

Information geometry of perturbed gradient flow systems on hypergraphs: A perspective towards nonequilibrium physics

Dimitri Loutchko^{1*}, Keisuke Sugie² and Tetsuya J Kobayashi^{1,2*}

^{1*}Institute of Industrial Science, The University of Tokyo, 4-6-1, Komaba, Meguro-ku, 153-8505, Tokyo, Japan.

²Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, 113-8656, Tokyo, Japan.

*Corresponding author(s). E-mail(s): dimitri.loutchko@gmail.com; tetsuya@sat.t.u-tokyo.ac.jp;

Abstract

This article serves to concisely review the link between gradient flow systems on hypergraphs and information geometry which has been established within the last five years. Gradient flow systems describe a wealth of physical phenomena and provide powerful analytical techniques which are based on the variational energy-dissipation principle. Modern nonequilibrium physics has complemented this classical principle with thermodynamic uncertainty relations, speed limits, entropy production rate decompositions, and many more. In this article, we formulate these modern principles within the framework of perturbed gradient flow systems on hypergraphs. In particular, we discuss the geometry induced by the Bregman divergence, the physical implications of dual foliations, as well as the corresponding infinitesimal Riemannian geometry for gradient flow systems. Through the geometrical perspective, we are naturally led to new concepts such as moduli spaces for perturbed gradient flow systems and thermodynamical area which is crucial for understanding speed limits. We hope to encourage the readers working in either of the two fields to further expand on and foster the interaction between the two fields.

Keywords: Gradient flow systems, Moduli spaces, Entropy production decompositions, Thermodynamic uncertainty relations, Thermodynamic speed limits, Stochastic thermodynamics, Chemical reaction networks

1 Introduction

The gradient flow formalism provides a wealth of analytical techniques and physical insight to study evolution equations of the type

$$\dot{\mathbf{x}} = D_{\mu}\mathcal{R}^*(\mathbf{x}, -D_{\mathbf{x}}\mathcal{E}(\mathbf{x})), \quad (1.1)$$

where $\mathcal{E} : \mathbf{X} \rightarrow \mathbb{R}$ is a driving functional such as the system energy or entropy, and $\mathcal{R}^* : \mathcal{T}^*\mathbf{X} \rightarrow \mathbb{R}$ is the dissipation potential.

If a system has a gradient flow structure, it allows to deal with stability and convergence properties [1, 2], to construct weak solutions [3–7], and to implement various limiting procedures while retaining thermodynamical consistency [8–12]. Applications include classical reaction diffusion-systems [13], the Boltzmann equation [14], Cahn-Hilliard equations [15], the physics of thin layers [16], diodes [17] and semiconductor devices [18], among others. We refer to the article [19] for an introduction and an exensive review of gradient flow systems.

In a generalized sense, the system (1.1) describes the evolution of an equilibrium system as it minimizes the driving functional \mathcal{E} . In this article, we deal with perturbed gradient flow systems on hypergraphs [20–22] as they can accomodate an external driving of the system via an external force term $\mathbf{f}_x^{\text{ext}}$ in the continuity equation

$$\begin{aligned} \dot{\mathbf{x}} &= -\text{div}_{\mathbf{H}}\mathbf{j}, \\ \mathbf{j} &= D_{\mathbf{f}}\Psi^*(\mathbf{x}, \text{grad}_{\mathbf{H}}[-D_{\mathbf{x}}\mathcal{E}(\mathbf{x})] + \mathbf{f}_x^{\text{ext}}) \end{aligned} \quad (1.2)$$

and thus can describe genuine nonequilibrium situations. Thereby, $\text{div}_{\mathbf{H}}$ and $\text{grad}_{\mathbf{H}}$ denote discrete divergence and gradient operators associated to a hypergraph, and $\Psi^* : \mathcal{F} \rightarrow \mathbb{R}$ is the dissipation potential on the space of thermodynamical forces. In particular, jump processes and chemical reaction networks (CRNs) have been widely investigated within the gradient flow formalism [11, 23–26].

In recent decades, in nonequilibrium and statistical physics, and especially within the field of stochastic thermodynamics [27–29], the concepts of free energy minimization in equilibrium systems have been extended to the nonequilibrium situation, and have been complemented with newly discovered principles which include, among others, fluctuation theorems [30–32], thermodynamic uncertainty relations (TURs) [33–38], speed limits [39–41], and entropy production rate (EPR) decompositions [42–44].

Starting with the pioneering work [45], the importance and utility of information geometry to concisely derive and analyze such principles for Markov jump processes and chemical reaction networks has become ever more apparent and successful [22, 43, 46–55]. In [22], it was noted that, to a great extent, such results do not depend on the specifics of physical system but all that is mathematically required in the derivations is the existence of a gradient flow structure (1.2). As an application, information geometrical EPR decompositions have been treated from this general point of view [43, 56]. Moreover, within the statistics community, works which combine the two fields have started to emerge [57, 58].

In this article, we continue and expand on building the mathematical framework of modern nonequilibrium physics based on the information geometry of gradient flow systems. Following [22], we introduce the information geometry of gradient flow systems and demonstrate how it can be used to concisely derive the above mentioned modern physical principles purely from the existence of a gradient flow structure.

In Section 2.1, we introduce and discuss the basics of gradient flow systems, then extend them to gradient flow systems on hypergraphs in Section 2.2, and add the perturbation in Section 2.3. Thereby we focus on the energy-dissipation principle and how it is influenced by the perturbation.

In Section 3.1, we develop the corresponding information geometry, and reformulate the energy-dissipation principle by using the Bregman divergence. We also discuss the large deviations theoretical interpretation of this quantity. In Section 3.2, this stochastic interpretation is used to relate the Fisher metric from information geometry to the covariance matrix on the flux space. We conclude this section by deriving the dual orthogonal foliations of the flux and force spaces, discuss their physical significance, and derive a new theorem on the moduli spaces of perturbed gradient flow systems on a hypergraph in Section 3.3.

Section 4 is the centerpiece of this article as we provide concise formulations of modern results in nonequilibrium physics through the setup from Section 3. In particular, we do not assume any specific model or structure of the system but only require it to be a gradient flow system of the form (1.2). In Section 4.1 we present the information geometrical dissipation rate decomposition from [43]. It is worth noting that this decomposition does not require the force-flux pair to be conjugate and thus is valid for fluctuating systems. In Section 4.2, we present a package of techniques to obtain TURs, thermodynamic speed limits and another information geometrical dissipation rate decomposition which is based on the Riemannian aspects of the theory.

Finally, we discuss future perspectives in Section 5.

2 Gradient flow systems

In this section, we introduce and discuss the formalism of gradient flow systems. We closely follow the excellent review article [19].

2.1 Basics of gradient flow systems

A gradient flow system is given by an evolution equation with an additional variational structure. In its basic form, the evolution equation for the state $\mathbf{x} \in \mathbf{X}$ of the system with the state space \mathbf{X} (which can be discrete, continuous, or a mixture of both) is given by

$$\dot{\mathbf{x}} = D_{\boldsymbol{\mu}} \mathcal{R}^*(\mathbf{x}, -D_{\mathbf{x}} \mathcal{E}(\mathbf{x})). \quad (2.1)$$

Here, $\mathcal{E} : \mathbf{X} \rightarrow \mathbb{R}$ is the *driving functional* such as the system energy or entropy, and $\mathcal{R}^* : \mathcal{T}^* \mathbf{X} \rightarrow \mathbb{R}$ is the *dissipation potential* on the cotangent bundle with coordinates $(\mathbf{x}, \boldsymbol{\mu}) \in \mathcal{T}^* \mathbf{X}$. The operators $D_{\mathbf{x}}$ and $D_{\boldsymbol{\mu}}$ denote derivatives with respect to \mathbf{x} and $\boldsymbol{\mu}$. This gradient system is denoted by the triple $(\mathbf{X}, \mathcal{R}^*, \mathcal{E})$.

This structure has a clear physical meaning: The term $-D_{\mathbf{x}}\mathcal{E}(\mathbf{x})$ corresponds to a thermodynamic force, and the operation $D_{\boldsymbol{\mu}}\mathcal{R}^*(\mathbf{x}, \cdot)$ converts this force to a velocity. This is a generalization of the Onsager relations [59] which constitute a linear relation between forces and velocities, and are recovered when $\mathcal{R}^*(\mathbf{x}, \boldsymbol{\mu})$ is quadratic in the second argument [6, 13, 60, 61]. Analogous to the Onsager relations, $\mathcal{R}^*(\mathbf{x}, \boldsymbol{\mu})$ and its Legendre transform in the second argument $\mathcal{R}(\mathbf{x}, \mathbf{v})$ are required to satisfy

- (a) $\mathcal{R}^*(\mathbf{x}, \cdot)$ and $\mathcal{R}(\mathbf{x}, \cdot)$ are strictly convex in the second argument,
- (b) $\mathcal{R}^*(\mathbf{x}, \cdot)$ and $\mathcal{R}(\mathbf{x}, \cdot)$ are 1-coercive,
- (c) $\mathcal{R}^*(\mathbf{x}, \cdot)$ and $\mathcal{R}(\mathbf{x}, \cdot)$ are even in the second argument,
- (d) $\min \mathcal{R}^*(\mathbf{x}, \cdot) = \min \mathcal{R}(\mathbf{x}, \cdot) = 0$, and $\mathcal{R}^*(\mathbf{x}, \mathbf{0}) = \mathcal{R}(\mathbf{x}, \mathbf{0}) = 0$

for all $\mathbf{x} \in \mathbf{X}$. The Young-Fenchel relation reads

$$\mathcal{R}(\mathbf{x}, \mathbf{v}) + \mathcal{R}^*(\mathbf{x}, \boldsymbol{\mu}) \geq \langle \mathbf{v}, \boldsymbol{\mu} \rangle \quad (2.2)$$

with equality iff $\mathbf{v} = D_{\boldsymbol{\mu}}\mathcal{R}^*(\mathbf{x}, \boldsymbol{\mu})$ and $\boldsymbol{\mu} = D_{\mathbf{v}}\mathcal{R}(\mathbf{x}, \mathbf{v})$ hold. A central quantity of a gradient flow system is the *dissipation rate* σ^1 at a point $(\mathbf{x}, \mathbf{v}) \in \mathcal{TX}$, defined as

$$\sigma(\mathbf{x}, \mathbf{v}) := \mathcal{R}(\mathbf{x}, \mathbf{v}) + \mathcal{R}^*(\mathbf{x}, -D_{\mathbf{x}}\mathcal{E}(\mathbf{x})). \quad (2.3)$$

The dissipation rate is lower bounded by $\langle \mathbf{v}, -D_{\mathbf{x}}\mathcal{E}(\mathbf{x}) \rangle$ according to (2.2), and if $\mathbf{v} = \dot{\mathbf{x}}$ is the gradient flow vector field, then equality holds and the dissipation rate becomes $\sigma(\mathbf{x}, \dot{\mathbf{x}}) = \langle \dot{\mathbf{x}}, -D_{\mathbf{x}}\mathcal{E}(\mathbf{x}) \rangle = -\dot{\mathcal{E}}(\mathbf{x})$. Together with the non-negativity of the dissipation functions, this implies that $\dot{\mathcal{E}}(\mathbf{x}) = -[\mathcal{R}(\mathbf{x}, \dot{\mathbf{x}}) + \mathcal{R}^*(\mathbf{x}, -D_{\mathbf{x}}\mathcal{E}(\mathbf{x}))] \leq 0$ with equality iff $\dot{\mathbf{x}} = \mathbf{0}$ and $-D_{\mathbf{x}}\mathcal{E}(\mathbf{x}) = \mathbf{0}$ and thus \mathcal{E} is a Lyapunov function for the dynamics.

Globally, the *total dissipation functional* $\Sigma[\cdot]$ is defined for a differentiable curve $\mathbf{x} : [0, T] \rightarrow \mathbf{X}$ as

$$\Sigma[\mathbf{x}(\cdot)] := \int_0^T \mathcal{R}(\mathbf{x}, \dot{\mathbf{x}}) + \mathcal{R}^*(\mathbf{x}, -D_{\mathbf{x}}\mathcal{E}(\mathbf{x})) dt. \quad (2.4)$$

Following the discussion of the dissipation rate, one verifies the energy-dissipation relation

$$\mathcal{E}(\mathbf{x}(T)) + \int_0^T \mathcal{R}(\mathbf{x}, \dot{\mathbf{x}}) + \mathcal{R}^*(\mathbf{x}, -D_{\mathbf{x}}\mathcal{E}(\mathbf{x})) dt \geq \mathcal{E}(\mathbf{x}(0)) \quad (2.5)$$

with the equality if and only if $\dot{\mathbf{x}}$ and $-D_{\mathbf{x}}\mathcal{E}(\mathbf{x})$ are Legendre dual, i.e., if and only if the gradient flow equation (2.1) holds true. This equivalence is known as De Giorgi's energy-dissipation principle (EDP).

¹ The dissipation rate σ is often referred to as the entropy production rate (EPR) in the physics literature. In particular, the concepts discussed in Section 4 are usually stated with σ being identified with the EPR. However, the dissipation rate has a more general meaning in the context of gradient flow systems as the driving functional \mathcal{E} can be any thermodynamic potential and thus σ can describe the dissipation of various thermodynamic potentials, depending on the setup.

In many applications, the state space \mathbf{X} is associated to an underlying discrete structure. For example, it can be the space of positive measures on the vertices of a hypergraph. In such cases, it is more natural to consider the dynamics being generated by forces which live on the edges of the hypergraph rather than in the cotangent space of \mathbf{X} . This approach is now formalized.

2.2 Gradient flow systems on hypergraphs

The gradient flow system $(\mathbf{X}, \mathcal{R}^*, \mathcal{E})$ can be generalized to take into account an underlying structure of a directed weighted hypergraph. Such systems are known as gradient flow systems in the continuity-equation format according to [19], and as the process-flux space formulation of gradient flow systems [21, 26]. A finite² oriented hypergraph $H = (V, E)$ consists of a finite vertex set V and finite edge set E . An edge $e = (h(e, \cdot), t(e, \cdot)) \in E$ is defined by a pair of integer valued maps $h(e, \cdot), t(e, \cdot) : V \rightarrow \mathbb{Z}$, called the head and the tail of the edge. This gives, analogously to the graph setting, the discrete divergence operator $\text{div}_H : \mathbb{R}^E \rightarrow \mathbb{R}^V$ and discrete gradient operator $\text{grad}_H : (\mathbb{R}^V)^* \rightarrow (\mathbb{R}^E)^*$, which are represented by the matrices $[\text{div}_H]_{ve} = t(e, v) - h(e, v)$ and $[\text{grad}_H]_{ev} = h(e, v) - t(e, v)$, respectively. They pose a pair of negative adjoints:

$$\text{div}_H = -\text{grad}_H^*. \quad (2.6)$$

It is natural to treat \mathbb{R}^E and $(\mathbb{R}^E)^*$ as fibers of a flux bundle \mathcal{J} and a force bundle \mathcal{F} over \mathbf{X} , with coordinates $(\mathbf{x}, \mathbf{j}) \in \mathcal{J}$ and $(\mathbf{x}, \mathbf{f}) \in \mathcal{F}$. Assuming that \mathbf{X} is the space of positive measures on V , the spaces \mathbb{R}^V and $(\mathbb{R}^V)^*$ are identified with the fibers of $\mathcal{T}\mathbf{X}$ and $\mathcal{T}^*\mathbf{X}$, respectively. The divergence and gradient operators give bundle morphisms $\text{div}_H : \mathcal{J} \rightarrow \mathcal{T}\mathbf{X}$ and $\text{grad}_H : \mathcal{T}^*\mathbf{X} \rightarrow \mathcal{F}$ over \mathbf{X} .

A dissipation function on a hypergraph is a function $\Psi^* : \mathcal{F} \rightarrow \mathbb{R}$ which satisfies the properties (a)-(d). It induces the Legendre duality between forces and fluxes as $\mathbf{j} = D_{\mathbf{f}}\Psi^*(\mathbf{x}, \mathbf{f})$ and $\mathbf{f} = D_{\mathbf{j}}\Psi(\mathbf{x}, \mathbf{j})$, where $\Psi : \mathcal{J} \rightarrow \mathbb{R}$ is the Legendre transform of Ψ^* with respect to the second argument. Analogously to \mathcal{R} , the dissipation function Ψ generalizes the Onsager reciprocity relations. A gradient flow system on a hypergraph is the tuple $(\mathbf{X}, \Psi^*, \mathcal{E}, H)$ with the dynamics given by the continuity equation

$$\dot{\mathbf{x}} = -\text{div}_H \mathbf{j}, \quad (2.7)$$

$$\mathbf{j} = D_{\mathbf{f}}\Psi^*(\mathbf{x}, \text{grad}_H[-D_{\mathbf{x}}\mathcal{E}(\mathbf{x})]). \quad (2.8)$$

The discussion regarding the gradient flow system $(\mathbf{X}, \mathcal{R}^*, \mathcal{E})$ remains valid in this context:

$$\text{The pair } \Psi \text{ and } \Psi^* \text{ satisfies the inequality } \Psi(\mathbf{x}, \mathbf{j}) + \Psi^*(\mathbf{x}, \mathbf{f}) \geq \langle \mathbf{j}, \mathbf{f} \rangle. \quad (2.9)$$

$$\text{The dissipation rate is given by } \sigma(\mathbf{x}, \mathbf{j}) := \Psi(\mathbf{x}, \mathbf{j}) + \Psi^*(\mathbf{x}, \text{grad}_H[-D_{\mathbf{x}}\mathcal{E}(\mathbf{x})]) \quad (2.10)$$

$$\text{The energy-dissipation principle holds.} \quad (2.11)$$

²We make the finiteness assumption for notational simplicity and remark that the theory for the infinite case is well-established.

Remark 2.1 Note that, if the evolution equations (2.7)-(2.8) are satisfied, then $\sigma(\mathbf{x}, \mathbf{j}) = \langle \mathbf{j}, \text{grad}_{\mathbf{H}}[-\text{D}_{\mathbf{x}}\mathcal{E}(\mathbf{x})] \rangle = \langle -\text{div}_{\mathbf{H}}\mathbf{j}, -\text{D}_{\mathbf{x}}\mathcal{E}(\mathbf{x}) \rangle = -\dot{\mathcal{E}}(\mathbf{x})$ which coincides with the dissipation rate $\sigma(\mathbf{x}, \mathbf{v})$ of the $(\mathbf{X}, \mathcal{R}^*, \mathcal{E})$ gradient flow system in case the gradient flow equation (2.1) is satisfied. Again, \mathcal{E} is a Lyapunov function for the dynamics. We also remark that in the physics literature it is customary to compute the dissipation rate via the bilinear product $\langle \mathbf{j}, \mathbf{f} \rangle$ for a Legendre dual pair \mathbf{f} and \mathbf{j} . The definition (2.10) is more general as it accounts for possible fluctuations of the flux vector.

The advantage of moving to the hypergraph setting is the possibility to describe perturbations which often arise on the edges of the hypergraph in physical contexts and which cannot be captured by $(\mathbf{X}, \mathcal{R}^*, \mathcal{E})$ type systems.

2.3 Perturbed gradient flow systems on hypergraphs and their energetics

We describe a nonequilibrium system by a gradient flow system on a hypergraph $(\mathbf{X}, \Psi^*, \mathcal{E}, \mathbf{H})$ where the driving force in Equation (2.7) is perturbed by an \mathbf{x} -dependent (and possibly time-dependent) external force $\mathbf{f}_{\mathbf{x}}^{\text{ext}} \in \mathbb{R}^E$. In other words, the perturbation is given by a (possibly time-dependent) continuous section $\mathbf{f}^{\text{ext}} : \mathbf{X} \rightarrow \mathcal{F}$. This results in the perturbed evolution equation

$$\dot{\mathbf{x}} = -\text{div}_{\mathbf{H}}\mathbf{j}, \quad (2.12)$$

$$\mathbf{j} = \text{D}_{\mathbf{f}}\Psi^*(\mathbf{x}, \text{grad}_{\mathbf{H}}[-\text{D}_{\mathbf{x}}\mathcal{E}(\mathbf{x})] + \mathbf{f}_{\mathbf{x}}^{\text{ext}}). \quad (2.13)$$

This perturbed gradient flow system is denoted by the quintuple $(\mathbf{X}, \Psi^*, \mathcal{E}, \mathbf{H}, \mathbf{f}^{\text{ext}})$. For any $(\mathbf{x}, \mathbf{j}) \in \mathcal{J}$, the dissipation rate now acquires an additional contribution $\sigma^{\text{ext}}(\mathbf{x}, \mathbf{j}) = \langle \mathbf{j}, \mathbf{f}_{\mathbf{x}}^{\text{ext}} \rangle$ due to the external force:

$$\sigma(\mathbf{x}, \mathbf{j}) = \Psi(\mathbf{x}, \mathbf{j}) + \Psi^*(\mathbf{x}, \text{grad}_{\mathbf{H}}[-\text{D}_{\mathbf{x}}\mathcal{E}(\mathbf{x})] + \mathbf{f}_{\mathbf{x}}^{\text{ext}}) \quad (2.14)$$

$$\geq \langle \mathbf{j}, \text{grad}_{\mathbf{H}}[-\text{D}_{\mathbf{x}}\mathcal{E}(\mathbf{x})] + \mathbf{f}_{\mathbf{x}}^{\text{ext}} \rangle = -\dot{\mathcal{E}}(\mathbf{x}) + \sigma^{\text{ext}}(\mathbf{x}, \mathbf{j}), \quad (2.15)$$

where the inequality becomes an equality iff (2.13) is satisfied and the last equality holds true iff (2.12) is satisfied. For the dynamics of the system, this is a drastic modification as \mathcal{E} is no longer a Lyapunov function and phenomena such as oscillations and multistability can occur even when $\mathbf{f}_{\mathbf{x}}^{\text{ext}}$ is time- and \mathbf{x} -independent. The energy-dissipation balance for a differentiable curve $(\mathbf{x}, \mathbf{j}) : [0, T] \rightarrow \mathcal{J}$ with $\dot{\mathbf{x}} = -\text{div}_{\mathbf{H}}\mathbf{j}$ also acquires a perturbation in the form of the external dissipation $\Sigma^{\text{ext}}[(\mathbf{x}, \mathbf{j})(\cdot)] := \int_0^T \sigma^{\text{ext}}(\mathbf{x}, \mathbf{j})dt$ as

$$\mathcal{E}(\mathbf{x}(T)) + \Sigma[(\mathbf{x}, \mathbf{j})(\cdot)] \geq \mathcal{E}(\mathbf{x}(0)) + \Sigma^{\text{ext}}[(\mathbf{x}, \mathbf{j})(\cdot)], \quad (2.16)$$

where the system's dissipation is $\Sigma[(\mathbf{x}, \mathbf{j})(\cdot)] = \int_0^T \Psi(\mathbf{x}, \mathbf{j}) + \Psi^*(\mathbf{x}, \text{grad}_{\mathbf{H}}[-\text{D}_{\mathbf{x}}\mathcal{E}(\mathbf{x})] + \mathbf{f}_{\mathbf{x}}^{\text{ext}})dt$. The equality holds if and only if \mathbf{j} satisfies the Legendre duality (2.13). We refer to [20] for more details on perturbed gradient flow systems and now focus on the information geometrical aspects.

3 Information geometric aspects of gradient flow systems

The setup discussed so far lends itself naturally to an information geometrical description. We establish the basic information geometric notions of the Bregman divergence, the associated Riemannian metric structure, and dual foliations, and comment on their physical relevance.

The information geometry takes place over a fixed point $\mathbf{x} \in \mathbf{X}$, within the fibers $\mathcal{J}_{\mathbf{x}}$ and $\mathcal{F}_{\mathbf{x}}$ which we treat as a pair of dually flat affine spaces (in the coordinates \mathbf{j} and \mathbf{f} , respectively) equipped with the strictly convex functions $\Psi(\mathbf{x}, \cdot)$ and $\Psi^*(\mathbf{x}, \cdot)$ and the natural bilinear pairing $\mathcal{J}_{\mathbf{x}} \times \mathcal{F}_{\mathbf{x}} \rightarrow \mathbb{R}$ between dual vector spaces. The Legendre duality between $\Psi(\mathbf{x}, \cdot)$ and $\Psi^*(\mathbf{x}, \cdot)$ induces a pair of mutually inverse diffeomorphisms $D_{\mathbf{j}}\Psi(\mathbf{x}, \cdot) : \mathcal{J}_{\mathbf{x}} \rightarrow \mathcal{F}_{\mathbf{x}}$ and $D_{\mathbf{f}}\Psi^*(\mathbf{x}, \cdot) : \mathcal{F}_{\mathbf{x}} \rightarrow \mathcal{J}_{\mathbf{x}}$ between the fibers.

3.1 The Bregman divergence and large deviations theory

The Bregman divergence between two points $\mathbf{j}, \mathbf{j}' \in \mathcal{J}_{\mathbf{x}}$ is given by

$$\mathcal{D}_{\mathbf{x}}^{\mathcal{J}}[\mathbf{j} \parallel \mathbf{j}'] := \Psi(\mathbf{x}, \mathbf{j}) - \Psi(\mathbf{x}, \mathbf{j}') - \langle \mathbf{j} - \mathbf{j}', D_{\mathbf{j}}\Psi(\mathbf{x}, \mathbf{j}') \rangle. \quad (3.1)$$

The Bregman divergence satisfies $\mathcal{D}_{\mathbf{x}}^{\mathcal{J}}[\mathbf{j} \parallel \mathbf{j}'] \geq 0$ and $\mathcal{D}_{\mathbf{x}}^{\mathcal{J}}[\mathbf{j} \parallel \mathbf{j}'] = 0$ iff $\mathbf{j} = \mathbf{j}'$. However, it is not symmetric in the two arguments, and therefore it is often referred to as pseudo-distance. Analogously, for $\mathbf{f}, \mathbf{f}' \in \mathcal{F}_{\mathbf{x}}$, one defines $\mathcal{D}_{\mathbf{x}}^{\mathcal{F}}[\mathbf{f} \parallel \mathbf{f}'] := \Psi^*(\mathbf{x}, \mathbf{f}) - \Psi^*(\mathbf{x}, \mathbf{f}') - \langle D_{\mathbf{f}}\Psi^*(\mathbf{x}, \mathbf{f}'), \mathbf{f} - \mathbf{f}' \rangle$. For a Legendre dual pair $\mathbf{f} = D_{\mathbf{j}}\Psi(\mathbf{x}, \mathbf{j})$ and $\mathbf{f}' = D_{\mathbf{j}'}\Psi(\mathbf{x}, \mathbf{j}')$, one verifies

$$\mathcal{D}_{\mathbf{x}}^{\mathcal{J}}[\mathbf{j} \parallel \mathbf{j}'] = \mathcal{D}_{\mathbf{x}}^{\mathcal{F}}[\mathbf{f}' \parallel \mathbf{f}]. \quad (3.2)$$

Moreover, the mixed representation

$$\begin{aligned} \mathcal{D}_{\mathbf{x}}[\mathbf{j} \parallel \mathbf{f}'] &:= \Psi(\mathbf{x}, \mathbf{j}) + \Psi^*(\mathbf{x}, \mathbf{f}') - \langle \mathbf{j}, \mathbf{f}' \rangle \\ &= \mathcal{D}_{\mathbf{x}}^{\mathcal{J}}[\mathbf{j} \parallel \mathbf{j}']. \end{aligned} \quad (3.3)$$

for $\mathbf{f}' = D_{\mathbf{j}'}\Psi(\mathbf{x}, \mathbf{j}')$ is convenient to use. This representation shows that the Bregman divergence quantifies the failure of the Young-Fenchel relation (2.9) to be an equality and has an immediate physical meaning through its relation to the energetic quantities of the gradient flow system as one computes just like in (2.15):

$$\mathcal{D}_{\mathbf{x}}[\mathbf{j} \parallel \text{grad}_{\mathbf{H}}[-D_{\mathbf{x}}\mathcal{E}(\mathbf{x})] + \mathbf{f}_{\mathbf{x}}^{\text{ext}}] = \sigma(\mathbf{x}, \mathbf{j}) + \dot{\mathcal{E}}(\mathbf{x}) - \sigma^{\text{ext}}(\mathbf{x}, \mathbf{j}). \quad (3.4)$$

Hence, the EDP inequality (2.16) can be refined to yield the information geometric EDP

$$\mathcal{E}(\mathbf{x}(T)) + \Sigma[(\mathbf{x}, \mathbf{j})(\cdot)]$$

$$= \mathcal{E}(\mathbf{x}(0)) + \Sigma^{\text{ext}}[(\mathbf{x}, \mathbf{j})(\cdot)] + \int_0^T \mathcal{D}_{\mathbf{x}}[\mathbf{j} \|\text{grad}_{\mathbb{H}}[-\mathbf{D}_{\mathbf{x}}\mathcal{E}(\mathbf{x})] + \mathbf{f}_{\mathbf{x}}^{\text{ext}}] dt \quad (3.5)$$

for any differentiable curve $(\mathbf{x}, \mathbf{j}) : [0, T] \rightarrow \mathcal{J}$ with $\dot{\mathbf{x}} = -\text{div}_{\mathbb{H}}\mathbf{j}$.

Furthermore, the Bregman divergence (3.4) has a probabilistic interpretation as the large deviations rate function of a stochastic jump process on the hypergraph in the large volume limit [21, 24–26]. In other words, it quantifies the exponential unlikeliness of the flux \mathbf{j} to deviate from the macroscopic flux given by (2.12). Hence, the last term in (3.5) is the quantifier for the exponential “unlikeliness” of the curve $(\mathbf{x}, \mathbf{j}) : [0, T] \rightarrow \mathcal{J}$ to be realized by a stochastic system.

3.2 Riemannian metric structure and covariance

Continuing with this probabilistic interpretation of the Bregman divergence (3.4) as a rate function, it is a standard fact in large deviations theory that the inverse covariance matrix of the flux vector \mathbf{j} is given by the Hessian of the rate function, which can be explicitly computed as

$$[\text{Cov}_{\mathbf{x}}(\mathbf{j})]^{-1} = \frac{\partial^2 \mathcal{D}_{\mathbf{x}}[\mathbf{j} \|\text{grad}_{\mathbb{H}}[-\mathbf{D}_{\mathbf{x}}\mathcal{E}(\mathbf{x})] + \mathbf{f}_{\mathbf{x}}^{\text{ext}}]}{\partial \mathbf{j} \partial \mathbf{j}} = \frac{\partial^2 \sigma(\mathbf{x}, \mathbf{j})}{\partial \mathbf{j} \partial \mathbf{j}} = \frac{\partial^2 \Psi(\mathbf{x}, \mathbf{j})}{\partial \mathbf{j} \partial \mathbf{j}}. \quad (3.6)$$

We recognize this as the natural Riemannian metric on the flux space $\mathcal{J}_{\mathbf{x}}$, known as the Fisher metric, and denote it by $\mathbf{g}_{\mathbf{x}, \mathbf{j}}^{\mathcal{J}} : \mathcal{T}_{\mathbf{j}}\mathcal{J}_{\mathbf{x}} \times \mathcal{T}_{\mathbf{j}}\mathcal{J}_{\mathbf{x}} \rightarrow \mathbb{R}$. The Legendre duality between $\mathcal{J}_{\mathbf{x}}$ and $\mathcal{F}_{\mathbf{x}}$, induces an isometry with respect to the metric $\mathbf{g}_{\mathbf{x}, \mathbf{f}}^{\mathcal{F}} : \mathcal{T}_{\mathbf{f}}\mathcal{F}_{\mathbf{x}} \times \mathcal{T}_{\mathbf{f}}\mathcal{F}_{\mathbf{x}} \rightarrow \mathbb{R}$ on $\mathcal{F}_{\mathbf{x}}$, which is given by the Hessian

$$\mathbf{g}_{\mathbf{x}, \mathbf{f}}^{\mathcal{F}} := \frac{\partial^2 \Psi^*(\mathbf{x}, \mathbf{f})}{\partial \mathbf{f} \partial \mathbf{f}} = \left[\frac{\partial^2 \Psi(\mathbf{x}, \mathbf{j})}{\partial \mathbf{j} \partial \mathbf{j}} \right]^{-1} = \text{Cov}_{\mathbf{x}}(\mathbf{j}) \quad (3.7)$$

in the \mathbf{f} -coordinates. Note that this Riemannian metric structure can be thought of as an infinitesimal version of the pseudo-distance function provided by the Bregman divergence due to the relation $\mathcal{D}_{\mathbf{x}}^{\mathcal{J}}[\mathbf{j} \|\mathbf{j} + \delta\mathbf{j}] = \frac{1}{2} \sum_{r, r'=1}^{|\mathbb{E}|} [\mathbf{g}_{\mathbf{x}, \mathbf{j}}^{\mathcal{J}}]_{rr'} \delta j_r \delta j_{r'} + \mathcal{O}(\|\delta\mathbf{j}\|^3)$.

3.3 Dual foliations and moduli for perturbed gradient flow systems

The combination of the hypergraph structure and gradient flow structure imposes natural foliations of $\mathcal{J}_{\mathbf{x}}$ and $\mathcal{F}_{\mathbf{x}}$, which correspond to mixture families in information geometry. Here, we focus on the foliation of $\mathcal{F}_{\mathbf{x}}$ in more detail as it reveals the moduli structure of the perturbed gradient flow system, and note that duality yields the analogous statements for $\mathcal{J}_{\mathbf{x}}$. The two linear subspaces $\ker[\text{div}_{\mathbb{H}}] \subset \mathcal{J}_{\mathbf{x}}$ and $\text{img}[\text{grad}_{\mathbb{H}}] \subset \mathcal{F}_{\mathbf{x}}$ are orthogonal with respect to the natural bilinear pairing between $\mathcal{J}_{\mathbf{x}}$ and $\mathcal{F}_{\mathbf{x}}$, and any linear isomorphism between the vector spaces $\mathcal{J}_{\mathbf{x}}$ and $\mathcal{F}_{\mathbf{x}}$ will yield an orthogonal decomposition of $\mathcal{F}_{\mathbf{x}}$ into $\text{img}[\text{grad}_{\mathbb{H}}]$ and its orthogonal complement which is isomorphic to $\ker[\text{div}_{\mathbb{H}}]$. The mixture families in information geometry generalize this situation by using Legendre duality instead of the linear isomorphism

which results in an orthogonal decomposition of \mathcal{F}_x into $\text{img}[\text{grad}_H]$ and the Legendre transform of $\ker[\text{div}_H]$ which is a non-linear space in general. The orthogonality results from a generalized Pythagorean relation using the Bregman divergence. We now recall the construction: For any $\mathbf{f} \in \mathcal{F}_x$, define the isoperturbation space of \mathbf{f} as

$$\mathcal{F}_x^b(\mathbf{f}) := \{\mathbf{f}' \in \mathcal{F}_x | \mathbf{f} - \mathbf{f}' \in \text{img}[\text{grad}_H]\} \quad (3.8)$$

and the isovelocity space of \mathbf{f} as

$$\mathcal{F}_x^c(\mathbf{f}) := \{\mathbf{f}' \in \mathcal{F}_x | D_{\mathbf{f}}\Psi^*(x, \mathbf{f}) - D_{\mathbf{f}'}\Psi^*(x, \mathbf{f}') \in \ker[\text{div}_H]\}. \quad (3.9)$$

For any $\mathbf{j} \in \mathcal{J}_x$, one can analogously define the isovelocity space $\mathcal{J}_x^c(\mathbf{j}) := \{\mathbf{j}' \in \mathcal{J}_x | \mathbf{j} - \mathbf{j}' \in \ker[\text{div}_H]\}$ and the isoperturbation space $\mathcal{J}_x^b(\mathbf{j}) := \{\mathbf{j}' \in \mathcal{J}_x | D_{\mathbf{j}}\Psi(x, \mathbf{j}) - D_{\mathbf{j}'}\Psi(x, \mathbf{j}') \in \text{img}[\text{grad}_H]\}$ on the flux space. The Legendre transformations $D_{\mathbf{j}}\Psi(x, \cdot)$ and $D_{\mathbf{f}}\Psi^*(x, \cdot)$ restrict to diffeomorphisms between $\mathcal{F}_x^b(\mathbf{f})$ and $\mathcal{J}_x^b(\mathbf{j})$ as well as between $\mathcal{F}_x^c(\mathbf{f})$ and $\mathcal{J}_x^c(\mathbf{j})$ iff \mathbf{f} and \mathbf{j} are Legendre dual.

The following generalized Pythagorean relation holds for any $\mathbf{f}' \in \mathcal{F}_x^b(\mathbf{f})$ and $\mathbf{f}'' \in \mathcal{F}_x^c(\mathbf{f})$:

$$\mathcal{D}_x^{\mathcal{F}}[\mathbf{f}' \| \mathbf{f}''] = \mathcal{D}_x^{\mathcal{F}}[\mathbf{f}' \| \mathbf{f}] + \mathcal{D}_x^{\mathcal{F}}[\mathbf{f} \| \mathbf{f}''], \quad (3.10)$$

which follows from the definition (3.1) and the orthogonality condition $\langle \mathbf{f}' - \mathbf{f}, D_{\mathbf{f}}\Psi^*(x, \mathbf{f}) - D_{\mathbf{f}'}\Psi^*(x, \mathbf{f}') \rangle = 0$ between $\ker[\text{div}_H]$ and $\text{img}[\text{grad}_H]$. The generalized Pythagorean relation implies that \mathbf{f} is the unique intersection point³ of $\mathcal{F}_x^b(\mathbf{f})$ and $\mathcal{F}_x^c(\mathbf{f})$. Thus, the space \mathcal{F}_x is foliated by the spaces $\{\mathcal{F}_x^c(\mathbf{f}') | \mathbf{f}' \in \mathcal{F}_x^b(\mathbf{f})\}$ for any $\mathbf{f} \in \mathcal{F}_x$, i.e., any $\mathcal{F}_x^b(\mathbf{f})$ serves as the base of the foliation. Dually, any $\mathcal{F}_x^c(\mathbf{f}')$ is a base for the foliation $\{\mathcal{F}_x^b(\mathbf{f}) | \mathbf{f} \in \mathcal{F}_x^c(\mathbf{f}')\}$ of \mathcal{F}_x . The Legendre transform gives the analogous foliations $\{\mathcal{J}_x^b(\mathbf{j}') | \mathbf{j}' \in \mathcal{J}_x^c(\mathbf{j})\}$ and $\{\mathcal{J}_x^c(\mathbf{j}) | \mathbf{j} \in \mathcal{J}_x^b(\mathbf{j}')\}$ of \mathcal{J}_x which can be obtained from the ones of \mathcal{F}_x via the Legendre transformation. This geometry is illustrated in Figure 1.

This geometrical setup can be used to construct the moduli space for the perturbed gradient flow system (2.12)-(2.13) as follows: First, we require the following definitions.

Definition 3.1 For a given gradient flow system on a hypergraph $(\mathbf{X}, \Psi^*, \mathcal{E}, H)$, two continuous perturbation fields $\mathbf{f}^{ext}, \bar{\mathbf{f}}^{ext} : \mathbf{X} \rightarrow \mathcal{F}$ are equivalent if and only if

$$\mathcal{F}_x^b(\mathbf{f}_x^{ext}) = \mathcal{F}_x^b(\bar{\mathbf{f}}_x^{ext}) \quad (3.11)$$

for all $x \in \mathbf{X}$. In other words, if and only if the perturbations \mathbf{f}^{ext} and $\bar{\mathbf{f}}^{ext}$ differ by a section $\mathbf{h}_x := \mathbf{f}_x^{ext} - \bar{\mathbf{f}}_x^{ext}$, which can be written as a discrete gradient, i.e., $\mathbf{h}_x \in \text{img}[\text{grad}_H]$.

Remark 3.1 This definition of equivalence is rather coarse as we do not require the existence of a potential $\tilde{\mathcal{E}} : \mathbf{X} \rightarrow \mathbb{R}$ such that $\mathbf{h}_x = \text{grad}_H[-D_x\tilde{\mathcal{E}}(x)]$. This additional requirement leads to finer moduli problems which we do not address here.

³To be precise, the relation (3.10) implies that \mathbf{f} has the variational characterization $\mathbf{f} = \text{argmin}_{\mathbf{f}' \in \mathcal{F}_x^b(\mathbf{f})} \mathcal{D}_x^{\mathcal{F}}[\mathbf{f}' \| \mathbf{f}']$ for a fixed $\mathbf{f}'' \in \mathcal{F}_x^c(\mathbf{f})$ and thus is unique.

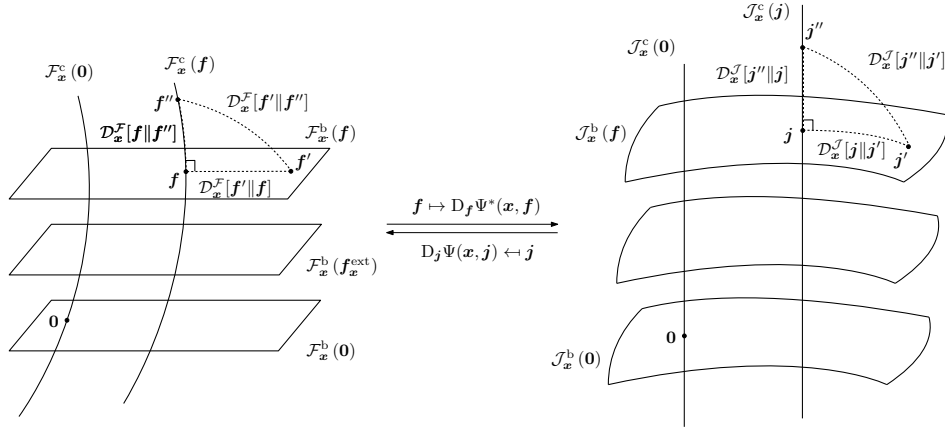


Fig. 1 Illustration of the dual foliations of \mathcal{F}_x and \mathcal{J}_x resulting from the interplay between the hypergraph structure and information geometry. Each space $\mathcal{F}_x^c(f')$ acts as a base for a foliation of \mathcal{F}_x with leaves $\{\mathcal{F}_x^b(f)|f \in \mathcal{F}_x^c(f')\}$, and dually, each space $\mathcal{F}_x^b(f)$ is a base space for a foliation with leaves $\{\mathcal{F}_x^c(f')|f' \in \mathcal{F}_x^b(f)\}$. The analogous situation holds for the space \mathcal{J}_x . The spaces $\mathcal{F}_x^b(f)$ and $\mathcal{J}_x^c(j)$ are affine while $\mathcal{F}_x^c(f)$ and $\mathcal{J}_x^b(j)$ are curved. The leaves intersect the base orthogonally according to the Pythagorean relation (3.10), which is also illustrated in the figure.

Definition 3.2 A moduli space for continuous perturbation fields $\mathbf{f}^{ext} : \mathbf{X} \rightarrow \mathcal{F}$ of a gradient flow system $(\mathbf{X}, \Psi^*, \mathcal{E}, \mathbf{H})$ is a set that is in one-to-one correspondence with the equivalence classes (according to Definition 3.1) of continuous perturbation fields.

The Definition 3.1 states that, for a given $\mathbf{x} \in \mathbf{X}$, the equivalence classes of perturbation fields \mathbf{f}_x^{ext} are given by the leaves of the foliation $\{\mathcal{F}_x^b(f)|f \in \mathcal{F}_x^c(f')\}$, which we call the *local* moduli space for the moduli problem 3.2. In particular, choosing the base space $\mathcal{F}_x^c(0)$ for the foliation and using the Legendre duality between $\mathbf{f} = 0$ and $\mathbf{j} = 0$, one sees that $\mathcal{J}_x^b(0)$, which is the linear subspace $\ker[\text{div}_{\mathbf{H}}] \subset \mathcal{J}_x$, is a local moduli space. This gives the solution to the (global) moduli problem:

Theorem 3.1 The moduli space for continuous perturbation fields $\mathbf{f}^{ext} : \mathbf{X} \rightarrow \mathcal{F}$ of a gradient flow system $(\mathbf{X}, \Psi^*, \mathcal{E}, \mathbf{H})$ is given by the set of continuous sections $\{s : \mathbf{X} \rightarrow \mathcal{K}\}$, where \mathcal{K} denotes the subbundle of \mathcal{J} with the fibers $\mathcal{K}_x = \ker[\text{div}_{\mathbf{H}}] \subset \mathcal{J}_x$.

This is a generalized reformulation of the Schnakenberg theory for the decomposition of nonequilibrium forces at steady states into cycle components [62]. Note that the classical Schnakenberg theory treats the case of \mathbf{x} -independent perturbation fields \mathbf{f}^{ext} where one obtains one fiber $\mathcal{K}_x = \ker[\text{div}_{\mathbf{H}}]$ as the corresponding moduli space.

In the next section, we present some typical physical applications which utilize information and differential geometrical techniques. We show how information geometry can be used to streamline known results from statistical physics and to derive new ones.

4 Nonequilibrium physics via the information geometry to gradient flow systems

Within the recent decades, nonequilibrium physics has seen the discovery of modern principles such as fluctuation theorems [30–32], thermodynamic uncertainty relations (TURs) [33–38, 63], speed limits [39–41], and entropy production rate (EPR) decompositions [42–44].

We give two example applications which have been developed in the context of chemical reaction network (CRN) theory [43, 64]. However, the results can be phrased more generally in the setting of perturbed gradient flow systems without requiring the particular details from CRN theory for the most part, as we demonstrate here. We point out, however, that the assumption of Lemma 4.1 requires more detail as we do not have a proof in the most general case. We introduce the gradient flow system formulation for CRNs and Markov jump processes in Appendix A, and prove the assumption of the lemma.

4.1 Dissipation rate decomposition

The decomposition of the dissipation rate into separate contributions from cycle and boundary components of the force and flux vectors has a long standing history in the theory of nonequilibrium physics, where it is known as entropy production rate (EPR) decomposition [42, 44, 62, 65–69]. As pointed out in the Footnote 1, the concept of dissipation rate is more general and we hence we keep using this terminology. Such decompositions have also been treated in the gradient flow literature [70, 71]. We refer to [43] for an extended discussion.

The cycle components are naturally identified as fluxes \mathbf{j}^c in the linear subspace $\mathcal{J}_{\mathbf{x}}^c(\mathbf{0}) = \ker[\text{div}_{\mathbb{H}}]$ whereas the boundaries are forces \mathbf{f}^b in the linear space $\mathcal{F}_{\mathbf{x}}^b(\mathbf{0}) = \text{img}[\text{grad}_{\mathbb{H}}]$. Recall from Eq. (2.14) the definition of the dissipation rate as

$$\sigma(\mathbf{x}, \mathbf{j}) = \Psi(\mathbf{x}, \mathbf{j}) + \Psi^*(\mathbf{x}, \text{grad}_{\mathbb{H}}[-\mathbf{D}_{\mathbf{x}}\mathcal{E}(\mathbf{x})] + \mathbf{f}_{\mathbf{x}}^{\text{ext}}). \quad (4.1)$$

The great advantage of this gradient flow systems representation is the additive splitting of the contributions from the flux \mathbf{j} and the force $\text{grad}_{\mathbb{H}}[-\mathbf{D}_{\mathbf{x}}\mathcal{E}(\mathbf{x})] + \mathbf{f}_{\mathbf{x}}^{\text{ext}}$ through the dual dissipation potentials. This allows to consider dissipation rate decomposition where the flux is not given by Eq. (2.13), i.e., by $\mathbf{j} = \mathbf{D}_{\mathbf{f}}\Psi^*(\mathbf{x}, \text{grad}_{\mathbb{H}}[-\mathbf{D}_{\mathbf{x}}\mathcal{E}(\mathbf{x})] + \mathbf{f}_{\mathbf{x}}^{\text{ext}})$, but is allowed to fluctuate - this situation is not treated within the classical approaches.

We recall the decomposition for a pair of quadratic dissipation potentials $\Psi(\mathbf{x}, \mathbf{j}) = \frac{1}{2}\mathbf{j}^t\mathbf{L}(\mathbf{x})\mathbf{j}$ and $\Psi^*(\mathbf{x}, \mathbf{j}) = \frac{1}{2}\mathbf{f}^t\mathbf{L}(\mathbf{x})^{-1}\mathbf{f}$ (where $\mathbf{L}(\mathbf{x})$ is a symmetric positive definite matrix) [42]. Usually, this decomposition is derived under the condition that \mathbf{j} is the Legendre dual of \mathbf{f} . However, here, we do not require this condition. The quadratic dissipation potentials naturally equip the spaces with $\mathbf{L}(\mathbf{x})$, respectively $\mathbf{L}(\mathbf{x})^{-1}$ -weighted bilinear product structures⁴ and the respective vector norms $\|\cdot\|_{\mathbf{L}(\mathbf{x})}$ and $\|\cdot\|_{\mathbf{L}(\mathbf{x})^{-1}}$. The dissipation rate is then given by the sum of squared norms

⁴The bilinear product $\langle \cdot, \cdot \rangle_{\mathbf{L}(\mathbf{x})}$ on $\mathcal{J}_{\mathbf{x}}$ is defined as $\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{L}(\mathbf{x})} := \sum_{i,j=1}^{|\mathbb{E}|} u_i \mathbf{L}(\mathbf{x})_{ij} v_j$ for $\mathbf{u}, \mathbf{v} \in \mathcal{J}_{\mathbf{x}}$, and the bilinear product $\langle \cdot, \cdot \rangle_{\mathbf{L}(\mathbf{x})^{-1}}$ is defined analogously on $\mathcal{F}_{\mathbf{x}}$.

$\sigma(\mathbf{x}, \mathbf{j}) = \frac{1}{2} \|\mathbf{j}\|_{\mathcal{L}(\mathbf{x})}^2 + \frac{1}{2} \|\mathbf{f}\|_{\mathcal{L}(\mathbf{x})^{-1}}^2$, and a natural decomposition is obtained by letting \mathbf{j}^c be the $\langle \cdot, \cdot \rangle_{\mathcal{L}(\mathbf{x})}$ -orthogonal projection of \mathbf{j} to $\mathcal{J}_x^c(\mathbf{0})$, and \mathbf{f}^b be the $\langle \cdot, \cdot \rangle_{\mathcal{L}(\mathbf{x})^{-1}}$ -orthogonal projection of \mathbf{f} to $\mathcal{F}_x^b(\mathbf{0})$. Let $\mathbf{j}^b = \mathbf{j} - \mathbf{j}^c$ and $\mathbf{f}^c = \mathbf{f} - \mathbf{f}^b$ be the respective orthogonal complements. Note that we have $D_{\mathbf{j}}\Psi(\mathbf{x}, \mathbf{j}^b) \in \mathcal{F}_x^b(\mathbf{0})$ as well as $D_{\mathbf{f}}\Psi^*(\mathbf{x}, \mathbf{f}^c) \in \mathcal{J}_x^c(\mathbf{0})$. This yields the desired decomposition into the housekeeping and excess dissipation rates $\sigma_{\text{IG}}^{\text{hk}}(\mathbf{x}, \mathbf{j})$ and $\sigma_{\text{IG}}^{\text{ex}}(\mathbf{x}, \mathbf{j})$:

$$\sigma(\mathbf{x}, \mathbf{j}) = \underbrace{\frac{1}{2} \left[\|\mathbf{j}^c\|_{\mathcal{L}(\mathbf{x})}^2 + \|\mathbf{f}^c\|_{\mathcal{L}(\mathbf{x})^{-1}}^2 \right]}_{\sigma_{\text{MN}}^{\text{hk}}(\mathbf{x}, \mathbf{j})} + \underbrace{\frac{1}{2} \left[\|\mathbf{j}^b\|_{\mathcal{L}(\mathbf{x})}^2 + \|\mathbf{f}^b\|_{\mathcal{L}(\mathbf{x})^{-1}}^2 \right]}_{\sigma_{\text{MN}}^{\text{ex}}(\mathbf{x}, \mathbf{j})}, \quad (4.2)$$

which separates the effect of the boundary and cycle components. When \mathbf{j} and \mathbf{f} are Legendre dual, one verifies the equalities $\|\mathbf{j}^c\|_{\mathcal{L}(\mathbf{x})}^2 = \|\mathbf{f}^c\|_{\mathcal{L}(\mathbf{x})^{-1}}^2$ and $\|\mathbf{j}^b\|_{\mathcal{L}(\mathbf{x})}^2 = \|\mathbf{f}^b\|_{\mathcal{L}(\mathbf{x})^{-1}}^2$.

We now discuss the generalization of this decomposition to the case of arbitrary dissipation potentials following [43]. The orthogonal splitting of the respective dissipation potentials uses the dual orthogonal foliations introduced in Section 3.3, which generalize the orthogonal decompositions from linear algebra. The following discussion applies for arbitrary \mathbf{f} , i.e., not necessarily of the form $\mathbf{f} = \text{grad}_{\mathcal{H}}[-D_{\mathbf{x}}\mathcal{E}(\mathbf{x})] + \mathbf{f}_x^{\text{ext}}$, and arbitrary \mathbf{j} . Let $\mathbf{f}^c \in \mathcal{F}_x$ be the unique intersection point of $\mathcal{F}_x^b(\mathbf{f})$ and $\mathcal{F}_x^c(\mathbf{0})$. The dissipation potential for arbitrary $\mathbf{f}' \in \mathcal{F}_x$ can be represented as $\Psi^*(\mathbf{x}, \mathbf{f}') = \mathcal{D}_x^{\mathcal{F}}[\mathbf{f}' \|\mathbf{0}]$, and thus the generalized Pythagorean relation (3.10) yields

$$\begin{aligned} \Psi^*(\mathbf{x}, \mathbf{f}) &= \mathcal{D}_x^{\mathcal{F}}[\mathbf{f} \|\mathbf{0}] = \mathcal{D}_x^{\mathcal{F}}[\mathbf{f} \|\mathbf{f}^c] + \mathcal{D}_x^{\mathcal{F}}[\mathbf{f}^c \|\mathbf{0}] \\ &= \mathcal{D}_x^{\mathcal{F}}[\mathbf{f} \|\mathbf{f}^c] + \Psi^*(\mathbf{x}, \mathbf{f}^c). \end{aligned} \quad (4.3)$$

Analogously, let $\mathbf{j}^b \in \mathcal{J}_x$ be the unique intersection point of $\mathcal{J}_x^c(\mathbf{j})$ and $\mathcal{J}_x^b(\mathbf{0})$. The dual Pythagorean relation yields the decomposition

$$\Psi(\mathbf{x}, \mathbf{j}) = \mathcal{D}_x^{\mathcal{J}}[\mathbf{j} \|\mathbf{j}^b] + \Psi(\mathbf{x}, \mathbf{j}^b). \quad (4.4)$$

This generalizes the decomposition (4.2) as

$$\sigma(\mathbf{x}, \mathbf{j}) = \underbrace{\left[\Psi^*(\mathbf{x}, \mathbf{f}^c) + \mathcal{D}_x^{\mathcal{J}}[\mathbf{j} \|\mathbf{j}^b] \right]}_{\sigma_{\text{IG}}^{\text{hk}}(\mathbf{x}, \mathbf{j})} + \underbrace{\left[\Psi(\mathbf{x}, \mathbf{j}^b) + \mathcal{D}_x^{\mathcal{F}}[\mathbf{f} \|\mathbf{f}^c] \right]}_{\sigma_{\text{IG}}^{\text{ex}}(\mathbf{x}, \mathbf{j})}. \quad (4.5)$$

This is the decomposition proposed in [43]. The geometry of this decomposition is illustrated in Fig. 2.

We briefly discuss the interpretation of information geometrical decomposition 4.5 for the case that the gradient flow equations (2.12)-(2.13) hold true. If the system is unperturbed, then $\mathcal{F}_x^b(\mathbf{f}) = \mathcal{F}_x^b(\mathbf{0})$ and $\mathbf{f}^c = \mathcal{F}_x^b(\mathbf{0}) \cap \mathcal{F}_x^c(\mathbf{0}) = \mathbf{0}$ hold. This yields $\mathbf{f}^b = \mathbf{f}$ and, by Legendre duality, $\mathbf{j}^b = \mathbf{j}$. Thus, only the excess dissipation rate survives and we have $\sigma(\mathbf{x}, \mathbf{j}) = \sigma_{\text{IG}}^{\text{ex}}(\mathbf{x}, \mathbf{j})$ for the unperturbed case.

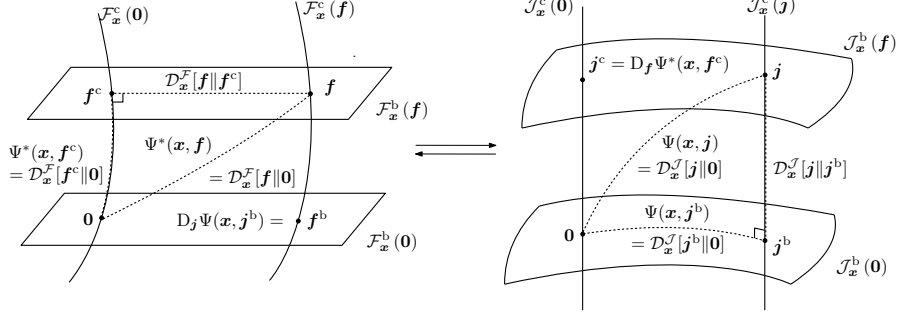


Fig. 2 Illustration of the geometry underlying the dissipation rate decomposition (4.5). On \mathcal{F}_x , the composition is achieved via the cycle force $f^c = \mathcal{F}_x^b(f) \cap \mathcal{F}_x^c(0)$ and the corresponding orthogonal decomposition of $\Psi^*(x, f)$, and dually, on \mathcal{J}_x via the boundary flux $j^b = \mathcal{J}_x^b(0) \cap \mathcal{J}_x^c(j)$. Note that one can also define directly $f^b := \mathcal{F}_x^b(0) \cap \mathcal{F}_x^c(f)$ to get the Legendre dual of j^b , and analogously for $j^c := \mathcal{J}_x^b(j) \cap \mathcal{J}_x^c(0)$.

At a steady state x^{ss} , which is defined by $\dot{x}^{ss} = 0$, the continuity equation $\dot{x} = -\text{div}_H j$ implies that $j \in \mathcal{J}_x^c(0)$ and thus $j^b = \mathcal{J}_x^b(0) \cap \mathcal{J}_x^c(0) = 0$. Moreover, the Legendre duality between j and f yields $f \in \mathcal{F}_x^c(0)$ and therefore $f^c = \mathcal{F}_x^c(f) \cap \mathcal{F}_x^c(0) = f$. This shows that the excess dissipation rate $\sigma_{IG}^{ex}(x^{ss}, j)$ vanishes and we obtain $\sigma(x^{ss}, j) = \sigma_{IG}^{hk}(x^{ss}, j)$.

Hence, for a perturbed gradient flow system, the excess term generalizes the dissipation away from equilibrium to the dissipation away from the steady state, and relates it to the boundary components of fluxes and forces. The housekeeping term quantifies the dissipation due to the nonequilibrium nature of the system, i.e., due to the presence of the perturbation term, and relates it to the cycle components.

4.2 Riemmanian geometry of thermodynamic uncertainty relations and speed limits

In this section, we present a coherent information geometrical toolbox to deal with thermodynamic uncertainty relations (TURs) and thermodynamic speed limits.

4.2.1 Thermodynamic uncertainty relations

Thermodynamic uncertainty relations have been extensively studied in the physics literature for over a decade, and various formulations have been obtained based on explicit models [33, 34, 36–38]. Throughout this section, we assume that the gradient flow equations (2.12)–(2.13) hold true, set $f = \text{grad}_H[-D_x \mathcal{E}(x)] + f_x^{ext}$, and denote the corresponding flux by $j := D_f \Psi^*(x, f_x)$.

The main idea of a TUR is to bound the relative precision $p_x(j)$ of a flux, defined as

$$p_x(j) := j^t [\text{Cov}_x(j)]^{-1} j = \sum_{r, r'=1}^{|E|} \frac{\partial^2 \Psi(x, j)}{\partial j_r \partial j_{r'}} j_r j_{r'} = \|j\|_{g_{x, j}^{\mathcal{J}}}^2, \quad (4.6)$$

through the dissipation rate $\sigma(\mathbf{x}, \mathbf{j})$. Here, using the notation from Section 3.2 we denote by $\|\cdot\|_{\mathbf{g}_{\mathbf{x}, \mathbf{j}}^{\mathcal{J}}}$ the norm associated to the inner product $\mathbf{g}_{\mathbf{x}, \mathbf{j}}^{\mathcal{J}}$ on $\mathcal{T}_{\mathbf{j}}\mathcal{J}_{\mathbf{x}}$ ⁵. By using the symmetrized Bregman divergence

$$\mathcal{D}_{\mathbf{x}}^{\mathcal{J}, \text{sym}}[\mathbf{j} \|\mathbf{j}'] := \frac{1}{2} [\mathcal{D}_{\mathbf{x}}^{\mathcal{J}}[\mathbf{j} \|\mathbf{j}'] + \mathcal{D}_{\mathbf{x}}^{\mathcal{J}}[\mathbf{j}' \|\mathbf{j}]] \quad (4.7)$$

one can represent the dissipation rate, following (2.14), via

$$\sigma(\mathbf{x}, \mathbf{j}) = \Psi(\mathbf{x}, \mathbf{j}) + \Psi^*(\mathbf{x}, \mathbf{f}) = 2\mathcal{D}_{\mathbf{x}}^{\mathcal{J}, \text{sym}}[\mathbf{j} \|\mathbf{0}]. \quad (4.8)$$

Then, using the integral representation of $\mathcal{D}_{\mathbf{x}}^{\mathcal{J}, \text{sym}}[\mathbf{j} \|\mathbf{j}']$ from [72], Theorem 4.2,⁶ one obtains the integral representation of the dissipation rate via

$$\sigma(\mathbf{x}, \mathbf{j}) = 2 \int_0^1 \|\mathbf{j}\|_{\mathbf{g}_{\mathbf{x}, \tau \mathbf{j}}^{\mathcal{J}}}^2 d\tau \quad (4.9)$$

With the definition (4.6), the formula (4.9) gives the following general result on TUR:

Lemma 4.1 *If $\|\mathbf{j}\|_{\mathbf{g}_{\mathbf{x}, \tau \mathbf{j}}^{\mathcal{J}}} \geq \|\mathbf{j}\|_{\mathbf{g}_{\mathbf{x}, \mathbf{j}}^{\mathcal{J}}}$ holds for all $\tau \in [0, 1]$, then the relative precision of the flux is bounded by the dissipation rate as*

$$\frac{1}{2}\sigma(\mathbf{x}, \mathbf{j}) \geq \mathfrak{p}_{\mathbf{x}}(\mathbf{j}). \quad (4.10)$$

In Section 4.2.3, we show how this implies generalizations of some known TURs. Note that the condition $\|\mathbf{j}\|_{\mathbf{g}_{\mathbf{x}, \tau \mathbf{j}}^{\mathcal{J}}} \geq \|\mathbf{j}\|_{\mathbf{g}_{\mathbf{x}, \mathbf{j}}^{\mathcal{J}}}$ for $\tau \in [0, 1]$ holds true for chemical reaction networks and Markov jump processes, as shown in Appendix A. The same argument also goes through more generally for cosh-type gradient flow systems [19]. We leave it as an open question to establish this condition under more general assumptions.

4.2.2 Speed limits

In addition to the TUR, the integral representation (4.9) allows to obtain a clean geometrization of thermodynamic speed limits. Speed limits relate the time T to transition between two points \mathbf{x}_0 and \mathbf{x}_T to the dissipation and have been extensively studied in the physics literature [39–41].

Let $(\mathbf{x}, \mathbf{j}_{\mathbf{x}}) : [0, T] \rightarrow \mathcal{J}$ be a curve which obeys the gradient flow equations (2.12)–(2.13). Recall from Section 2.3 that the total dissipation is given by $\Sigma[(\mathbf{x}, \mathbf{j}_{\mathbf{x}})(\cdot)] =$

⁵Note that this notation requires to treat \mathbf{j} as an element of the tangent space $\mathcal{T}_{\mathbf{j}}\mathcal{J}_{\mathbf{x}}$ by making the canonical identification $\mathcal{J}_{\mathbf{x}} \simeq \mathcal{T}_{\mathbf{j}}\mathcal{J}_{\mathbf{x}}$ of a vector space with its tangent space at a point via $e_r \mapsto \partial_r$. The same is required in the integral representation of the dissipation rate in formula (4.9).

⁶The representation is given by

$$\mathcal{D}_{\mathbf{x}}^{\mathcal{J}, \text{sym}}[\mathbf{j} \|\mathbf{j}'] = \int_0^1 \|\dot{\gamma}(\tau)\|_{\mathbf{g}_{\mathbf{x}, \gamma(\tau)}^{\mathcal{J}}}^2 d\tau,$$

for the curve $\gamma : [0, 1] \rightarrow \mathcal{J}_{\mathbf{x}}$ defined by $\gamma(\tau) = \tau \mathbf{j} + (1 - \tau) \mathbf{j}'$.

$\int_0^T \sigma(\mathbf{x}, \mathbf{j}_\mathbf{x}) dt$ which yields, via the formula (4.9):

$$\Sigma[(\mathbf{x}, \mathbf{j}_\mathbf{x})(\cdot)] = 2 \int_0^T \int_0^1 \|\mathbf{j}_\mathbf{x}\|_{\mathbf{g}_{\mathbf{x}, \tau \mathbf{j}_\mathbf{x}}^\mathcal{J}}^2 d\tau dT. \quad (4.11)$$

By using the Cauchy-Schwarz inequality one obtains

$$\Sigma[(\mathbf{x}, \mathbf{j}_\mathbf{x})(\cdot)] \geq \frac{2}{T} \left[\int_0^T \int_0^1 \|\mathbf{j}_\mathbf{x}\|_{\mathbf{g}_{\mathbf{x}, \tau \mathbf{j}_\mathbf{x}}^\mathcal{J}}^2 d\tau dT \right]^2. \quad (4.12)$$

Then, by defining the *thermodynamic area* $\mathcal{A}_{\mathbf{x}_0: \mathbf{x}_T}$ between two points \mathbf{x}_0 and \mathbf{x}_T via

$$\mathcal{A}_{\mathbf{x}_0: \mathbf{x}_T} := \min_{\substack{(\mathbf{x}, \mathbf{j}_\mathbf{x}): [0, T] \rightarrow \mathcal{J} \\ \mathbf{x}(0) = \mathbf{x}_0 \\ \mathbf{x}(T) = \mathbf{x}_T \\ \dot{\mathbf{x}} = -\text{div}_\mathbf{H} \mathbf{j}}} \int_0^T \int_0^1 \|\mathbf{j}_\mathbf{x}\|_{\mathbf{g}_{\mathbf{x}, \tau \mathbf{j}_\mathbf{x}}^\mathcal{J}}^2 d\tau dT, \quad (4.13)$$

we obtain the information geometric speed limit:

$$T \geq \frac{2\mathcal{A}_{\mathbf{x}_0: \mathbf{x}_T}^2}{\Sigma[(\mathbf{x}, \mathbf{j}_\mathbf{x})(\cdot)]}. \quad (4.14)$$

The speed limit relates the transition time between \mathbf{x}_0 and \mathbf{x}_T to the inverse dissipation via the thermodynamic area. This area term is mathematically similar to the concept of thermodynamic length [51, 73–76]. However, the area describes the dissipation in a genuine nonequilibrium situation whereas the length treats slowly driven equilibrium systems.

4.2.3 Infinitesimal decompositions

We end this subsection by combining the discussion of information geometric dissipation rate decompositions presented in Section 4.1 with TURs and speed limits which are best treated with the infinitesimal methods of Riemannian geometry presented in this subsection.

The relative precision of the flux vector \mathbf{j} splits into the relative precision of the velocity $\dot{\mathbf{x}} = -\text{div}_\mathbf{H} \mathbf{j}$ and the one of the cycle flux component $\mathbf{j}^c \in \mathcal{J}_\mathbf{x}^c(\mathbf{j})$ as follows: For any flux vector $\mathbf{j} \in \mathcal{J}_\mathbf{x}$, one verifies that the spaces $\mathcal{J}_\mathbf{x}^c(\mathbf{j})$ and $\mathcal{J}_\mathbf{x}^b(\mathbf{j})$ intersect $\mathbf{g}_{\mathbf{x}, \mathbf{j}}^\mathcal{J}$ -orthogonally at \mathbf{j} , i.e., that the tangent spaces $\mathcal{T}_\mathbf{j} \mathcal{J}_\mathbf{x}^c(\mathbf{j})$ and $\mathcal{T}_\mathbf{j} \mathcal{J}_\mathbf{x}^b(\mathbf{j})$ are orthogonal with respect to the inner product on $\mathcal{T}_\mathbf{j} \mathcal{J}_\mathbf{x}$ defined by the Riemannian metric $\mathbf{g}_{\mathbf{x}, \mathbf{j}}^\mathcal{J}$. Moreover, the tangent spaces $\mathcal{T}_\mathbf{j} \mathcal{J}_\mathbf{x}^b(\mathbf{j})$ and $\mathcal{T}_\mathbf{j} \mathcal{J}_\mathbf{x}^c(\mathbf{j})$ span the whole space $\mathcal{T}_\mathbf{j} \mathcal{J}_\mathbf{x}$. We write

$$\mathcal{T}_\mathbf{j} \mathcal{J}_\mathbf{x} = \mathcal{T}_\mathbf{j} \mathcal{J}_\mathbf{x}^b(\mathbf{j}) \oplus_{\mathbf{g}_{\mathbf{x}, \mathbf{j}}^\mathcal{J}} \mathcal{T}_\mathbf{j} \mathcal{J}_\mathbf{x}^c(\mathbf{j}) \quad (4.15)$$

for this decomposition. Denote the respective $\mathbf{g}_{\mathbf{x},j}^{\mathcal{J}}$ -orthogonal projections $\mathcal{T}_j\mathcal{J}_{\mathbf{x}} \rightarrow \mathcal{T}_j\mathcal{J}_{\mathbf{x}}$ onto the first factor by $\pi_{\mathbf{x}}^b$ and onto the second factor by $\pi_{\mathbf{x}}^c$, respectively. One verifies that the boundary component of the flux is the velocity, i.e.,

$$\pi_{\mathbf{x}}^b(j) = \dot{\mathbf{x}} \quad (4.16)$$

and that its orthogonal complement $j^c := \pi_{\mathbf{x}}^c(j)$ lies in $\ker[\operatorname{div}_{\mathbb{H}}]$ and is thus a cycle component. The orthogonal decomposition (4.15) leads to the decomposition of the relative precision as

$$\mathfrak{p}_{\mathbf{x}}(j) = \|j\|_{\mathbf{g}_{\mathbf{x},j}^{\mathcal{J}}}^2 = \underbrace{\|\pi_{\mathbf{x}}^c(j)\|_{\mathbf{g}_{\mathbf{x},j}^{\mathcal{J}}}^2}_{\mathfrak{p}_{\mathbf{x}}(j^c)} + \underbrace{\|\pi_{\mathbf{x}}^b(j)\|_{\mathbf{g}_{\mathbf{x},j}^{\mathcal{J}}}^2}_{\mathfrak{p}_{\mathbf{x}}(\dot{\mathbf{x}})} \quad (4.17)$$

where $\mathfrak{p}_{\mathbf{x}}(\dot{\mathbf{x}})$ is the relative precision of the velocity and $\mathfrak{p}_{\mathbf{x}}(j^c)$ is the relative precision of the cycle component j^c . Then, combining this with Lemma 4.1 (which requires the gradient flow system to satisfy the assumption of the lemma) yields the inequality $\frac{1}{2}\sigma(\mathbf{x}, j) \geq \mathfrak{p}_{\mathbf{x}}(\dot{\mathbf{x}})$ which sharpens and generalizes the TUR derived in [40], as well as the inequality $\frac{1}{2}\sigma(\mathbf{x}, j) \geq \mathfrak{p}_{\mathbf{x}}(j^c)$ which recovers the TUR from [77].

Finally, we remark that the decomposition (4.17) can be used in the integral formula (4.9) to obtain a Riemannian geometric variation of the dissipation rate decomposition (4.5) as

$$\sigma(\mathbf{x}, j) = \underbrace{2 \int_0^1 \|\pi_{\mathbf{x}}^c(j)\|_{\mathbf{g}_{\mathbf{x},\tau j}^{\mathcal{J}}}^2 d\tau}_{\sigma_{\text{RG}}^{\text{hk}}(\mathbf{x}, j)} + \underbrace{2 \int_0^1 \|\pi_{\mathbf{x}}^b(j)\|_{\mathbf{g}_{\mathbf{x},\tau j}^{\mathcal{J}}}^2 d\tau}_{\sigma_{\text{RG}}^{\text{ex}}(\mathbf{x}, j)} \quad (4.18)$$

As discussed in last paragraph of Section 4.1, the term $\sigma_{\text{RG}}^{\text{hk}}(\mathbf{x}, j)$ captures the dissipation due the relaxation dynamics whereas the term $\sigma_{\text{RG}}^{\text{ex}}(\mathbf{x}, j)$ is related to the nonequilibrium nature of the system due to the perturbation. One also can tighten the TURs above to obtain

$$\frac{1}{2}\sigma_{\text{RG}}^{\text{ex}}(\mathbf{x}, j) \geq \mathfrak{p}_{\mathbf{x}}(\dot{\mathbf{x}}) \quad (4.19)$$

$$\frac{1}{2}\sigma_{\text{RG}}^{\text{hk}}(\mathbf{x}, j) \geq \mathfrak{p}_{\mathbf{x}}(j^c). \quad (4.20)$$

Note that by summing both the inequalities (4.19) and (4.20) one recovers the original TUR (4.10). From the geometrical point of view, according to (4.15) this is the only possible orthogonal splitting of the TUR because any other decomposition - when applied to the integral expression (4.9) would necessarily introduce mixture terms.

5 Future perspective

In this article, we have shown how the gradient flow system formalism naturally can be transferred to an information geometrical setting: The Bregman divergence is tightly

connected to the dissipation, the Fisher metric to covariance, and the dual orthogonal foliations provide natural cycle-boundary decompositions and moduli spaces. We have then shown how to use the geometrical formulation to derive dissipation rate decompositions, TURs, and speed limits. It is worth stressing that these modern principles discovered in nonequilibrium physics are derived without any specific physical but only require the gradient flow structures. As such, the results presented here make such principles applicable to any system which can be described by a gradient flow evolution equation. In particular, it will be interesting to discuss such principles within the gradient flow formulation of machine learning [78].

This work should be thought of as first foray into the information geometry of gradient flow systems, and we hope to stimulate further developments. We briefly mention some possible directions:

We have extensively used the finite dimensionality of the state space and of the vector bundles on that space. The information geometry on infinite dimensional state spaces is well-developed [79], and gradient flow systems on infinite dimensional state space are well-known. Albeit the merging of these two formalisms is technically demanding, it will be interesting to develop this theory.

There are coordinate-free formulations of both gradient-flow systems [80] and of information geometry [81], and it will be rewarding to formulate the nonequilibrium principles in this setting. Connected to this is the study of dissipation potentials which are not strictly convex, as the corresponding information geometry for singular models has been recently developed [82].

Most importantly, we have not touched the bread and butter of gradient flow systems which are thermodynamically consistent and mathematically rigorous limit procedures (discrete to continuous, time-scale separations). Such limits should have nice geometrical counterparts, and should allow to study the nonequilibrium physical principles in such situations.

Acknowledgements. We thank Matthias Liero for fruitful discussions and for educating us on technical details as well as on the intuition of gradient flow systems.

Declarations

Funding. This research is supported by JST (JPMJCR2011) and JSPS (24K00542, 25H01365). The second author is financially supported by JST SPRING, Grant No. JPMJSP2108.

Conflict of interest. The authors have no conflict of interest to declare.

Code and data availability. Data sharing is not applicable to this article as no new data was created or analyzed.

Appendix A Chemical reaction networks as perturbed gradient flow systems

We use chemical reaction network (CRN) theory as an example to flesh out and illustrate the interplay between information geometry and gradient flow systems. Originally, the gradient flow formulation for CRNs with quadratic dissipation functions has been used in [13]. However, in [24] it has been discovered that it is more natural to use cosh-type dissipation as they provide a link to large deviations theory as presented in Section 3.1. For an extensive overview of the relation between the gradient flow formulation CRN theory and information geometry, we refer to [22] and references therein.

A.1 Chemical reaction networks as gradient flow systems

A reversible chemical reaction network (CRN) is a hypergraph with its vertices x_1, x_2, \dots, x_n corresponding to the chemical species and its edges r_1, r_2, \dots, r_m corresponding to reactions. The reactions $r_r, r = 1, \dots, m$ are commonly written as maps between formal linear combinations of chemicals

$$\sum_{i=1}^n t(r_r, x_i) x_i \mapsto \sum_{i=1}^n h(r_r, x_i) x_i \quad (\text{A1})$$

with the coefficients $h(r_r, x_i)$ and $t(r_r, x_i)$ determining the hypergraph structure in accordance with Section 2.2.

We now make the corresponding gradient flow system precise. The state space is the space of positive concentration vectors $\mathbf{x} = (x_1, x_2, \dots, x_n)^t \in \mathbf{X} := \mathbb{R}_{>0}^n$. The kinetics of CRNs is a classical subject, where the kinetics in the large volume limit for non-interacting particles is given by the law of mass action. The corresponding flux vector $\mathbf{j} = (j_1, j_2, \dots, j_m)^t \in \mathcal{J}_{\mathbf{x}}$ has the components

$$j_r = k_r^+ \prod_{i=1}^n x_i^{t(r_r, x_i)} - k_r^- \prod_{i=1}^n x_i^{h(r_r, x_i)} \quad (\text{A2})$$

for constant reaction rate vectors $\mathbf{k}^+, \mathbf{k}^- \in \mathbb{R}_{>0}^m$, and the concentration dynamics is given by the continuity equation $\dot{\mathbf{x}} = -\text{div}_{\mathbf{H}} \mathbf{j}$.

This evolution equation can be recovered by a perturbed gradient flow system as follows: The energy function is given by

$$\mathcal{E}(\mathbf{x}) = \sum_{i=1}^n x_i \log \frac{x_i}{x_i^*} - x_i + x_i^*, \quad (\text{A3})$$

where $\mathbf{x}^* \in \mathbf{X}$ is a fixed reference state which satisfies the detailed balancing condition

$$\text{grad}_{\mathbf{H}} \log \mathbf{x}^* = \log \frac{\boldsymbol{\kappa}^+}{\boldsymbol{\kappa}^-}. \quad (\text{A4})$$

Here, $\kappa^+, \kappa^- \in \mathbb{R}_{>0}^m$ are reaction rate vectors which are different from $\mathbf{k}^+, \mathbf{k}^-$ in general, and the quotients and logarithms of vectors are taken componentwise⁷. The dissipation function $\Psi^*(\mathbf{x}, \mathbf{f})$ has the form

$$\Psi^*(\mathbf{x}, \mathbf{f}) = 4 \sum_{r=1}^m \omega_r(\mathbf{x}) \left(\cosh \frac{f_r}{2} - 1 \right). \quad (\text{A5})$$

where we have used the \mathbf{x} -dependent frenetic activity

$$\omega_r(\mathbf{x}) = \sqrt{k_r^+ k_r^- \prod_{i=1}^n x_i^{t(r, \mathbf{x}_i) + h(r, \mathbf{x}_i)}} \quad (\text{A6})$$

for notational convenience. The gradient flow system is given by (2.12)-(2.13) with the external driving force

$$\mathbf{f}_{\mathbf{x}}^{\text{ext}} = \log \frac{\mathbf{k}^+}{\mathbf{k}^-} - \log \frac{\kappa^+}{\kappa^-}. \quad (\text{A7})$$

A direct calculation of \mathbf{j} via (2.13), i.e., $\mathbf{j} = \mathbf{D}_{\mathbf{f}} \Psi^*(\mathbf{x}, \text{grad}_{\mathbb{H}}[-\mathbf{D}_{\mathbf{x}} \mathcal{E}(\mathbf{x})] + \mathbf{f}_{\mathbf{x}}^{\text{ext}})$, yields the mass action kinetics (A2). Note that for $\mathbf{f}_{\mathbf{x}}^{\text{ext}} = \mathbf{0}$, one obtains mass action kinetics with the rate constant vectors κ^+, κ^- for the detailed balanced system and that the external driving force rescales these rate vectors by \mathbf{k}^+/κ^+ and \mathbf{k}^-/κ^- , respectively (the quotients of vectors are taken componentwise). Physically, this rescaling can be thought of as the result of the presence of external chemical reservoirs. The concentrations of the reservoirs do not directly appear in the rate equation in this case. This can be derived in the limit of infinite reservoir size by using the limiting procedure introduced in [12].

Remark A.1 The energy function \mathcal{E} of the gradient flow system given in (A3) is the generalized Kullback-Leibler divergence between \mathbf{x} and \mathbf{x}^* . It is strictly convex in \mathbf{x} and therefore can be used to do information geometry on the space \mathbf{X} . The dual coordinates are given by the Legendre transform $\mathbf{D}_{\mathbf{x}} \mathcal{E}(\cdot)$, the Riemannian metric by the Hessian of \mathcal{E} , etc. This geometry has been thoroughly developed in [49, 50], and has been applied to characterize growing systems [52], finite time driving [51], and chemical sensitivity [83]. However, we do not discuss this geometry in this article.

A.2 Proof of the assumption of Lemma 4.1

We show that the assumption

$$\|\mathbf{j}\|_{\mathbf{g}_{\mathbf{x}, \tau \mathbf{j}}^{\mathcal{J}}} = \sum_{r, r'=1}^m \frac{\partial^2 \Psi(\mathbf{x}, \tau \mathbf{j})}{\partial(\tau j_r) \partial(\tau j_{r'})} j_r j_{r'} \geq \sum_{r, r'=1}^m \frac{\partial^2 \Psi(\mathbf{x}, \mathbf{j})}{\partial j_r \partial j_{r'}} j_r j_{r'} = \|\mathbf{j}\|_{\mathbf{g}_{\mathbf{x}, \mathbf{j}}^{\mathcal{J}}} \quad (\text{A8})$$

⁷This condition can be equivalently written as $\kappa_r^+ \prod_{i=1}^n (x_i^*)^{t(r, \mathbf{x}_i)} = \kappa_r^- \prod_{i=1}^n (x_i^*)^{h(r, \mathbf{x}_i)}$.

for $\tau \in [0, 1]$ of Lemma 4.1 is satisfied for CRNs with mass action kinetics by a direct calculation. Following [22, 25], the dual dissipation potential $\Psi(\mathbf{x}, \mathbf{j})$ can be directly computed as

$$\Psi(\mathbf{x}, \mathbf{j}) = \sum_{r=1}^m \left[2j_r \operatorname{arcsinh} \frac{j_r}{2\omega_r(\mathbf{x})} - 2\sqrt{4\omega_r(\mathbf{x})^2 + j_r^2} + 4\omega_r(\mathbf{x}) \right]. \quad (\text{A9})$$

This gives the Legendre transform

$$f_r = \frac{\partial}{\partial j_r} \Psi(\mathbf{x}, \mathbf{j}) = 2 \operatorname{arcsinh} \frac{j_r}{2\omega_r(\mathbf{x})} \quad (\text{A10})$$

and the Riemannian metric $\mathbf{g}_{\mathbf{x}, \mathbf{j}}^{\mathcal{J}}$ via the Hessian

$$\frac{\partial^2 \Psi(\mathbf{x}, \mathbf{j})}{\partial j_r \partial j_{r'}} = \frac{2\delta_{rr'}}{\sqrt{j_r^2 + 4\omega_r(\mathbf{x})^2}}, \quad (\text{A11})$$

where $\delta_{rr'} = 1$ iff $r = r'$ and $\delta_{rr'} = 0$ otherwise. This leads to the desired relation

$$\|\mathbf{j}\|_{\mathbf{g}_{\mathbf{x}, \tau \mathbf{j}}^{\mathcal{J}}} = \sum_{r=1}^m \frac{2j_r^2}{\sqrt{(\tau j_r)^2 + 4\omega_r(\mathbf{x})^2}} \geq \sum_{r=1}^m \frac{2j_r^2}{\sqrt{j_r^2 + 4\omega_r(\mathbf{x})^2}} = \|\mathbf{j}\|_{\mathbf{g}_{\mathbf{x}, \mathbf{j}}^{\mathcal{J}}} \quad (\text{A12})$$

for all $\tau \in [0, 1]$.

References

- [1] Otto, F., Villani, C.: Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality. *Journal of Functional Analysis* **173**(2), 361–400 (2000)
- [2] Bakry, D., Gentil, I., Ledoux, M.: *Analysis and Geometry of Markov Diffusion Operators* vol. 348. Springer, Heidelberg, Germany (2013)
- [3] Mielke, A., Theil, F., Levitas, V.I.: A variational formulation of rate-independent phase transformations using an extremum principle. *Archive for rational mechanics and analysis* **162**(2), 137–177 (2002)
- [4] Maso, G.D., DeSimone, A., Mora, M.G.: Quasistatic evolution problems for linearly elastic–perfectly plastic materials. *Archive for rational mechanics and analysis* **180**(2), 237–291 (2006)
- [5] Ambrosio, L., Gigli, N., Savaré, G.: *Gradient Flows: in Metric Spaces and in the Space of Probability Measures*. Springer, Heidelberg, Germany (2005)
- [6] Mielke, A., Roubíček, T.: Rate-independent systems. *Applied Mathematical Sciences* **193** (2015)

- [7] Peletier, M.A., Rossi, R., Savaré, G., Tse, O.: Jump processes as generalized gradient flows. *Calculus of Variations and Partial Differential Equations* **61**(1), 33 (2022)
- [8] Peletier, M.A., Schlottke, M.C.: Gamma-convergence of a gradient-flow structure to a non-gradient-flow structure. *Calculus of Variations and Partial Differential Equations* **61**(3), 103 (2022)
- [9] Disser, K., Liero, M.: On gradient structures for markov chains and the passage to wasserstein gradient flows. *Networks Heterog. Media* **10**(2), 233–253 (2015)
- [10] Gladbach, P., Kopfer, E., Maas, J.: Scaling limits of discrete optimal transport. *SIAM Journal on Mathematical Analysis* **52**(3), 2759–2802 (2020)
- [11] Mielke, A., Peletier, M.A., Stephan, A.: Edp-convergence for nonlinear fast–slow reaction systems with detailed balance. *Nonlinearity* **34**(8), 5762 (2021)
- [12] Mielke, A.: Non-equilibrium steady states as saddle points and edp-convergence for slow-fast gradient systems. *Journal of Mathematical Physics* **64**(12) (2023)
- [13] Mielke, A.: A gradient structure for reaction–diffusion systems and for energy-drift-diffusion systems. *Nonlinearity* **24**(4), 1329 (2011)
- [14] Erbar, M.: A gradient flow approach to the boltzmann equation. *Journal of the European Mathematical Society* **26**(11), 4441–4490 (2023)
- [15] Garcke, H., Knopf, P.: Weak solutions of the cahn–hilliard system with dynamic boundary conditions: A gradient flow approach. *SIAM Journal on Mathematical Analysis* **52**(1), 340–369 (2020)
- [16] Glitzky, A., Liero, M.: A drift-diffusion based electrothermal model for organic thin-film devices including electrical and thermal environment. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik* **104**(3), 202300376 (2024)
- [17] Kirch, A., Fischer, A., Liero, M., Fuhrmann, J., Glitzky, A., Reineke, S.: Electrothermal tristability causes sudden burn-in phenomena in organic leds. *Advanced Functional Materials* **31**(47), 2106716 (2021)
- [18] Sawatzki, F.M., Doan, D.H., Kleemann, H., Liero, M., Glitzky, A., Koprucki, T., Leo, K.: Balance of horizontal and vertical charge transport in organic field-effect transistors. *Physical Review Applied* **10**(3), 034069 (2018)
- [19] Peletier, M.A., Schlichting, A.: Cosh gradient systems and tilting. *Nonlinear Analysis* **231**, 113094 (2023). *Variational Models for Discrete Systems*
- [20] Mielke, A.: Deriving effective models for multiscale systems via evolutionary γ -convergence. In: *Control of Self-organizing Nonlinear Systems*, pp. 235–251.

Springer, Heidelberg, Germany (2016)

- [21] Renger, D.M.: Gradient and generic systems in the space of fluxes, applied to reacting particle systems. *Entropy* **20**(8), 596 (2018)
- [22] Kobayashi, T.J., Loutchko, D., Kamimura, A., Horiguchi, S.A., Sughiyama, Y.: Information geometry of dynamics on graphs and hypergraphs. *Information Geometry* **7**(1), 97–166 (2024)
- [23] Adams, S., Dirr, N., Peletier, M., Zimmer, J.: Large deviations and gradient flows. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**(2005), 20120341 (2013)
- [24] Mielke, A., Peletier, M.A., Renger, D.M.: On the relation between gradient flows and the large-deviation principle, with applications to markov chains and diffusion. *Potential Analysis* **41**, 1293–1327 (2014)
- [25] Mielke, A., Patterson, R.I., Peletier, M.A., Michiel Renger, D.: Non-equilibrium thermodynamical principles for chemical reactions with mass-action kinetics. *SIAM Journal on Applied Mathematics* **77**(4), 1562–1585 (2017)
- [26] Renger, D.M.: Flux large deviations of independent and reacting particle systems, with implications for macroscopic fluctuation theory. *Journal of Statistical Physics* **172**(5), 1291–1326 (2018)
- [27] Peliti, L., Pigolotti, S.: *Stochastic Thermodynamics: An Introduction*
- [28] Seifert, U.: *Stochastic Thermodynamics*. Cambridge University Press, Cambridge, England (2025)
- [29] Falasco, G., Esposito, M.: Macroscopic stochastic thermodynamics. *Reviews of Modern Physics* **97**(1), 015002 (2025)
- [30] Seifert, U.: Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on progress in physics* **75**(12), 126001 (2012)
- [31] Korbel, J., Wolpert, D.H.: Stochastic thermodynamics and fluctuation theorems for non-linear systems. *New Journal of Physics* **23**(3), 033049 (2021)
- [32] Martins, T.T., Malavazi, A.H., Kamizaki, L.P., Petrosyan, A., Besga, B., Ciliberto, S., Muniz, S.R.: A brief introduction to fluctuation theorems: from theory to experiments. *arXiv preprint arXiv:2503.20894* (2025)
- [33] Barato, A.C., Seifert, U.: Thermodynamic uncertainty relation for biomolecular processes. *Physical review letters* **114**(15), 158101 (2015)
- [34] Horowitz, J.M., Gingrich, T.R.: Thermodynamic uncertainty relations constrain non-equilibrium fluctuations. *Nature Physics* **16**(1), 15–20 (2020)

- [35] Falasco, G., Esposito, M., Delvenne, J.-C.: Unifying thermodynamic uncertainty relations. *New Journal of Physics* **22**(5), 053046 (2020)
- [36] Koyuk, T., Seifert, U.: Thermodynamic uncertainty relation in interacting many-body systems. *Physical Review Letters* **129**(21), 210603 (2022)
- [37] Kamijima, T., Ito, S., Dechant, A., Sagawa, T.: Thermodynamic uncertainty relations for steady-state thermodynamics. *Physical Review E* **107**(5), 052101 (2023)
- [38] Ray, K.J., Boyd, A.B., Guarnieri, G., Crutchfield, J.P.: Thermodynamic uncertainty theorem. *Physical Review E* **108**(5), 054126 (2023)
- [39] Vo, V.T., Van Vu, T., Hasegawa, Y.: Unified approach to classical speed limit and thermodynamic uncertainty relation. *Physical Review E* **102**(6), 062132 (2020)
- [40] Yoshimura, K., Ito, S.: Thermodynamic uncertainty relation and thermodynamic speed limit in deterministic chemical reaction networks. *Physical review letters* **127**(16), 160601 (2021)
- [41] Nagayama, R., Yoshimura, K., Ito, S.: Infinite variety of thermodynamic speed limits with general activities. *Physical Review Research* **7**(1), 013307 (2025)
- [42] Dechant, A., Sasa, S.-I., Ito, S.: Geometric decomposition of entropy production in out-of-equilibrium systems. *Physical Review Research* **4**(1), 012034 (2022)
- [43] Kobayashi, T.J., Loutchko, D., Kamimura, A., Sughiyama, Y.: Hessian geometry of nonequilibrium chemical reaction networks and entropy production decompositions. *Physical Review Research* **4**(3), 033208 (2022)
- [44] Yoshimura, K., Kolchinsky, A., Dechant, A., Ito, S.: Housekeeping and excess entropy production for general nonlinear dynamics. *Physical Review Research* **5**(1), 013017 (2023)
- [45] Ito, S.: Stochastic thermodynamic interpretation of information geometry. *Physical review letters* **121**(3), 030605 (2018)
- [46] Ito, S., Oizumi, M., Amari, S.-i.: Unified framework for the entropy production and the stochastic interaction based on information geometry. *Physical Review Research* **2**(3), 033048 (2020)
- [47] Ito, S., Dechant, A.: Stochastic time evolution, information geometry, and the cramer-rao bound. *Physical Review X* **10**(2), 021056 (2020)
- [48] Yoshimura, K., Ito, S.: Information geometric inequalities of chemical thermodynamics. *Physical Review Research* **3**(1), 013175 (2021)
- [49] Sughiyama, Y., Loutchko, D., Kamimura, A., Kobayashi, T.J.: Hessian geometric

- structure of chemical thermodynamic systems with stoichiometric constraints. *Physical Review Research* **4**(3), 033065 (2022)
- [50] Kobayashi, T.J., Loutchko, D., Kamimura, A., Sughiyama, Y.: Kinetic derivation of the hessian geometric structure in chemical reaction networks. *Physical Review Research* **4**(3), 033066 (2022)
 - [51] Loutchko, D., Sughiyama, Y., Kobayashi, T.J.: Riemannian geometry of optimal driving and thermodynamic length and its application to chemical reaction networks. *Physical Review Research* **4**(4), 043049 (2022)
 - [52] Sughiyama, Y., Kamimura, A., Loutchko, D., Kobayashi, T.J.: Chemical thermodynamics for growing systems. *Physical Review Research* **4**(3), 033191 (2022)
 - [53] Ito, S.: Information geometry, trade-off relations, and generalized glansdorff–prigogine criterion for stability. *Journal of Physics A: Mathematical and Theoretical* **55**(5), 054001 (2022)
 - [54] Ito, S.: Geometric thermodynamics for the fokker–planck equation: stochastic thermodynamic links between information geometry and optimal transport. *Information geometry* **7**(Suppl 1), 441–483 (2024)
 - [55] Mizohata, T., Kobayashi, T.J., Bouchard, L.-S., Miyahara, H.: Information geometric bound on general chemical reaction networks. *Physical Review E* **109**(4), 044308 (2024)
 - [56] Kolchinsky, A., Dechant, A., Yoshimura, K., Ito, S.: Information geometry of excess and housekeeping entropy production. *arXiv preprint arXiv:2206.14599* (2022)
 - [57] Li, W.: Transport information geometry: Riemannian calculus on probability simplex. *Information Geometry* **5**(1), 161–207 (2022)
 - [58] Ay, N.: Information geometry of the otto metric. *Information geometry*, 1–24 (2024)
 - [59] Onsager, L.: Reciprocal relations in irreversible processes. i. *Physical review* **37**(4), 405 (1931)
 - [60] Mielke, A.: Energetic formulation of multiplicative elasto-plasticity using dissipation distances. *Continuum Mechanics and Thermodynamics* **15**(4), 351–382 (2003)
 - [61] Mielke, A.: Evolution of rate-independent systems. In: *Handbook of Differential Equations: Evolutionary Equations* vol. 2, pp. 461–559. Elsevier, Amsterdam, Netherlands (2005)
 - [62] Schnakenberg, J.: Network theory of microscopic and macroscopic behavior of

- master equation systems. *Reviews of Modern Physics* **48**(4), 571–585 (1976)
- [63] Van Vu, T., Saito, K.: Thermodynamic unification of optimal transport: Thermodynamic uncertainty relation, minimum dissipation, and thermodynamic speed limits. *Physical Review X* **13**(1), 011013 (2023)
 - [64] Loutchko, D., Sughiyama, Y., Kobayashi, T.J.: The geometry of thermodynamic uncertainty relations in chemical reaction networks. *arXiv preprint arXiv:2308.04806* (2023)
 - [65] Oono, Y., Paniconi, M.: Steady state thermodynamics. *Progress of Theoretical Physics Supplement* **130**, 29–44 (1998)
 - [66] Hatano, T., Sasa, S.-I.: Steady-state thermodynamics of langevin systems. *Physical review letters* **86**(16), 3463 (2001)
 - [67] Komatsu, T.S., Nakagawa, N., Sasa, S.-I., Tasaki, H.: Entropy and nonlinear nonequilibrium thermodynamic relation for heat conducting steady states. *Journal of Statistical Physics* **142**(1), 127–153 (2011)
 - [68] Maes, C., Netočný, K.: A nonequilibrium extension of the clausius heat theorem. *Journal of Statistical Physics* **154**(1), 188–203 (2014)
 - [69] Rao, R., Esposito, M.: Nonequilibrium thermodynamics of chemical reaction networks: Wisdom from stochastic thermodynamics. *Physical Review X* **6**(4), 041064 (2016)
 - [70] Renger, D.R.M., Zimmer, J.: Orthogonality of fluxes in general nonlinear reaction networks. *Discrete and Continuous Dynamical Systems - S* **14**(1), 205–217 (2021)
 - [71] Gao, Y., Liu, J.-G.: Revisit of macroscopic dynamics for some non-equilibrium chemical reactions from a hamiltonian viewpoint. *Journal of Statistical Physics* **189**(2), 22 (2022)
 - [72] Amari, S.-I.: *Information Geometry and Its Applications* vol. 194. Springer, ??? (2016)
 - [73] Salamon, P., Berry, R.S.: Thermodynamic length and dissipated availability. *Physical Review Letters* **51**(13), 1127 (1983)
 - [74] Crooks, G.E.: Measuring thermodynamic length. *Physical Review Letters* **99**(10), 100602 (2007)
 - [75] Sivak, D.A., Crooks, G.E.: Thermodynamic metrics and optimal paths. *Physical review letters* **108**(19), 190602 (2012)
 - [76] Mandal, D., Jarzynski, C.: Analysis of slow transitions between nonequilibrium steady states. *Journal of Statistical Mechanics: Theory and Experiment* **2016**(6),

063204 (2016)

- [77] Polettini, M., Falasco, G., Esposito, M.: Tight uncertainty relations for cycle currents. *Physical Review E* **106**(6), 064121 (2022)
- [78] Liero, M., Mielke, A., Tse, O., Zhu, J.-J.: Evolution of gaussians in the hellinger-kantorovich-boltzmann gradient flow. *arXiv preprint arXiv:2504.20400* (2025)
- [79] Ay, N., Jost, J., Vân Lê, H., Schwachhöfer, L.: *Information Geometry* vol. 64. Springer, Heidelberg, Germany (2017)
- [80] Mielke, A.: An eulerian formulation for dissipative materials using lie derivatives and generic. *arXiv preprint arXiv:2502.21157* (2025)
- [81] Amari, S.-I., Nagaoka, H.: *Methods of Information Geometry* vol. 191. American Mathematical Soc., Providence, Rhode Island (2000)
- [82] Nakajima, N., Ohmoto, T.: The dually flat structure for singular models. *Information Geometry* **4**(1), 31–64 (2021)
- [83] Loutchko, D., Sughiyama, Y., Kobayashi, T.J.: Information geometry of chemical reaction networks: Cramer-rao bound and absolute sensitivity revisited. *arXiv preprint arXiv:2503.19384* (2025)