{KL}}(P_t \| P_t^{\mathrm{pcan}})], \tag{94}$$

where $D_{\mathrm{KL}}(P_t \| P_t^{\mathrm{pcan}})$ is the Kullback–Leibler divergence defined as

$$D_{\mathrm{KL}}(P_t \| P_t^{\mathrm{pcan}}) = \int dx \left[P_t(x) \ln \frac{P_t(x)}{P_t^{\mathrm{pcan}}(x)} - P_t(x) + P_t^{\mathrm{pcan}}(x) \right], \tag{95}$$

and $\mathsf{D}[\cdot]$ *stands for the weighted Laplacian operator defined as*

$$\mathsf{D}[\cdot] = \nabla \cdot (\mu T P_t(\boldsymbol{x}) \nabla[\cdot]). \tag{96}$$

Proof The functional derivative $\partial_{P_t(\boldsymbol{x})} D_{\mathrm{KL}}(P_t \| P_t^{\mathrm{pcan}})$ is calculated as

$$
\begin{aligned}
\partial_{P_t(\boldsymbol{x})} D_{\mathrm{KL}}(P_t \| P_t^{\mathrm{pcan}}) &= \partial_{P_t(\boldsymbol{x})} \int d\boldsymbol{x} \left[P_t(\boldsymbol{x}) \ln \frac{P_t(\boldsymbol{x})}{P_t^{\mathrm{pcan}}(\boldsymbol{x})} - P_t(\boldsymbol{x}) + P_t^{\mathrm{pcan}}(\boldsymbol{x}) \right] \\
&= \ln \frac{P_t(\boldsymbol{x})}{P_t^{\mathrm{pcan}}(\boldsymbol{x})},
\end{aligned}
\tag{97}
$$

and its gradient is calculated as

$$
\begin{aligned}
\nabla(\partial_{P_t(\boldsymbol{x})} D_{\mathrm{KL}}(P_t \| P_t^{\mathrm{pcan}})) &= \nabla \ln \frac{P_t(\boldsymbol{x})}{P_t^{\mathrm{pcan}}(\boldsymbol{x})} \\
&= \frac{1}{\mu T} \nabla(\mu U_t^*(\boldsymbol{x}) + \mu T \ln P_t(\boldsymbol{x})) \\
&= -\frac{1}{\mu T} \nabla \phi_t(\boldsymbol{x}),
\end{aligned}
\tag{98}
$$

where we used $\nabla \left[\int d\boldsymbol{x} \exp\left[-U_t^*(\boldsymbol{x})/T \right] \right] = 0$. Thus, the optimal mean local velocity provides Eq. (94) as follows,

$$
\begin{aligned}
\partial_t P_t(\boldsymbol{x}) &= -\nabla \cdot ((\nabla \phi_t(\boldsymbol{x})) P_t(\boldsymbol{x})) = \nabla \cdot (\mu T P_t(\boldsymbol{x}) \nabla(\partial_{P_t(\boldsymbol{x})} D_{\mathrm{KL}}(P_t \| P_t^{\mathrm{pcan}}))) \\
&= \mathsf{D}[\partial_{P_t(\boldsymbol{x})} D_{\mathrm{KL}}(P_t \| P_t^{\mathrm{pcan}})],
\end{aligned}
\tag{99}
$$

$\square$

Remark 26 The Kullback–Leibler divergence is calculated as $D_{\mathrm{KL}}(P_t \| P_t^{\mathrm{pcan}}) = \int d\boldsymbol{x} P_t(\boldsymbol{x}) \ln[P_t(\boldsymbol{x})/P_t^{\mathrm{pcan}}(\boldsymbol{x})]$ by using the normalization of the probability $\int d\boldsymbol{x} P_t(\boldsymbol{x}) = \int d\boldsymbol{x} P_t^{\mathrm{pcan}}(\boldsymbol{x}) = 1$. We take care that the functional derivative for $\int d\boldsymbol{x} P_t(\boldsymbol{x}) \ln[P_t(\boldsymbol{x})/P_t^{\mathrm{pcan}}(\boldsymbol{x})]$ is different from the functional derivative for Eq. (95), and the functional derivative $\partial_{P_t(\boldsymbol{x})} D_{\mathrm{KL}}(P_t \| P_t^{\mathrm{pcan}}) = \ln[P_t(\boldsymbol{x})/P_t^{\mathrm{pcan}}(\boldsymbol{x})]$ is defined for Eq. (95).

Remark 27 If a force is given by a potential force $\boldsymbol{F}_t(\boldsymbol{x}) = -\nabla U(\boldsymbol{x})$ with a time-independent potential energy $U(\boldsymbol{x}) \in \mathbb{R}$, a pseudo energy becomes a potential energy $U_t^*(\boldsymbol{x}) = U(\boldsymbol{x})$ and a pseudo canonical distribution $P_t^{\mathrm{pcan}}(\boldsymbol{x})$ becomes an equilibrium distribution $P_t^{\mathrm{pcan}}(\boldsymbol{x}) = P^{\mathrm{eq}}(\boldsymbol{x})$, which satisfies the condition that $P_t(\boldsymbol{x}) \to P^{\mathrm{eq}}(\boldsymbol{x})$ in the limit $t \to \infty$. In this case, the gradient flow expression Eq. (94) is given by $\partial_t P_t(\boldsymbol{x}) = \mathsf{D}[\partial_{P_t(\boldsymbol{x})} D_{\mathrm{KL}}(P_t \| P^{\mathrm{eq}})]$ which describes a relaxation to an equilibrium distribution $P^{\mathrm{eq}}(\boldsymbol{x})$.

Remark 28 If a pseudo canonical distribution is given by a time-independent distribution $P_t^{\mathrm{pcan}}(\boldsymbol{x}) = P^{\mathrm{st}}(\boldsymbol{x})$, the Fokker–Planck equation can be rewritten as the heat

equation,

$$\partial_t [\partial_{P_t(\boldsymbol{x})} D_{\mathrm{KL}}(P_t \| P^{\mathrm{st}})] = \mu T \operatorname{div}_{P_t} \left(\nabla [\partial_{P_t(\boldsymbol{x})} D_{\mathrm{KL}}(P_t \| P^{\mathrm{st}})] \right), \tag{100}$$

or equivalently,

$$\partial_t \ln \frac{P_t(\boldsymbol{x})}{P^{\mathrm{st}}(\boldsymbol{x})} = \mu T \operatorname{div}_{P_t} \left(\nabla \ln \frac{P_t(\boldsymbol{x})}{P^{\mathrm{st}}(\boldsymbol{x})} \right), \tag{101}$$

where div_{P_t} is the operator defined as

$$\operatorname{div}_{\sqrt{|g|}}(\cdots) = (\sqrt{|g|})^{-1} \nabla \cdot [\sqrt{|g|}(\cdots)], \tag{102}$$

with $\sqrt{|g|} = P_t(\boldsymbol{x})$. Because the operator $\operatorname{div}_{\sqrt{|g|}}$ is a generalization of the divergence operator for non-Euclidean space with the absolute value of the determinant of the metric tensor $|g|$, the Fokker–Planck equation may be regarded as a kind of the diffusion equation for $\partial_{P_t(\boldsymbol{x})} D_{\mathrm{KL}}(P_t \| P^{\mathrm{st}}) = \ln P_t(\boldsymbol{x}) - \ln P^{\mathrm{st}}(\boldsymbol{x})$. As a consequence of the diffusion process, we may obtain $\partial_{P_t(\boldsymbol{x})} D_{\mathrm{KL}}(P_t \| P^{\mathrm{st}}) = \ln P_t(\boldsymbol{x}) - \ln P^{\mathrm{st}}(\boldsymbol{x}) \to 0$ in the limit $t \to \infty$, which implies the relaxation to a steady state $P_t(\boldsymbol{x}) \to P^{\mathrm{st}}(\boldsymbol{x})$.

Based on the gradient flow expression (94), we obtain the following expression of the excess entropy production rate discussed in Ref. [40].

Theorem 15 *We assume that $P_\tau(\boldsymbol{x})$ decays sufficiently rapidly at infinity. The excess entropy production rate is given by*

$$\sigma_\tau^{\mathrm{ex}} = -\partial_\tau D_{\mathrm{KL}}(P_\tau \| P_s^{\mathrm{pcan}})|_{s=\tau}. \tag{103}$$

Proof The excess entropy production rate is calculated as

$$\begin{aligned}
\sigma_\tau^{\mathrm{ex}} &= \frac{1}{\mu T} \int d\boldsymbol{x} \| \nabla \phi_\tau(\boldsymbol{x}) \|^2 P_\tau(\boldsymbol{x}) \\
&= -\int d\boldsymbol{x} \nabla (\partial_{P_\tau(\boldsymbol{x})} D_{\mathrm{KL}}(P_\tau \| P_\tau^{\mathrm{pcan}})) \cdot (\nabla \phi_\tau(\boldsymbol{x}) P_\tau(\boldsymbol{x})) \\
&= \int d\boldsymbol{x} (\partial_{P_\tau(\boldsymbol{x})} D_{\mathrm{KL}}(P_\tau \| P_\tau^{\mathrm{pcan}})) \nabla \cdot (\nabla \phi_\tau(\boldsymbol{x}) P_\tau(\boldsymbol{x})) \\
&= -\int d\boldsymbol{x} (\partial_\tau P_\tau(\boldsymbol{x}) (\partial_{P_\tau(\boldsymbol{x})} D_{\mathrm{KL}}(P_\tau \| P_\tau^{\mathrm{pcan}})) \\
&= -\partial_\tau D_{\mathrm{KL}}(P_\tau \| P_s^{\mathrm{pcan}})|_{s=\tau},
\end{aligned} \tag{104}$$

where we used $\int d\boldsymbol{x} \nabla \cdot (\partial_{P_\tau(\boldsymbol{x})} D_{\mathrm{KL}}(P_\tau \| P_\tau^{\mathrm{pcan}})(\nabla \phi_\tau(\boldsymbol{x})) P_\tau(\boldsymbol{x})) = 0$ because of the assumption that $P_\tau(\boldsymbol{x})$ decays sufficiently rapidly at infinity. $\qquad \square$

Remark 29 If the pseudo distribution does not depend on time $P_\tau^{\mathrm{pcan}}(\boldsymbol{x}) = P^{\mathrm{st}}(\boldsymbol{x})$, the non-negativity of the excess entropy production rate is related to the monotonicity of the Kullback–Leibler divergence $\partial_\tau D_{\mathrm{KL}}(P_\tau \| P^{\mathrm{st}}) \le 0$, where $P^{\mathrm{st}}(\boldsymbol{x})$ is a steady-state

distribution that satisfies $\partial_\tau P^{\mathrm{st}}(\boldsymbol{x}) = 0$. Because $\sigma_\tau^{\mathrm{ex}} = 0$ if and only if the system is in a steady-state $P_\tau(\boldsymbol{x}) = P^{\mathrm{st}}(\boldsymbol{x})$, the excess entropy production rate can be a Lyapunov function and this monotonicity $\partial_\tau D_{\mathrm{KL}}(P_\tau \| P^{\mathrm{st}}) \leq 0$ gives the relaxation to a steady state distribution $P_\tau(\boldsymbol{x}) \to P^{\mathrm{st}}(\boldsymbol{x})$ in the limit $\tau \to \infty$.

Remark 30 This expression of the excess entropy production rate in terms of the Kullback–Leibler divergence provides a link between optimal transport theory and information geometry. As discussed in Ref. [41], the excess entropy production can be expressed using the dual coordinate systems for the Kullback–Leibler divergence. By using the dual coordinate systems with an affine transformation, the Kullback–Leibler divergence between P and Q is given by $D_{\mathrm{KL}}(P\|Q) = \varphi(\eta_P(\boldsymbol{x})) + \psi(\theta_Q(\boldsymbol{x})) - \int d\boldsymbol{x}\, \eta_P(\boldsymbol{x})\theta_Q(\boldsymbol{x})$, where $\eta_P(\boldsymbol{x}) = P(\boldsymbol{x}) - P^{\mathrm{st}}(\boldsymbol{x})$ and $\theta_Q(\boldsymbol{x}) = \ln Q(\boldsymbol{x}) - \ln P^{\mathrm{st}}(\boldsymbol{x})$ are the eta and theta coordinate systems that satisfy $\eta_{P^{\mathrm{st}}}(\boldsymbol{x}) = \theta_{P^{\mathrm{st}}}(\boldsymbol{x}) = 0$, and $\varphi(\eta_P(\boldsymbol{x})) = D_{\mathrm{KL}}(P\|P^{\mathrm{st}})$ and $\psi(\theta_Q(\boldsymbol{x})) = D_{\mathrm{KL}}(P^{\mathrm{st}}\|Q)$ are the dual convex functions, respectively. Thus, if the pseudo distribution does not depend on time $P_\tau^{\mathrm{pcan}}(\boldsymbol{x}) = P^{\mathrm{st}}(\boldsymbol{x})$, the excess entropy production is given by $\sigma_\tau^{\mathrm{ex}} = -\partial_\tau \varphi(\eta_{P_\tau}(\boldsymbol{x}))$.

Remark 31 The relaxation to an equilibrium distribution $P^{\mathrm{eq}}(\boldsymbol{x})$ for a time-independent potential force $\boldsymbol{F}_\tau(\boldsymbol{x}) = -\nabla U(\boldsymbol{x})$ was discussed from the viewpoint of information geometry based on the expression of the entropy production rate $\sigma_\tau^{\mathrm{ex}} = \sigma_\tau = -\partial_\tau D_{\mathrm{KL}}(P_\tau \| P^{\mathrm{eq}})$ in Refs. [30, 70, 71].

Near steady state, the Fisher metric for a probability density is also related to the entropy production rate. A thermodynamic interpretation of the Fisher metric was discussed in Ref. [11] as a generalization of the Weinhold geometry [7] or the Ruppeiner geometry [8, 10] in a stochastic system near equilibrium. We also examined this Fisher metric for a far-from-equilibrium system in Refs. [22, 24, 29]. To discuss a thermodynamic interpretation of the Fisher metric, we start with the definition of the Fisher information of time for a probability density.

Definition 9 Let $P_t(\boldsymbol{x})$ be a probability density of $\boldsymbol{x} \in \mathbb{R}^d$. *The Fisher information of time* is defined as

$$\frac{ds^2}{dt^2} = \int d\boldsymbol{x}\, P_t(\boldsymbol{x})(\partial_t \ln P_t(\boldsymbol{x}))^2 = 4\int d\boldsymbol{x}\,(\partial_t \sqrt{P_t(\boldsymbol{x})})^2. \tag{105}$$

The positive square root $v_{\mathrm{info}}(t) = \sqrt{ds^2/dt^2}$ is called *the intrinsic speed*.

Remark 32 The Fisher information of time is given by the Taylor expansion of the Kullback–Leibler divergence as follows.

$$2\frac{D_{\mathrm{KL}}(P_t \| P_{t+dt})}{dt^2} = \frac{ds^2}{dt^2} + O(dt). \tag{106}$$

If we consider a time-dependent parameter $\theta \in \mathbb{R}$, the Fisher information of time is given by

$$\frac{ds^2}{dt^2} = g_{\theta\theta}(P_t)\left(\frac{d\theta}{dt}\right)^2, \tag{107}$$

where $g_{\theta\theta}(P_t)$ is the Fisher metric defined as $g_{\theta\theta}(P_t) = \int d\mathbf{x}\, P_t(\mathbf{x})(\partial_\theta \ln P_t(\mathbf{x}))^2$. Thus, the intrinsic speed $v_{\text{info}}(t) = \sqrt{ds^2/dt^2}$ means the speed on the manifold of the probability simplex, where the metric is given by the Fisher metric. This differential geometry is well discussed in information geometry [12].

Remark 33 The relaxation to a steady state can be discussed in terms of the monotonicity of the intrinsic speed (or the monotonicity of the Fisher information of time) $\partial_t v_{\text{info}}(t) \leq 0$ (or $\partial_t (v_{\text{info}}(t))^2 \leq 0$), which is valid if a force $\mathbf{F}_t(\mathbf{x})$ is time-independent. When a force $\mathbf{F}_t(\mathbf{x})$ depends on time, the upper bound on $\partial_t (v_{\text{info}}(t))^2$ cannot be zero. The upper bound on $\partial_t (v_{\text{info}}(t))^2$ for the general case $\partial_t \mathbf{F}_t(\mathbf{x}) \neq \mathbf{0}$ was discussed in Ref. [24].

If a pseudo canonical distribution is given by a time-independent steady-state distribution $P_\tau^{\text{pcan}} = P^{\text{st}}$, the Fisher information of time is related to the excess entropy production rate. This was discussed in Ref. [29] for the entropy production rate with a general rate equation on chemical reaction networks. For the Fokker–Planck equation, we newly propose the following relation between the Fisher information of time and the excess entropy production rate as a correspondence of the result in Ref. [29].

Proposition 16 *We assume that a pseudo canonical distribution does not depend on time $P_\tau^{\text{pcan}}(\mathbf{x}) = P^{\text{st}}(\mathbf{x})$. We assume that $P_\tau(\mathbf{x})$ decays sufficiently rapidly at infinity. Let $\eta_{P_t}(\mathbf{x}) = P_t(\mathbf{x}) - P^{\text{st}}(\mathbf{x})$ be a difference from a steady-state distribution. The Fisher information of time is given by*

$$(v_{\text{info}}(t))^2 = -\frac{1}{2}\partial_t \sigma_t^{\text{ex}} + O(\eta_{P_t}^3), \tag{108}$$

where $O(\eta_{P_t}^3)$ stands for $O(\eta_{P_t}^3)/(\eta_{P_t}(\mathbf{x}))^2 \to 0$ in the limit $P_t \to P^{\text{st}}$.

Proof Let $\theta_{P_t}(\mathbf{x}) = \ln P_t(\mathbf{x}) - \ln P^{\text{st}}(\mathbf{x})$ be the theta coordinate that satisfies $\theta_{P^{\text{st}}}(\mathbf{x}) = 0$. The Fisher information of time is given by

$$(v_{\text{info}}(t))^2 = \int d\mathbf{x}[\partial_t \eta_{P_t}(\mathbf{x})][\partial_t \theta_{P_t}(\mathbf{x})]$$

$$= -\int d\mathbf{x}\nabla \cdot (-\mu T (\nabla \theta_{P_t}(\mathbf{x})) P_t(\mathbf{x}))[\partial_t \theta_{P_t}(\mathbf{x})]$$

$$= -\int d\mathbf{x}(\mu T P^{\text{st}}(\mathbf{x}))[(\nabla \theta_{P_t}(\mathbf{x})) \cdot \partial_t (\nabla \theta_{P_t}(\mathbf{x}))] + O(\eta_{P_t}^3), \tag{109}$$

where we used $\int d\mathbf{x}\nabla \cdot ((\nabla \theta_{P_t}(\mathbf{x})) P_t(\mathbf{x})[\partial_t \theta_{P_t}(\mathbf{x})]) = 0$ because of the assumption that $P_\tau(\mathbf{x})$ decays sufficiently rapidly at infinity. The excess entropy production rate is given by

$$\sigma_t^{\text{ex}} = -\int d\mathbf{x}[\partial_t \eta_{P_t}(\mathbf{x})]\theta_{P_t}(\mathbf{x})$$

$$= \int d\mathbf{x}\nabla \cdot (-\mu T \nabla(\theta_{P_t}(\mathbf{x})) P_t(\mathbf{x}))\theta_{P_t}(\mathbf{x})$$

$$= \int d\mathbf{x}(\mu T P^{\text{st}}(\mathbf{x}))\|\nabla \theta_{P_t}(\mathbf{x})\|^2 + O(\eta_{P_t}^3), \tag{110}$$

where we used $\int d\boldsymbol{x} \nabla \cdot ((\nabla \theta_{P_t}(\boldsymbol{x})) P_t(\boldsymbol{x}) \theta_{P_t}(\boldsymbol{x})) = 0$ because of the assumption that $P_\tau(\boldsymbol{x})$ decays sufficiently rapidly at infinity. Thus, we obtain

$$
\begin{aligned}
(v_{\text{info}}(t))^2 &= -\frac{1}{2} \int d\boldsymbol{x} (\mu T P^{\text{st}}(\boldsymbol{x})) \partial_t \|\nabla \theta_{P_t}(\boldsymbol{x})\|^2 + O(\eta_{P_t}^3) \\
&= -\frac{1}{2} \partial_t \sigma_t^{\text{ex}} + O(\eta_{P_t}^3).
\end{aligned}
\tag{111}
$$

$\square$

Remark 34 Because the excess entropy production rate is defined in terms of the L^2-Wasserstein distance $\sigma_t^{\text{ex}} = [\mathcal{W}_2(P_\tau, P_{\tau+dt})]^2/(\mu T dt^2)$ up to $O(dt)$, Eq. (108) implies a relation between the L^2-Wasserstein distance and the Fisher information of time near steady state $(v_{\text{info}}(\tau))^2 = -\partial_\tau [\mathcal{W}_2(P_\tau, P_{\tau+dt})]^2/(2\mu T dt^2)$ up to $O(\eta_{P_t}^3)$ and $O(dt)$.

Remark 35 We discussed a relation between the Fisher information and the excess entropy production rate proposed by Glansdorff and Prigogine near steady state in Ref. [38] from the viewpoint of the Glansdorff–Prigogine criterion for stability [5, 6, 72, 73]. We remark that the definition of the excess entropy production rate by Glansdorff and Prigogine is slightly different from the definition based on L^2-Wasserstein distance in this paper.

Remark 36 As discussed in Ref. [41], the expression of $\sigma_t^{\text{ex}} = -\int d\boldsymbol{x} [\partial_t \eta_{P_t}(\boldsymbol{x})] \theta_{P_t}(\boldsymbol{x})$ in Eq. (110) implies that the time derivative of the eta coordinate system $\partial_t \eta_{P_t}(\boldsymbol{x})$ corresponds to the thermodynamic flow and the theta coordinate $-\theta_{P_t}(\boldsymbol{x})$ corresponds to the conjugated thermodynamic force, respectively. The expression of the Fisher information of time in terms of the thermodynamic flow and the conjugated thermodynamic force $(v_{\text{info}}(t))^2 = \int d\boldsymbol{x} [\partial_t \eta_{P_t}(\boldsymbol{x})][\partial_t \theta_{P_t}(\boldsymbol{x})]$ has been substantially obtained in Ref. [22]. The gradient of the thermodynamic force $\nabla \theta_{P_t}(\boldsymbol{x})$ is also regarded as the thermodynamic force because the gradient is given by the linear combination of the thermodynamic force at position $\boldsymbol{x} + \Delta\boldsymbol{x}$ and position $\boldsymbol{x}$ for the infinitesimal distance $\Delta\boldsymbol{x}$. The quantity $\mu T P^{\text{st}}(\boldsymbol{x})$ is also regarded as the Onsager coefficient near equilibrium because Eq. (110) is the quadratic function of the thermodynamic force $\nabla \theta_{P_t}(\boldsymbol{x})$ with proportionality coefficient $\mu T P^{\text{st}}(\boldsymbol{x})$. The gradient flow expression of the Fokker–Planck equation Eq. (94) is given by the weighted Laplacian operator Eq. (96) where this weight is regarded as the Onsager coefficient near equilibrium. Based on the quadratic expression in Eq. (110), we also can consider a geometry where the weight of the Onsager coefficient is a metric. We used the weight of the generalized Onsager coefficient in Ref. [29] to define the excess entropy production rate for general Markov processes based on optimal transport theory, and discussed a geometric interpretation of the excess entropy production rate.

By using the Fisher information of time, the information-geometric speed limit discussed in Refs. [11, 22, 24, 29] can be obtained in parallel with the derivation of the thermodynamic speed limit Eq. (68). The information-geometric speed limit provides a lower bound on the quantity $\int_\tau^{\tau+\Delta\tau} dt (v_{\text{info}}(t))^2$. The quantity $\int_\tau^{\tau+\Delta\tau} dt (v_{\text{info}}(t))^2$ can

be regarded as the thermodynamic cost because this quantity is related to the change of the excess entropy production rate $\int_\tau^{\tau+\Delta\tau} dt (v_{\text{info}}(t))^2 = (\sigma_\tau^{\text{ex}} - \sigma_{\tau+\Delta\tau}^{\text{ex}})/2 + O(\eta_{P_t}^3)$ near steady state by using Eq. (108).

Theorem 17 *The quantity $\int_\tau^{\tau+\Delta\tau} dt (v_{\text{info}}(t))^2$ is bounded by*

$$\int_\tau^{\tau+\Delta\tau} dt (v_{\text{info}}(t))^2 \geq \frac{\left(\int_\tau^{\tau+\Delta t} dt v_{\text{info}}(t)\right)^2}{\Delta\tau} \geq \frac{[\mathcal{D}(P_\tau, P_{\tau+\Delta\tau})]^2}{\Delta\tau}, \quad (112)$$

where $\mathcal{D}(P_\tau, P_{\tau+\Delta\tau})$ is the twice of the Bhattacharyya angle ζ^{B} defined as

$$\mathcal{D}(P_\tau, P_{\tau+\Delta\tau}) = 2\arccos\left(\int dx \sqrt{P_\tau(x)P_{\tau+\Delta\tau}(x)}\right) = 2\zeta^{\text{B}}. \quad (113)$$

Proof From the Cauchy–Schwartz inequality, we obtain

$$\left(\int_\tau^{\tau+\Delta\tau} dt (v_{\text{info}}(t))^2\right)\left(\int_\tau^{\tau+\Delta\tau} dt\right) \geq \left(\int_\tau^{\tau+\Delta\tau} dt v_{\text{info}}(t)\right)^2. \quad (114)$$

Thus, the tighter lower bound is obtained as

$$\int_\tau^{\tau+\Delta\tau} dt (v_{\text{info}}(t))^2 \geq \frac{\left(\int_\tau^{\tau+\Delta\tau} dt v_{\text{info}}(t)\right)^2}{\Delta\tau}. \quad (115)$$

To solve the minimization of $\int_\tau^{\tau+\Delta t} dt (v_{\text{info}}(t))^2$ under the constraint $\int dx P_t(x) = 1$ with fixed P_τ and $P_{\tau+\Delta\tau}$, we consider the Euler–Lagrange equation

$$\partial_{\sqrt{P_t(x)}} \mathbb{L}'(\{\sqrt{P_t}\}, \{\partial_t \sqrt{P_t}\}, \phi) = \partial_t \left[\partial_{(\partial_t \sqrt{P_t(x)})} \mathbb{L}'(\{\sqrt{P_t}\}, \{\partial_t \sqrt{P_t}\}, \phi)\right], \quad (116)$$

for $\tau < t < \tau + \Delta\tau$ with the Lagrangian

$$\mathbb{L}'(\{\sqrt{P_t}\}, \{\partial_t \sqrt{P_t}\}, \phi)$$
$$= \int_\tau^{\tau+\Delta\tau} dt \left[4 \int dx (\partial_t \sqrt{P_t(x)})^2 - \phi\left[\int dx (\sqrt{P_t(x)})^2 - 1\right]\right]. \quad (117)$$

The Euler–Lagrange equation can be rewritten as

$$(\partial_t)^2 \sqrt{P_t(x)} = -\frac{\phi}{4}\sqrt{P_t(x)}, \quad (118)$$

which solution is generally given by $\sqrt{P_t(x)} = \alpha(x)\cos(\sqrt{\phi}/2(t - \beta(x)))$ for $\alpha(x) \in \mathbb{R}$ and $\beta(x) \in \mathbb{R}$. The constraint $\int dx P_t(x) = 1$ with fixed P_τ and $P_{\tau+\Delta\tau}$ for this

solution provides the optimal solution that minimizes $\int_\tau^{\tau+\Delta t} dt\,(v_{\text{info}}(t))^2$ under the constraint,

$$\sqrt{P_t^*(x)} = \frac{\sqrt{P_\tau(x)}\sin\left[\zeta^{\mathrm{B}}\left(1 - \frac{t-\tau}{\Delta\tau}\right)\right] + \sqrt{P_{\tau+\Delta\tau}(x)}\sin\left[\zeta^{\mathrm{B}}\frac{t-\tau}{\Delta\tau}\right]}{\sin\left[\zeta^{\mathrm{B}}\right]}, \quad (119)$$

where the normalization of the probability is satisfied for $\tau \leq t \leq \tau + \Delta\tau$,

$$\int dx(\sqrt{P_t^*(x)})^2$$
$$= \frac{\sin^2\left[\zeta^{\mathrm{B}}\left(1 - \frac{t-\tau}{\Delta\tau}\right)\right] + \sin^2\left[\zeta^{\mathrm{B}}\frac{t-\tau}{\Delta\tau}\right] + 2\cos\left[\zeta^{\mathrm{B}}\right]\sin\left[\zeta^{\mathrm{B}}\left(1 - \frac{t-\tau}{\Delta\tau}\right)\right]\sin\left[\zeta^{\mathrm{B}}\frac{t-\tau}{\Delta\tau}\right]}{\sin^2\left[\zeta^{\mathrm{B}}\right]}$$
$$= \frac{\sin^2\left[\zeta^{\mathrm{B}}\right]\left(\sin^2\left[\zeta^{\mathrm{B}}\frac{t-\tau}{\Delta\tau}\right] + \cos^2\left[\zeta^{\mathrm{B}}\frac{t-\tau}{\Delta\tau}\right]\right)}{\sin^2\left[\zeta^{\mathrm{B}}\right]} = 1. \quad (120)$$

Thus, the weaker lower bound is calculated as

$$\int_\tau^{\tau+\Delta\tau} dt\,(v_{\text{info}}(t))^2$$
$$\geq 4\int_\tau^{\tau+\Delta\tau} dt \int dx\,(\partial_t\sqrt{P_t^*(x)})^2$$
$$= \frac{[\mathcal{D}(P_\tau, P_{\tau+\Delta\tau})]^2}{\Delta\tau^2}\int_\tau^{\tau+\Delta\tau} dt\left[\frac{\cos^2\left[\zeta^{\mathrm{B}}\left(1 - \frac{t-\tau}{\Delta\tau}\right)\right] + \cos^2\left[\zeta^{\mathrm{B}}\frac{t-\tau}{\Delta\tau}\right]}{\sin^2\left[\zeta^{\mathrm{B}}\right]}\right.$$
$$\left. -2\frac{\cos\left[\zeta^{\mathrm{B}}\right]\cos\left[\zeta^{\mathrm{B}}\left(1 - \frac{t-\tau}{\Delta\tau}\right)\right]\cos\left[\zeta^{\mathrm{B}}\frac{t-\tau}{\Delta\tau}\right]}{\sin^2\left[\zeta^{\mathrm{B}}\right]}\right]$$
$$= \frac{[\mathcal{D}(P_\tau, P_{\tau+\Delta\tau})]^2}{(\Delta\tau)^2}\int_\tau^{\tau+\Delta\tau} dt = \frac{[\mathcal{D}(P_\tau, P_{\tau+\Delta\tau})]^2}{\Delta\tau}. \quad (121)$$

<div align="right">□</div>

Remark 37 $\mathcal{D}(P_\tau, P_{\tau+\Delta\tau})$ is regarded as the geodesic on the hyper-sphere surface of radius 2. An interpretation of the Bhattacharyya angle as the geodesic on the hyper-sphere surface is related to the fact that information geometry can be regarded as the geometry of a hyper-sphere surface of radius 2 because the square of the line element is obtained from the Fisher metric as $ds^2 = \int dx(2d\sqrt{P_t(x)})^2$ with the constraint $\int dx(\sqrt{P_t(x)})^2 = 1$. The Bhattacharyya angle ζ^{B} is given by the inner product for a unit vector on the hyper-sphere $\cos\zeta^{\mathrm{B}} = \int dx(\sqrt{P_\tau(x)}\sqrt{P_{\tau+\Delta\tau}(x)})$.

Remark 38 The quantity $\int_\tau^{\tau+\Delta\tau} dt\,v_{\text{info}}(t)$ is called the thermodynamic length proposed in Ref. [11] as a generalization of the result in Ref. [9]. The thermodynamic length is minimized as $\int_\tau^{\tau+\Delta\tau} dt\,v_{\text{info}}(t) \geq \mathcal{D}(P_\tau, P_{\tau+\Delta\tau})$ for the fixed initial distribution P_τ and the final distribution $P_{\tau+\Delta\tau}$. The minimization of the thermodynamic

length near equilibrium for large time interval $\Delta \tau$ is related to an optimal protocol to minimize the quadratic cost representing an observable fluctuation [11, 20, 74].

From the Cramèr–Rao bound, the intrinsic speed is also related to the speed of the observable. From the viewpoint of thermodynamics, this fact was discussed in Ref. [24] for a time-independent observable, and in Ref. [26] for a time-dependent observable.

Definition 10 Let $r(\boldsymbol{x}) \in \mathbb{R}$ be time-independent $\partial_t r(\boldsymbol{x}) = 0$. *The speed of the observable $v_r(t)$ is defined as*

$$v_r(t) = \sqrt{\frac{\left(\partial_t \mathbb{E}_{P_t}[r]\right)^2}{\mathrm{Var}_{P_t}[r]}} = \frac{|\,\partial_t \mathbb{E}_{P_t}[r]\,|}{\sqrt{\mathrm{Var}_{P_t}[r]}}, \tag{122}$$

where $\mathbb{E}_{P_t}[r] = \int d\boldsymbol{x}\, r(\boldsymbol{x}) P_t(\boldsymbol{x})$ and $\mathrm{Var}_{P_t}[r] = \mathbb{E}_{P_t}[(\Delta r)^2]$ with $\Delta r(\boldsymbol{x}) = r(\boldsymbol{x}) - \mathbb{E}_{P_t}[r]$.

Remark 39 The speed of the observable can be regarded as the degree of the expected value's change $|\,\partial_t \mathbb{E}_{P_t}[r]\,|$, which is normalized by its standard deviation $\sqrt{\mathrm{Var}_{P_t}[r]}$.

Lemma 18 *For any $r(\boldsymbol{x}) \in \mathbb{R}$, the speed of the observable $v_r(t)$ is generally bounded by the intrinsic speed $v_{\mathrm{info}}(t)$,*

$$v_{\mathrm{info}}(t) \geq v_r(t). \tag{123}$$

Proof The Fisher information of time $[v_{\mathrm{info}}(t)]^2$ is the Fisher metric for the parameter $\theta = t$. As discussed in Lemma 7, the Cramér–Rao bound for the parameter $\theta = t$ is given by

$$\begin{aligned}
[v_{\mathrm{info}}(t)]^2 &= \int d\boldsymbol{x}\, P_t(\boldsymbol{x})(\partial_t \ln P_t(\boldsymbol{x}))^2 \\
&\geq \frac{\left(\int d\boldsymbol{x}\, P_t(\boldsymbol{x})(\Delta r(\boldsymbol{x}))(\partial_t \ln P_t(\boldsymbol{x}))\right)^2}{\int d\boldsymbol{x}\, P_t(\boldsymbol{x})(\Delta r(\boldsymbol{x}))^2} = (v_r(t))^2, \tag{124}
\end{aligned}$$

where we used the Cauchy–Schwartz inequality and $\int d\boldsymbol{x}\, \partial_t P_t(\boldsymbol{x}) = 0$. By taking the square root of each side, we obtain Eq. (7). □

We newly propose that the intrinsic speed $v_{\mathrm{info}}(t)$ also provides an upper bound on the excess entropy production rate. This fact was substantially proposed in Refs. [29, 38].

Proposition 19 *The excess entropy production rate σ_t^{ex} is bounded as follows.*

$$v_{\mathrm{info}}(t)\sqrt{\mathrm{Var}_{P_t}[\theta_{P_t}]} \geq \sigma_t^{\mathrm{ex}}, \tag{125}$$

where $\theta_{P_t}(\boldsymbol{x})$ is the theta coordinate system defined as $\theta_{P_t}(\boldsymbol{x}) = \ln P_t(\boldsymbol{x}) - \ln P_t^{\mathrm{pcan}}(\boldsymbol{x})$.

Proof The excess entropy production is given by

$$\sigma_t^{\text{ex}} = -\int d\boldsymbol{x}\, P_t(\boldsymbol{x})[\partial_t \ln P_t(\boldsymbol{x})]\theta_{P_t}(\boldsymbol{x}) \tag{126}$$

$$= -\int d\boldsymbol{x}\, P_t(\boldsymbol{x})[\partial_t \ln P_t(\boldsymbol{x})][\theta_{P_t}(\boldsymbol{x}) - \mathbb{E}_{P_t}[\theta_{P_t}]], \tag{127}$$

where we used $\int d\boldsymbol{x}\,\partial_t P_t(\boldsymbol{x}) = 0$. From the Cauchy–Schwartz inequality, we obtain

$$(\sigma_t^{\text{ex}})^2 = \left(\int d\boldsymbol{x}\, P_t(\boldsymbol{x})[\partial_t \ln P_t(\boldsymbol{x})](\theta_{P_t}(\boldsymbol{x}) - \mathbb{E}_{P_t}[\theta_{P_t}]) \right)^2$$

$$\leq \left(\int d\boldsymbol{x}\, P_t(\boldsymbol{x})(\partial_t \ln P_t(\boldsymbol{x}))^2 \right) \left(\int d\boldsymbol{x}\, P_t(\boldsymbol{x})(\theta_{P_t}(\boldsymbol{x}) - \mathbb{E}_{P_t}[\theta_{P_t}])^2 \right)$$

$$= [v_{\text{info}}(t)]^2 \text{Var}_{P_t}[\theta_{P_t}]. \tag{128}$$

By taking the square root of each side, we obtain Eq. (125). □

Remark 40 Proposition 19 implies that the excess entropy production rate σ_t^{ex} is generally bounded by the intrinsic speed $v_{\text{info}}(t)$. From the bound (125), the excess entropy production rate is zero $\sigma_t^{\text{ex}} = 0$ if the intrinsic speed is zero $v_{\text{info}}(t) = 0$. This result is consistent with the fact that the excess entropy production rate is zero in a steady state.

5.2 Excess entropy production and information geometry in space of path probability densities

Here we newly propose that the excess entropy production rate, which is given by the L^2-Wasserstein distance in optimal transport theory, can also be obtained from the projection theorem in the space of path probability densities as analogous to the entropy production rate. This projection theorem for the excess entropy production rate was substantially obtained in Ref. [44] for the general Markov process. This result also gives another link between information geometry in the space of path probability densities and optimal transport theory.

We start with the expressions of the entropy production rate, the excess entropy production rate and the housekeeping entropy production rate by the Kullback–Leibler divergence between the path probability densities of interpolated dynamics $\mathbb{P}_{\boldsymbol{\nu}_\tau'}^\theta$ defined in Definition 4.

Proposition 20 *The entropy production rate σ_τ, the excess entropy production rate σ_τ^{ex} and the housekeeping entropy production rate σ_τ^{hk} for original Fokker–Planck dynamics (1) are given by*

$$\sigma_\tau = \lim_{dt \to 0} \frac{4D_{\text{KL}}(\mathbb{P}_{\boldsymbol{\nu}_\tau}^1 \| \mathbb{P}_{\boldsymbol{0}}^1)}{dt}, \tag{129}$$

$$\sigma_\tau^{\text{ex}} = \lim_{dt \to 0} \frac{4D_{\text{KL}}(\mathbb{P}_{v_\tau}^1 \| \mathbb{P}_{v_\tau - v_\tau^*}^1)}{dt} = \lim_{dt \to 0} \frac{4D_{\text{KL}}(\mathbb{P}_{v_\tau^*}^1 \| \mathbb{P}_0^1)}{dt}, \tag{130}$$

$$\sigma_\tau^{\text{hk}} = \lim_{dt \to 0} \frac{4D_{\text{KL}}(\mathbb{P}_{v_\tau}^1 \| \mathbb{P}_{v_\tau^*}^1)}{dt} = \lim_{dt \to 0} \frac{4D_{\text{KL}}(\mathbb{P}_{v_\tau - v_\tau^*}^1 \| \mathbb{P}_0^1)}{dt}, \tag{131}$$

where $v_\tau^*(x)$ is the optimal mean local velocity defined as $v_\tau^*(x) = \nabla \phi_\tau(x)$, and $v_\tau(x_\tau)$ is the mean local velocity $v_\tau(x) = \mu(F_\tau(x) - T \nabla \ln P_\tau(x))$ for the original Fokker–Planck equation $\partial_\tau P_\tau(x) = -\nabla \cdot (v_\tau(x) P_\tau(x))$.

Proof For any $v(x) \in \mathbb{R}^d$ and $v'(x) \in \mathbb{R}^d$, we obtain

$$\ln \frac{\mathbb{T}_{\tau;v}^1(x_{\tau+dt} \mid x_\tau)}{\mathbb{T}_{\tau;v'}^1(x_{\tau+dt} \mid x_\tau)} = -\frac{\|x_{\tau+dt} - x_\tau - \mu F_\tau(x_\tau)dt - v(x)dt - v_\tau(x_\tau)dt\|^2}{4\mu T dt}$$

$$+ \frac{\|x_{\tau+dt} - x_\tau - \mu F_\tau(x_\tau)dt - v'(x)dt - v_\tau(x_\tau)dt\|^2}{4\mu T dt}$$

$$= \frac{(x_{\tau+dt} - x_\tau - \mu F_\tau(x_\tau)dt - v_\tau(x_\tau)dt) \cdot (v(x_\tau) - v(x_\tau))}{2\mu T}$$

$$+ \frac{\|v'^2(x_\tau) - v^2(x_\tau)\| dt}{4\mu T}. \tag{132}$$

Thus, the Kullback–Leibler divergecne is calculated as

$$D_{\text{KL}}(\mathbb{P}_v^1 \| \mathbb{P}_{v'}^1) = \int dx_\tau P_\tau(x_\tau) \int dx_{\tau+dt} \mathbb{T}_{\tau;v}^1(x_{\tau+dt} \mid x_\tau) \ln \frac{\mathbb{T}_{\tau;v}^1(x_{\tau+dt} \mid x_\tau)}{\mathbb{T}_{\tau;v'}^1(x_{\tau+dt} \mid x_\tau)}$$

$$= dt \frac{1}{4\mu T} \int dx_\tau P_\tau(x_\tau) \|v(x_\tau) - v(x_\tau)\|^2. \tag{133}$$

Therefore, by plugging $(v, v') = (v_\tau, 0)$, $(v, v') = (v_\tau, v_\tau - v_\tau^*)$, $(v, v') = (v_\tau^*, 0)$, $(v, v') = (v_\tau, v_\tau^*)$, $(v, v') = (v_\tau - v_\tau^*, 0)$ into Eq. (133), we obtain

$$\lim_{dt \to 0} \frac{4D_{\text{KL}}(\mathbb{P}_{v_\tau}^1 \| \mathbb{P}_0^1)}{dt} = \frac{1}{\mu T} \int dx_\tau P_\tau(x_\tau) \|v_\tau(x_\tau)\|^2 = \sigma_\tau, \tag{134}$$

$$\lim_{dt \to 0} \frac{4D_{\text{KL}}(\mathbb{P}_{v_\tau}^1 \| \mathbb{P}_{v_\tau - v_\tau^*}^1)}{dt} = \lim_{dt \to 0} \frac{4D_{\text{KL}}(\mathbb{P}_{v_\tau^*}^1 \| \mathbb{P}_0^1)}{dt}$$

$$= \frac{1}{\mu T} \int dx_\tau P_\tau(x_\tau) \|v_\tau^*(x_\tau)\|^2 = \sigma_\tau^{\text{ex}}, \tag{135}$$

$$\lim_{dt \to 0} \frac{4D_{\text{KL}}(\mathbb{P}_{v_\tau}^1 \| \mathbb{P}_{v_\tau^*}^1)}{dt} = \lim_{dt \to 0} \frac{4D_{\text{KL}}(\mathbb{P}_{v_\tau - v_\tau^*}^1 \| \mathbb{P}_0^1)}{dt}$$

$$= \frac{1}{\mu T} \int dx_\tau P_\tau(x_\tau) \|v_\tau(x_\tau) - v_\tau^*(x_\tau)\|^2 = \sigma_\tau^{\text{hk}}, \tag{136}$$

$\square$

Remark 41 Proposition 20 implies that the origin of the decomposition $\sigma_\tau = \sigma_\tau^{\text{ex}} + \sigma_\tau^{\text{hk}}$ comes from the generalized Pythagorean theorems

$$D_{\text{KL}}(\mathbb{P}^1_{\boldsymbol{v}_\tau} \| \mathbb{P}^1_{\boldsymbol{0}}) = D_{\text{KL}}(\mathbb{P}^1_{\boldsymbol{v}_\tau} \| \mathbb{P}^1_{\boldsymbol{v}_\tau^*}) + D_{\text{KL}}(\mathbb{P}^1_{\boldsymbol{v}_\tau^*} \| \mathbb{P}^1_{\boldsymbol{0}}), \tag{137}$$

$$D_{\text{KL}}(\mathbb{P}^1_{\boldsymbol{v}_\tau} \| \mathbb{P}^1_{\boldsymbol{0}}) = D_{\text{KL}}(\mathbb{P}^1_{\boldsymbol{v}_\tau} \| \mathbb{P}^1_{\boldsymbol{v}_\tau - \boldsymbol{v}_\tau^*}) + D_{\text{KL}}(\mathbb{P}^1_{\boldsymbol{v}_\tau - \boldsymbol{v}_\tau^*} \| \mathbb{P}^1_{\boldsymbol{0}}), \tag{138}$$

that is consistent with the Pythagorean theorem in Remark 22.

Based on the projection theorem for the Pythagorean theorem in Remark 41, we obtain expressions of the excess entropy production rate and the housekeeping entropy production rate by the minimization problem of the Kullback–Leibler divergence.

Proposition 21 *We assume that $P_\tau(\boldsymbol{x}_\tau)$ decays sufficiently rapidly at infinity. The excess entropy production rate and the housekeeping entropy production rate are given by*

$$\sigma_\tau^{\text{ex}} = \lim_{dt \to 0} \inf_{\mathbb{Q} \in \mathcal{M}_{\text{ZD}}(\mathbb{P})} \frac{4 D_{\text{KL}}(\mathbb{P} \| \mathbb{Q})}{dt}, \tag{139}$$

$$\sigma_\tau^{\text{hk}} = \lim_{dt \to 0} \inf_{\mathbb{Q} \in \mathcal{M}_{\text{G}}(\mathbb{P})} \frac{4 D_{\text{KL}}(\mathbb{P} \| \mathbb{Q})}{dt}, \tag{140}$$

where $\mathcal{M}_{\text{ZD}}(\mathbb{P})$ is the zero-divergence manifold defined as

$$\mathcal{M}_{\text{ZD}}(\mathbb{P}) = \{\mathbb{P}^1_{\boldsymbol{v}} \mid \nabla \cdot (\boldsymbol{v}(\boldsymbol{x}_\tau) P_\tau(\boldsymbol{x}_\tau)) = 0\}, \tag{141}$$

and $\mathcal{M}_{\text{G}}(\mathbb{P})$ is the gradient manifold defined as

$$\mathcal{M}_{\text{G}}(\mathbb{P}) = \{\mathbb{P}^1_{\nabla r} \mid r(\boldsymbol{x}_\tau) \in \mathbb{R}\}. \tag{142}$$

Proof Let $\boldsymbol{v}_\tau^*(\boldsymbol{x}_\tau) = \nabla \phi_\tau(\boldsymbol{x}_\tau)$ be the optimal mean local velocity. For any $\mathbb{P}^1_{\boldsymbol{v}} \in \mathcal{M}_{\text{ZD}}(\mathbb{P})$, we obtain the generalized Pythagorean theorem,

$$
\begin{aligned}
&D_{\text{KL}}(\mathbb{P} \| \mathbb{P}^1_{\boldsymbol{v}}) \\
&= \frac{dt \left[\int d\boldsymbol{x}_\tau P_\tau(\boldsymbol{x}_\tau) \| \boldsymbol{v}_\tau(\boldsymbol{x}_\tau) - \boldsymbol{v}_\tau^*(\boldsymbol{x}_\tau) - \boldsymbol{v} + \nabla \phi_\tau(\boldsymbol{x}_\tau) \|^2 \right]}{4 \mu T} \\
&= \frac{dt \left[\int d\boldsymbol{x}_\tau P_\tau(\boldsymbol{x}_\tau) \| \boldsymbol{v}_\tau^*(\boldsymbol{x}_\tau) \|^2 + \int d\boldsymbol{x}_\tau P_\tau(\boldsymbol{x}_\tau) \| \boldsymbol{v}_\tau(\boldsymbol{x}_\tau) - \boldsymbol{v}_\tau^*(\boldsymbol{x}_\tau) - \boldsymbol{v} \|^2 \right]}{4 \mu T} \\
&= D_{\text{KL}}(\mathbb{P} \| \mathbb{P}^1_{\boldsymbol{v}_\tau - \boldsymbol{v}_\tau^*}) + D_{\text{KL}}(\mathbb{P}^1_{\boldsymbol{v}_\tau - \boldsymbol{v}_\tau^*} \| \mathbb{P}^1_{\boldsymbol{v}}), \tag{143}
\end{aligned}
$$

where we used Eq. (133) and

$$
\begin{aligned}
&\int d\boldsymbol{x}_\tau P_\tau(\boldsymbol{x}_\tau)(\boldsymbol{v}_\tau(\boldsymbol{x}_\tau) - \boldsymbol{v}_\tau^*(\boldsymbol{x}_\tau) - \boldsymbol{v}(\boldsymbol{x}_\tau)) \cdot \nabla \phi_\tau(\boldsymbol{x}_\tau) \\
&= - \int d\boldsymbol{x}_\tau \nabla \cdot (P_\tau(\boldsymbol{x}_\tau)(\boldsymbol{v}_\tau(\boldsymbol{x}_\tau) - \boldsymbol{v}_\tau^*(\boldsymbol{x}_\tau) - \boldsymbol{v}(\boldsymbol{x}_\tau))(\phi_\tau(\boldsymbol{x}_\tau)) = 0, \tag{144}
\end{aligned}
$$

because $\nabla \cdot (P_\tau(\boldsymbol{x}_\tau)(\boldsymbol{v}_\tau(\boldsymbol{x}_\tau) - \boldsymbol{v}_\tau^*(\boldsymbol{x}_\tau) - \boldsymbol{v}(\boldsymbol{x}_\tau))) = 0$ and $P_\tau(\boldsymbol{x}_\tau)$ decays sufficiently rapidly at infinity. Because $D_{\mathrm{KL}}(\mathbb{P}^1_{\boldsymbol{v}_\tau - \boldsymbol{v}_\tau^*} \| \mathbb{P}^1_{\boldsymbol{v}}) \geq 0$, we obtain $\inf_{\mathbb{Q} \in \mathcal{M}_{\mathrm{ZD}}(\mathbb{P})} D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{Q}) = D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^1_{\boldsymbol{v} - \boldsymbol{v}_\tau^*})$ and Eq. (139) from the generalized Pythagorean theorem Eq. (143).

For any $\mathbb{P}^1_{\nabla r} \in \mathcal{M}_{\mathrm{G}}(\mathbb{P})$, we obtain the generalized Pythagorean theorem,

$$
\begin{aligned}
& D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^1_{\nabla r}) \\
&= \frac{dt \left[\int d\boldsymbol{x}_\tau P_\tau(\boldsymbol{x}_\tau) \| \boldsymbol{v}_\tau(\boldsymbol{x}_\tau) - \boldsymbol{v}_\tau^*(\boldsymbol{x}_\tau) - \nabla(r(\boldsymbol{x}_\tau) - \phi_\tau(\boldsymbol{x}_\tau)) \|^2 \right]}{4\mu T} \\
&= \frac{dt \left[\int d\boldsymbol{x}_\tau P_\tau(\boldsymbol{x}_\tau) \| \boldsymbol{v}_\tau(\boldsymbol{x}_\tau) - \boldsymbol{v}_\tau^*(\boldsymbol{x}_\tau) \|^2 + \int d\boldsymbol{x}_\tau P_\tau(\boldsymbol{x}_\tau) \| \nabla(r(\boldsymbol{x}_\tau) - \phi_\tau(\boldsymbol{x}_\tau)) \|^2 \right]}{4\mu T} \\
&= D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^1_{\boldsymbol{v}_\tau^*}) + D_{\mathrm{KL}}(\mathbb{P}^1_{\boldsymbol{v}_\tau^*}\|\mathbb{P}^1_{\nabla r}),
\end{aligned} \tag{145}
$$

where we used Eq. (133) and

$$
\begin{aligned}
& \int d\boldsymbol{x}_\tau P_\tau(\boldsymbol{x}_\tau)(\boldsymbol{v}_\tau(\boldsymbol{x}_\tau) - \boldsymbol{v}_\tau^*(\boldsymbol{x}_\tau)) \cdot \nabla(r(\boldsymbol{x}_\tau) - \phi_\tau(\boldsymbol{x}_\tau)) \\
&= -\int d\boldsymbol{x}_\tau \nabla \cdot (P_\tau(\boldsymbol{x}_\tau)(\boldsymbol{v}_\tau(\boldsymbol{x}_\tau) - \boldsymbol{v}_\tau^*(\boldsymbol{x}_\tau)))(r(\boldsymbol{x}_\tau) - \phi_\tau(\boldsymbol{x}_\tau)) = 0, \quad (146)
\end{aligned}
$$

because $\nabla \cdot [P_\tau(\boldsymbol{x}_\tau)(\boldsymbol{v}_\tau(\boldsymbol{x}_\tau) - \boldsymbol{v}_\tau^*(\boldsymbol{x}_\tau))] = 0$ and $P_\tau(\boldsymbol{x}_\tau)$ decays sufficiently rapidly at infinity. Because $D_{\mathrm{KL}}(\mathbb{P}^1_{\boldsymbol{v}_\tau^*}\|\mathbb{P}^1_{\nabla r}) \geq 0$, we obtain $\inf_{\mathbb{Q} \in \mathcal{M}_{\mathrm{G}}(\mathbb{P})} D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{Q}) = D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{P}^1_{\boldsymbol{v}_\tau^*})$ and Eq. (140) from the generalized Pythagorean theorem Eq. (145). □

Remark 42 Because $\mathbb{P}^1_0 \in \mathcal{M}_{\mathrm{G}}(\mathbb{P})$ and $\mathbb{P}^1_0 \in \mathcal{M}_{\mathrm{ZD}}(\mathbb{P})$, the path probability $\mathbb{P}^1_0$ is on the intersection of two manifolds $\mathcal{M}_{\mathrm{G}}(\mathbb{P})$ and $\mathcal{M}_{\mathrm{ZD}}(\mathbb{P})$.

Remark 43 The path probability density $\mathbb{P}^\theta_{\boldsymbol{v}_\tau^*}$ corresponds to the e-geodesic between $\mathbb{P}$ and $\mathbb{P}^1_{\boldsymbol{v}_\tau^*}$. The path probability density $\mathbb{P}^\theta_{\boldsymbol{v}_\tau - \boldsymbol{v}_\tau^*}$ corresponds to the e-geodesic between $\mathbb{P}$ and $\mathbb{P}^1_{\boldsymbol{v}_\tau - \boldsymbol{v}_\tau^*}$.

The excess entropy production rate and the housekeeping entropy production rate are also regarded as the Fisher metric for the path probability density $\mathbb{P}$. This fact implies that optimal transport can be discussed from the viewpoint of information geometry in the space of path probability densities.

Proposition 22 *The entropy production rate, the excess entropy production rate and the housekeeping entropy production rate are given by*

$$
\sigma_\tau(\mathbb{P}) = \frac{2}{dt} g_{\theta(\boldsymbol{0})\theta(\boldsymbol{0})}(\mathbb{P}), \tag{147}
$$

$$
\sigma_\tau^{\mathrm{ex}}(\mathbb{P}) = \frac{2}{dt} g_{\theta(\boldsymbol{v}_\tau - \boldsymbol{v}_\tau^*)\theta(\boldsymbol{v}_\tau - \boldsymbol{v}_\tau^*)}(\mathbb{P}), \tag{148}
$$

$$
\sigma_\tau^{\mathrm{hk}}(\mathbb{P}) = \frac{2}{dt} g_{\theta(\boldsymbol{v}_\tau^*)\theta(\boldsymbol{v}_\tau^*)}(\mathbb{P}). \tag{149}
$$

Proof By plugging $\mathbf{v}'_\tau = \mathbf{0}$, $\mathbf{v}'_\tau = \mathbf{v}_\tau - \mathbf{v}^*_\tau$ and $\mathbf{v}'_\tau = \mathbf{v}^*_\tau$ into Eq. (39), we obtain Eqs. (147), (148) and (149), respectively. □

Remark 44 The expressions of the excess entropy production rate and the housekeeping entropy production rate by the Fisher metric lead to the thermodynamic uncertainty relations for the excess entropy production rate and the housekeeping entropy production rate as a consequence of the Cramér–Rao inequality. The thermodynamic uncertainty relation for the excess entropy production rate had been discussed in Theorem. 13. These thermodynamic uncertainty relations for the excess entropy production rate and the housekeeping entropy production rate have been substaintially obtained in Ref. [40]. The thermodynamic uncertainty relation for the excess entropy production rate and the housekeeping entropy production rate can be generalized based on the orthogonality as discussed in Ref. [75].

Remark 45 Because the excess entropy production rate is defined in terms of the L^2-Wasserstein distance $\sigma^{\text{ex}}_t = [\mathcal{W}_2(P_\tau, P_{\tau+dt})]^2/(\mu T dt^2)$ up to $O(dt)$, Eq. (139) implies a relation between the L^2-Wasserstein distance and the Kullback–Leibler divergence, and Eq. (148) implies a relation between the L^2-Wasserstein distance and the Fisher metric, respectively. Information-geometrically, the square of the line element for path probability densities can be defined as $ds^2_{\text{path}} = g_{\theta(\mathbf{v}_\tau - \mathbf{v}^*_\tau)\theta(\mathbf{v}_\tau - \mathbf{v}^*_\tau)}d\theta^2 = (1/2)\sigma^{\text{ex}}_\tau dt d\theta^2$ for the interpolation parameter θ in $\mathbb{P}^\theta_{\mathbf{v}_\tau - \mathbf{v}^*_\tau}$. Thus, the L^2-Wasserstein distance can be information-geometrically interpreted as $[\mathcal{W}_2(P_\tau, P_{\tau+dt})]^2 = (2\mu T dt)(ds^2_{\text{path}}/d\theta^2)$.

6 Conclusion and discussion

We discuss stochastic thermodynamic links between information geometry and optimal transport theory via the excess entropy production rate. We can discuss a link between information geometry in the space of probability densities and optimal transport theory and a link between information geometry in the space of path probability densities and optimal transport theory because the excess entropy production rate is related to the L^2-Wasserstein distance, the time derivative of the Kullback–Leibler divergence between probability densities, and the Kullback–Leibler divergence between the path probability densities. These links are useful for studying the mathematical properties of the entropy production rate in stochastic thermodynamics. For example, thermodynamic trade-off relations, namely the thermodynamic uncertainty relations and the thermodynamic speed limit, can be obtained from geometric inequalities such as the Cauchy–Schwartz inequality and the triangle inequality. The optimal protocol to minimize the thermodynamic cost can also be discussed in terms of the geodesic.

We also remark on possible generalizations of the results in this paper. In this paper, we only focus on the stochastic dynamics described by the Fokker–Planck equation. Because stochastic thermodynamics has been discussed for the general Markov process described by the master equation, the generalization of the proposed results for the general Markov process is interesting. For example, generalizations of the results

in this paper for the master equation have been seen in Refs. [22, 25, 28, 38, 41, 43, 44, 60, 76]. Because generalizations are not unique, rather different approaches of optimal transport theory for stochastic thermodynamics in the Markov jump process have also been seen in Refs. [77–80]. Unlike these generalizations, our generalizations [43, 44] are related to the gradient flow expression and information-geometric projection discussed in this paper. The generalization for the deterministic chemical rate equation is also interesting to consider information geometry and optimal transport theory for chemical thermodynamics, which was proposed in Refs. [29, 42–44, 80–83]. For the deterministic chemical rate equation, we do not need to use stochasticity to obtain generalized results, and geometric properties play a crucial role in the derivation of generalized results, similarly as in the stochastic case. This is the reason why we call our framework geometric thermodynamics instead of stochastic thermodynamics.

Finally, we point out that geometric thermodynamics is related to several fascinating topics, and has the potential to clarify the geometric properties of these topics. The classical correspondence of the control protocol called *shortcuts to adiabatically* for the stochastic process [84–88] is related to the geometry of the probability distribution. Remarkably, the link between shortcuts and information geometry has been proposed in Ref. [89]. Indeed, a generalization of our framework for general Markov jump processes [43] is related to the stochastic correspondence of shortcuts called *shortcuts in stochastic systems* [86]. An application of shortcuts or our framework to a low-power electronic circuits called *adiabatic circuits* [90] is promising. A connection between geometric thermodynamics and a geometrical interpretation of another excess entropy production rate proposed in Ref. [91] in terms of *the Berry phase* [92], which is related to the geometry of the cyclic path, is interesting. The cyclic path in information geometry and optimal transport theory was discussed in the optimal heat engine [27, 31, 32, 93, 94] and the geometric pump [95]. A geometric interpretation of the restricted path may also be interesting in the context of *optimal limited control* [96–99]. The dual coordinate systems in stochastic thermodynamics provide the duality in stochastic thermodynamics [41, 42, 44, 82, 83, 100, 101], which is related to the variational calculus such as *the maximum caliber principle* [44, 101–103] and *the Schrödinger bridge* [104, 105]. Our results may be also useful for a machine learning technique based on non-equilibrium thermodynamics called *the diffusion model* [106] or *the score-based generative modeling*, because our result provides a link between machine learning technique based on optimal transport and non-equilibrium thermodynamics for diffusion dynamics described by the Fokker–Planck equation. Because *the denoising matching* [107, 108] or *the reversed stochastic differential equation* [109] can be described by proposed interpolated dynamics, our framework may be helpful in learning process by the diffusion-based generative model. Because our framework also can provides a link between the reversed stochastic differential equation and optimal transport theory, which is well used in generative models such as *the Wasserstein generative adversarial networks* [110], it might be interesting to consider a link between the diffusion model and learning based on the L^2-Wasserstein distance in our framework. A connection to *information thermodynamics* [111] is also interesting. For example, the information-thermodynamic quantities, called the partial entropy production and the transfer entropy [112–116] can be information-geometrically treated by the projection theorem [25, 117], and the optimal transport for the subsystem is related

to the problem of the finite bit erasure [118–121] and the problem of the minimum partial entropy production in the subsystem [31, 122]. Applications to the evolutionary process [45, 123–125] is also interesting because information geometry provides a geometric interpretation of *the Price equation* [126, 127]. As a generalization of the gradient flow expression, an approach based on *the general equation for non-equilibrium reversible-irreversible coupling (GENERIC)* [128–131] or *the transport Hessian Hamiltonian flow* [132] might be promising, The experimental application of geometric thermodynamics to biological dynamics is interesting [37, 133, 134], and the quantitative discussion on the design principle of the complex biological system from the viewpoint of geometric thermodynamics could be a significant topic in the near future.

Data Availability Data sharing is not applicable to this article as no data sets were generated or analyzed during the current study.

Declarations

Conflict of interest On behalf of all authors, Sosuke Ito states that there is no conflict of interest.

References

1. Callen, H.B.: Thermodynamics and an Introduction to Thermostatistics. American Association of Physics Teachers, New York (1998)
2. Einstein, A.: Investigations on the Theory of the Brownian Movement. Courier Corporation, Chelmsford (1956)
3. Onsager, L.: Reciprocal relations in irreversible processes. I. Phys. Rev. **37**(4), 405 (1931)
4. Onsager, L.: Reciprocal relations in irreversible processes. II. Phys. Rev. **38**(12), 2265 (1931)
5. Glansdorff, P., Nicolis, G., Prigogine, I.: The thermodynamic stability theory of non-equilibrium states. Proc. Natl. Acad. Sci. **71**(1), 197–199 (1974)
6. Schnakenberg, J.: Network theory of microscopic and macroscopic behavior of master equation systems. Rev. Mod. Phys. **48**(4), 571 (1976)
7. Weinhold, F.: Metric geometry of equilibrium thermodynamics. J. Chem. Phys. **63**(6), 2479–2483 (1975)
8. Ruppeiner, G.: Thermodynamics: A Riemannian geometric model. Phys. Rev. A **20**(4), 1608 (1979)

9. Salamon, P., Berry, R.S.: Thermodynamic length and dissipated availability. Phys. Rev. Lett. **51**(13), 1127 (1983)
10. Ruppeiner, G.: Riemannian geometry in thermodynamic fluctuation theory. Rev. Mod. Phys. **67**(3), 605 (1995)
11. Crooks, G.E.: Measuring thermodynamic length. Phys. Rev. Lett. **99**(10), 100602 (2007)
12. Amari, S.-I., Nagaoka, H.: Methods of information geometry. American Mathematical Society, Providence (2000)
13. Amari, S.-i: Information Geometry and Its Applications. Springer, Berlin (2016)
14. Rao, C.R.: Information and the accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc. **37**, 81–91 (1945)
15. Sekimoto, K.: Stochastic Energetics. Springer, Berlin (2010)
16. Seifert, U.: Stochastic thermodynamics, fluctuation theorems and molecular machines. Rep. Prog. Phys. **75**(12), 126001 (2012)
17. Ge, H., Qian, H.: Nonequilibrium thermodynamic formalism of nonlinear chemical reaction systems with Waage-Guldberg's law of mass action. Chem. Phys. **472**, 241–248 (2016)
18. Rao, R., Esposito, M.: Nonequilibrium thermodynamics of chemical reaction networks: wisdom from stochastic thermodynamics. Phys. Rev. X **6**(4), 041064 (2016)
19. Aurell, E., Mejía-Monasterio, C., Muratore-Ginanneschi, P.: Optimal protocols and optimal transport in stochastic thermodynamics. Phys. Rev. Lett. **106**(25), 250601 (2011)
20. Sivak, D.A., Crooks, G.E.: Thermodynamic metrics and optimal paths. Phys. Rev. Lett. **108**(19), 190602 (2012)
21. Mandal, D., Jarzynski, C.: Analysis of slow transitions between nonequilibrium steady states. J. Stat. Mech. Theory Exp. **2016**(6), 063204 (2016)
22. Ito, S.: Stochastic thermodynamic interpretation of information geometry. Phys. Rev. Lett. **121**(3), 030605 (2018)
23. Chen, Y., Georgiou, T.T., Tannenbaum, A.: Stochastic control and nonequilibrium thermodynamics: Fundamental limits. IEEE Trans. Autom. Control **65**(7), 2979–2991 (2019)
24. Ito, S., Dechant, A.: Stochastic time evolution, information geometry, and the cramér-rao bound. Phys. Rev. X **10**(2), 021056 (2020)
25. Ito, S., Oizumi, M., Amari, S.-i: Unified framework for the entropy production and the stochastic interaction based on information geometry. Phys. Rev. Res. **2**(3), 033048 (2020)
26. Nicholson, S.B., García-Pintos, L.P., del Campo, A., Green, J.R.: Time-information uncertainty relations in thermodynamics. Nat. Phys. **16**(12), 1211–1215 (2020)
27. Brandner, K., Saito, K.: Thermodynamic geometry of microscopic heat engines. Phys. Rev. Lett. **124**(4), 040602 (2020)
28. Van Vu, T., Hasegawa, Y.: Geometrical bounds of the irreversibility in Markovian systems. Phys. Rev. Lett. **126**(1), 010601 (2021)
29. Yoshimura, K., Ito, S.: Information geometric inequalities of chemical thermodynamics. Phys. Rev. Res. **3**(1), 013175 (2021)
30. Kolchinsky, A., Wolpert, D.H.: Work, entropy production, and thermodynamics of information under protocol constraints. Phys. Rev. X **11**(4), 041024 (2021)
31. Nakazato, M., Ito, S.: Geometrical aspects of entropy production in stochastic thermodynamics based on wasserstein distance. Phys. Rev. Res. **3**(4), 043093 (2021)
32. Fu, R., Taghvaei, A., Chen, Y., Georgiou, T.T.: Maximal power output of a stochastic thermodynamic engine. Automatica **123**, 109366 (2021)
33. Van Kampen, N.G.: Stochastic Processes in Physics and Chemistry. Elsevier, Berlin (1992)
34. Van den Broeck, C., Esposito, M.: Three faces of the second law. II. Fokker-Planck formulation. Phys. Rev. E **82**(1), 011144 (2010)
35. Villani, C.: Optimal Transport: Old and New. Springer, Berlin (2009)
36. Villani, C.: Topics in Optimal Transportation. American Mathematical Society, Providence (2003)
37. Ashida, K., Aoki, K., Ito, S.: Experimental evaluation of thermodynamic cost and speed limit in living cells via information geometry. BioRxiv, 2020–11 (2020)
38. Ito, S.: Information geometry, trade-off relations, and generalized Glansdorff-Prigogine criterion for stability. J. Phys. A Math. Theor. **55**(5), 054001 (2022)
39. Dechant, A., Sasa, S.-I., Ito, S.: Geometric decomposition of entropy production in out-of-equilibrium systems. Phys. Rev. Res. **4**(1), 012034 (2022)

40. Dechant, A., Sasa, S.-I., Ito, S.: Geometric decomposition of entropy production into excess, housekeeping, and coupling parts. Phys. Rev. E **106**(2), 024125 (2022)
41. Ohga, N., Ito, S.: Inferring nonequilibrium thermodynamics from tilted equilibrium using information-geometric legendre transform. arXiv:2112.11008 (arXiv preprint) (2021)
42. Ohga, N., Ito, S.: Information-geometric structure for chemical thermodynamics: An explicit construction of dual affine coordinates. Phys. Rev. E **106**(4), 044131 (2022)
43. Yoshimura, K., Kolchinsky, A., Dechant, A., Ito, S.: Housekeeping and excess entropy production for general nonlinear dynamics. Phys. Rev. Res. **5**(1), 013017 (2023)
44. Kolchinsky, A., Dechant, A., Yoshimura, K., Ito, S.: Information geometry of excess and housekeeping entropy production. arXiv:2206.14599 (arXiv preprint) (2022)
45. Hoshino, M., Nagayama, R., Yoshimura, K., Yamagishi, J.F., Ito, S.: A geometric speed limit for acceleration by natural selection in evolutionary processes. arXiv:2207.04640 (arXiv preprint) (2022)
46. Amari, S.-I., Karakida, R., Oizumi, M.: Information geometry connecting Wasserstein distance and Kullback-Leibler divergence via the entropy-relaxed transportation problem. Inf. Geometry **1**, 13–37 (2018)
47. Li, W.: Transport information geometry: Riemannian calculus on probability simplex. Inf. Geometry **5**(1), 161–207 (2022)
48. Khan, G., Zhang, J.: When optimal transport meets information geometry. Inf. Geometry **5**(1), 47–78 (2022)
49. Wong, T.-K.L., Yang, J.: Pseudo-Riemannian geometry encodes information geometry in optimal transport. Inf. Geometry **5**(1), 131–159 (2022)
50. Cover, T.M.: Elements of Information Theory. Wiley, New York (1999)
51. Crooks, G.E.: Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. Phys. Rev. E **60**(3), 2721 (1999)
52. Chernyak, V.Y., Chertkov, M., Jarzynski, C.: Path-integral analysis of fluctuation theorems for general Langevin processes. J. Stat. Mech. Theory Exp. **2006**(08), 08001 (2006)
53. Kawai, R., Parrondo, J.M., Van den Broeck, C.: Dissipation: The phase-space perspective. Phys. Rev. Lett. **98**(8), 080602 (2007)
54. Dechant, A., Sasa, S.-i: Continuous time reversal and equality in the thermodynamic uncertainty relation. Phys. Rev. Res. **3**(4), 042012 (2021)
55. Beghi, A.: On the relative entropy of discrete-time Markov processes with given end-point densities. IEEE Trans. Inf. Theory **42**(5), 1529–1535 (1996)
56. Barato, A.C., Seifert, U.: Thermodynamic uncertainty relation for biomolecular processes. Phys. Rev. Lett. **114**(15), 158101 (2015)
57. Gingrich, T.R., Horowitz, J.M., Perunov, N., England, J.L.: Dissipation bounds all steady-state current fluctuations. Phys. Rev. Lett. **116**(12), 120601 (2016)
58. Dechant, A.: Multidimensional thermodynamic uncertainty relations. J. Phys. A Math. Theor. **52**(3), 035001 (2018)
59. Hasegawa, Y., Van, Vu., T.: Uncertainty relations in stochastic processes: An information inequality approach. Phys. Rev. E **99**(6), 062126 (2019)
60. Liu, K., Gong, Z., Ueda, M.: Thermodynamic uncertainty relation for arbitrary initial states. Phys. Rev. Lett. **125**(14), 140602 (2020)
61. Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numer. Math. **84**(3), 375–393 (2000)
62. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal. **29**(1), 1–17 (1998)
63. Aurell, E., Gawdzki, K., Mejía-Monasterio, C., Mohayaee, R., Muratore-Ginanneschi, P.: Refined second law of thermodynamics for fast random processes. J. Stat. Phys. **147**, 487–505 (2012)
64. Dechant, A., Sakurai, Y.: Thermodynamic interpretation of wasserstein distance. arXiv:1912.08405 (arXiv preprint) (2019)
65. Hatano, T., Sasa, S.-i: Steady-state thermodynamics of Langevin systems. Phys. Rev. Lett. **86**(16), 3463 (2001)
66. Maes, C., Netočnỳ, K.: A nonequilibrium extension of the Clausius heat theorem. J. Stat. Phys. **154**, 188–203 (2014)
67. Dechant, A.: Current fluctuations and transport efficiency for general Langevin systems. J. Stat. Mech. Theory Exp. **2018**(6), 063209 (2018)
68. Li, W., Zhao, J.: Wasserstein information matrix. Inf. Geom. (2023)

69. Maas, J.: Gradient flows of the entropy for finite Markov chains. J. Funct. Anal. **261**(8), 2250–2292 (2011)
70. Nakamura, T., Hasegawa, H., Driebe, D.: Reconsideration of the generalized second law based on information geometry. J. Phys. Commun. **3**(1), 015015 (2019)
71. Shiraishi, N., Saito, K.: Information-theoretical bound of the irreversibility in thermal relaxation processes. Phys. Rev. Lett. **123**(11), 110603 (2019)
72. Maes, C., Netočnỳ, K.: Revisiting the Glansdorff-Prigogine criterion for stability within irreversible thermodynamics. J. Stat. Phys. **159**(6), 1286–1299 (2015)
73. Qian, H.: Entropy production and excess entropy in a nonequilibrium steady-state of single macro-molecules. Phys. Rev. E **65**(2), 021111 (2002)
74. Rotskoff, G.M., Crooks, G.E., Vanden-Eijnden, E.: Geometric approach to optimal nonequilibrium control: minimizing dissipation in nanomagnetic spin systems. Phys. Rev. E **95**(1), 012148 (2017)
75. Kamijima, T., Ito, S., Dechant, A., Sagawa, T.: Thermodynamic uncertainty relations for steady-state thermodynamics. arXiv:2207.05332 (arXiv preprint) (2022)
76. Otsubo, S., Ito, S., Dechant, A., Sagawa, T.: Estimating entropy production by machine learning of short-time fluctuating currents. Phys. Rev. E **101**(6), 062106 (2020)
77. Dechant, A.: Minimum entropy production, detailed balance and wasserstein distance for continuous-time Markov processes. J. Phys. A: Math. Theor. **55**(9), 094001 (2022)
78. Hamazaki, R.: Speed limits for macroscopic transitions. PRX Quantum **3**(2), 020319 (2022)
79. Van Vu, T., Saito, K.: Thermodynamic unification of optimal transport: thermodynamic uncertainty relation, minimum dissipation, and thermodynamic speed limits. Phys. Rev. X **13**(1), 011013 (2023)
80. Van Vu, T., Saito, K.: Topological speed limit. Phys. Rev. Lett. **130**(1), 010402 (2023)
81. Yoshimura, K., Ito, S.: Thermodynamic uncertainty relation and thermodynamic speed limit in deterministic chemical reaction networks. Phys. Rev. Lett. **127**(16), 160601 (2021)
82. Sughiyama, Y., Loutchko, D., Kamimura, A., Kobayashi, T.J.: Hessian geometric structure of chemical thermodynamic systems with stoichiometric constraints. Phys. Rev. Res. **4**(3), 033065 (2022)
83. Kobayashi, T.J., Loutchko, D., Kamimura, A., Sughiyama, Y.: Hessian geometry of nonequilibrium chemical reaction networks and entropy production decompositions. Phys. Rev. Res. **4**(3), 033208 (2022)
84. Patra, A., Jarzynski, C.: Shortcuts to adiabaticity using flow fields. New J. Phys. **19**(12), 125009 (2017)
85. Li, G., Chen, J.-F., Sun, C., Dong, H.: Geodesic path for the minimal energy cost in shortcuts to isothermality. Phys. Rev. Lett. **128**(23), 230603 (2022)
86. Ilker, E., Güngör, Ö., Kuznets-Speck, B., Chiel, J., Deffner, S., Hinczewski, M.: Shortcuts in stochastic systems and control of biophysical processes. Phys. Rev. X **12**(2), 021048 (2022)
87. Guéry-Odelin, D., Jarzynski, C., Plata, C.A., Prados, A., Trizac, E.: Driving rapidly while remaining in control: classical shortcuts from hamiltonian to stochastic dynamics. Rep. Progress Phys. **20**, 20 (2022)
88. Patrón, A., Prados, A., Plata, C.A.: Thermal brachistochrone for harmonically confined Brownian particles. Eur. Phys. J. Plus **137**(9), 1–20 (2022)
89. Takahashi, K.: Shortcuts to adiabaticity applied to nonequilibrium entropy production: an information geometry viewpoint. New J. Phys. **19**(11), 115007 (2017)
90. Koller, J.G., Athas, W.C.: Adiabatic switching, low energy computing, and the physics of storing and erasing information. In: Workshop on Physics and Computation, pp. 267–270 (1992). IEEE
91. Komatsu, T.S., Nakagawa, N., Sasa, S.-I., Tasaki, H.: Steady-state thermodynamics for heat conduction: microscopic derivation. Phys. Rev. Lett. **100**(23), 230602 (2008)
92. Sagawa, T., Hayakawa, H.: Geometrical expression of excess entropy production. Phys. Rev. E **84**(5), 051110 (2011)
93. Miangolarra, O.M., Taghvaei, A., Chen, Y., Georgiou, T.T.: Geometry of finite-time thermodynamic cycles with anisotropic thermal fluctuations. IEEE Control Syst. Lett. **6**, 3409–3414 (2022)
94. Frim, A.G., DeWeese, M.R.: Geometric bound on the efficiency of irreversible thermodynamic cycles. Phys. Rev. Lett. **128**(23), 230601 (2022)
95. Takahashi, K., Fujii, K., Hino, Y., Hayakawa, H.: Nonadiabatic control of geometric pumping. Phys. Rev. Lett. **124**(15), 150602 (2020)
96. Blaber, S., Louwerse, M.D., Sivak, D.A.: Steps minimize dissipation in rapidly driven stochastic systems. Phys. Rev. E **104**(2), 022101 (2021)

97. Zhong, A., DeWeese, M.R.: Limited-control optimal protocols arbitrarily far from equilibrium. Phys. Rev. E **106**(4), 044135 (2022)
98. Blaber, S., Sivak, D.A.: Optimal control with a strong harmonic trap. Phys. Rev. E **106**(2), 022103 (2022)
99. Chennakesavalu, S., Rotskoff, G.M.: A unified, geometric framework for nonequilibrium protocol optimization. arXiv:2205.01205 (arXiv preprint) (2022)
100. Lu, Z., Qian, H.: Emergence and breaking of duality symmetry in generalized fundamental thermodynamic relations. Phys. Rev. Lett. **128**(15), 150603 (2022)
101. Yang, Y.-J., Qian, H.: Statistical thermodynamics and data infinitum: Conjugate variables as forces, and their statistical variations. arXiv:2205.09321 (arXiv preprint) (2022)
102. Pressé, S., Ghosh, K., Lee, J., Dill, K.A.: Principles of maximum entropy and maximum caliber in statistical physics. Rev. Mod. Phys. **85**(3), 1115 (2013)
103. Léonard, C.: A survey of the schrödinger problem and some of its connections with optimal transport. Discret. Contin. Dyn. Syst. **34**(4), 1533 (2014)
104. Chen, Y., Georgiou, T.T., Pavon, M.: On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint. J. Optim. Theory Appl. **169**, 671–691 (2016)
105. Kamiya, S., Kawakita, G., Sasai, S., Kitazono, J., Oizumi, M.: Optimal control costs of brain state transitions in linear stochastic systems. J. Neurosci. **43**(2), 270–281 (2023)
106. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning, pp. 2256–2265 (2015). PMLR
107. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Adv. Neural. Inf. Process. Syst. **33**, 6840–6851 (2020)
108. Kingma, D., Salimans, T., Poole, B., Ho, J.: Variational diffusion models. Adv. Neural. Inf. Process. Syst. **34**, 21696–21707 (2021)
109. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (2021)
110. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: International Conference on Machine Learning, pp. 214–223 (2017). PMLR
111. Parrondo, J.M., Horowitz, J.M., Sagawa, T.: Thermodynamics of information. Nat. Phys. **11**(2), 131–139 (2015)
112. Ito, S., Sagawa, T.: Information thermodynamics on causal networks. Phys. Rev. Lett. **111**(18), 180603 (2013)
113. Hartich, D., Barato, A.C., Seifert, U.: Stochastic thermodynamics of bipartite systems: transfer entropy inequalities and a Maxwell's demon interpretation. J. Stat. Mech. Theory Exp. **2014**(2), 02016 (2014)
114. Horowitz, J.M., Esposito, M.: Thermodynamics with continuous information flow. Phys. Rev. X **4**(3), 031015 (2014)
115. Ito, S.: Backward transfer entropy: informational measure for detecting hidden Markov models and its interpretations in thermodynamics, gambling and causality. Sci. Rep. **6**(1), 36831 (2016)
116. Schreiber, T.: Measuring information transfer. Phys. Rev. Lett. **85**(2), 461 (2000)
117. Oizumi, M., Tsuchiya, N., Amari, S.-I.: Unified framework for information integration based on information geometry. Proc. Natl. Acad. Sci. **113**(51), 14817–14822 (2016)
118. Proesmans, K., Ehrich, J., Bechhoefer, J.: Finite-time landauer principle. Phys. Rev. Lett. **125**(10), 100602 (2020)
119. Proesmans, K., Ehrich, J., Bechhoefer, J.: Optimal finite-time bit erasure under full control. Phys. Rev. E **102**(3), 032105 (2020)
120. Zhen, Y.-Z., Egloff, D., Modi, K., Dahlsten, O.: Universal bound on energy cost of bit reset in finite time. Phys. Rev. Lett. **127**(19), 190602 (2021)
121. Lee, J.S., Lee, S., Kwon, H., Park, H.: Speed limit for a highly irreversible process and tight finite-time Landauer's bound. Phys. Rev. Lett. **129**(12), 120603 (2022)
122. Fujimoto, Y., Ito, S.: Game-theoretical approach to minimum entropy productions in information thermodynamics. arXiv:2112.14035 (arXiv preprint) (2021)
123. Zhang, Z., Guan, S., Shi, H.: Information geometry in the population dynamics of bacteria. J. Stat. Mech. Theory Exp. **2020**(7), 073501 (2020)
124. Adachi, K., Iritani, R., Hamazaki, R.: Universal constraint on nonlinear population dynamics. Commun. Phys. **5**(1), 129 (2022)

125. García-Pintos, L.P.: Diversity and fitness uncertainty allow for faster evolutionary rates. arXiv:2202.07533 (arXiv preprint) (2022)
126. Frank, S.A.: The price equation program: simple invariances unify population dynamics, thermodynamics, probability, information and inference. Entropy **20**(12), 978 (2018)
127. Frank, S.A., Bruggeman, F.J.: The fundamental equations of change in statistical ensembles and biological populations. Entropy **22**(12), 1395 (2020)
128. Grmela, M., Öttinger, H.C.: Dynamics and thermodynamics of complex fluids. I. Development of a general formalism. Phys. Rev. E **56**(6), 6620 (1997)
129. Öttinger, H.C., Grmela, M.: Dynamics and thermodynamics of complex fluids. II. Illustrations of a general formalism. Phys. Rev. E **56**(6), 6633 (1997)
130. Esen, O., Grmela, M., Pavelka, M.: On the role of geometry in statistical mechanics and thermodynamics. I. Geometric perspective. J. Math. Phys. **63**(12), 122902 (2022)
131. Esen, O., Grmela, M., Pavelka, M.: On the role of geometry in statistical mechanics and thermodynamics. II. Thermodynamic perspective. J. Math. Phys. **63**(12), 123305 (2022)
132. Li, W.: Hessian metric via transport information geometry. J. Math. Phys. **62**(3), 033301 (2021)
133. Tafoya, S., Large, S.J., Liu, S., Bustamante, C., Sivak, D.A.: Using a system's equilibrium behavior to reduce its energy dissipation in nonequilibrium processes. Proc. Natl. Acad. Sci. **116**(13), 5920–5924 (2019)
134. Brown, A.I., Sivak, D.A.: Theory of nonequilibrium free energy transduction by molecular machines. Chem. Rev. **120**(1), 434–459 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.