RESEARCH PAPER

Information geometry of dynamics on graphs and hypergraphs

Tetsuya J. Kobayashi[1,2,3] (iD) · Dimitri Loutchko[1] · Atsushi Kamimura[1] ·
Shuhei A. Horiguchi[2] · Yuki Sughiyama[1]

Abstract

We introduce a new information-geometric structure associated with the dynamics on discrete objects such as graphs and hypergraphs. The presented setup consists of two dually flat structures built on the vertex and edge spaces, respectively. The former is the conventional duality between density and potential, e.g., the probability density and its logarithmic form induced by a convex thermodynamic function. The latter is the duality between flux and force induced by a convex and symmetric dissipation function, which drives the dynamics of the density. These two are connected topologically by the homological algebraic relation induced by the underlying discrete objects. The generalized gradient flow in this doubly dual flat structure is an extension of the gradient flows on Riemannian manifolds, which include Markov jump processes and nonlinear chemical reaction dynamics as well as the natural gradient. The information-geometric projections on this doubly dual flat structure lead to information-geometric extensions of the Helmholtz–Hodge decomposition and the Otto structure in L^2-Wasserstein geometry. The structure can be extended to non-gradient nonequilibrium flows, from which we also obtain the induced dually flat structure on cycle spaces. This abstract but general framework can broaden the applicability of information geometry to various problems of linear and nonlinear dynamics.

✉ Tetsuya J. Kobayashi
  tetsuya@mail.crmind.net

1   Institute of Industrial Science, The University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo
    153-8505, Japan

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

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

🙋 Springer

Contents

1 Introduction

Information geometry is finding and establishing a firm position as a geometric language in various scientific disciplines [1, 2]. Information geometry enables us to gain an intuitive understanding of the structures behind complicated problems of inference and estimation, for which Euclidean or Riemannian geometry may not be sufficient. In addition, it can provide ways to devise new solutions and approaches for the problems [1]. While information geometry was originally developed for statistics, its applicability now reaches far beyond statistical problems. Whenever the notions of probability, information, or positive density appear in a problem, it is natural to consider its information–geometric structure.

1.1 Information geometry of dynamics

Dynamical systems and phenomena can be naturally analyzed with information geometric methods, as conventionally one considers the dynamics of probability distributions [3–5], e.g., via the Fokker–Planck equations (FPE) and the Master equation, or those of positive densities, e.g, via population dynamics, epidemic models, diffusion dynamics on networks, and chemical reaction dynamics [6–8]. Although the application of information geometry to dynamical systems has been attempted almost since its birth, information geometry for dynamics is much less organized and principled compared with those for static problems in statistics, optimization, and others [1]. In connection with statistical inference, information geometry was employed by Amari and others to investigate Gaussian time series and autoregressive moving average (ARMA) models by representing their power spectrum as parametric manifolds [9–11]. This idea was also used to investigate linear systems [12]. Markov jump processes on finite states[1] were investigated information-geometrically by considering the hierarchical structure of joint or conditional probabilities at different time points, e.g., $\mathbb{P}_{\boldsymbol{\theta}}(x_1, x_2, \ldots, x_t)$ [13], or by introducing exponential families of Markov kernels (transition matrices), $\mathbb{T}_{\boldsymbol{\theta}}(x|x')$, via exponential tilting of the kernels [14–20]. Furthermore, information geometry was applied to studies of random walks, nonlinear diffusion equations of porous media, and networks [21–23]. In relation to mechanics, integrable systems were associated with the dualistic gradient flow of information geometry in the seminal works [24, 25], and other connections of information geometry with Lagrangian or Hamiltonian mechanics have been pursued [26–28].

1.2 Information measures for dynamics

Concurrently with and almost independently of these attempts within the community of information geometry, information measures relevant to information geometry have been employed in various problems of dynamical systems and stochastic processes in information theory [29], filtering theory [30, 31], control theory [32–34], and nonequilibrium physics and chemistry [35–37]. The Kullback–Leibler (KL) divergence

---

[1] Also known as Markov chains.

[38] for probabilities and positive densities was shown to be a Lyapunov function of Markov jump processes (MJP) [5], FPE [3, 39], deterministic chemical reaction networks (CRN) [40, 41], and other dynamical systems [42, 43], the origin of which can be dated back to Gibbs' H-theorem [44]. Among those topics, since the establishment of chemical thermodynamics by Gibbs [44] and chemical kinetics by Guldberg and Waage [45], CRN has played the role of a seedbed for cultivating the theory between dynamics and divergences owing to its close connection with thermodynamics [46–48]. More recently, it was also clarified that the divergences and information geometry are fundamental in stochastic thermodynamics [49–53].

In addition to the KL divergence, the Fisher-information-like quantity

$$\mathbb{I}_F[p] := \int p(\boldsymbol{r})(\nabla_{\boldsymbol{r}} \ln p(\boldsymbol{r}))^2 \mathrm{d}\boldsymbol{r} \in \mathbb{R}_{\geq 0} \tag{1}$$

was also revealed to play an important role in characterizing dynamics for densities on a continuous space, e.g., Gaussian convolution, diffusion processes, and FPE [54–56]. Various governing equations in physics were claimed to be derived in a unified way from this quantity [36]. The quantity $\mathbb{I}_F$ looks like the Fisher information [57] but is different from the conventional Fisher information matrix [58–60] because the derivative $\nabla_{\boldsymbol{r}} \ln p(\boldsymbol{r})$ is not for the parameters but for the base space variable of $p(\boldsymbol{r})$.[2] Because $\mathbb{I}_F$ is a scalar, we follow [59] and call it Fisher information number. The Fisher information number $\mathbb{I}_F$ is related to the KL divergence in additive Gaussian channels [54] and other systems [56, 62], which is known as the De Bruijn identity [54]. In addition, the logarithmic Sobolev inequality also provides a relation between the Fisher information number and the KL divergence (or Shannon information) [63, 64]. These results have recently been associated with the formal Riemannian geometric structure induced by the L^2-Wasserstein geometry [65, 66].

1.3 Information geometry and dynamics in machine learning

On top of these traditional trends, information geometry is now playing a pivotal role in machine learning for designing and evaluating online optimization algorithms (dynamics) in the space of model parameters such as natural gradient [67] and mirror descent [68, 69] as well as evolutionary computation (information-geometric optimization) [70]. Geometric interpretation allows us to understand the behaviors and efficiency of algorithms and their dynamics more intuitively in a principled manner [69–71].

1.4 Aim and contributions of this work

Despite the wide applicability and the long history of information geometry, we still lack a solid theoretical framework to unify these outcomes that spread across different

---

[2] The relation between the two forms of Fisher information has been explained in multiple ways. For example, they are related as the shift of the base space via parameters [36, 61]. The Fisher information number was introduced by Rao [58].
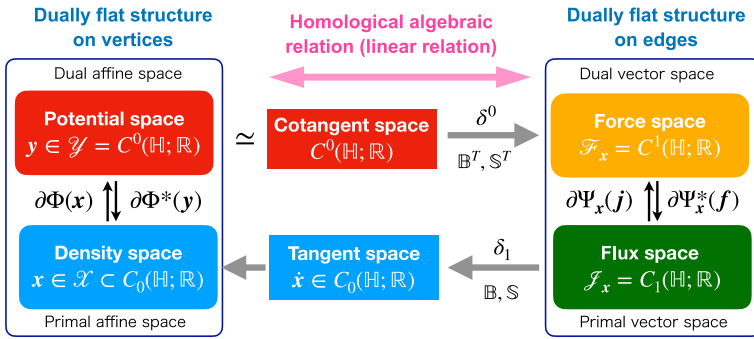
Fig. 1 Diagrammatic illustration of the doubly dual flat structure established on the vertex affine spaces (left) and the edge spaces (right), which are topologically related via the underlying graph or hypergraph

fields from the viewpoint of information geometry. In this work, we introduce a new information geometric structure for the dynamics of probability and positive densities. In this structure, we consider not only the single dually flat structure built on the space of densities as in [24, 25] but also another structure constructed on the space of fluxes. These two structures are linked algebraically and topologically via the continuity equation and the gradient equation as illustrated in Fig. 1.

Under this doubly dual flat structure, we can consider the dynamics of densities as a generalized flow, and various previous results can be unified in this framework. We exclusively consider dynamics of densities on finite-dimensional discrete manifolds, i.e., finite graphs or hypergraphs, because the structure introduced here can be explicitly manifested in this setup and also because we do not need the mathematically elaborated setup for infinite-dimensional information geometry on a smooth manifold [72]. For the case of FPE in a continuous state space, the dually flat structure built on the flux space can be reduced to the formal Riemannian geometric structure of L^2 Wasserstein geometry where the convex functions that induce the dually flat structure become quadratic. Our structure generalizes the linear inner product on the tangent and cotangent spaces with the nonlinear Legendre transform, thereby requiring information geometry. By elucidating this information geometric structure, we can easily see that some quantities such as the bilinear product, convex thermodynamic potential functions, the Fisher information matrix, and the Fisher information number are consolidated into one quantity for FPE with the quadratic convex functions (see Sect. 5.3 and Sect. 5.4). Therefore, our structure provides a way to unify the dualistic gradient flow mentioned in Sect. 1.1 and also the information-number related topics in Sect. 1.2.

From the viewpoint of homological algebra, the structure we work on is a modification of the chain and cochain complexes of graphs or hypergraphs, which replace the usual inner product duality [73] on each pair of chains and cochains with Legendre duality. Moreover, the dually flat space built on the flux space is linked to a finite-dimensional version of Orlicz spaces [74], which have been employed for constructing infinite-dimensional information geometry [72]. From the nice properties of the doubly dual flat structures, we can obtain information-geometric extensions of

the Helmholtz-Hodge-Kodaira (HHK) decomposition (Theorem 1), the Otto calculus (Theorem 2), and its induction to cycle spaces(Theorem 3).

Our construction of an information geometry for dynamics is heavily based on the idea of using Legendre duality for the force and flux relation, proposed in the recent work of large deviations theory and the macroscopic fluctuation theorem for MJP and CRN led by A.Mieleke, R.I.A.Petterson, M.A.Peletier, D.R.M. Renger, J.Zimmer, and others[3] [75–82]. We clarified its information-geometric aspects in the context of CRN and thermodynamics in our previous work [83]. We also concurrently elucidated the intimate link of equilibrium chemical thermodynamics and information geometry on the density state space [48, 84, 85]. In light of those, the contribution of this work is three-fold. First, we integrate these results in terms of information geometry, which clarifies the underlying geometric nature of the problem, provides transparent interpretations for known results, and leads to new information geometric results and insights (Theorem 1–Theorem 3); Second, this structure substantially extends the applicability of information geometry to a wide variety of dynamical problems; Lastly, the structure links information geometry to algebraic graph theory, discrete calculus, and homological algebra, which were not fully appreciated yet but provides a versatile way to consider the topology of the base manifold in information geometry.

1.5 Organization of this paper

This work is organized as follows: In Sect. 2, we introduce a range of models of dynamics on graphs and hypergraphs. In Sect. 3, we outline the homological algebra of graphs and hypergraphs. In Sect. 4, we abstractly introduce the doubly dual flat structures on the density and flux spaces and define the generalized flow associated with these structures. In Sect. 5, we clarify that the introduced structures include a wide class of dynamics on graphs and hypergraphs. In Sect. 6 and Sect. 7, we further define information-geometric objects and quantities, which naturally appear from this setup and play an integral role in the subsequent analysis of dynamics. In Sect. 8 and Sect. 9, we derive several results for equilibrium and nonequilibrium flows, respectively. Finally, we provide a summary and prospects of our work in Sect. 10. The notations and symbols are listed in the appendix.

2 Classes of models for density dynamics on graphs and hypergraphs

In this work, we focus on linear and nonlinear dynamics defined on graphs [86] and hypergraphs [87].

The linear dynamics of densities on graphs (LDG) includes Markov jump processes (MJP) [88], monomolecular chemical reaction networks [89], and others [86]. We consider an extension of LDG to hypergraphs and nonlinear dynamics, common instances of which are chemical reaction networks (CRN) with the law of mass action (LMA) kinetics [8] and polynomial dynamical systems (PDS) [90]. Because the extension we

---

[3] Ordered alphabetically.
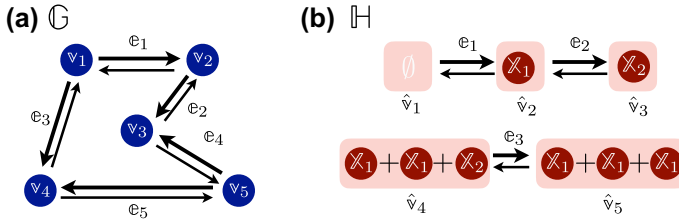
Fig. 2 Schematic diagrams of a reversible finite graph $\mathbb{G}$ (a) and a CRN-hypergraph $\mathbb{H}$ (b). Each pair of thick and thin arrows represents the pair of forward and reverse orientations of the corresponding edge. The CRN-hypergraph $\mathbb{H}$ in (b) corresponds to the simplified Brusselator reaction. The hypervertex $\hat{\mathbb{v}}_1$ contains no vertices

deal with in this work is a subclass of nonlinear dynamical systems on hypergraphs, we use CRN to designate this subclass.

In the following subsections, LDG and CRN are introduced using the language of algebraic graph theory [86, 91]. Then, we also give a brief and formal introduction of the Fokker-Planck equation (FPE) [3], a linear dynamics of probability densities defined in Euclidean space. We use the FPE throughout this paper only to contrast our results with the previous ones obtained for the FPE.

2.1 Reversible linear dynamics of densities on graphs

Definition 1 (*Edge-weighted finite graph* $\mathbb{G}_{k^{\pm}}$) A finite graph $\mathbb{G} := (\{\mathbb{v}_i\}, \{\mathbb{e}_e\}, \mathbb{B})$ consists of $N_{\mathbb{v}} \in \mathbb{Z}_{>0}$ vertices, $\{\mathbb{v}_i\}_{i \in [1, N_{\mathbb{v}}]}$, and $N_{\mathbb{e}} \in \mathbb{Z}_{>0}$ oriented edges, $\{\mathbb{e}_e\}_{e \in [1, N_{\mathbb{e}}]}$, each of which connects two different vertices[4] (Fig. 2a). The incidence relation is represented by the incidence matrix $\mathbb{B} \in \{0, \pm 1\}^{N_{\mathbb{v}} \times N_{\mathbb{e}}}$ where, for $\mathbb{B} = (b_{i,e})$,

$$
\begin{aligned}
b_{i,e} &:= +1 && \text{if } \mathbb{v}_i \text{ is the tail of edge } \mathbb{e}_e, \\
b_{i,e} &:= -1 && \text{if } \mathbb{v}_i \text{ is the head of edge } \mathbb{e}_e, \\
b_{i,e} &:= 0 && \text{otherwise.}
\end{aligned}
$$

An edge-weighted finite graph $\mathbb{G}_{k^{\pm}} := (\{\mathbb{v}_i\}, \{\mathbb{e}_e\}, \mathbb{B}, \{k_e^{\pm}\})$ has two positive weighting parameters $k_e^{\pm} = (k_e^+, k_e^-) \in \mathbb{R}_{>0}$ for each edge $\mathbb{e}_e$. The parameters k_e^+ and k_e^- are denoted as forward and reverse rates or weights of edge $\mathbb{e}_e$, respectively.

A reversible linear dynamics (rLDG) on a graphs is defined on the edge-weighted finite graph $\mathbb{G}_{k^{\pm}}$:

Definition 2 (*Reversible linear dynamics of density on graph* $\mathbb{G}_{k^{\pm}}$) The reversible linear dynamics of non-negative density $\boldsymbol{x}(t) = (x_1(t), \cdots, x_{N_{\mathbb{v}}}(t))^T \in \mathbb{R}_{\geq 0}^{N_{\mathbb{v}}}$ on $\mathbb{G}_{k^{\pm}}$ is defined by the continuity equation

$$
\dot{\boldsymbol{x}} = -\mathbb{B}\boldsymbol{j}(\boldsymbol{x}) = -\mathbb{B}[\boldsymbol{j}^+(\boldsymbol{x}) - \boldsymbol{j}^-(\boldsymbol{x})], \tag{2}
$$

---

[4] This means that we exclude self loops.

and linear forward and reverse one-way fluxes $\boldsymbol{j}^{\pm}(\boldsymbol{x}) = (j_1^{\pm}(\boldsymbol{x}), \cdots, j_{N_e}^{\pm}(\boldsymbol{x}))^T \in \mathbb{R}_{\geq 0}^{N_e}$ with the following specific functional form[5]:

$$j^{\pm}(x) = k^{\pm} \circ (\mathbb{B}^{\pm})^T x, \qquad (3)$$

where $\boldsymbol{j}(\boldsymbol{x}) := \boldsymbol{j}^+(\boldsymbol{x}) - \boldsymbol{j}^-(\boldsymbol{x})$ is the total flux, the symbol $\circ$ denotes the component-wise product of two vectors,[6] and $\mathbb{B}^+$ and $\mathbb{B}^-$ are the head and tail incidence matrices defined respectively as $\mathbb{B}^+ := \max[\mathbb{B}, 0]$ and $\mathbb{B}^- := \max[-\mathbb{B}, 0]$. The incidence matrix $\mathbb{B}$ in Eq. 2 is often regarded as the discrete divergence operator on a graph [73] and denoted also by $\mathrm{div}_{\mathbb{B}} = \mathbb{B}$ to emphasize this interpretation in this work.[7]

Reversible Markov jump processes (rMJP) are a representative class of the rLDG describing random jumps of noninteracting particles on $\mathbb{G}_{\boldsymbol{k}^{\pm}}$.[8] The weighting parameter k_e^+ is interpreted as the forward jump rate from the tail of the oriented edge $\mathfrak{e}_e$ to its head, whereas k_e^- is the reverse jump rate from the head to the tail of $\mathfrak{e}_e$.[9] For infinitely many such particles, we consider $p_i(t) \in [0, 1]$, the fraction of particles on vertex v_i at time t, which is a non-negative density on vertices. Then, the forward and reverse one-way fluxes on the eth edge defined by Eq. 3 are represented as

$$j_e^+(\boldsymbol{p}) = k_e^+ p_{\mathsf{v}_e^+} \in \mathbb{R}_{\geq 0}, \qquad j_e^-(\boldsymbol{p}) = k_e^- p_{\mathsf{v}_e^-} \in \mathbb{R}_{\geq 0}, \qquad (4)$$

where v_e^+ and v_e^- are the head and tail vertices of edge $\mathfrak{e}_e$[10]. The linearity of $j_e^{\pm}(\boldsymbol{p})$ with respect to $\boldsymbol{p}$ comes from the independence of particles on the graph. Then, the continuity equation (Eq. 2) with the state vector $\boldsymbol{p}(t) := (p_1(t), \cdots, p_{N_{\mathsf{v}}}(t))^T \in \mathbb{R}_{\geq 0}^{N_{\mathsf{v}}}$ is reduced to the master equation: $\dot{\boldsymbol{p}} = -\mathbb{B}\boldsymbol{j}(\boldsymbol{p})$.

Definition 3 (*Weighted asymmetric graph Laplacian* [91, 92]) For $\mathbb{G}_{\boldsymbol{k}^{\pm}}$, the corresponding weighted asymmetric graph Laplacian is defined by

$$\mathcal{L}_{\boldsymbol{\theta}} := \mathbb{B} \left[\mathrm{diag}[\boldsymbol{k}^+](\mathbb{B}^+)^T - \mathrm{diag}[\boldsymbol{k}^-](\mathbb{B}^-)^T \right], \qquad (5)$$

---

[5] We may consider other functional forms for $\boldsymbol{j}(\boldsymbol{x})$, which can induce nonlinear dynamics on the graph. In this work, we focus mainly on the linear case.

[6] Also known as the Hadamard product or Schur product of vectors.

[7] This interpretation is because Eq. 2 is associated with the continuity equation on a Euclidean space or on a Riemannian manifold where we have the divergence operator $\nabla\cdot$ instead of $\mathbb{B}$. However, divergence on a Riemannian manifold implicitly includes the information of the metric via the Hodge operator. On the contrary, $\mathbb{B}$ does not. From the viewpoint of homological algebra, $\mathbb{B}$ should be regarded as the adjoint (transpose) of the discrete exterior derivative operator $\delta^0 := \mathbb{B}^T$, which is also often called a discrete gradient operator [73]. For a Euclidean space, they are the same.

[8] We use the word 'reversible' in this work to mean that each edge allows both forward and reverse jumps while reversible Markov jump processes sometimes mean that the detailed balance condition is satisfied. We introduce the notion of equilibrium later to designate the detailed balanced situation.

[9] If we allow $k_e^{\pm}$ to be 0, we can include the irreversible MJP and also LDG in this formulation. We leave this extension for future work because it should require additional assumptions on the Legendre duality introduced in the subsequent sections.

[10] Here, we have abused the notation v_e^+ to indicate the index of the vertex v_e^+. Equation 3 is reduced to Eq. 4 because $b_{i,e}^{\pm} = +1$ only when i is the index of the tail (head) vertex $\mathsf{v}_e^{\pm}$ of $\mathfrak{e}_e$ and 0 otherwise.

where $\boldsymbol{\theta} := (\boldsymbol{k}^+, \boldsymbol{k}^-)$ and $\mathrm{diag}[\boldsymbol{k}^+]$ is the diagonal matrix whose diagonal elements are $\boldsymbol{k}^+$. Using $\mathcal{L}_{\boldsymbol{\theta}}$, Eq. 2 and Eq. 3 are represented as

$$\dot{\boldsymbol{x}} = -\mathbb{B}\boldsymbol{j}(\boldsymbol{x}) = -\mathbb{B}\left[\mathrm{diag}[\boldsymbol{k}^+](\mathbb{B}^+)^T - \mathrm{diag}[\boldsymbol{k}^-](\mathbb{B}^-)^T\right]\boldsymbol{x} = -\mathcal{L}_{\boldsymbol{\theta}}\boldsymbol{x}. \qquad (6)$$

The operator $\mathcal{L}_{\boldsymbol{\theta}}$ is reduced to the weighted symmetric graph Laplacian if $\boldsymbol{k}^+ = \boldsymbol{k}^-$ and also to the conventional graph Laplacian if $\boldsymbol{k}^+ = \boldsymbol{k}^- = \boldsymbol{1}$ [91, 92]. Equation 6 can also cover linear transport on graphs, a class of linear electric circuits [93], consensus dynamics on graphs [94], and other linear dynamics on graphs [86, 95].[11]

2.2 Chemical reaction network and polynomial dynamical systems on hypergraphs

Next, we introduce a class of nonlinear dynamics on hypergraphs, which includes the rLDG (Eq. 2 and Eq. 3) as a special case. The most common instance is deterministic chemical reaction networks (CRN) with the law of mass action (LMA) kinetics [7, 8, 45, 96], and this class is sometimes referred to as polynomial dynamical systems (PDS). Because the major part of the PDS theory has been developed for CRN, we use CRN to introduce and specify this class in this work.

Definition 4 (*Reversible edge-weighted CRN hypergraph* $\mathbb{H}_{\boldsymbol{k}^\pm}$) The reversible CRN hypergraph $\mathbb{H} := (\{\mathbb{X}_i\}, \{\mathbb{e}_e\}, \Gamma, \mathbb{B})$ consists of a finite number of vertices $\{\mathbb{X}_i\}_{i \in [1, N_{\mathbb{X}}]}$ and hyperedges $\{\mathbb{e}_e\}_{e \in [1, N_{\mathbb{e}}]}$ where $N_{\mathbb{X}}, N_{\mathbb{e}} \in \mathbb{Z}_{>0}$ (Fig. 2b). Each hyperedge $\mathbb{e}_e$ connects two different hypervertices $\hat{\mathbb{v}}_e^+$ and $\hat{\mathbb{v}}_e^-$ where $\hat{\mathbb{v}}_e^+ \neq \hat{\mathbb{v}}_e^-$.[12] The hypervertices are multisets of vertices $\{\mathbb{X}_i\}_{i \in [1, N_{\mathbb{X}}]}$, each of which is defined as $\hat{\mathbb{v}}_\ell = \sum_{i=1}^{N_{\mathbb{X}}} \gamma_{i,\ell} \mathbb{X}_i$ where $\gamma_{i,\ell} \in \mathbb{Z}_{\geq 0}$ is the number of the ith vertex included in the ℓth hypervertex.[13] Thus, the nonnegative integer vector $\boldsymbol{\gamma}_\ell := (\gamma_{1,\ell}, \cdots, \gamma_{N_{\mathbb{X}},\ell})^T \in \mathbb{Z}_{\geq 0}^{N_{\mathbb{X}}}$ defines the ℓth hypervertex. Let $N_{\hat{\mathbb{v}}} \in \mathbb{Z}_{>0}$ be the total number of the hypervertices and $\Gamma := (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_{N_{\hat{\mathbb{v}}}}) \in \mathbb{Z}_{\geq 0}^{N_{\mathbb{X}} \times N_{\hat{\mathbb{v}}}}$ be the hypervertex matrix. The matrix $\mathbb{B} \in \{0, \pm 1\}^{N_{\hat{\mathbb{v}}} \times N_{\mathbb{e}}}$ is the incidence matrix encoding the incidence relations among the hypervertices and the hyperedges. The hypergraph incidence matrix $\mathbb{S} \in \mathbb{Z}^{N_{\mathbb{X}} \times N_{\mathbb{e}}}$ is then defined as

$$\mathbb{S} := \Gamma\mathbb{B}. \qquad (7)$$

If $\Gamma = I$ where I is the identity matrix, then $\mathbb{H} = (\{\mathbb{X}_i\}_{i \in [1, N_{\mathbb{X}}]}, \{\mathbb{e}_e\}_{e \in [1, N_{\mathbb{e}}]}, \Gamma, \mathbb{B})$ is reduced to $\mathbb{G} = (\{\mathbb{v}_\ell\}_{\ell \in [1, N_{\mathbb{X}}]}, \{\mathbb{e}_e\}_{e \in [1, N_{\mathbb{e}}]}, \mathbb{B})$ where $\mathbb{v}_\ell = \mathbb{X}_\ell$. An edge-weighted CRN hypergraph $\mathbb{H}_{\boldsymbol{k}^\pm} := (\{\mathbb{X}_i\}, \{\mathbb{e}_e\}, \Gamma, \mathbb{B}, \{k_e^\pm\})$ has forward and reverse rates $k_e^\pm > 0$ as the weights of edge $\mathbb{e}_e$.

---

[11] For some of these applications, the relevant state space is $\mathbb{R}^{N_{\hat{\mathbb{v}}}}$ instead of $\mathbb{R}_{\geq 0}^{N_{\hat{\mathbb{v}}}}$.

[12] This means that we exclude self loop hyperedges. However, the head and tail hypervertices are allowed to contain the same vertices as long as $\hat{\mathbb{v}}_e^+ \neq \hat{\mathbb{v}}_e^-$ holds.

[13] Our definition of CRN hypergraph differs in a couple of aspects from the conventional definition because of the additional information required to define CRN. For example, while the definition of edges is usually extended from those of graphs [87], our definition extends vertices instead.

In the context of CRN theory, the vertices $\{\mathbb{X}_i\}$ correspond to the molecular species involved in a CRN, and each hyperedge $\mathfrak{e}_e$ represents a pair of forward and reverse reactions:

$$\gamma_{1,e}^+ \mathbb{X}_1 + \cdots + \gamma_{N_\mathbb{X},e}^+ \mathbb{X}_{N_\mathbb{X}} \rightleftharpoons \gamma_{1,e}^- \mathbb{X}_1 + \cdots + \gamma_{N_\mathbb{X},e}^- \mathbb{X}_{N_\mathbb{X}}, \tag{8}$$

where the forward and reverse reactions are from left to right and from right to left, respectively. Head and tail hypervertices $\hat{v}_e^+ := (\gamma_{1,e}^+ \mathbb{X}_1 + \cdots + \gamma_{N_\mathbb{X},e}^+ \mathbb{X}_{N_\mathbb{X}})$ and $\hat{v}_e^- := (\gamma_{1,e}^- \mathbb{X}_1 + \cdots + \gamma_{N_\mathbb{X},e}^- \mathbb{X}_{N_\mathbb{X}})$ in Eq. 8 are the sets of reactants and products of the eth forward reaction, respectively. More specifically, $\gamma_{i,e}^+ \in \mathbb{Z}_{\geq 0}$ and $\gamma_{i,e}^- \in \mathbb{Z}_{\geq 0}$ are the numbers of the molecule $\mathbb{X}_i$ involved as the reactants and products of the eth forward reaction, respectively. For the reverse reaction, $\hat{v}_e^-$ and $\hat{v}_e^+$ are the reactants and products. Some head and tail hypervertices are overlapping among different reactions (hyperedges) as in Fig. 2b. As a result, $\{\hat{v}_\ell\}_{\ell \in N_{\hat{v}}}$ is the union of the head and tail hypervertices, $\{\hat{v}_\ell\}_{\ell \in N_{\hat{v}}} = \bigcup_{e \in N_\mathfrak{e}} \{\hat{v}_e^+, \hat{v}_e^-\}$.

The hypervertices are called complexes in CRN theory [8][14]. From $\{\gamma_{i,e}^+\}$ and $\{\gamma_{i,e}^-\}$, we can define

$$s_e := (\gamma_{1,e}^+ - \gamma_{1,e}^-, \cdots, \gamma_{N_\mathbb{X},e}^+ - \gamma_{N_\mathbb{X},e}^-)^T \in \mathbb{Z}^{N_\mathbb{X}}, \tag{9}$$

where $\mp s_e$ specify the change in the number of molecules induced when the eth forward and reverse reaction occurs just once, respectively. The hypergraph incidence matrix $\mathbb{S}$ defined in Eq. 7 is represented as $\mathbb{S} = (s_1, \cdots, s_{N_\mathfrak{e}}) \in \mathbb{Z}^{N_\mathbb{X} \times N_\mathfrak{e}}$. In chemistry, the negative of s_e and $\mathbb{S}$, i.e., $-s_e$ and $-\mathbb{S}$, are called the stoichiometric vector and matrix, respectively [8].

Remark 1 To define a reversible CRN hypergraph, the hypergraph matrix $\mathbb{S}$ is not sufficient. If the head and tail hypervertices of a hyperedge contain the same vertex (molecule), the corresponding element in $\mathbb{S}$ of such a shared vertex becomes 0 by canceling out. Thus, the existence of shared vertices (molecules) is invisible in $\mathbb{S}$, and the pair $(\Gamma, \mathbb{B})$ is required to define $\mathbb{H}$. Such shared molecules are called catalysts in CRN.

For a CRN hypergraph, the continuity equation for CRN is defined:

Definition 5 (*CRN continuity equation*) Let a vector of nonnegative densities $x = (x_1, \cdots, x_{N_\mathbb{X}})^T \in \mathbb{R}_{\geq 0}^{N_\mathbb{X}}$ represents the concentration of molecules $\{\mathbb{X}_i\}$. The CRN continuity equation is defined as

$$\dot{x} = -\mathbb{S} j(x) = -\text{div}_\mathbb{S} j(x), \tag{10}$$

where $j_e^+(x) \in \mathbb{R}_{\geq 0}$ and $j_e^-(x) \in \mathbb{R}_{\geq 0}$ are the one-way fluxes of the eth forward and reverse reactions, $j^\pm(x) := (j_1^\pm(x), \cdots, j_{N_\mathfrak{e}}^\pm(x))^T \in \mathbb{R}_{\geq 0}^{N_\mathfrak{e}}$ are their vector representations, and $j(x) := j^+(x) - j^-(x) \in \mathbb{R}^{N_\mathfrak{e}}$ is the total reaction flux [7, 8, 96]. The hypergraph divergence operator $\text{div}_\mathbb{S} := \mathbb{S}$ is defined accordingly.

---

[14] Because we reserve the term *complex* for the cell complex in homological algebra, we use hypervertices to indicate the complexes of the CRN theory.

To define the dynamics of a CRN, the functional form of $j_e^{\pm}(x)$ is required.[15] Before introducing specific forms, we define two important properties of the fluxes and also other functions defined on edges:

Definition 6 (*Consistency of fluxes $j^{\pm}(x)$ with hypergraph $\mathbb{H}$*) One-way fluxes $j^{\pm}(x)$ are consistent with the hypergraph $\mathbb{H}$ if, for all $e \in [1, N_{\mathbb{e}}]$, $j_e^{\pm}(x)$ becomes 0 when $x_i = 0$ where $\mathbb{X}_i$ is any reactant of $j_e^{\pm}(x)$, respectively. In other words, $j_e^{\pm}(x)$ satisfies $\gamma_{i,e}^{\pm} j_e^{\pm}(x) = 0$ if $x_i = 0$ for any $i \in [1, N_{\mathbb{X}}]$.

Definition 7 (*Locality of function on edges over $\mathbb{H}$*) A vector function $g(x) \in \mathbb{R}^{N_{\mathbb{e}}}$ defined on edges is local on $\mathbb{H}$ if, for all $e \in [1, N_{\mathbb{e}}]$, $g_e(x)$ is a function only of the elements of x incident to the edge $\mathbb{e}_e$ on $\mathbb{H}$, i.e., $g_e(x) = g_e(\bar{\gamma}_{1,e}^+ x_1, \cdots, \bar{\gamma}_{N_{\mathbb{X}},e}^+ x_{N_{\mathbb{X}}}, \bar{\gamma}_{1,e}^- x_1, \cdots, \bar{\gamma}_{N_{\mathbb{X}},e}^- x_{N_{\mathbb{X}}})$ where $\bar{\gamma}_{i,e}^{\pm} := \min[1, \gamma_{i,e}^{\pm}] \in \{0, 1\}$.

The consistency condition is indispensable to prohibit a reaction that can decrease x_i from occurring when $x_i = 0$. For $j^{\pm}(x)$, the locality means that the fluxes of the eth reaction depend only on the concentrations of their reactants and products. The local flux is determined solely by the information stored on the vertices incident to the edge and plays a crucial role when we regard the structure introduced in this work as an extension of differential forms on continuous manifolds to graphs and hypergraphs. When we work on specific forms of fluxes in this work, we consider only local fluxes consistent with the given hypergraph $\mathbb{H}$.

In chemistry, we have a variety of candidates for the functional form of flux, e.g., the Michaelis-Menten function, Hill's function, and others [7, 97]. Among others, the LMA kinetics is the most basic and well-established one.

Definition 8 (*Waage–Guldberg's law of mass action kinetics (LMA kinetics)*) A CRN follows the LMA kinetics if, for all $e \in [1, N_{\mathbb{e}}]$, the eth forward and reverse reaction fluxes are represented as

$$j_e^{\pm}(x) = k_e^{\pm} \prod_{j=1}^{N_{\mathbb{X}}} x_j^{\gamma_{j,e}^{\pm}} = k_e^{\pm} \sum_{\ell=1}^{N_{\hat{\mathbb{v}}}} b_{\ell,e}^{\pm} \prod_{j=1}^{N_{\mathbb{X}}} x_j^{\gamma_{j,\ell}}, \tag{11}$$

where $k_e^+ \in \mathbb{R}_{>0}$ and $k_e^- \in \mathbb{R}_{>0}$ are the reaction rate constants of the eth forward and reverse reactions, respectively. The fluxes under LMA kinetics can be compactly represented as

$$j_{\mathrm{MA}}^{\pm}(x) = k^{\pm} \circ (\mathbb{B}^{\pm})^T x^{\Gamma^T}, \tag{12}$$

where $x^{\gamma} := \prod_{j=1}^{N_{\mathbb{X}}} x_j^{\gamma_j} \in \mathbb{R}_{\geq 0}$ and $x^{\Gamma^T} := (x^{\gamma_1}, \cdots, x^{\gamma_{N_{\hat{\mathbb{v}}}}})^T \in \mathbb{R}_{\geq 0}^{N_{\hat{\mathbb{v}}}}$.[16] We use the subscript MA as in $j_{\mathrm{MA}}^{\pm}(x)$ to discriminate this specific form of the fluxes from

---

[15] Even though the functional form of $j_e^{\pm}(x)$ is automatically determined in the case of LDG because of the linearity, we have multiple possibilities to define nonlinear $j_e^{\pm}(x)$.

[16] We should note an important relation, $(\mathbb{B}^{\pm})^T x^{\Gamma^T} = x^{(\Gamma \mathbb{B}^{\pm})^T}$, which holds because every column vector of $\mathbb{B}^{\pm}$ contains only one $+1$ and the others are 0.

others. We can easily observe that $\boldsymbol{j}_{\mathrm{MA}}^{\pm}(\boldsymbol{x})$ is consistent and local with respect to $\mathbb{H}$. Furthermore, $\boldsymbol{j}_{\mathrm{MA}}^{\pm}(\boldsymbol{x})$ is specified by the edge-weighted CRN hypergraph $\mathbb{H}_{k^{\pm}} :=$ $(\{\mathbb{X}_i\}, \{\mathbb{e}_e\}, \mathbb{\Gamma}, \mathbb{B}, \{k_e^{\pm}\})$.

Remark 2 (Algebraic aspect of LMA kinetics) Because $\boldsymbol{x}^{\mathbb{\Gamma}^T}$ is a vector of monomials of $\boldsymbol{x}$, each one-way flux, $j_e^{\pm}(\boldsymbol{x})$, is a monomial of $\boldsymbol{x}$ under Eq. 12 and thus the total flux $j_e(\boldsymbol{x}) = j_e^+(\boldsymbol{x}) - j_e^-(\boldsymbol{x})$ is a binomial. This fact links the real algebraic geometry of toric varieties [98, 99] to CRN [84, 100] as it does in algebraic statistics [48, 101].

Remark 3 (Extended LMA kinetics) While we mainly work on the normal LMA kinetics, we can extend it. The extended LMA kinetics defined on $\mathbb{H}$ is defined as

$$\boldsymbol{j}_{\mathrm{eMA}}^{\pm}(\boldsymbol{x}) = \boldsymbol{k}^{\pm} \circ \boldsymbol{g}(\boldsymbol{x}) \circ (\mathbb{B}^{\pm})^T \boldsymbol{x}^{\mathbb{\Gamma}^T}, \tag{13}$$

where $\boldsymbol{g}(\boldsymbol{x}) \in \mathbb{R}_{>0}^{N_{\mathbb{e}}}$ and is local with respect to $\mathbb{H}$.[17] An example of the extended LMA kinetics is reversible Michaelis-Menten kinetics [103].

By combining the continuity equation (Eq. 10) and the LMA kinetics (Eq. 12), we have the following chemical rate equation:

$$\dot{\boldsymbol{x}} = -\mathbb{S}\boldsymbol{j}_{\mathrm{MA}}(\boldsymbol{x}) = -\mathbb{\Gamma}\mathbb{B}\left[\mathrm{diag}[\boldsymbol{k}^+](\mathbb{B}^+)^T - \mathrm{diag}[\boldsymbol{k}^-](\mathbb{B}^-)^T\right]\boldsymbol{x}^{\mathbb{\Gamma}^T}$$
$$= -\mathbb{\Gamma}\mathcal{L}_{\boldsymbol{\theta}}\boldsymbol{x}^{\mathbb{\Gamma}^T}, \tag{14}$$

where $\mathcal{L}_{\boldsymbol{\theta}}$ is the weighted asymmetric graph Laplacian defined as in Eq. 5. Now, we can see that CRN contains rLDG (Eq. 6) as a special case if $\mathbb{\Gamma} = I$. Owing to this inclusion relation, CRN with LMA kinetics is a mathematically sound generalization of rLDG. Because LDG has been used in various fields of social science, network science, machine learning, and so on, CRN theory is potentially important for extending the results there.

Example 1 (Simplified Brusselator CRN [8, 104]) The Brusselator is a representative CRN, which can generate non-trivial dynamic behaviors such as oscillations. We use a reversible CRN version of the simplified Brusselator [8, 104], whose CRN-hypergraph depicted in Fig. 2b has the following structural information:

$$\mathbb{\Gamma} = \begin{array}{c} \\ \mathbb{X}_1 \\ \mathbb{X}_2 \end{array}\!\!\begin{array}{c} \hat{v}_1 \;\; \hat{v}_2 \;\; \hat{v}_3 \;\; \hat{v}_4 \;\; \hat{v}_5 \\ \begin{pmatrix} 0 & 1 & 0 & 2 & 3 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix} \end{array}, \quad \mathbb{B} = \begin{array}{c} \\ \hat{v}_1 \\ \hat{v}_2 \\ \hat{v}_3 \\ \hat{v}_4 \\ \hat{v}_5 \end{array}\!\!\begin{array}{c} \mathbb{e}_1 \;\;\; \mathbb{e}_2 \;\;\; \mathbb{e}_3 \\ \begin{pmatrix} +1 & 0 & 0 \\ -1 & +1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & +1 \\ 0 & 0 & -1 \end{pmatrix} \end{array}, \quad \mathbb{S} = \begin{array}{c} \\ \mathbb{X}_1 \\ \mathbb{X}_2 \end{array}\!\!\begin{array}{c} \mathbb{e}_1 \;\;\; \mathbb{e}_2 \;\;\; \mathbb{e}_3 \\ \begin{pmatrix} -1 & +1 & -1 \\ 0 & -1 & +1 \end{pmatrix} \end{array}$$
$$\tag{15}$$

,

---

[17] There exists another type of extension known as generalized LMA kinetics where the monomials $\boldsymbol{x}^{\mathbb{\Gamma}^T}$ are replaced with fractional monomials, i.e., powers of $\boldsymbol{x}$ with nonnegative real-valued exponents [102].

The rate equation (Eq. 14) can be represented as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -\overbrace{\begin{pmatrix} -1 & +1 & -1 \\ 0 & -1 & +1 \end{pmatrix}}^{\mathbb{S}}\left[\overbrace{\begin{pmatrix} k_1^+ \\ k_2^+ x_1 \\ k_3^+ x_1^2 x_2 \end{pmatrix}}^{\boldsymbol{j}^+(\boldsymbol{x})} - \overbrace{\begin{pmatrix} k_1^- x_1 \\ k_2^- x_2 \\ k_3^- x_1^3 \end{pmatrix}}^{\boldsymbol{j}^-(\boldsymbol{x})} \right]. \tag{16}$$

2.3 Fokker Planck equations

While our main focus is the dynamics on graphs and hypergraphs, we use FPE as a representative class of density dynamics on a continuous Euclidean space. Specifically, we use FPE only to demonstrate the relation of our results with previous ones obtained for FPE in various contexts. Because FPE is infinite-dimensional, we treat it here only formally.

Let $\boldsymbol{r} \in \mathbb{R}^d$ be a vector in a d dimensional Euclidean space. We consider infinitely many noninteracting particles randomly walking in the space and describe the dynamics by a probability density $p_t(\boldsymbol{r}) \in \mathbb{R}_{\geq 0}$ of the particles. The continuity equation for $p_t(\boldsymbol{r})$ is

$$\partial_t p_t(\boldsymbol{r}) = -\nabla \cdot \boldsymbol{j}_{\mathrm{FP}}[p_t(\boldsymbol{r})] \tag{17}$$

where $\boldsymbol{j}_{\mathrm{FP}}[p_t(\boldsymbol{r})] \in \mathbb{R}^d$ is the probability flux, $\nabla := (\partial/\partial r_1, \cdots, \partial/\partial r_d)^T$ is the gradient operator on the Euclidean space, and $(\nabla\cdot) : \nabla \cdot \boldsymbol{F}(\boldsymbol{r}) := \sum_{i=1}^d \partial F_i(\boldsymbol{r})/\partial r_i \in \mathbb{R}$ is the divergence. The flux of the FPE is defined as

$$\boldsymbol{j}_{\mathrm{FP}}[p(\boldsymbol{r})] = [\boldsymbol{F}(\boldsymbol{r})p(\boldsymbol{r}) - D_0 \nabla p(\boldsymbol{r})], \tag{18}$$

where $\boldsymbol{F}(\boldsymbol{r}) \in \mathbb{R}^d$ is the drift force, and $D_0 \in \mathbb{R}_{>0}$ is the diffusion constant.

3 Discrete calculus and homological algebra of graphs and hypergraphs

The algebraic and topological structure of the dynamics on graphs and hypergraphs can be explicitly and abstractly treated using the language of discrete calculus and homological algebra. The discrete version of the gradient and divergence mentioned in Sect. 2 is also characterized. In this section, we briefly introduce the chain and cochain complexes defined for a finite graph or a hypergraph and discrete calculus [73, 91, 105, 106]. We first introduce the complexes for a graph $\mathbb{G}$ and then extend them to a hypergraph $\mathbb{H}$ algebraically.[18] It should be noted that the conventional discrete calculus (the discrete version of the theory for differential forms) presumes the Riemannian metric in the dual space of chains and cochains or that of cochains on primal and dual

---

[18] The complex used here should not be confused with complexes used in CRN theory [8]

complexes [107, 108]. However, we are going to introduce Legendre duality instead. For this purpose, our introduction of chain and cochain complexes depends only on the topological (algebraic) information of the underlying graph and hypergraph [73] without specifying the metric information.

3.1 Chain and cochain complexes on graphs

The elements of a graph $\mathbb{G}$ are called cells in discrete calculus.[19] A vertex and an edge are, respectively, called 0-cell and 1-cell, and the graph $\mathbb{G}$ is denoted as a cell-complex.[20] For each type of the cells, we consider vectors (chains and cochains) defined on the cells. For $\mathbb{G}$, a 0-chain with field $\mathbb{R}$ is an $N_{\mathbb{v}}$-tuple of real scalars, each of which is assigned to a vertex, i.e., a 0 cell. Thus, a 0-chain is a real vector defined on the vertices of $\mathbb{G}$ with the basis $\{\mathbb{v}_i\}$. This basis is called the standard basis. The vector space of real 0-chains is called the vertex space here and denoted as $C_0(\mathbb{G}) = \mathbb{R}^{N_{\mathbb{v}}}$ [91].[21] The components of the vector $\boldsymbol{x} \in C_0(\mathbb{G})$ are given as $\boldsymbol{x}(\mathbb{v}_i) := x_i$. Similarly, a real 1-chain is a real vector defined on the edges of $\mathbb{G}$. The real vector space of 1-chains is called the edge space and denoted as $C_1(\mathbb{G}) = \mathbb{R}^{N_{\mathbb{e}}}$. The standard basis is introduced by using edges $\{\mathbb{e}_e\}$, accordingly. A flux $\boldsymbol{j}$ is a 1-chain: $\boldsymbol{j}(\mathbb{e}_e) := j_e$. The graph incidence matrix $\mathbb{B}$ induces the discrete differential $\delta_1 : C_1(\mathbb{G}) \to C_0(\mathbb{G})$ as $\delta_1 \boldsymbol{j} := \mathbb{B}\boldsymbol{j}$.[22]

To obtain an exact sequence, we algebraically define the (-1) and 2 chains and the corresponding differentials δ_0 and δ_2. Let $C_2(\mathbb{G}) = \mathbb{R}^{N_{\mathbb{z}}}$ where $N_{\mathbb{z}} = \dim[\text{Ker}\mathbb{B}]$ and $\{\boldsymbol{v}_i\}_{i \in [1,N_{\mathbb{z}}]}$ is a set of complete basis of $\text{Ker}\mathbb{B}$ where $\boldsymbol{v}_i \in \{0, +1, -1\}^{N_{\mathbb{e}}}$[23]. In algebraic graph theory, $\text{Ker}\mathbb{B}$ is called a cycle subspace [86, 91, 109]. For a graph $\mathbb{G}$, we can construct $\{\boldsymbol{v}_i\}_{i \in [1,N_{\mathbb{z}}]}$ by, for example, using the fundamental cycle basis of $\mathbb{G}$ obtained from a fixed spanning tree of $\mathbb{G}$[24] [86]. Thus, $C_2(\mathbb{G})$ is the vector space defined on the cycles of $\mathbb{G}$ and isomorphic to the cycle subspace. We define a matrix, $\mathbb{V} := (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{N_{\mathbb{z}}})$[25], and the differential $\delta_2 : C_2(\mathbb{G}) \to C_1(\mathbb{G})$ as $\delta_2 := \mathbb{V}$. From the construction, $\mathbb{B}\mathbb{V} = \delta_1 \delta_2 = 0$ and $\text{Im}[\delta_2] = \text{Ker}[\delta_1]$ hold. Similarly, let $C_{-1}(\mathbb{G}) = \mathbb{R}^{N_{\mathbb{l}}}$ where $N_{\mathbb{l}} = \dim[\text{Ker}\mathbb{B}^T]$ and $\{\boldsymbol{u}_\ell\}_{\ell \in [1,N_{\mathbb{l}}]}$ is a set of complete basis of $\text{Ker}\mathbb{B}^T$ where $\boldsymbol{u}_\ell \in \{0, +1, -1\}^{N_{\mathbb{v}}}$. The subspace $\text{Ker}\mathbb{B}^T$ is related to the connected components of

---

[19] We follow the terminology in [73]. While we use "cell", we do not presume any N-dimensional topological manifold underlying the graph. The graph is just treated algebraically as in algebraic graph theory and homological algebra.

[20] Depending on the choice of which elements of a graph are considered, the content of the complex changes. For example, vertices and edges are the major ingredients of the complex of a graph. The faces of a graph are often included in the complex. The definition of the higher-order elements than edges requires additional structural information to the incidence matrix of the graph, e.g., the edge-face incidence matrix.

[21] In algebraic graph theory, the chain of a graph $\mathbb{G}$ is defined as an integer-valued vector space $\mathbb{Z}^{N_{\mathbb{v}}}$ to represent the discrete and combinatorial nature of $\mathbb{G}$ and also to specify the domain of integration. Here, we use $\mathbb{R}$ as the field of the vector space.

[22] In algebraic graph theory, $\mathbb{B}$ is also identical to the discrete boundary operator from $C_1(\mathbb{G}; \mathbb{Z})$ to $C_0(\mathbb{G}; \mathbb{Z})$.

[23] $N_{\mathbb{z}} = 0$ when $\mathbb{G}$ is a set of trees.

[24] The spanning tree chosen specifies a fundamental cycle and cocycle bases.

[25] $\mathbb{V}^T$ is called the fundamental tieset matrix in graph theory

$\mathbb{G}$ and $\boldsymbol{u}_i$ can be chosen such that $u_{i,\ell} = +1$ if the ith vertex is included in the ℓth connected component and $u_{i,\ell} = 0$, otherwise. Thus, $C_{-1}(\mathbb{G})$ is the vector space on the connected components. From the matrix $\mathbb{U} := (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{N_{\mathbb{I}}})^T$, the differential $\delta_0 : C_0(\mathbb{G}) \to C_{-1}(\mathbb{G})$ is defined as $\delta_0 := \mathbb{U}$. From the construction, $\mathbb{U}\mathbb{B} = \delta_0\delta_1 = 0$ and $\text{Im}[\delta_1] = \text{Ker}[\delta_0]$ hold. Then, we obtain the exact chain sequence[26][27]:

$$0 \leftarrow C_{-1}(\mathbb{G}) \xleftarrow{\delta_0=\mathbb{U}} C_0(\mathbb{G}) \xleftarrow{\delta_1=\mathbb{B}} C_1(\mathbb{G}) \xleftarrow{\delta_2=\mathbb{V}} C_2(\mathbb{G}) \leftarrow 0. \qquad (19)$$

Because $C_p(\mathbb{G})$ is a vector space for each $p \in \{-1, 0, 1, 2\}$, we can consider its dual vector space $C^p(\mathbb{G}) := C_p^*(\mathbb{G})$ consisting of the linear functions on $C_p(\mathbb{G})$. An element of $C^p(\mathbb{G})$ is called p-cochain. Let $\langle \cdot, \cdot \rangle : C_p(\mathbb{G}) \times C^p(\mathbb{G}) \to \mathbb{R}$ be the standard bilinear pairing of the p-chain and p-cochain defined with the standard basis. The transposes of $\mathbb{U}$, $\mathbb{B}$, and $\mathbb{V}$ induce the differentials between cochains as $\delta^{-1} := \mathbb{U}^T : C^{-1}(\mathbb{G}) \to C^0(\mathbb{G})$, $\delta^0 := \mathbb{B}^T : C^0(\mathbb{G}) \to C^1(\mathbb{G})$, and $\delta^1 := \mathbb{V}^T : C^1(\mathbb{G}) \to C^2(\mathbb{G})$. The differentials δ^p on cochains are the adjoints of the differentials δ_p on chains, which induce the exact cochain sequence:

$$0 \to C^{-1}(\mathbb{G}) \xrightarrow{\delta^{-1}=\mathbb{U}^T} C^0(\mathbb{G}) \xrightarrow{\delta^0=\mathbb{B}^T} C^1(\mathbb{G}) \xrightarrow{\delta^1=\mathbb{V}^T} C^2(\mathbb{G}) \to 0. \qquad (20)$$

Note that the definition of chains, cochains, and differential operators are topological in the sense that we do not include any metric information.

3.2 Chain and cochain complexes on hypergraphs

The definitions of chain and cochain complexes introduced above are algebraically extended to hypergraphs $\mathbb{H}$ simply by replacing the graph incidence matrix $\mathbb{B}$ with the hypergraph incidence matrix $\mathbb{S}$.

Definition 9 (*Exact chain and cochain sequences on a hypergraph*) The chain and cochain complexes on a hypergraph are defined by the following diagram:

$$
\begin{array}{ccccccccc}
0 \to & C^{-1}(\mathbb{H}) & \xrightarrow{\delta^{-1}=\mathbb{U}^T} & C^0(\mathbb{H}) & \xrightarrow{\delta^0=\mathbb{S}^T} & C^1(\mathbb{H}) & \xrightarrow{\delta^1=\mathbb{V}^T} & C^2(\mathbb{H}) \to & 0 \\
0 \leftarrow & C_{-1}(\mathbb{H}) & \xleftarrow{\delta_0=\mathbb{U}} & C_0(\mathbb{H}) & \xleftarrow{\delta_1=\mathbb{S}} & C_1(\mathbb{H}) & \xleftarrow{\delta_2=\mathbb{V}} & C_2(\mathbb{H}) \leftarrow & 0.
\end{array}
$$

where $C^{-1}(\mathbb{H}) \simeq C_{-1}(\mathbb{H}) \simeq \mathbb{R}^{N_{\mathbb{I}}}$, $C^0(\mathbb{H}) \simeq C_0(\mathbb{H}) \simeq \mathbb{R}^{N_{\mathbb{X}}}$, $C^1(\mathbb{H}) \simeq C_1(\mathbb{H}) \simeq \mathbb{R}^{N_{\mathbb{e}}}$, and $C^2(\mathbb{H}) \simeq C_2(\mathbb{H}) \simeq \mathbb{R}^{N_{\mathbb{z}}}$.

The bases, $\mathbb{V}$ and $\mathbb{U}$, are obtained as integral bases, i.e., the components of $\mathbb{V}$ and $\mathbb{U}$ can be chosen from $\mathbb{Z}$ because $\mathbb{S}$ is an integer-valued matrix.[28] As we will explain in

---

[26] We should note that the sequence is not canonical because $\mathbb{U}$ and $\mathbb{V}$ depend on the choice of bases.

[27] Upon necessity, we can consider the harmonic components by employing an under-complete basis for $\mathbb{V}$.

[28] As far as we know, there is not a systematic and widely-appreciated way to define these bases because we have multiple ways to extend the notion of spanning tree of a graph to a hypergraph.

Sect. 6 and Sect. 9, the meaning of $C_2(\mathbb{H})$ can be retained as the space on generalized cycles. The meaning of $C_{-1}(\mathbb{H})$ becomes the space of conserved quantities under the dynamics (Eq. 10).

3.3 Discrete calculus on graphs and hypergraphs

The p-cochain and p-chain introduced above are an algebraic abstraction of the p-differential form and its Hodge dual on a differential manifold [73]. Accordingly, the discrete versions of gradient, divergence, and curl are associated with the differentials (exterior derivative).

Definition 10 (*Discrete gradients, divergences, and curls*) The discrete gradient is defined as $\text{grad}_{\mathbb{B}} := \delta^0 = \mathbb{B}^T$ for $\mathbb{G}$ and also as $\text{grad}_{\mathbb{S}} := \delta^0 = \mathbb{S}^T$ for $\mathbb{H}$. The adjoints of the gradients are defined with the corresponding adjoint differentials: $\text{grad}_{\mathbb{B}}^* := \delta_1 = \mathbb{B}$ and $\text{grad}_{\mathbb{S}}^* := \delta_1 = \mathbb{S}$. They are called discrete divergences and denoted also as $\text{div}_{\mathbb{B}} = \text{grad}_{\mathbb{B}}^*$ and $\text{div}_{\mathbb{S}} = \text{grad}_{\mathbb{S}}^*$.[29] The discrete curl and its adjoint are defined as $\text{curl}_{\mathbb{V}} := \delta^1 = \mathbb{V}^T$ and $\text{curl}_{\mathbb{V}}^* := \delta_2 = \mathbb{V}$, respectively.

3.4 Linear graph Laplacian dynamics and metric structure in discrete calculus

In the theory of graph Laplacian, a metric matrix M_p and its associated inner product are typically endowed for each p. To contrast it with the Legendre duality introduced later, we briefly outline it here. For an edge-weighted graph $\mathbb{G}_{k^\pm}$ and for the case that $k^+ = k^- = k \in \mathbb{R}_{>0}^{N_e}$, $M_0 = I$ and $M_1 = \text{diag}[1/k]$ are conventionally employed. With these metric matrices, the graph Laplacian introduced in Eq. 5 can be described as

$$\mathcal{L}_k = \text{div}_{\mathbb{B}} M^1 \text{grad}_{\mathbb{B}} M_0 \tag{21}$$

where $M^p := M_p^{-1}$. By including such metric information, the following pair of metric gradient and divergence is often used in graph theory and network theory: $\text{grad}_M := \sqrt{M^1} \mathbb{B}^T$ and $\text{div}_M := \mathbb{B}\sqrt{M^1}$ where $\sqrt{M^1} := \text{diag}[\sqrt{k}]$. This symmetric graph Laplacian $\mathcal{L}_k$ induces a linear dynamics of $x \in \mathbb{R}^{N_v}$ on graph via Eq. 6[30]:

$$\dot{x} = -\mathcal{L}_k x. \tag{22}$$

The eigenvalues and eigenvectors of $\mathcal{L}_k$ enable us to obtain spectral information of the underlying graph [92]. Even for nonlinear dynamics on a hypergraph as in Eq. 14, the same symmetric Laplacian can provide some information when $k^+ = k^- = k$. We can also include other information in the metric matrices such as the degree of vertices [110]. Various normalizations of the graph Laplacian can be attributed to the choice of metrics.

---

[29] These notations are consistent with those in Sect. 2

[30] Here x is not density but a vector in $\mathbb{R}^{N_v}$.

However, such a choice of metric matrices ends up only with linear dynamics on $\mathbb{R}^{N_{\mathbb{v}}}$ and is relevant only when the weighting is symmetric: $\boldsymbol{k}^+ = \boldsymbol{k}^- = \boldsymbol{k}$. In addition, it may not always capture important aspects of the density dynamics such as gradient flow properties and information–theoretic properties, because nonlinear terms such as $\ln \boldsymbol{p}$ appear in information–theoretic quantities. To extend the class of dynamics being covered and to enable the information-geometric characterization of dynamics, we have to generalize the conventional inner product structure by replacing it with the Legendre dual structure induced by convex functions.

4 Dually flat spaces on vertices and edges and generalized flow

In this section, we introduce two pairs of dually flat spaces (Fig. 1): one is associated with the vertex spaces, i.e., the dual spaces of 0-chains and 0-cochains. The other corresponds to the edge spaces, i.e., the dual spaces of 1-chains and 1-cochains. By combining them, the dynamics on graphs and hypergraphs are characterized as a generalized flow.

4.1 Dually flat spaces on vertices and thermodynamic functions

We work on the density $\boldsymbol{x}$ and the vertex space for CRN because its reduction to rLDG is straightforward. For a probability vector $\boldsymbol{p}$, the introduction of dually flat spaces of $\boldsymbol{p}$ and $\ln \boldsymbol{p}$ is natural from the information-geometric viewpoint. In CRN, $\boldsymbol{x}$ is the vector of concentrations of molecular species. As we recently clarified [48], the dually flat spaces, in this case, result from the Legendre duality between extensive and intensive variables in thermodynamics, which is also natural from the physical viewpoint.

Definition 11 (*Density space (primal vertex affine space)*) The density space (also called primal affine vertex space) is the positive orthant $\mathcal{X} := \mathbb{R}_{>0}^{N_{\mathbb{X}}}$ of a vector space $\mathbb{R}^{N_{\mathbb{X}}}$, which is isomorphic to $C_0(\mathbb{H})$; $\mathbb{R}^{N_{\mathbb{X}}} \simeq C_0(\mathbb{H})$ (Fig. 1, lower left).

Remark 4 The density space $\mathcal{X}$ is defined as the positive orthant rather than as $\mathbb{R}_{\geq 0}^{N_{\mathbb{X}}}$. This excludes the cases where some elements of $\boldsymbol{x}$ become 0. From the viewpoint of information geometry, this restriction is necessary to consider densities with the same support (all $\boldsymbol{x}$ in $\mathcal{X}$ should be equivalent in terms of absolute continuity of measures). From the viewpoint of dynamical systems, depending on the specific functional form of the flux $\boldsymbol{j}(\boldsymbol{x})$, the trajectory $\boldsymbol{x}(t)$ may not be restricted within $\mathcal{X}$. The property $\boldsymbol{x}(t)$ in $\mathcal{X}$ for $t \in [0, \infty]$ is known as persistence.[31] Without going into this intricate problem, we simply assume that $\boldsymbol{x}(t) \in \mathcal{X}$ for $t \in [0, \infty]$. We call $\partial \mathcal{X} := \mathbb{R}_{\geq 0}^{N_{\mathbb{X}}} \backslash \mathcal{X}$ the boundary of $\mathcal{X}$.

We define the dual of the density space by the Legendre transformation via the thermodynamic function:

---

[31] The persistence of a dynamical system is a hard problem, and the persistence for a subclass of CRN is an open problem [111, 112], which goes by the name of Global Attractor Conjecture since 1974.

Definition 12 (*Primal thermodynamic function*) A strictly convex differentiable function $\Phi : \mathcal{X} \to \mathbb{R}$ is called the primal thermodynamic function[32][33] if the following two conditions are satisfied: (1) the associated Legendre transformation

$$\partial \Phi : \mathcal{X} \to \mathbb{R}^{N_{\times}} \tag{23}$$

$$\boldsymbol{x} \longmapsto \boldsymbol{y} := \partial_{\boldsymbol{x}} \Phi(\boldsymbol{x}) = \left(\frac{\partial \Phi(\boldsymbol{x})}{\partial x_1}, \cdots, \frac{\partial \Phi(\boldsymbol{x})}{\partial x_{N_{\times}}} \right)^T \tag{24}$$

has the image $\mathcal{Y} := \{ \boldsymbol{y} | \boldsymbol{y} = \partial \Phi(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{X} \}$ being equal to $\mathbb{R}^{N_{\times}}$, i.e., $\mathcal{Y} = \mathbb{R}^{N_{\times}}$; (2) for any $\boldsymbol{x}_{in} \in \mathcal{X}$ and any point on the boundary $\boldsymbol{x}_{bd} \in \partial \mathcal{X}$,

$$\lim_{\lambda \to +0} \frac{\mathrm{d}\Phi(\boldsymbol{x}_\lambda)}{\mathrm{d}\lambda} = -\infty \tag{25}$$

holds where $\boldsymbol{x}_\lambda := \lambda \boldsymbol{x}_{in} + (1 - \lambda) \boldsymbol{x}_{bd}$ for $\lambda \in [0, 1]$.

Definition 13 (*Potential space (dual affine vertex space) and dual thermodynamic function*) The potential (field) space $\mathcal{Y} = \mathbb{R}^{N_{\times}}$ (also called the dual affine vertex space) is an affine space dual to $\mathcal{X}$ with the associated vector space $C^0(\mathbb{H})((\text{Fig. } 1,$ upper left)).[34] The dual thermodynamic function $\Phi^* : \mathcal{Y} \to \mathbb{R}$ is the Legendre-Fenchel conjugate of the primal thermodynamic function:

$$\Phi^* : \mathcal{Y} \to \mathbb{R}, \quad \boldsymbol{y} \mapsto \Phi^*(\boldsymbol{y}) := \max_{\boldsymbol{x}' \in \mathcal{X}} \left[\langle \boldsymbol{x}', \boldsymbol{y} \rangle - \Phi(\boldsymbol{x}') \right], \tag{26}$$

where $\langle \cdot, \cdot \rangle : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is the bilinear pairing under the standard basis. From the properties of the primal function, $\Phi^*(\boldsymbol{y})$ is also a strictly convex differentiable function. From $\Phi^*(\boldsymbol{y})$, we have the inverse Legendre transformation $\partial \Phi^* : \mathcal{Y} \to \mathcal{X}, \boldsymbol{y} \mapsto \boldsymbol{x} = \partial_{\boldsymbol{y}} \Phi^*(\boldsymbol{y})$.

The Legendre transformations, $\partial \Phi$ and $\partial \Phi^*$, are continuous and establish a bijection between $\mathcal{X}$ and $\mathcal{Y}$, where $\partial \Phi^* = \partial \Phi^{-1}$. In the following, we regard a pair $(\boldsymbol{x}, \boldsymbol{y})$ with the same decoration as a Legendre dual pair satisfying $\boldsymbol{y} = \partial \Phi(\boldsymbol{x})$. For a pair, the Legendre-Fenchel-Young identity holds:

$$\Phi(\boldsymbol{x}) + \Phi^*(\boldsymbol{y}) = \langle \boldsymbol{x}, \boldsymbol{y} \rangle. \tag{27}$$

---

[32] In information geometry, the convex function inducing duality is often called a potential function. We avoid using the word "potential" to discriminate it with an element of the dual vertex affine space, which is called a potential (field) or chemical potential in physics and chemistry.

[33] We may consider a convex function $\Phi(\boldsymbol{x})$, which does not induce a bijection between $\mathcal{X}$ and $\mathcal{Y}$, e.g., the one which is not strictly convex. Such a situation can arise if a phase transition occurs. It would be an important direction to include this class of functions in this framework.

[34] $\mathcal{Y}$ is not only associated with but also isomorphic to the 0-cochain. This condition is important when we consider information-geometric projections in the later sections. In the theory of differential forms, a 0-form is often described as a potential field on a manifold. Our choice of the potential space is consistent with this convention.

Different pairs are discriminated with the difference of decorations as (x', y') or (x_p, y_p).

For a thermodynamic function, the Bregman divergence can be defined:

Definition 14 (*Bregman divergence* [1, 113]) The Bregman divergence on $\mathcal{X}$ with the generating thermodynamic function $\Phi(x)$ is defined as

$$\mathcal{D}_\Phi^\mathcal{X}[x\|x'] := \Phi(x) - \Phi(x') - \langle x - x', \partial\Phi(x')\rangle \in \mathbb{R}_{\geq 0}. \tag{28}$$

The non-negativity of the Bregman divergence follows from the Fenchel-Young inequality for products [114, 115]. Furthermore, from the strict convexity of the thermodynamic function, $\mathcal{D}_\Phi^\mathcal{X}[x\|x']$ is also strictly convex with respect to x and $\mathcal{D}_\Phi^\mathcal{X}[x\|x'] = 0$ if and only if $x = x'$. Bregman divergences are defined for (y, y') and also for (x, y') as

$$\mathcal{D}_{\Phi^*}^\mathcal{Y}[y'\|y] := \Phi^*(y') - \Phi^*(y) - \langle \partial\Phi^*(y), y' - y\rangle, \tag{29}$$

$$\mathcal{D}_{\Phi,\Phi^*}^{\mathcal{X},\mathcal{Y}}[x; y'] := \Phi(x) + \Phi^*(y') - \langle x, y'\rangle. \tag{30}$$

Because (x, y) and (x', y') are Legendre pairs, all the three representations are equivalent[35]: $\mathcal{D}_\Phi^\mathcal{X}[x\|x'] = \mathcal{D}_{\Phi^*}^\mathcal{Y}[y'\|y] = \mathcal{D}_{\Phi,\Phi^*}^{\mathcal{X},\mathcal{Y}}[x; y']$.[36] $\mathcal{D}_\Phi^\mathcal{X}[x\|x']$, $\mathcal{D}_{\Phi^*}^\mathcal{Y}[y'\|y]$, and $\mathcal{D}_{\Phi,\Phi^*}^{\mathcal{X},\mathcal{Y}}[x; y']$ are abbreviated as $\mathcal{D}^\mathcal{X}[x\|x'], \mathcal{D}^\mathcal{Y}[y'\|y]$, and $\mathcal{D}^{\mathcal{X},\mathcal{Y}}[x; y']$, respectively.

Finally, the Hessian matrices of the primal and dual thermodynamic functions are defined when they are twice differentiable[37]:

Definition 15 (*Hessian matrices*) The primal and dual Hessian matrices, $G_x \in \mathbb{R}^{N_v \times N_v}$ and $G_y^* \in \mathbb{R}^{N_v \times N_v}$, of thermodynamic functions, $\Phi(x)$ and $\Phi^*(y)$, are defined as

$$(G_x)_{i,j} := \frac{\partial^2 \Phi(x)}{\partial x_i \partial x_j}, \qquad\qquad (G_y^*)_{i,j} := \frac{\partial^2 \Phi^*(y)}{\partial y_i \partial y_j}. \tag{31}$$

In addition, they are positive definite and $G_x^{-1} = G_y^*$ holds for a Legendre dual pair x and y.

The Hessian matrices induce a Riemannian metric over $\mathcal{X}$. The tangent and cotangent spaces $\mathcal{T}_x\mathcal{X}$ and $\mathcal{T}_x^*\mathcal{X}$ are isomorphic to the corresponding tangent and cotangent spaces $\mathcal{T}_y^*\mathcal{Y}$ and $\mathcal{T}_y\mathcal{Y}$ over $\mathcal{Y}$ and also to $C_0(\mathbb{H})$ and $C^0(\mathbb{H})$: $\mathcal{T}_x\mathcal{X} \cong \mathcal{T}_y^*\mathcal{Y} \cong C_0(\mathbb{H})$ and $\mathcal{T}_x^*\mathcal{X} \cong \mathcal{T}_y\mathcal{Y} \cong C^0(\mathbb{H})$.

The typical example of the duality between x and y in statistics is that between probability p and its logarithm $\ln p$. Other than this typical one, depending on the purpose, we adopt different forms of thermodynamic functions $(\Phi(x), \Phi^*(y))$, associated dual variables, and Bregman divergence to endow different properties to inference or

---

[35] $\mathcal{D}_{\Phi,\Phi^*}^{\mathcal{X},\mathcal{Y}}[x; y']$ is also called Fenchel-Young divergence [116].

[36] We here used the Legendre-Fenchel-Young identity (Eq. 27).

[37] When we work on Hessian matrices, we always suppose additionally that they are twice-differentiable.

estimation methods that we are designing [1]. In the case of CRN, the thermodynamic functions and Legendre duality are associated with the equilibrium thermodynamics [117]. Specifically, as we recently demonstrated [48], $\mathcal{X}$ and $\mathcal{Y}$ are the conjugate spaces of the extensive and intensive thermodynamic variables (density of molecules and their chemical potential), $\Phi(\boldsymbol{x})$ is the thermodynamic potential function of the system, and the Bregman divergence becomes the difference of the total entropy. These correspondences are derived directly from the axiomatic formulation of thermodynamics [48, 117]. The explicit functional form of $\Phi(\boldsymbol{x})$ is then determined by the physical details of the thermodynamic system that we work on.

Before closing this subsection, we introduce the notion of separability, which will be linked to the locality of the flux.

Definition 16 (*Separability of a thermodynamic function*) A thermodynamic function $\Phi(\boldsymbol{x})$ is separable if it can be represented as

$$\Phi(\boldsymbol{x}) = \sum_{i=1}^{N_{\mathrm{v}}} c_i \phi(x_i/x_i^o), \tag{32}$$

where $c_i > 0$, $x_i^o > 0$, and $\phi(x) : \mathbb{R}_{>0} \to \mathbb{R}$ is a scalar primal thermodynamic function.

If $\Phi(\boldsymbol{x})$ is separable, then its conjugate $\Phi^*(\boldsymbol{y})$ is also separable as $\Phi^*(\boldsymbol{y}) = \sum_{i=1}^{N_{\mathrm{v}}} c_i \phi^*(\frac{x_i^o}{c_i} y_i)$ where $\phi^*(y) : \mathbb{R} \to \mathbb{R}$ is the Legendre conjugate of $\phi(x)$.[38] If a thermodynamic function is separable, then the corresponding Bregman divergence is separable. The Hessian matrices become diagonal for a separable thermodynamic function. Most of our results can hold without the separability, but common thermodynamic functions and related quantities are typically separable. For example, the Kullback-Leibler divergence is an example of separable Bregman divergences.

4.2 Dually flat spaces on edges and dissipation functions

Next, we introduce another dually flat structure onto the edge space of graphs and hypergraphs based on the flux-force relation.

Definition 17 (*Flux and force spaces (primal and dual edge spaces)*) The flux and force spaces on the edges, $\mathcal{J}_{\boldsymbol{x}} = \mathbb{R}^{N_{\mathrm{e}}}$ and $\mathcal{F}_{\boldsymbol{x}} = \mathbb{R}^{N_{\mathrm{e}}}$, are a pair of the primal and dual vector spaces defined for each $\boldsymbol{x} \in \mathcal{X}$, which are isomorphic to $C_1(\mathbb{H})$ and $C^1(\mathbb{H})$, respectively (Fig. 1, right). The bilinear pairing under the standard basis $\langle \cdot, \cdot \rangle : C_1(\mathbb{H}) \times C^1(\mathbb{H}) \to \mathbb{R}$ is inherited to $(\mathcal{J}_{\boldsymbol{x}}, \mathcal{F}_{\boldsymbol{x}})$.

To introduce Legendre duality on $(\mathcal{J}_{\boldsymbol{x}}, \mathcal{F}_{\boldsymbol{x}})$, we use the dissipation functions:

Definition 18 (*Dissipation function*[39]) A dissipation function on $\mathcal{F}_{\boldsymbol{x}}$, $\Psi_{\boldsymbol{x}}^* : \mathcal{F}_{\boldsymbol{x}} \to \mathbb{R}$, $\boldsymbol{f} \mapsto \Psi_{\boldsymbol{x}}^*(\boldsymbol{f})$, is a strictly convex and continuously differentiable function with

---

[38] One may further generalize the separability so that $\phi(x)$ depends on i as $\phi_i(x)$.

[39] The definition of dissipation functions is more strict than those used in the previous works, e.g., [75]. This is because we define extended projections in this space as in [83].

respect to f for all $x \in \mathcal{X}$ that also satisfies the following additional conditions:

$$\text{1-coercive:} \qquad \frac{\Psi_x^*(f)}{\|f\|} \to \infty \quad \text{as } \|f\| \to \infty, \qquad (33)$$

$$\text{Symmetric:} \qquad \Psi_x^*(f) = \Psi_x^*(-f) \qquad (34)$$

$$\text{Bounded below by 0 :} \qquad \Psi_x^*(0) = 0, \qquad (35)$$

Proposition 1 (Duality of dissipation functions) *The Legendre-Fenchel conjugate of* $\Psi_x^*(f)$, *i.e.,* $\Psi_x(j) := \max_f \left[\langle j, f \rangle - \Psi_x^*(f) \right]$, *is also the dissipation function on* $\mathcal{J}_x$. $\Psi_x(j)$ *and* $\Psi_x^*(f)$ *are called primal and dual dissipation functions.*

Proof For each $x \in \mathcal{X}$, the function $\Psi_x(j)$ is strictly convex, continuously differentiable, 1-coercive, and $\Psi_x(j) < +\infty$ for all $j \in \mathcal{J}_x$ because $\Psi_x^*(f)$ is (see Corollary 4.1.4 in [118]). For $j \in \mathcal{J}_x$, the symmetry holds as $\Psi_x(-j) = \max_f \left[\langle -j, f \rangle - \Psi_x^*(f) \right] = \max_f \left[\langle j, -f \rangle - \Psi_x^*(f) \right] = \max_f \left[\langle j, f \rangle - \Psi_x^*(-f) \right] = \max_f \left[\langle j, f \rangle - \Psi_x^*(f) \right] = \Psi_x(j)$. From the convexity and symmetry, the minimum of $\Psi_x(j)$ is attained at $j = 0$ and $\min_j \Psi_x(j) = \Psi_x(0) = \max_f \left[\langle 0, f \rangle - \Psi_x^*(f) \right] = -\min_f \Psi_x^*(f) = 0$. $\qquad \square$

From these properties, for each $x \in X$, the one-to-one Legendre duality via Legendre transformations is established for all over $(\mathcal{J}_x, \mathcal{F}_x)$:

$$j = \partial_f \Psi_x^*(f), \qquad\qquad f = \partial_j \Psi_x(j). \qquad (36)$$

In the following, we abbreviate the Legendre transformations as $\partial_f \Psi_x^*(f) = \partial \Psi_x^*(f)$ and $\partial_j \Psi_x(j) = \partial \Psi_x(j)$ [40]. Similarly to the Legendre dual pair (x, y) in $\mathcal{X}$ and $\mathcal{Y}$, a pair of flux and force with the same decoration, e.g., $(j, f)_x$ or $(j_0, f_0)_x$, represents a Legendre dual pair linked by Eq. 36 at x. We omit the x-dependency for simplicity. The Legendre dual pair (j, f) satisfies the Legendre-Fenchel-Young identity for each $x \in \mathcal{X}$:

$$\Psi_x^*(f) + \Psi_x(j) - \langle j, f \rangle = 0. \qquad (37)$$

Furthermore, the additional conditions of dissipation functions enable the Legendre duality to work as an extension of a Riemannian metric structure:

Proposition 2 ([75]) *The Legendre transformations satisfy the following properties:*

$$\text{Pairing of } 0 \in \mathcal{J} \text{ and } 0 \in \mathcal{F} : 0 = \partial \Psi_x^*(0), \quad 0 = \partial \Psi_x(0) \qquad (38)$$

$$\text{Symmetry:} -f = \partial \Psi_x(-j), \quad -j = \partial \Psi_x^*(-f). \qquad (39)$$

$$\text{Nonnegativity of bilinear pairing:} \langle j, f \rangle = \Psi_x^*(f) + \Psi_x(j) \geq 0. \qquad (40)$$

The first property means that zero force $f = 0$ and zero flux $j = 0$ are always Legendre dual regardless of x, and the second one indicates that if (j, f) is a Legendre dual

---

[40] We do not use differentiation of $\Psi_x^*(f)$ and $\Psi_x(j)$ with respect to x in this work.

pair, then $(-j, -f)$ is as well.[41] The third property, as well as the nonnegativity of the dissipation functions, enables them to play the similar roles to the metric-induced norm in Riemannian geometry.[42]

With the dissipation functions, $\Psi_x(j)$ and $\Psi_x^*(f)$, we now have the second dually flat structure on the edge spaces $(\mathcal{J}_x, \mathcal{F}_x)$. On these dually flat spaces, we define the Bregman divergence and Hessian matrices:

Definition 19 (*Bregman divergence and Hessian matrices on the edge spaces*) For each $x \in \mathcal{X}$, the Bregman divergence between $j \in \mathcal{J}_x$ and $f' \in \mathcal{F}_x$ is defined as

$$\mathcal{D}_x^{\mathcal{J},\mathcal{F}}[j; f'] := \Psi_x(j) + \Psi_x^*(f') - \langle j, f' \rangle. \tag{41}$$

$\mathcal{D}_x^{\mathcal{J}}[j \| j']$ and $\mathcal{D}_x^{\mathcal{F}}[f \| f']$ are also defined analogously to the Bregman divergence on the vertex space $(\mathcal{X}, \mathcal{Y})$. For a Legendre conjugate pair of twice differentiable dissipation functions, the Hessian matrices, $G_{x,j}$ and $G_{x,f}^*$, are defined as

$$(G_{x,j})_{e,e'} := \frac{\partial^2 \Psi_x(j)}{\partial j_e \partial j_{e'}}, \qquad (G_{x,f}^*)_{e,e'} := \frac{\partial^2 \Psi_x^*(f)}{\partial f_e \partial f_{e'}}. \tag{42}$$

These matrices are positive-definite.

The Legendre dual structure via the dissipation functions provides an extension of a Riemannian metric structure in the following sense. If the dissipation function is a quadratic function, i.e., a positive definite quadratic form as

$$\Psi_x^{q,*}(f) := \frac{1}{2}\langle f, M_x^* f \rangle, \tag{43}$$

where M_x^* is a positive definite $N^{\mathrm{e}} \times N^{\mathrm{e}}$ matrix, the Legendre transformation is reduced to the linear mapping $j = \partial \Psi_x^{q,*}(f) = M_x^* f$[43]. Then, the bilinear pairing, $\langle j, f' \rangle = \langle j, M_x j' \rangle = \langle M_x^* f, f' \rangle$, becomes the inner product under the metric matrix M_x where $M_x = (M_x^*)^{-1}$. The dissipation functions are associated with the induced norms: $\Psi_x^*(f) = \frac{1}{2}\|f\|_{M_x^*}^2$, $\Psi_x(j) = \frac{1}{2}\|j\|_{M_x}^2$. The Bregman divergence is reduced to the norm-induced squared distance: $\mathcal{D}_x^{\mathcal{J},\mathcal{F}}[j; f'] = \frac{1}{2}\|j - j'\|_{M_x}^2 = \frac{1}{2}\|f - f'\|_{M_x^*}^2$.

Finally, we also introduce the notion of separability to the dissipation functions:

---

[41] From the physical point of view, these conditions are consistent with the thermodynamic requirement that, if the force is zero, the corresponding flux becomes zero, and vice versa and that a sign-reversed force induced the sign-reversed flux.

[42] In the context of thermodynamics, the nonnegativity of $\langle j, f \rangle$ is linked to the nonnegativity of the entropy production rate and thus the second law of thermodynamics.

[43] This correspondence illustrates that the dependency of $\Psi_x^{q,*}(f)$ on x is a formal generalization of the Riemannian metric. But for this case, the relevant state space for x is not the positive orthant but the vector space $\mathbb{R}^{N_{\mathrm{v}}}$.

Definition 20 (*Separability and locality of dissipation functions*) A dissipation function $\Psi_x^*(f)$ is separable if it can be represented as

$$\Psi_x^*(f) = \sum_{e=1}^{N_e} \omega_e(x)\psi^*(f_e/f_e^o(x)),\tag{44}$$

where $\omega_e(x) > 0$ and $f_e^o(x) > 0$ for $x \in \mathcal{X}$ are positive weights and $\psi^*(f) : \mathbb{R} \to \mathbb{R}$ is a scalar dissipation function, i.e., a strictly convex differentiable scalar function satisfying Eq. 34, Eq. 35, and Eq. 33. If $\omega_e(x)$ and $f_e^o(x)$ are additionally local, then the dissipation function is separable and local. If $\Psi_x^*(f)$ is separable, then its dual $\Psi_x(j)$ is also separable. The same is true for the locality.

Remark 5 (*Young functions and N functions*) The scalar dissipation function is a N function, which appears in the theory of Orlicz spaces. A function $\tilde{\psi}(j) : [0, \infty) \to [0, \infty]$ represented as $\tilde{\psi}(j) = \int_0^j \varsigma(j')dj'$ is called Young function where $\varsigma(j) : [0, \infty) \to [0, \infty]$ is a non-decreasing function satisfying $\varsigma(0) = 0$ and being left-continuous on $(0, \infty)$. If $\varsigma(j)$ additionally satisfies $0 < \varsigma(j) < +\infty(0 < j < \infty)$, $\lim_{j\to+0}\varsigma(j) = 0$, and $\lim_{j\to\infty}\varsigma(j) = +\infty$, then $\tilde{\psi}(j)$ is called an N-function. If we define a function $\psi(j)$ with a N-function $\tilde{\psi}(j)$ as $\psi(j) = \tilde{\psi}(|j|)$, this becomes a scalar dissipation function [119]. A separable dissipation function (Eq. 44) is often called a weighted N-function [120, 121]. The dissipation function and induced Legendre duality are, therefore, related to Birnbaum-Orlicz spaces, which are an extension of L^p spaces.

4.3 Generalized flow on graphs and hypergraphs and its steady state

Because of the one-to-one Legendre duality between $(j, f)_x$, the continuity equation (Eq. 10) can be represented as a generalized flow driven by the force $f(x)$ dual to $j(x)$ [77]:

Definition 21 (*Generalized flow*) A curve $x(t)$ is a generalized flow on $\mathbb{H}$ driven by force $f(x)$ under the dissipation function Ψ_x^* if it can be represented as

$$\dot{x} = -\text{div}_\mathbb{S}\,j(x) = -\text{div}_\mathbb{S}\partial\Psi_x^*[f(x)].\tag{45}$$

This representation is independent of the specific functional form of $f(x)$ and $\Psi_x^*(f)$ and also on the definition of $\text{div}_\mathbb{S}$ as long as the generated $j(x)$ is consistent with $\mathbb{H}$[44,45]. Thus, we can potentially apply this framework to various systems by choosing these functions appropriately depending on the system or the problem we work on.

The generalized flow naturally encompasses three types of steady states:

Definition 22 (*Steady state, complex-balanced state, and detailed-balanced state*) We define the manifolds of steady state $\mathcal{M}^{\text{ST}}$, complex-balanced (CB) state $\mathcal{M}^{\text{CB}}$, and detailed-balanced (DB) state $\mathcal{M}^{\text{DB}}$, respectively, as follows:

---

[44] The consistency is required because of our choice of $\mathbb{R}_{\geq 0}^{N_\mathcal{X}}$ as the density space.

[45] The consistency with $\mathbb{H}$ is assumed to hold.

$$\mathcal{M}^{\text{ST}} := \{x \in \mathcal{X} | \mathbb{S}j(x) = 0\}, \tag{46}$$

$$\mathcal{M}^{\text{CB}} := \{x \in \mathcal{X} | \mathbb{B}j(x) = 0\}, \tag{47}$$

$$\mathcal{M}^{\text{DB}} := \{x \in \mathcal{X} | j(x) = 0\} = \{x \in \mathcal{X} | f(x) = 0\}, \tag{48}$$

where we used $j(x) = 0$ iff $f(x) = 0$ from the properties of the dissipation functions. The relations $j(x) = 0$ and $\mathbb{B}j(x) = 0$ are called the detail-balanced (DB) condition and the complex-balanced (CB) condition, respectively. From the decomposition $\mathbb{S} = \mathbb{\Gamma}\mathbb{B}$, an inclusion relation holds: $\mathcal{M}^{\text{DB}} \subseteq \mathcal{M}^{\text{CB}} \subseteq \mathcal{M}^{\text{ST}}$. It should be noted that, depending on the details of $j(x)$, these manifolds can be empty.

A steady state is a state at which $\dot{x} = 0$ holds. The DB condition $j(x) = \mathbf{0}$ means that all the fluxes are zero at x. In other words, all the forward and reverse fluxes are balanced at x, i.e., $j_e^+(x) = j_e^-(x)$. The CB condition is equivalent to the balance of all influx and outflux at each hypervertex of $\mathbb{H}$. As we will see later, DB states are tightly linked to the equilibrium state and equilibrium flow. The CB state is relevant as an extension of the equilibrium state to nonequilibrium flows.

4.4 Generalized gradient flow and De Giorgi's formulation

When $f(x)$ can be represented as a gradient, i.e., $f(x) = \text{grad}_{\mathbb{S}}\partial\mathcal{F}(x)$ of a function $\mathcal{F}(x) \in \mathbb{R}$ on the density space, Eq. 45 is reduced to the generalized gradient flow of $\mathcal{F}(x)$.

Definition 23 (*Generalized gradient flow*) $x(t)$ is a generalized gradient flow when it is a generalized flow driven by a gradient force of $\mathcal{F}(x)$, i.e., $f(x) = \text{grad}_{\mathbb{S}}\partial\mathcal{F}(x)$ and

$$\dot{x} = -\text{div}_{\mathbb{S}}j(x) = -\text{div}_{\mathbb{S}}\partial\Psi_x^*[\text{grad}_{\mathbb{S}}\partial\mathcal{F}(x)]. \tag{49}$$

The following proposition ensures that the generalized gradient flow behaves like the conventional gradient flow:

Proposition 3 (*$\mathcal{F}(x)$ is non-increasing along the trajectory of generalized gradient flow*) *For a trajectory $\{x_t\}_{t \in [0,\tau]}$ of the generalized gradient flow of $\mathcal{F}(x)$, $\mathcal{F}(x_t)$ is always decreasing except at the DB states $\mathcal{M}^{\text{DB}}$. In addition, all the steady states of the generalized gradient flow are the DB states, i.e., $\mathcal{M}^{\text{ST}} = \mathcal{M}^{\text{DB}}$*[46].

Proof $\mathcal{F}(x_t)$ is non-increasing over time as follows:

$$\begin{aligned}
\dot{\mathcal{F}}(x_t) &= \langle \dot{x}, \partial_x\mathcal{F}(x) \rangle = -\langle \text{div}_{\mathbb{S}}\partial\Psi_x^*[f(x)], \partial_x\mathcal{F}(x) \rangle \\
&= -\langle \partial\Psi_x^*[f(x)], \text{grad}_{\mathbb{S}}\partial_x\mathcal{F}(x) \rangle \\
&= -\langle j(x), f(x) \rangle = -\left(\Psi_x[j(x)] + \Psi_x^*[f(x)]\right) \leq 0,
\end{aligned} \tag{50}$$

where Eq. 40 is used. The equality holds iff $f(x) = \text{grad}_{\mathbb{S}}\partial\mathcal{F}(x) = 0$ because $\Psi_x^*[f(x)] = \Psi_x[j(x)] = 0$ iff $f(x) = j(x) = 0$. Thus, $\dot{\mathcal{F}}(x_t) = 0$ iff $x_t \in \mathcal{M}^{\text{DB}}$. Because $\dot{x}_t = 0 \Rightarrow \dot{\mathcal{F}}(x_t) = 0$, $\mathcal{M}^{\text{ST}} = \mathcal{M}^{\text{DB}}$. □

---

[46] $\mathcal{M}^{\text{ST}} = \mathcal{M}^{DB} = \emptyset$ can hold, e.g., when $\mathcal{F}(x)$ is a strictly monotonous function.

It should be noted that, even if $\mathcal{F}(\boldsymbol{x})$ has a single minimum, the steady state $\boldsymbol{x}_{st} := \lim_{t \to \infty} \boldsymbol{x}(t)$ may not be the minimum, because $\dot{\mathcal{F}}(\boldsymbol{x}_t) = 0$ holds for any $\boldsymbol{x} \in \mathcal{M}^{\mathrm{DB}}$.[47]

The generalized gradient flow of this form (Eq. 49) was devised in the process to extend the conventional gradient flow to metric spaces [122, 123].[48] Furthermore, dissipation functions have been recognized since the seminal work of Onsager [124–126]. However, only quadratic dissipation functions have been investigated until very recently [75–82]. This may be partly because we lack an adequate geometric language to handle the non-quadratic cases, i.e., information geometry. Actually, if the dissipation function is quadratic $\Psi_{\boldsymbol{x}}^{q,*}[\boldsymbol{f}]$ as in Eq. 43, then the generalized flow (Eq. 45) formally reduces to the flow on a Riemannian manifold with the metric $(\mathbb{S}M_{\dot{\boldsymbol{x}}}^{*}\mathbb{S}^{T})^{-1}$.

The non-negativity of $\dot{\mathcal{F}}(\boldsymbol{x}_t)$ is essentially attributed to the fact that $\dot{\mathcal{F}}(\boldsymbol{x}_t) = -\langle \boldsymbol{j}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x}) \rangle$ holds in Eq. 50 for the generalized gradient flow. The converse also holds.

Proposition 4 (De Giorgi's formulation of generalized gradient flow *[75, 79]*) *Let $\boldsymbol{x}_t$ be a generalized flow induced by a force $\boldsymbol{f}(\boldsymbol{x})$. $\boldsymbol{x}_t$ is the generalized gradient flow of $\mathcal{F}(\boldsymbol{x})$ iff*

$$\dot{\mathcal{F}}(\boldsymbol{x}_t) = -\langle \boldsymbol{j}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x}) \rangle = -\left(\Psi_{\boldsymbol{x}}[\boldsymbol{j}(\boldsymbol{x})] + \Psi_{\boldsymbol{x}}^{*}[\boldsymbol{f}(\boldsymbol{x})] \right). \tag{51}$$

holds. The integral form of Eq. 51,

$$\mathcal{F}(\boldsymbol{x}_0) - \mathcal{F}(\boldsymbol{x}_t) = \int_0^t \left[\Psi_{\boldsymbol{x}_{t'}}^{*}(\boldsymbol{f}(\boldsymbol{x}_{t'})) + \Psi_{\boldsymbol{x}_{t'}}(\boldsymbol{j}(\boldsymbol{x}_{t'})) \right] \mathrm{d}t', \tag{52}$$

is called De Giorgi's (Ψ, Ψ^*)-formulation of generalized gradient flow.

Proof For a generalized flow $\boldsymbol{x}_t$ driven by force $\boldsymbol{f}(\boldsymbol{x})$ as in Eq. 45 and for any $\mathcal{F}(\boldsymbol{x})$, the following inequality holds:

$$\dot{\mathcal{F}}(\boldsymbol{x}_t) = \left\langle \dot{\boldsymbol{x}}, \frac{\partial \mathcal{F}(\boldsymbol{x}_t)}{\partial \boldsymbol{x}} \right\rangle = \left\langle -\boldsymbol{j}(\boldsymbol{x}_t), \mathrm{grad}_{\mathbb{S}} \frac{\partial \mathcal{F}(\boldsymbol{x}_t)}{\partial \boldsymbol{x}} \right\rangle \tag{53}$$

$$= -\left[\Psi_{\boldsymbol{x}_t}(\boldsymbol{j}(\boldsymbol{x}_t)) + \Psi_{\boldsymbol{x}_t}^{*}(\boldsymbol{f}'(\boldsymbol{x}_t)) \right]$$
$$+ \mathcal{D}_{\boldsymbol{x}_t}^{\mathcal{J}, \mathcal{F}}[\boldsymbol{j}(\boldsymbol{x}_t); \boldsymbol{f}'(\boldsymbol{x}_t)] \tag{54}$$

$$\geq -\left[\Psi_{\boldsymbol{x}_t}(\boldsymbol{j}(\boldsymbol{x}_t)) + \Psi_{\boldsymbol{x}_t}^{*}(\boldsymbol{f}'(\boldsymbol{x}_t)) \right], \tag{55}$$

where we define $\boldsymbol{f}'(\boldsymbol{x}) := \mathrm{grad}_{\mathbb{S}} \partial \mathcal{F}(\boldsymbol{x})$. The last inequality becomes an equality if and only if $\boldsymbol{f}'(\boldsymbol{x}_t)$ is the Legendre dual of $\boldsymbol{j}(\boldsymbol{x}_t)$,[49] i.e.,

$$\mathcal{D}_{\boldsymbol{x}_t}^{\mathcal{J}, \mathcal{F}}[\boldsymbol{j}(\boldsymbol{x}_t); \boldsymbol{f}'(\boldsymbol{x}_t)] = 0 \iff \boldsymbol{j}(\boldsymbol{x}_t) = \partial \Psi_{\boldsymbol{x}_t}^{*}[\boldsymbol{f}'(\boldsymbol{x}_t)] \tag{56}$$

---

[47] In addition, there exists the possibility that $\boldsymbol{x}(t)$ converges to the boundary of $\mathcal{X}$.

[48] The metric here means a general metric, which is not restricted to one associated with the inner product.

[49] These inequality and equality conditions are usually derived by using Cauchy-Schwarz inequality [127]. From the information-geometric framework, they are trivially attributed to the non-negativity of Bregman divergence.

Thus, Eq. 51 holds only when x_t is the generalized gradient flow of $\mathcal{F}(x)$. □

De Giorgi's formulation is a well-established approach for defining gradient flow in metric spaces [122].

4.5 Equilibrium and nonequilibrium flow

In this work, we mainly focus on the case that $\mathcal{F}(x) = \mathcal{D}_\Phi^{\mathcal{X}}[x \| \tilde{x}]$ where $\mathcal{D}_\Phi^{\mathcal{X}}$ is the Bregman divergence associated with a thermodynamic function Φ.

Definition 24 (*Equilibrium force, equilibrium flux, and equilibrium flow*) The force generated by the gradient of Bregman divergence associated with a thermodynamic function Φ is called the (thermodynamic) equilibrium force, and the following equation is denoted as the thermodynamic gradient equation:

$$f(x) = \mathrm{grad}_{\mathbb{S}} \partial \mathcal{D}_\Phi^{\mathcal{X}}[x \| \tilde{x}], \tag{57}$$

where $\tilde{x} \in \mathcal{X}$ is a parameter. The dual of $f(x)$, i.e., $j(x) = \partial \Psi_x^*[f(x)]$, is called the equilibrium flux: A generalized flow $x(t)$ is an equilibrium flow if it is driven by the equilibrium force:

$$\dot{x} = -\mathrm{div}_{\mathbb{S}} \partial \Psi_x^*[\mathrm{grad}_{\mathbb{S}} \partial \mathcal{D}_\Phi^{\mathcal{X}}[x \| \tilde{x}]]. \tag{58}$$

Using the relation $\partial \mathcal{D}_\Phi^{\mathcal{X}}[x \| \tilde{x}] = \partial \Phi(x) - \tilde{y}$ where $\tilde{y} = \partial \Phi(\tilde{x})$, Eq. 58 can be rewritten as

$$\dot{x} = -\mathrm{div}_{\mathbb{S}} \left[\partial \Psi_x^* \left[\mathrm{grad}_{\mathbb{S}} \left\{ \partial \Phi(x) - \tilde{y} \right\} \right] \right], \tag{59}$$

which explicitly shows the contribution of both the thermodynamic function and the dissipation function to the dynamics (Fig. 3a).

Various properties of the equilibrium flow (Eq. 58) can be obtained from the doubly dual flat structure as we will see in the following sections. In addition, the equilibrium flow captures the properties that the dynamics of thermodynamic equilibrium systems should have. In this sense, the equilibrium flow is the mathematical representation of the dynamics of equilibrium systems.

Beyond the gradient equilibrium flow, we also consider the non-gradient nonequilibrium flow of the following type:

Definition 25 (*Nonequilibrium force and nonequilibrium flow*) The force generated by a shift of the equilibrium force

$$f(x) = \mathrm{grad}_{\mathbb{S}} \partial \mathcal{D}_\Phi^{\mathcal{X}}[x \| \tilde{x}] + f_{NE}, \tag{60}$$

is called nonequilibrium force if $f_{NE} \notin \mathrm{Im}[\mathbb{S}^T]$.[50] If the shift f_{NE} satisfies $f_{NE} \in \mathrm{Im}[\mathbb{S}^T]$, then $f(x)$ is reduced to the equilibrium force $f_{NE} = 0$ by appropriately

---

[50] In physics, such f_{NE} can be identified with a nonequilibrium force applied externally to the system.
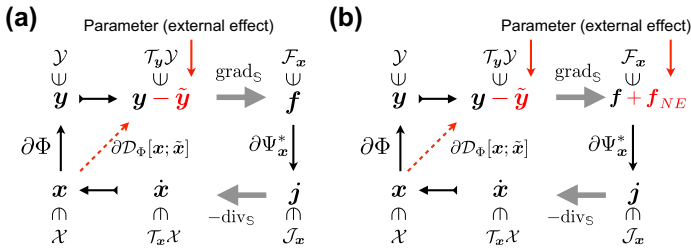
Fig. 3 Schematic representation of the equilibrium (a) and nonequilibrium flow (b)

changing $\tilde{x}$. The nonequilibrium flow is the flow induced by the nonequilibrium force (Fig. 3b):

$$\dot{x} = -\mathrm{div}_{\mathbb{S}} \partial \Psi_x^* \left[\left[\mathrm{grad}_{\mathbb{S}} \partial_x \mathcal{D}_\Phi^{\mathcal{X}}[x \| \tilde{x}] \right] + f_{NE} \right]. \tag{61}$$

In the next section, we show that this equation can cover a sufficiently wide class of models, e.g., all types of rLDG and CRN with extended LMA kinetics. Equation 61 can also be associated with nonequilibrium dynamics with a constant environmental force. The techniques in information geometry, Hessian geometry, and convex analysis enable us to investigate such non-gradient dynamics.

Remark 6 (Variational modeling [128]) We introduced and characterized dynamics based on the thermodynamic functions and dissipation functions. While we employed a restricted definition in order to link dynamics to information geometry, we may further generalize this approach by appropriately choosing the state space, $f(x)$, $\Psi_x^*(f)$, and $\mathrm{div}_{\mathbb{S}}$. For example, we may consider a x-dependent and noninteger-valued matrix for $\mathbb{S}(x)$. The equilibrium flow may not be restricted to $\mathcal{F}(x) = \mathcal{D}_\Phi^{\mathcal{X}}[x \| \tilde{x}]$, and the nonequilibrium flow may be defined for x-dependent $f_{NE}(x)$. This type of approach for modeling dissipative dynamics has been known as variational modeling.

Before closing this section, we mention that the existence of DB states, i.e., $\mathcal{M}^{DB} \neq \emptyset$, is necessary and sufficient for a nonequilibrium flow to be an equilibrium flow.

Proposition 5 (Detailed balance condition and equilibrium flow*) Consider a flow given by Eq. 61. If $\mathcal{M}^{DB} \neq \emptyset$, then the flow is equilibrium, i.e., $f_{NE} \in \mathrm{Im}\mathbb{S}^T$.*

Proof $\mathcal{M}^{DB} \neq \emptyset$ means that there exists x_{DB} satisfying $j(x_{DB}) = 0$. Then we have $j(x_{DB}) = 0 \Leftrightarrow f(x_{DB}) = 0$. If $f_{NE} \notin \mathrm{Im}[\mathbb{S}^T]$, $f_{NE} \neq 0$ and thus $f(x) \neq 0$ for all $x \in \mathcal{X}$. Thus, $f_{NE} \in \mathrm{Im}[\mathbb{S}^T]$ if $\mathcal{M}^{DB} \neq \emptyset$.

The necessity follows basically from Prop. 3, but we have to show $\mathcal{M}^{ST} \neq \emptyset$. This will be shown in the following section (Lemma 1).

5 Explicit form of thermodynamic and dissipation functions

Before investigating the dynamics of the equilibrium (Eq. 58) and nonequilibrium (Eq. 61) flows, we show how the flows can be associated with the dynamics on graphs

and hypergraphs via specific forms of the thermodynamic and dissipation functions. The forms of functions depend on the functional form of the flux that we assume: Eq. 3 for rLDG, Eq. 12 for CRN with LMA kinetics, and Eq. 18 for FPE. It should be noted that the choice of the thermodynamic function and the dissipation function is not unique for a given dynamics in general. Depending on the purpose, we should choose or find an appropriate set of functions.

5.1 Explicit form of thermodynamic functions for rLDG and CRN

For rLDG (Eq. 3) and CRN with LMA kinetics (Eq. 12), the following pair of thermodynamic functions is particularly relevant[51]:

$$\Phi(\boldsymbol{x}) := \left[\ln \boldsymbol{x} - \ln \boldsymbol{x}^o - \mathbf{1}\right]^T \boldsymbol{x} = \sum_{i=1}^{N_{\mathbb{X}}} \left[\ln \frac{x_i}{x_i^o} - 1\right] x_i,$$

$$\Phi^*(\boldsymbol{y}) := (\boldsymbol{x}^o)^T e^{\boldsymbol{y}} = \sum_{i=1}^{N_{\mathbb{X}}} x_i^o e^{y_i}, \tag{62}$$

which induce the following Legendre transformation:

$$\boldsymbol{y} = \partial \Phi(\boldsymbol{x}) = \ln \boldsymbol{x} - \ln \boldsymbol{x}^o, \qquad \boldsymbol{x} = \partial \Phi^*(\boldsymbol{y}) = \boldsymbol{x}^o \circ e^{\boldsymbol{y}}. \tag{63}$$

Here, $\mathcal{Y} = \mathbb{R}^{N_{\mathbb{X}}}$, and $\boldsymbol{x}^o \in \mathcal{X}$ is a parameter determining the point in $\mathcal{X}$ that is associated with the origin of $\mathcal{Y}$ via the Legendre transformation. For these thermodynamic functions, the Bregman divergence is reduced to the generalized Kullback-Leibler divergence.

$$\mathcal{D}^{\mathcal{X}}[\boldsymbol{x}\|\boldsymbol{x}'] = \left(\ln \frac{\boldsymbol{x}}{\boldsymbol{x}'}\right)^T \boldsymbol{x} - \mathbf{1}^T (\boldsymbol{x} - \boldsymbol{x}'). \tag{64}$$

These thermodynamic functions and the generalized KL divergence are separable.

If we choose $\boldsymbol{x}^o = \mathbf{1}$, then the conventional dual representation for the probability density $\boldsymbol{p}$ on a discrete space is recovered:

$$\begin{aligned}\Phi(\boldsymbol{p}) &= \left[\ln \boldsymbol{p} - \mathbf{1}\right]^T \boldsymbol{p}, & \Phi^*(\boldsymbol{y}) &= \mathbf{1}^T e^{\boldsymbol{y}}, \\ \boldsymbol{y} &= \partial_{\boldsymbol{p}} \Phi(\boldsymbol{p}) = \ln \boldsymbol{p}, & \boldsymbol{p} &= \partial_{\boldsymbol{y}} \Phi^*(\boldsymbol{y}) = e^{\boldsymbol{y}}.\end{aligned} \tag{65}$$

---

[51] For CRN, these forms of the thermodynamic functions are derived from the conventional thermodynamics of ideal gas or dilute solution with non-reactive solvent [48]. Mathematically, we may employ other functions as we introduce different information geometric structures onto a family of probabilities depending on the purpose. Such exploitation is an interesting open problem.

In this case, $\mathcal{Y}$ is the space of the logarithm of $\boldsymbol{p}$. These representations hold even if $\boldsymbol{p}$ is not a probability density. If $\boldsymbol{p}$ satisfies $\mathbf{1}^T \boldsymbol{p} = 1$, the generalized KL divergence becomes the normal KL divergence $\mathcal{D}^{\mathcal{X}}[\boldsymbol{p} \| \boldsymbol{p}'] = \left(\ln \frac{\boldsymbol{p}}{\boldsymbol{p}'}\right)^T \boldsymbol{p}$.[52]

5.2 Explicit form of dissipation functions for rLDG and CRN

To determine the dissipation functions, we need the definition of force, which may depend on the phenomena and purpose.[53] In physics, the flux-force relations, which are also called constitutive equations [129], are central because they determine what kind of change is induced by an incurred force.[54] For rMJP and CRNs, the flux and force are conventionally defined using the one-way fluxes, $\boldsymbol{j}^+(\boldsymbol{x})$ and $\boldsymbol{j}^-(\boldsymbol{x})$ as

$$\boldsymbol{j} = \boldsymbol{j}^+ - \boldsymbol{j}^-, \qquad\qquad \boldsymbol{f} = \ln \boldsymbol{j}^+ - \ln \boldsymbol{j}^-, \qquad (66)$$

where the dependency of $\boldsymbol{j}^{\pm}(\boldsymbol{x})$ on $\boldsymbol{x}$ is abbreviated for notational simplicity. In physics, assuming this form of force-flux relation goes by the name of the local detailed balance (LDB) assumption,[55] or the generalized detailed balance assumption.[56] By defining the frenetic activity [132]:

$$\boldsymbol{\omega} := 2\sqrt{\boldsymbol{j}^+ \circ \boldsymbol{j}^-} \in \mathbb{R}_{\geq 0}^{N_{\mathrm{e}}}, \qquad (67)$$

we have a relation $\boldsymbol{j} = \boldsymbol{\omega} \circ \left[\frac{\exp(\boldsymbol{f}/2) - \exp(-\boldsymbol{f}/2)}{2}\right]$. For a fixed $\boldsymbol{\omega}$, this relation between the pair $(\boldsymbol{j}, \boldsymbol{f})$ is a one-to-one Legendre duality induced by the following specific form of dissipation functions:

$$\begin{aligned}
\Psi_{\boldsymbol{\omega}}^*(\boldsymbol{f}) &:= 2\boldsymbol{\omega}^T \left[\cosh(\boldsymbol{f}/2) - \mathbf{1}\right], \\
\Psi_{\boldsymbol{\omega}}(\boldsymbol{j}) &:= 2\boldsymbol{\omega}^T \left(\mathrm{diag}\left[\frac{\boldsymbol{j}}{\boldsymbol{\omega}}\right] \sinh^{-1}\left(\frac{\boldsymbol{j}}{\boldsymbol{\omega}}\right) - \left[\sqrt{1 + \left(\frac{\boldsymbol{j}}{\boldsymbol{\omega}}\right)^2} - \mathbf{1}\right]\right),
\end{aligned} \qquad (68)$$

which lead to the Legendre transformation:

$$\boldsymbol{j} = \partial \Psi_{\boldsymbol{\omega}}^*(\boldsymbol{f}) = \boldsymbol{\omega} \circ \sinh(\boldsymbol{f}/2), \qquad \boldsymbol{f} = \partial \Psi_{\boldsymbol{\omega}}(\boldsymbol{j}) = 2 \sinh^{-1}\left(\frac{\boldsymbol{j}}{\boldsymbol{\omega}}\right). \qquad (69)$$

---

[52] As we will see later, the condition $\mathbf{1}^T \boldsymbol{p}(t) = 1$ need not be assumed but is automatically satisfied due to the topological constraint of the graph and the initial condition $\mathbf{1}^T \boldsymbol{p}(0) = 1$ when we work on rMJP.

[53] This is parallel to the problem of how to define the dual of $\boldsymbol{x}$. The choice of logarithm is contingent on the domain and knowledge of physics and statistics.

[54] Some relations were obtained empirically through experiments and others were computed theoretically from microscopic models.

[55] LDB assumption is different from the DB condition in Def. 22.

[56] The validity of LDB was shown for rMJP and CRN with LMA kinetics via large deviation theory for the corresponding microscopic Markovian models or via its consistency with the macroscopic chemical thermodynamics [130, 131].

We can easily verify that these functions satisfy the conditions for dissipation functions, i.e., Eq. 34, Eq. 35, and Eq. 33.

For the flux $j_{MA}(x)$ of LMA kinetics (Eq. 12), the force and activity become[57]

$$f_{MA}(x; K) = \left[\ln K + \mathbb{S}^T \ln x \right], \qquad \omega_{MA}(x; \kappa) = 2\kappa \circ x^{[\lceil (B^+ + B^-) \rceil]^T / 2}. \qquad (70)$$

where we introduced a transformation of the kinetic parameters (k^+, k^-) into the force part K and activity part κ as $\kappa := \sqrt{k^+ \circ k^-}$ and $K := k^+ / k^-$.[58] Because $k^{\pm} = \kappa \circ K^{\pm 1/2}$ holds, (κ, K) has the same information as (k^+, k^-). Moreover, we can verify that the force and activity are dependent only on K and κ, respectively. The dissipation functions of the forms above and their relations to rLDG and CRN were derived from the large deviation function of the corresponding microscopic stochastic models [75, 133]. Actually, the Bregman divergence $\mathcal{D}_x^{\mathcal{J}}[j; j_{MA}(x)]$ of the dissipation functions is identical to the rate function of the flux for rMJP and CRN. Thus, these dissipation functions are keystones connecting macroscopic and microscopic dynamics.

If there exists $\tilde{y}$ satisfying $-\mathbb{S}^T \tilde{y} = \ln K$, i.e., $\ln K \in \text{Im} \mathbb{S}^T$, the force in Eq. 70 is represented as

$$f_{MA}(x; K) = \text{grad}_{\mathbb{S}} \left(\ln \frac{x}{\tilde{x}} \right) = \text{grad}_{\mathbb{S}} \partial_x \mathcal{D}^{\mathcal{X}}[x \| \tilde{x}] \in \text{Im} \mathbb{S}^T, \qquad (71)$$

where $\tilde{x}$ is the Legendre conjugate of $\tilde{y}$.[59] Thus, CRN (and rMJP) is an equilibrium flow of the generalized KL divergence $\mathcal{D}^{\mathcal{X}}[x \| \tilde{x}]$ when the parameter K satisfies $\ln K \in \text{Im} \mathbb{S}^T$. In chemistry, the condition $\ln K \in \text{Im} \mathbb{S}^T$ is called Wegscheider's equilibrium condition [47, 134], and the CRN satisfying this parametric condition is called equilibrium CRN.[60] Even if $\ln K \in \text{Im} \mathbb{S}^T$ is not satisfied, we can represent $\ln K = -\mathbb{S}^T \tilde{y} + f_{NE}$ with $f_{NE} \notin \text{Im} \mathbb{S}^T$. The force in Eq. 70 is always represented as

$$f_{MA}(x; K) = \text{grad}_{\mathbb{S}} \left(\ln \frac{x}{\tilde{x}} \right) + f_{NE} = \left[\text{grad}_{\mathbb{S}} \left[\partial_x \mathcal{D}^{\mathcal{X}}[x \| \tilde{x}] \right] + f_{NE} \right], \qquad (72)$$

which leads to the nonequilibrium flow (Eq. 61). Thus, CRN with LMA kinetics as well as rLDG are generally within the class of Eq. 61.

---

[57] The dissipation functions in Eq. 68 and the induced Legendre transformation in Eq. 69 are not necessarily restricted to these particular types of force and activity. Actually, the extended LMA kinetics (Eq. 13) can also be represented by replacing $\omega_{MA}(x)$ with $\omega_{eMA}(x) = g(x) \circ \omega_{MA}(x)$. Thus, Eq. 68 could be applied to a wider class of kinetics than Eq. 12.

[58] For CRN, K is referred as the equilibrium constant in chemistry.

[59] It should be noted that, while $\tilde{y}$ is not uniquely determined by K in general, it does not cause problems. This is clarified in the following section (Sect. 8) by introducing appropriate affine subspaces.

[60] Historically, the equilibrium chemical systems were characterized by macroscopic thermodynamics. The equilibrium condition was derived as the necessary and sufficient condition that the flux of the LMA kinetics (Eq. 12) should satisfy to have consistent properties with the thermodynamic equilibrium systems. It was found only recently that the equilibrium properties are mathematically attributed to the generalized gradient flow structure.

Example 2 (Simplified Brusselator CRN [8, 104] (continued)) For the Brusselator CRN introduced in Ex. 1, the force and activity defined in Eq. 70 can be explicitly represented as

$$
f(x) = \overbrace{\begin{pmatrix} \ln(k_1^+/k_1^-) \\ \ln(k_2^+/k_2^-) \\ \ln(k_3^+/k_3^-) \end{pmatrix}}^{\ln K} + \overbrace{\begin{pmatrix} -\ln x_1 \\ \ln x_1 - \ln x_2 \\ \ln x_2 - \ln x_1 \end{pmatrix}}^{\mathbb{S}^T \ln x},
$$

$$
\omega(x) = 2 \overbrace{\begin{pmatrix} \sqrt{k_1^+ k_1^-} \\ \sqrt{k_2^+ k_2^-} \\ \sqrt{k_3^+ k_3^-} \end{pmatrix}}^{\kappa} \circ \overbrace{\begin{pmatrix} \sqrt{x_1} \\ \sqrt{x_1 x_2} \\ \sqrt{x_1^5 x_2} \end{pmatrix}}^{x^{[\mathbb{\Gamma}(B^+ + B^-)]^T/2}}. \tag{73}
$$

Remark 7 (Wegscheider's equilibrium condition and Detailed balance condition) While we defined equilibrium flow by the specific functional form of force and obtained Wegscheider's equilibrium condition as the necessary and sufficient condition to have the equilibrium force under LMA kinetics, the equilibrium dynamics is often defined by the existence of the steady state satisfying the DB condition (Eq. 48) in CRN theory. In addition, the DB condition is also often assumed in statistics when we design or analyze a random walk in parameter spaces, e.g., in the Markov Chain Monte Carlo (MCMC) simulations or in other random-walk-based optimization schemes.[61] These two are equivalent for (extended) LMA kinetics. Actually, $\mathcal{M}^{\mathrm{DB}} \neq \emptyset$ means that there exists $x_{DB} \in \mathcal{X}$ such that $j_{MA}(x_{DB}) = 0 \Leftrightarrow -\mathbb{S}^T \ln x_{DB} = \ln K$. From the Fredholm alternative, we obtain the Wegscheider's equilibrium condition $\ln K \in \mathrm{Im}\mathbb{S}^T$ for the existence of x_{DB}.

Remark 8 (Linear graph Laplacian dynamics) The linear graph Laplacian dynamics defined by Eq. 22 can be formally regarded as a generalized flow. From the form of the graph Laplacian (Eq. 21),[62] it is easy to see that Eq. 22 coincides with Eq. 58 if

$$
\Phi(x) = \frac{1}{2}\langle x, M_0 x \rangle, \qquad \Psi_x^{*,q}(f) = \frac{1}{2}\langle f, M^1 f \rangle, \tag{74}
$$

where $M_0 = I$, $M^1 = \mathrm{diag}[k]$, and $\tilde{x} = 0$. In contrast to rLDG, the natural state space and the corresponding dual is $\mathcal{X} = \mathcal{Y} = \mathbb{R}^{N_v}$.[63] In [23], non-quadratic general $\Phi(x)$ is considered as a class of nonlinear diffusion on a network from information geometric viewpoint.

---

[61] The DB condition is conventionally adopted because, for example, it makes it easy to obtain an MCMC with a desirable stationary distribution. This nice property comes from the gradient-flow property of the equilibrium flow.

[62] It should be noted that this representation holds only when $k^+ = k^-$ holds.

[63] Even if we restrict the dynamics to $\mathcal{X} = \mathcal{Y} = \mathbb{R}^{N_v}_{>0}$, no problem arises for defining the generalized flow as long as we do not consider projections that we are going to introduce.

5.3 Some remarks on the dissipation functions for rLDG and CRN

The dissipation functions in Eq. 68 have several notable properties. First, they are separable:

$$\Psi^*_{\boldsymbol{\omega}(\boldsymbol{x})}(\boldsymbol{f}) = \sum_{e=1}^{N_\text{e}} \omega_e(\boldsymbol{x})\psi^*(f_e), \qquad \Psi_{\boldsymbol{\omega}(\boldsymbol{x})}(\boldsymbol{j}) = \sum_{e=1}^{N_\text{e}} \omega_e(\boldsymbol{x})\psi(j_e/\omega_e(\boldsymbol{x})), \qquad (75)$$

where

$$\psi^*(f) := 2\left[\cosh(f/2) - 1\right] \in [0, \infty), \qquad (76)$$

$$\psi(j) := 2\left(j\sinh^{-1}(j) - \left[\sqrt{1 + \bar{j}^2} - 1\right]\right) \in [0, \infty). \qquad (77)$$

and $\boldsymbol{\omega}(\boldsymbol{x})$ is local: $\omega_e(\boldsymbol{x}) = 2\kappa_e \prod_{i=1}^{N_\text{X}} x_i^{(\gamma_{i,e}^+ + \gamma_{i,e}^-)/2}$. The thermodynamic functions in Eq. 62 are also separable.[64]

Second, the scalar function $\psi^*(f)$ is the N-function. The N-function of the $(\cosh(f) - 1)$-type and the associated Orlicz space have been employed for establishing the infinite-dimensional information geometry by Pistone [72, 135, 136]. In functional analysis, the Orlicz space is a generalization of the L^p spaces, which arise naturally when we work on the $L\log^+ L$ space for the divergences and large deviation functions. Hence, the dissipation functions in Eq. 68 are tightly related to such topics.

Third, various information geometric measures and quantities are related to the dissipation functions in Eq. 68 and also to the associated quantities as follows:

$$\frac{1}{4}\Psi^*_{\boldsymbol{\omega}}(\boldsymbol{f}) = \frac{1}{2}\sum_{e=1}^{N_\text{e}}\left[\sqrt{j_e^+} - \sqrt{j_e^-}\right]^2 =: \mathcal{D}_{Hel}[\boldsymbol{j}^+; \boldsymbol{j}^-]^2 \qquad (78)$$

$$\frac{1}{2}\mathbf{1}^T\boldsymbol{\omega} = \sum_{e=1}^{N_\text{e}}\sqrt{j_e^+ j_e^+} =: \text{BC}[\boldsymbol{j}^+; \boldsymbol{j}^-] \qquad (79)$$

$$\langle \boldsymbol{j}, \boldsymbol{f} \rangle = \sum_{e=1}^{N_\text{e}}(j_e^+ - j_e^-)\ln\frac{j_e^+}{j_e^-} =: \mathcal{D}_{Jef}[\boldsymbol{j}^+; \boldsymbol{j}^-], \qquad (80)$$

where $\mathcal{D}_{Hel}[\boldsymbol{j}^+; \boldsymbol{j}^-]$, $\text{BC}[\boldsymbol{j}^+; \boldsymbol{j}^-]$, and $\mathcal{D}_{Jef}[\boldsymbol{j}^+; \boldsymbol{j}^-]$ are the Hellinger–Kakutani distance, the Bhattacharyya coefficient, and the Jeffreys divergence (symmetrized KL divergence) for $\boldsymbol{j}^+$ and $\boldsymbol{j}^-$, respectively. In addition, in physics, the bilinear pairing $\langle \boldsymbol{j}, \boldsymbol{f} \rangle$ of a Legendre dual pair and its approximation using the Hessian matrix are often referred to as the entropy production rate (EPR) $\dot{\Sigma}$ and pseudo-entropy production rate (pEPR) $\dot{\Sigma}^p$, respectively [83, 137]:

---

[64] The locality and separability may sound natural. However, from the physical viewpoint, the Onsager matrix can have nondiagonal components, which implies nonseparable dissipation functions. In addition, equilibrium thermodynamics does not preclude thermodynamic functions from being nonseparable.

$$\dot{\Sigma} := \langle \boldsymbol{j}, \boldsymbol{f} \rangle = \sum_{e=1}^{N_{\mathrm{e}}} (j_e^+ - j_e^-) \ln \frac{j_e^+}{j_e^-}. \tag{81}$$

$$\dot{\Sigma}^p := \langle \boldsymbol{j}, G_{\boldsymbol{\omega}, \boldsymbol{j}} \boldsymbol{j} \rangle = 2 \sum_{e=1}^{N_{\mathrm{e}}} \frac{(j_e^+ - j_e^-)^2}{j_e^+ + j_e^-}, \tag{82}$$

where we treat $\boldsymbol{j} \in \mathcal{J}_{\boldsymbol{x}}$ as a member of $\mathcal{T}_{\boldsymbol{j}} \mathcal{J}_{\boldsymbol{x}}$ by the isomorphism: $\mathcal{J}_{\boldsymbol{x}} \cong \mathcal{T}_{\boldsymbol{j}} \mathcal{J}_{\boldsymbol{x}} \cong C_1(\mathbb{H})$. The pEPR $\dot{\Sigma}^p$ is an approximation of EPR by replacing $\boldsymbol{f} = \partial \Psi_{\boldsymbol{\omega}}(\boldsymbol{j})$ with $G_{\boldsymbol{\omega}, \boldsymbol{j}} \boldsymbol{j}$ and works as a lower bound of $\dot{\Sigma}$: $\dot{\Sigma} \geq \dot{\Sigma}^p$ [137][65].

Finally, the dissipation functions in Eq. 68 are not the unique choice to reproduce the force-flux relation in Eq. 66. The quadratic dissipation functions $\Psi_{\boldsymbol{x}}^{q,*}(\boldsymbol{f}) := \frac{1}{2} \langle \boldsymbol{f}, M_{\boldsymbol{x}}^* \boldsymbol{f} \rangle$ in Eq. 43 with the following diagonal metric tensor can reproduce the relation in Eq. 66:

$$M_{\boldsymbol{x}}^* = \mathrm{diag} \left[\frac{\boldsymbol{j}^+(\boldsymbol{x}) - \boldsymbol{j}^-(\boldsymbol{x})}{\ln \boldsymbol{j}^+(\boldsymbol{x}) - \ln \boldsymbol{j}^-(\boldsymbol{x})} \right] = \mathrm{diag} \left[\left(\frac{j_e^+(\boldsymbol{x}) - j_e^-(\boldsymbol{x})}{\ln j_e^+(\boldsymbol{x}) - \ln j_e^-(\boldsymbol{x})} \right)_e \right]. \tag{83}$$

This type of quadratic dissipation function was proposed even earlier than the non-quadratic ones [138–140] and has been investigated [104, 141, 142]. Its advantage is that the induced geometry is Riemannian, and thus the information geometric argument is not necessarily required. In addition, this Riemannian geometric structure is analogous to the formal Riemannian geometric structure of FPE and other diffusion processes on continuous manifolds induced via the L^2-Wasserstein geometry [65, 66] (Fig. 4). Thus, this quadratic dissipation function provides a consistent extension of these results for FPE and diffusion processes to graphs and hypergraphs. Nevertheless, the doubly dual flat structure with the non-quadratic dissipation functions that we introduce is also another sound generalization of the formal Riemannian geometry of FPE, as we see in the next subsection.

As long as we focus only on the trajectory of the generalized flow (Eq. 45), the difference between the quadratic and non-quadratic functions does not matter because both induce the same dynamics. However, the Bregman divergence of the quadratic dissipation functions is not directly related to the rate function of the microscopic stochastic models, while that of nonquadratic ones in Eq. 68 is [133]. Thus, if we consider projections of fluxes and forces in the edge spaces, different choices of dissipation functions lead to different projections. In addition, for non-quadratic dissipation functions, the contributions of the kinetic parameters $\boldsymbol{k}^{\pm}$ can be clearly separated into the force part $\boldsymbol{K}$ and the activity part $\boldsymbol{\kappa}$ in the case of CRN with the LMA kinetics (Eq. 70). This separation enables a physical realization of the projected flux as we derive in the following section.

---

[65] This inequality is obtained directly from the inequality $2(a-b)^2/(a+b) \leq (a-b) \ln a/b$ for $a, b > 0$.
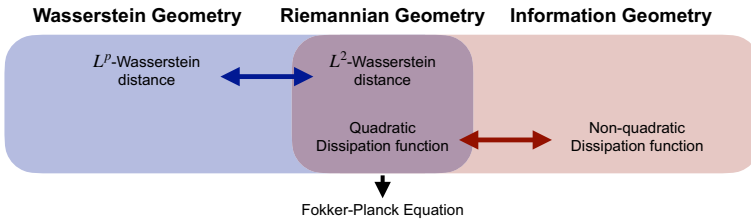
Fig. 4 A relationship between Wasserstein geometry and information geometry. The formal Riemannian geometric structure appears at their intersection. It should be noted that, while the regions of L^p-Wasserstein distance for $p \neq 2$ and nonquadratic dissipation function are not overlapping in this figure, this does not mean that they are unrelated. There may be undiscovered relations between these two regions

5.4 Explicit forms of thermodynamic and dissipation functions for FPE

For FPE, the dualistic representation of the density $p(\boldsymbol{r})$ and its logarithm $y(\boldsymbol{r}) = \ln p(\boldsymbol{r})$ is also relevant. This duality is induced formally by the following thermodynamic functions[66]:

$$\Phi[p] = \int [\ln p(\boldsymbol{r}) - \mathbf{1}]p(\boldsymbol{r})\mathrm{d}\boldsymbol{r}, \qquad \Phi^*[y] = \int e^{y(\boldsymbol{r})}\mathrm{d}\boldsymbol{r}, \qquad (84)$$

the Legendre transformations of which are

$$y(\boldsymbol{r}) = \frac{\delta \Phi[p]}{\delta p} = \ln p(\boldsymbol{r}), \qquad p(\boldsymbol{r}) = \frac{\delta \Phi^*[y]}{\delta y} = e^{y(\boldsymbol{r})}. \qquad (85)$$

The Bregman divergence becomes the KL divergence $\mathcal{D}_{\mathcal{X}}[p\|p'] = \int \mathrm{d}\boldsymbol{r}\, p(\boldsymbol{r}) \ln \frac{p(\boldsymbol{r})}{p'(\boldsymbol{r})}$. In physics, the flux and force for FPE are defined conventionally as

$$\boldsymbol{j}_{\mathrm{FP}}[p(\boldsymbol{r})] = D_0 p(\boldsymbol{r})\left\{\boldsymbol{F}(\boldsymbol{r})/D_0 - \nabla \ln p(\boldsymbol{r})\right\}, \qquad (86)$$

$$\boldsymbol{f}_{\mathrm{FP}}[p(\boldsymbol{r})] = D_0^{-1}\boldsymbol{F}(\boldsymbol{r}) - \nabla \ln p(\boldsymbol{r}). \qquad (87)$$

The dissipation functions associated with the force-flux relation above are

$$\Psi_{\boldsymbol{\omega}[p]}^{\mathrm{FP},*}[f] = \frac{1}{2}\int \boldsymbol{f}(p(\boldsymbol{r}))^T M_{p(\boldsymbol{r})}^* \boldsymbol{f}(p(\boldsymbol{r}))\mathrm{d}\boldsymbol{r},$$

$$\Psi_{\boldsymbol{\omega}[p]}^{\mathrm{FP}}[j] = \frac{1}{2}\int \boldsymbol{j}(p(\boldsymbol{r}))^T M_{p(\boldsymbol{r})} \boldsymbol{j}(p(\boldsymbol{r}))\mathrm{d}\boldsymbol{r}, \qquad (88)$$

where $M_{p(\boldsymbol{r})}^* := \mathrm{diag}[\boldsymbol{\omega}[p(\boldsymbol{r})]]$, $M_{p(\boldsymbol{r})} := (M_{p(\boldsymbol{r})}^*)^{-1}$, and $\omega_i[p(\boldsymbol{r})] = D_0 p(\boldsymbol{r})$. Thus, the dissipation functions are formally quadratic and positive definite. If $\boldsymbol{F}(\boldsymbol{r})$ is a gradient of $U(\boldsymbol{r})$ as $\boldsymbol{F}(\boldsymbol{r}) = D_0\nabla U(\boldsymbol{r})$, $\boldsymbol{f}_{\mathrm{FP}}[p(\boldsymbol{r})] = -\nabla \ln \frac{p(\boldsymbol{r})}{\tilde{p}(\boldsymbol{r})}$ holds where $\tilde{p}(\boldsymbol{r}) := \exp[U(\boldsymbol{r})]$. Then, the dissipation functions, the bilinear pairing $\langle \boldsymbol{j}_{\mathrm{FP}}, \boldsymbol{f}_{\mathrm{FP}}\rangle$, the EPR $\dot{\Sigma}_{\mathrm{FP}}$ in Eq. 81, and the pEPR $\dot{\Sigma}_{\mathrm{FP}}^p$ in Eq. 82 formally consolidate into the same quantity:

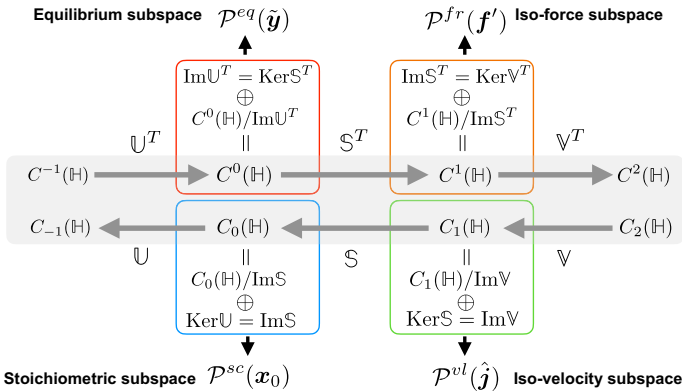[66] The base measure is omitted because this is just a formal one.

Fig. 5 Diagrammatic representation of the four subspaces and their relationship with the chain and cochain complexes of $\mathbb{H}$

$$2\Psi^{\text{FP},*}_{\omega[p]}[f_{\text{FP}}] = 2\Psi^{\text{FP}}_{\omega[p]}[j_{\text{FP}}] = \langle j_{\text{FP}}, f_{\text{FP}} \rangle = \dot{\Sigma}_{\text{FP}} = \dot{\Sigma}^{p}_{\text{FP}}$$

$$= D_0 \int p(r) \left(\nabla_r \ln \frac{p(r)}{\tilde{p}(r)} \right)^2 dr. \tag{89}$$

The last quantity without D_0 is known as relative Fisher information [66, 143] and Hyvärinen divergence [120, 144] between p and $\tilde{p}$. For $U(f) = 0$, it reduces to the Fisher information number $\mathbb{I}_F[p]$ in Eq. 1. This consolidation is a source of confusion, because the same quantity for FPE or linear diffusion processes has different names in different contexts and in different disciplines. However, they actually have different definitions, roles, and meanings, which become explicit in the information-geometric formulation.

6 Orthogonal subspaces, dual foliations, and Pythagorean relation

To investigate the behaviors and properties of the equilibrium (Eq. 58) and nonequilibrium (Eq. 61) flow, especially its topological and algebraic constraints from the graph or hypergraph structure, information geometry provides the ideal tools. In particular, the four affine subspaces associated with the cycle and cocycle subspaces of the chain and cochain complexes (Fig. 5) form dual foliations via the Legendre transformation, whose geometric properties are captured by information geometry [1, 145]. It should be noted that the results of this section do not assume the specific forms of the thermodynamic and dissipation functions introduced in Sect. 5.

6.1 Four affine subspaces

Two families of orthogonally complement affine subspaces are naturally introduced on $\mathcal{X}$ and $\mathcal{Y}$, respectively, from the topological structure of graph and hypergraph, i.e., $\mathbb{B}$ and $\mathbb{S}$.

Definition 26 (*Stoichiometric subspaces in $\mathcal{X}$*) The stoichiometric subspaces are defined as[67]

$$\mathcal{P}^{sc}(\boldsymbol{x}_0) := \{\boldsymbol{x} \in \mathcal{X} | \boldsymbol{x} - \boldsymbol{x}_0 \in \mathrm{Im}\mathbb{S}\}, \quad \boldsymbol{x}_0 \in \mathcal{X} \tag{90}$$

where $\boldsymbol{x}_0$ is a parameter to specify the position of the subspace (Fig. 5, lower left)[68].

Definition 27 (*Equilibrium subspaces in $\mathcal{Y}$*) The equilibrium subspaces (Fig. 5, upper left) are defined as

$$\mathcal{P}^{eq}(\tilde{\boldsymbol{y}}) := \left\{ \boldsymbol{y} \in \mathcal{Y} | \boldsymbol{y} - \tilde{\boldsymbol{y}} \in \mathrm{Ker}\mathbb{S}^T \right\}, \quad \tilde{\boldsymbol{y}} \in \mathcal{Y}. \tag{91}$$

$\mathcal{P}^{sc}(\boldsymbol{x}_0)$ and $\mathcal{P}^{eq}(\tilde{\boldsymbol{y}})$ are of orthogonal complement to each other: $\langle \boldsymbol{x} - \boldsymbol{x}_0, \boldsymbol{y}' - \tilde{\boldsymbol{y}} \rangle = 0$ for $\boldsymbol{x} \in \mathcal{P}^{sc}(\boldsymbol{x}_0)$ and $\boldsymbol{y}' \in \mathcal{P}^{eq}(\tilde{\boldsymbol{y}})$.[69] Because $\mathbb{S}$ and $\mathbb{S}^T$ are the discrete differentials, δ_1 and δ^0, $\mathcal{P}^{sc}(\boldsymbol{x}_0)$ and $\mathcal{P}^{eq}(\tilde{\boldsymbol{y}})$ are associated with the 0-cycle and 0-cocycle spaces, respectively.

Two other families of orthogonal-complement subspaces are introduced on $\mathcal{J}_{\boldsymbol{x}}$ and $\mathcal{F}_{\boldsymbol{x}}$.

Definition 28 (*Iso-velocity subspaces in $\mathcal{J}_{\boldsymbol{x}}$*) The iso-velocity subspaces (Fig. 5, lower right) are defined as

$$\mathcal{P}^{vl}(\hat{\boldsymbol{j}}) = \left\{ \boldsymbol{j} \in \mathcal{J}_{\boldsymbol{x}} | \boldsymbol{j} - \hat{\boldsymbol{j}} \in \mathrm{Ker}\mathbb{S} \right\}, \quad \hat{\boldsymbol{j}} \in \mathcal{J}_{\boldsymbol{x}}. \tag{92}$$

Definition 29 (*Iso-force subspaces in $\mathcal{F}_{\boldsymbol{x}}$*) The iso-external-force subspaces, iso-force subspaces in short, (Fig. 5, upper right) are defined as

$$\mathcal{P}^{fr}(\boldsymbol{f}') := \left\{ \boldsymbol{f} \in \mathcal{F}_{\boldsymbol{x}} | \boldsymbol{f} - \boldsymbol{f}' \in \mathrm{Im}\mathbb{S}^T \right\}, \quad \boldsymbol{f}' \in \mathcal{F}_{\boldsymbol{x}}. \tag{93}$$

Again, from the correspondence of $\delta_1 = \mathbb{S}$ and $\delta^0 = \mathbb{S}^T$, $\mathcal{P}^{vl}(\hat{\boldsymbol{j}})$ and $\mathcal{P}^{fr}(\boldsymbol{f}')$ are associated with the 1-cycle and 1-cocycle spaces, respectively. We specifically call $\mathcal{P}^{vl}(\boldsymbol{0})$ and $\mathcal{P}^{fr}(\boldsymbol{0})$ zero-velocity subspace and equilibrium force subspace, respectively.

6.2 Meaning of the subspaces

All four subspaces are natural constituents in the theory of algebraic graph theory and homological algebra. Here, we provide their meaning in terms of the dynamics on graphs and hypergraphs.

---

[67] In CRN theory, a stoichiometric subspace is called stoichiometric compatibility class [8].

[68] Because $\mathcal{P}^{sc}(\boldsymbol{x}_0)$ is restricted within the positive orthant $\mathcal{X}$, $\mathcal{P}^{sc}(\boldsymbol{x}_0)$ is a polyhedron. If bounded, it is called a polytope in discrete geometry and also in combinatorial optimization [146]. However, we abuse the word (affine) subspace for $\mathcal{P}^{sc}(\boldsymbol{x}_0)$, and use polyhedron or polytope when we care about the boundary.

[69] In this work, orthogonality always means the orthogonal complement in dual vector spaces except otherwise stated.

The stoichiometric and iso-velocity subspaces, $\mathcal{P}^{sc}(x_0)$ and $\mathcal{P}^{vl}(\hat{j})$, are related by the continuity equation (Eq. 10). From the continuity equation, $\mathcal{P}^{vl}(\hat{j})$ is the set of fluxes that induce the same velocity as a reference $\hat{j}$ does: $j \in \mathcal{P}^{vl}(\hat{j}) \iff \dot{x} = -\mathbb{S}\hat{j} = -\mathbb{S}j$. Thereby, $\mathcal{P}^{vl}(\hat{j})$ is parametrized as follows:

$$\mathcal{P}^{vl}(\dot{x}) = \{j \in \mathcal{J}_x \mid -\mathbb{S}j = \dot{x}\}, \quad \dot{x} \in \text{Im}[\mathbb{S}] = \text{Ker}[\mathbb{U}], \tag{94}$$

This subspace is crucial to characterize fluxes that can realize the same dynamics as the reference one.

The stoichiometric subspace $\mathcal{P}^{sc}(x_0)$ determines the subspace in which the dynamics are algebraically constrained via the topology of the underlying graph or hypergraph. Because $\dot{x} = -\mathbb{S}j(x(t))$, for an initial state $x(0) = x_0$, $x(t) - x_0 \in \text{Im}[\mathbb{S}]$ should hold, meaning that $x(t) \in \mathcal{P}^{sc}(x_0)$. Thus, $\mathcal{P}^{sc}(x_0)$ is the subspace in which the dynamics are restricted by the initial condition x_0. $\mathcal{P}^{sc}(x_0)$ can also be represented parametrically by the quantities which are conserved by the dynamics. For any vector $u \in \text{Ker}\mathbb{S}^T$, $\eta(t) := u^T x(t)$ is constant over time:

$$\dot{\eta}(t) = \frac{\mathrm{d}u^T x(t)}{\mathrm{d}t} = u^T \frac{\mathrm{d}x(t)}{\mathrm{d}t} = -u^T \mathbb{S}j(x) = 0. \tag{95}$$

In Sect. 3.1, we defined a matrix $\mathbb{U}$ by a complete basis of $\text{Ker}\mathbb{S}^T$ so that $\text{Im}\mathbb{U}^T = \text{Ker}\mathbb{S}^T$. Using $\mathbb{U}$, the conserved quantities for a given initial condition x_0 are obtained as $\eta = \mathbb{U}x_0 = \mathbb{U}x(t)$. Because $\text{Im}\mathbb{U}$ is isomorphic to $C_{-1}(\mathbb{H})$, the stoichiometric subspace is explicitly parametrized by the conserved quantities (an element of $C_{-1}(\mathbb{H})$):

$$\mathcal{P}^{sc}(\eta) = \{x \in \mathcal{X} \mid \mathbb{U}x = \eta\}, \quad \eta \in C_{-1}(\mathbb{H}). \tag{96}$$

For rMJP, the conserved quantity is reduced to the conservation of probability $\mathbf{1}^T p(t) = 1$ and $\mathcal{P}^{sc}(p_0)$ becomes the probability simplex. Because $\text{Ker}\mathbb{B}^T$ determines the connected components of the graph $\mathbb{G}$ and we conventionally assume that the underlying graph is connected in rMJP, we only have the one-dimensional cokernel space and one conserved quantity, which is $\eta = 1$. Thus, the conservation of probability or, equivalently, the restriction of p in the probability simplex is automatically guaranteed from the topological constraint of the dynamics if we start from an initial state satisfying $\mathbf{1}^T p_0 = 1$.

The iso-force subspace $\mathcal{P}^{fr}(f')$ and the equilibrium subspace $\mathcal{P}^{eq}(\tilde{y})$ are related to the equilibrium and nonequilibrium force equations, Eq. 57 and Eq. 60. The equilibrium force defined in Eq. 57 satisfies $f(x) \in \text{Im}\mathbb{S}^T = \mathcal{P}^{fr}(0)$. Thus, the equilibrium-force subspace $\mathcal{P}^{fr}(0)$ is literally the set of equilibrium forces. $\mathcal{P}^{fr}(f')$ is its shift by $f' \in \mathcal{F}_x$. Using $\mathbb{V}$ defined in Sect. 3.1, we can represent $\mathcal{P}^{fr}$ parametrically as

$$\mathcal{P}^{fr}(\zeta) = \{f \in \mathcal{F}_x \mid \mathbb{V}^T f = \zeta\}, \quad \zeta \in C^2(\mathbb{H}) \tag{97}$$

because $\mathcal{F}_x / \text{Im}\mathbb{S}^T \cong \mathcal{F}_x / \text{Ker}\mathbb{V}^T \cong \text{Im}\mathbb{V}^T \cong C^2(\mathbb{H})$. Thus, ζ characterizes the type of nonequilibrium forces quotient by the equilibrium forces.
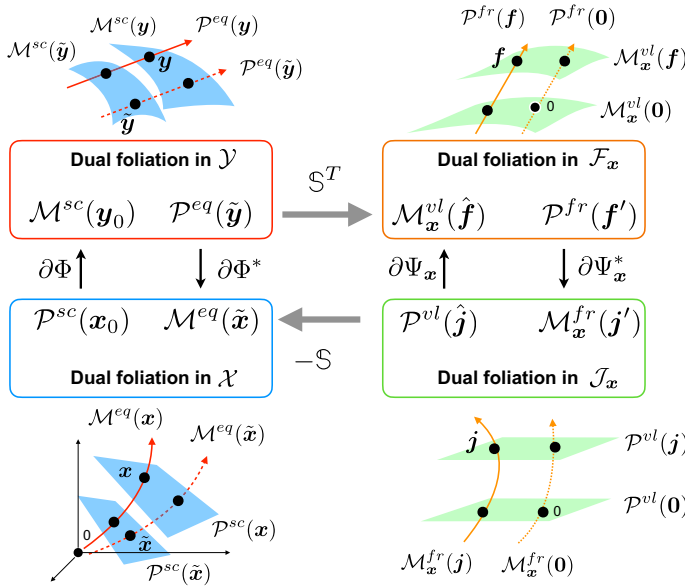
Fig. 6 Diagrammatic representation of the dual foliations in $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{J}_x$, and $\mathcal{F}_x$ spaces

Finally, the equilibrium subspace $\mathcal{P}^{eq}(\tilde{\boldsymbol{y}})$ can also be regarded as the set of potentials $\boldsymbol{y}$ that generate the same equilibrium force because any $\boldsymbol{y} \in \mathcal{P}^{eq}(\tilde{\boldsymbol{y}})$ satisfies $\boldsymbol{f}' = \mathbb{S}^T \boldsymbol{y} = \mathbb{S}^T \tilde{\boldsymbol{y}} \in \mathcal{P}^{fr}(\boldsymbol{0})$. Due to this, the equilibrium subspace $\mathcal{P}^{eq}$ is parameterized as

$$\mathcal{P}^{eq}(\boldsymbol{f}') = \{\boldsymbol{y} \in \mathcal{Y} | \mathbb{S}^T \boldsymbol{y} = \boldsymbol{f}'\}, \quad \boldsymbol{f}' \in \text{Im}[\mathbb{S}^T]. \tag{98}$$

The parametric forms of the subspaces are summarized as follows:

$$\mathcal{P}^{vl}(\dot{\boldsymbol{x}}) = \{\boldsymbol{j} \in \mathcal{J}_x | -\mathbb{S}\boldsymbol{j} = \dot{\boldsymbol{x}}\}, \qquad \dot{\boldsymbol{x}} \in \text{Im}[\mathbb{S}] = \text{Ker}[\mathbb{U}], \tag{99}$$

$$\mathcal{P}^{sc}(\boldsymbol{\eta}) = \{\boldsymbol{x} \in \mathcal{X} | \mathbb{U}\boldsymbol{x} = \boldsymbol{\eta}\}, \qquad \boldsymbol{\eta} \in \text{Im}[\mathbb{U}] = C_{-1}(\mathbb{H}), \tag{100}$$

$$\mathcal{P}^{fr}(\boldsymbol{\zeta}) = \{\boldsymbol{f} \in \mathcal{F}_x | \mathbb{V}^T \boldsymbol{f} = \boldsymbol{\zeta}\}, \qquad \boldsymbol{\zeta} \in \text{Im}[\mathbb{V}^T] = C^2(\mathbb{H}), \tag{101}$$

$$\mathcal{P}^{eq}(\boldsymbol{f}') = \{\boldsymbol{y} \in \mathcal{Y} | \mathbb{S}^T \boldsymbol{y} = \boldsymbol{f}'\}, \qquad \boldsymbol{f}' \in \text{Im}[\mathbb{S}^T] = \text{Ker}[\mathbb{V}^T]. \tag{102}$$

From these subspaces, we can obtain dual foliations on the vertex and edge spaces.

6.3 Dual manifolds, dual foliations, and Pythagorean relation in vertex spaces

For the subspaces $\mathcal{P}^{sc}$ and $\mathcal{P}^{eq}$ in the density and potential spaces, we introduce their Legendre transformation via the thermodynamic functions, $\Phi(\boldsymbol{x})$ and $\Phi^*(\boldsymbol{y})$, which form the dual foliation with the subspaces of orthogonal complement (Fig. 6, left).

Definition 30 (*Stoichiometric manifold in $\mathcal{Y}$ and equilibrium manifold in $\mathcal{X}$*) The stoichiometric and equilibrium manifolds (Fig. 6, left) are defined respectively as

$$\mathcal{M}^{sc}(\mathbf{y}_0) := \partial\Phi[\mathcal{P}^{sc}(\mathbf{x}_0)] \subset \mathcal{Y}, \quad \mathbf{y}_0 = \partial\Phi(\mathbf{x}_0), \tag{103}$$

$$\mathcal{M}^{eq}(\tilde{\mathbf{x}}) := \partial\Phi^*[\mathcal{P}^{eq}(\tilde{\mathbf{y}})] \subset \mathcal{X}, \quad \tilde{\mathbf{x}} = \partial\Phi^*(\tilde{\mathbf{y}}). \tag{104}$$

Lemma 1 (*Dual foliations in density and potential spaces* [48]) $\mathcal{P}^{sc}$ and $\mathcal{M}^{eq}$ are foliations of $\mathcal{X}$, and $\mathcal{M}^{sc}$ and $\mathcal{P}^{eq}$ are foliations of $\mathcal{Y}$. For each pair of $(\mathbf{x}_0, \tilde{\mathbf{x}})$, the intersection of $\mathcal{P}^{sc}(\mathbf{x}_0)$ and $\mathcal{M}^{eq}(\tilde{\mathbf{x}})$ is unique and transversal. The same applies to $\mathcal{M}^{sc}(\mathbf{y}_0)$ and $\mathcal{P}^{eq}(\tilde{\mathbf{y}})$. Then, $(\mathcal{P}^{sc}, \mathcal{M}^{eq})$ and $(\mathcal{M}^{sc}, \mathcal{P}^{eq})$ form dual foliations (nonlinear coordinate systems) in $\mathcal{X}$ and $\mathcal{Y}$ spaces, respectively.

Proof The polyhedron $\mathcal{P}^{sc}(\mathbf{x}_0)$ and the affine subspace $\mathcal{P}^{eq}(\tilde{\mathbf{y}})$ can cover the whole $\mathcal{X}$ and $\mathcal{Y}$ by changing $\mathbf{x}_0$ and $\tilde{\mathbf{y}}$, respectively. Similarly, $\mathcal{M}^{eq}(\tilde{\mathbf{x}})$ and $\mathcal{M}^{sc}(\mathbf{y}_0)$ can cover the whole $\mathcal{X}$ and $\mathcal{Y}$ because Legendre transformations by the thermodynamic functions are one-to-one between $\mathcal{X}$ and $\mathcal{Y}$. Consider the intersection of $\mathcal{P}^{sc}(\mathbf{x}_0)$ and $\mathcal{M}^{eq}(\tilde{\mathbf{x}})$ in $\mathcal{X}$ space. The condition that $\mathcal{P}^{sc}(\mathbf{x}_0) \cap \mathcal{M}^{eq}(\tilde{\mathbf{x}}) \neq \emptyset$ is related to the existence of $\mathbf{x}^\dagger$ defined by the following convex optimization problem:

$$\mathbf{x}^\dagger := \arg \min_{\mathbf{x} \in \overline{\mathcal{P}^{sc}(\mathbf{x}_0)}} \mathcal{D}_\Phi^{\mathcal{X}}[\mathbf{x}\|\tilde{\mathbf{x}}]. \tag{105}$$

Because of the properties of $\Phi(\mathbf{x})$, $\mathcal{D}_\Phi^{\mathcal{X}}[\mathbf{x}\|\tilde{\mathbf{x}}]$ and its restriction to $\overline{\mathcal{P}^{sc}(\mathbf{x}_0)}$ are strictly convex with respect to $\mathbf{x}$. Thus, $\mathbf{x}^\dagger$ is unique and either satisfies the stationarity condition $\mathbf{x}^\dagger \in \mathcal{P}^{sc}(\mathbf{x}_0) \cap \mathcal{M}^{eq}(\tilde{\mathbf{x}})$ if $\mathbf{x}^\dagger \in \mathcal{P}^{sc}(\mathbf{x}_0)$ or locates on the boundary $\partial\mathcal{X}$ if $\mathbf{x}^\dagger \notin \mathcal{P}^{sc}(\mathbf{x}_0)$, where we used $\mathbb{S}^T \frac{\partial\mathcal{D}^{\mathcal{X}}[\mathbf{x}\|\tilde{\mathbf{x}}]}{\partial\mathbf{x}} = \mathbf{0} \Leftrightarrow \mathbb{S}^T(\mathbf{y} - \tilde{\mathbf{y}}) = \mathbf{0} \Leftrightarrow \mathbf{y} \in \mathcal{P}^{eq}(\tilde{\mathbf{y}}) \Leftrightarrow \mathbf{x} \in \mathcal{M}^{eq}(\tilde{\mathbf{x}})$. Let $\mathbf{x}_{bd} \in \partial\mathcal{X}$ and $\mathbf{x}_{in} \in \mathcal{X}$ be arbitrary points on the boundary and interior of $\mathcal{X}$. From the condition Eq. 25 of the thermodynamic function, for $\mathbf{x}_\lambda := \lambda\mathbf{x}_{in} + (1-\lambda)\mathbf{x}_{bd}$ where $\lambda \in [0, 1]$,

$$\lim_{\lambda\to+0} \frac{d\mathcal{D}^{\mathcal{X}}[\mathbf{x}_\lambda\|\tilde{\mathbf{x}}]}{d\lambda} = \lim_{\lambda\to+0}\left[\frac{d\Phi(\mathbf{x}_\lambda)}{d\lambda} - \left\langle\tilde{\mathbf{y}}, \frac{d\mathbf{x}_\lambda}{d\lambda}\right\rangle\right] = -\infty. \tag{106}$$

Thus, $\mathbf{x}^\dagger \notin \mathcal{X}$ is excluded, and the intersection exists, i.e., $\mathbf{x}^\dagger \in \mathcal{P}^{sc}(\mathbf{x}_0) \cap \mathcal{M}^{eq}(\tilde{\mathbf{x}})$. The intersection $\mathbf{x}^\dagger$ is unique and transversal because $\langle\mathbf{x}_{sc} - \mathbf{x}^\dagger, \mathbf{y}_{eq} - \mathbf{y}^\dagger\rangle = 0$ holds for any $\mathbf{x}_{sc} \in \mathcal{P}^{sc}(\mathbf{x}_0)$ and $\mathbf{x}_{eq} \in \mathcal{M}^{eq}(\tilde{\mathbf{x}})$ and the dimensions of $\mathcal{P}^{sc}(\mathbf{x}_0)$ and $\mathcal{M}^{eq}(\tilde{\mathbf{x}})$ are complementary because $\mathcal{P}^{sc}(\mathbf{x}_0)$ and $\mathcal{P}^{eq}(\tilde{\mathbf{y}})$ are of orthogonal complement (see also the proof in [83]). As a result, $\mathbf{x}^\dagger \in \mathcal{P}^{sc}(\mathbf{x}_0) \cap \mathcal{M}^{eq}(\tilde{\mathbf{x}})$ always exists, and $(\mathcal{P}^{sc}, \mathcal{M}^{eq})$ forms a dual foliation in $\mathcal{X}$. Also $(\mathcal{M}^{sc}, \mathcal{P}^{eq})$ does in $\mathcal{Y}$ because they are bijective Legendre duals of $(\mathcal{P}^{sc}, \mathcal{M}^{eq})$. $\qquad\square$

This result is reduced to Birch's theorem [48, 100] and the seminal result by Horn and Jackson [41] when the thermodynamic function is the generalized KL divergence.

With the dual foliation, we can consider the generalized Pythagorean relations and orthogonal decomposition. For any three points satisfying $\mathbf{x} \in \mathcal{P}^{sc}(\mathbf{x}_0)$, $\mathbf{x}_q \in$

$\mathcal{M}^{eq}(\tilde{x})$, and $x^\dagger = \mathcal{P}^{sc}(x_0) \cap \mathcal{M}^{eq}(\tilde{x})$[70], we have the generalized Pythagorean relation:

$$\mathcal{D}^{\mathcal{X}}[x\|x_q] = \mathcal{D}^{\mathcal{X}}[x\|x^\dagger] + \mathcal{D}^{\mathcal{X}}[x^\dagger\|x_q]. \tag{107}$$

In $\mathcal{Y}$ space, we also have the dual version of the relations as

$$\mathcal{D}^{\mathcal{Y}}[y_q\|y] = \mathcal{D}^{\mathcal{Y}}[y_q\|y^\dagger] + \mathcal{D}^{\mathcal{Y}}[y^\dagger\|y]. \tag{108}$$

These relations are used to characterize the steady state of equilibrium and nonequilibrium flow geometrically and also variationally.

Remark 9 (Interpretation in terms of statistical inference) The meaning of the equilibrium manifold in statistics can be clarified more explicitly by considering the specific form of thermodynamic function (Eq. 65). For this thermodynamic function, the equilibrium manifold $\mathcal{M}^{eq}(\tilde{p})$ is represented as

$$\begin{aligned} \mathcal{M}^{eq}(\tilde{p}) &= \left\{ p \in \mathcal{X} | \ln p - \ln \tilde{p} \in \mathrm{Ker}\mathbb{S}^T \right\} \\ &= \left\{ p \in \mathcal{X} | p = \tilde{p} \circ \exp\left[\mathbb{U}^T \eta^* \right], \eta^* \in C^{-1}(\mathbb{H}) \right\} \end{aligned} \tag{109}$$

where we use the fact $\mathbb{S}^T \mathbb{U}^T = 0$. Thus, $\mathcal{M}^{eq}(\tilde{p})$ is an exponential family with algebraic constraints via $\mathbb{U}^T$. In contrast, $\mathcal{P}^{sc}(\eta)$ can be regarded as the data manifold, which constrains p by $\eta = \mathbb{U}p$, because $\mathbb{U}p$ can be interpreted as expectation of observables $\{u_\ell\}_{\ell \in [1, N_\emptyset]}$. Thus, the intersection $p^\dagger = \mathcal{P}^{sc}(\eta) \cap \mathcal{M}^{eq}(\tilde{p})$ is the maximum likelihood estimator. The exponential family with linear algebraic constraints as in Eq. 109 appears in algebraic statistics where $\mathbb{U}$ is sometimes called the design matrix [48].[71]

6.4 Dual manifolds, dual foliations in edge spaces and information-geometric extension of Helmholtz-Hodge-Kodaira decomposition

For the edge spaces, we similarly introduce the iso-velocity and iso-force manifolds, which are the duals of $\mathcal{P}^{fr}(f')$ and $\mathcal{P}^{vl}(\hat{j})$, respectively, via the Legendre transformations by the dissipation functions, $\Psi_x(j)$ and $\Psi_x^*(f)$ (Fig. 6, right):

Definition 31 (*Iso-velocity manifold in $\mathcal{F}_x$ and iso-force manifold in $\mathcal{J}_x$*) The iso-velocity and iso-force manifolds (Fig. 6, right) are defined as follows:

$$\mathcal{M}_x^{vl}(\hat{f}) := \partial \Psi_x[\mathcal{P}^{vl}(\hat{j})] \subset \mathcal{F}_x, \qquad \hat{f} = \partial \Psi_x(\hat{j}), \tag{110}$$

$$\mathcal{M}_x^{fr}(j') := \partial \Psi_x^*[\mathcal{P}^{fr}(f')] \subset \mathcal{J}_x, \qquad j' = \partial \Psi_x^*(f'). \tag{111}$$

---

[70] We abuse the notation $x^\dagger = \mathcal{P}^{sc}(x_0) \cap \mathcal{M}^{eq}(\tilde{x})$ because the intersection $\mathcal{P}^{sc}(x_0) \cap \mathcal{M}^{eq}(\tilde{x})$ is a unique point.

[71] In algebraic statistics, $\mathbb{U}$ is explicitly given as constraints of a statistical model. In the dynamics on graphs and hypergraphs, $\mathbb{S}$ explicitly given, and $\mathbb{U}$ is implicitly defined as a complete basis of $\mathrm{Ker}\mathbb{S}^T$. As a result, their connection is not apparently obvious.

It should be noted that $\mathcal{M}_x^{vl}(\hat{\boldsymbol{f}})$ and $\mathcal{M}_x^{fr}(\boldsymbol{j}')$ are dependent on $\boldsymbol{x}$ via the $\boldsymbol{x}$ dependence of the dissipation functions. We obtain the intersections in $\mathcal{J}_x$ and $\mathcal{F}_x$:

$$\boldsymbol{j}^\dagger := \mathcal{P}^{vl}(\hat{\boldsymbol{j}}) \cap \mathcal{M}_x^{fr}(\boldsymbol{j}'), \qquad \boldsymbol{f}^\dagger := \mathcal{M}_x^{vl}(\hat{\boldsymbol{f}}) \cap \mathcal{P}^{fr}(\boldsymbol{f}'), \tag{112}$$

which are also unique and transversal for each $\boldsymbol{x} \in \mathcal{X}$ because $\mathcal{J}_x$ and $\mathcal{F}_x$ are whole vector spaces and the Legendre transformations are one-to-one. Thus, similarly to the case of vertex space, we have the dual foliation:

Lemma 2 (Dual foliations in edge spaces [83]) *For each $\boldsymbol{x} \in X$, $(\mathcal{P}^{vl}, \mathcal{M}_x^{fr})$ and $(\mathcal{M}_x^{vl}, \mathcal{P}^{fr})$ form dual foliations in $\mathcal{J}_x$ and $\mathcal{F}_x$ spaces, respectively.*

For $\hat{\boldsymbol{j}}$ and $\boldsymbol{f}'$, and their intersections $\boldsymbol{j}^\dagger$ and $\boldsymbol{f}^\dagger$ defined in Eq. 112, $\langle \hat{\boldsymbol{j}} - \boldsymbol{j}^\dagger, \boldsymbol{f}^\dagger - \boldsymbol{f}' \rangle = 0$ holds. Thus, we have the generalized Pythagorean relations:

$$\begin{aligned} \mathcal{D}_x^{\mathcal{J}}[\hat{\boldsymbol{j}} \| \boldsymbol{j}'] &= \mathcal{D}_x^{\mathcal{J}}[\hat{\boldsymbol{j}} \| \boldsymbol{j}^\dagger] + \mathcal{D}_x^{\mathcal{J}}[\boldsymbol{j}^\dagger \| \boldsymbol{j}'], \\ \mathcal{D}_x^{\mathcal{F}}[\boldsymbol{f}' \| \hat{\boldsymbol{f}}] &= \mathcal{D}_x^{\mathcal{F}}[\boldsymbol{f}' \| \boldsymbol{f}^\dagger] + \mathcal{D}_x^{\mathcal{F}}[\boldsymbol{f}^\dagger \| \hat{\boldsymbol{f}}]. \end{aligned} \tag{113}$$

In contrast to the thermodynamic functions (Φ, Φ^*), the dissipation functions have symmetry, which makes the origins $\boldsymbol{0}$ in $\mathcal{J}_x$ and $\mathcal{F}_x$ special and leads to an extension of Helmholtz-Hodge-Kodaira decomposition.

Theorem 1 (Information-geometric extension of Helmholtz-Hodge-Kodaira (HHK) decomposition [83]) *For a given flux-force Legendre pair $(\boldsymbol{j}, \boldsymbol{f}) \in (\mathcal{J}_x, \mathcal{F}_x)$, we have their unique $\boldsymbol{x}$-dependent decompositions:*

$$\boldsymbol{j} = \boldsymbol{j}_{eq}(\boldsymbol{x}) + (\boldsymbol{j} - \boldsymbol{j}_{eq}(\boldsymbol{x})), \qquad \boldsymbol{f} = \boldsymbol{f}_{st}(\boldsymbol{x}) + (\boldsymbol{f} - \boldsymbol{f}_{st}(\boldsymbol{x})), \tag{114}$$

such that $\boldsymbol{f}_{eq}(\boldsymbol{x}) \in \mathcal{P}^{fr}(\boldsymbol{0})$, $\boldsymbol{j} - \boldsymbol{j}_{eq}(\boldsymbol{x}) \in \mathcal{P}^{vl}(\boldsymbol{0})$, $\boldsymbol{f} - \boldsymbol{f}_{st}(\boldsymbol{x}), \in \mathcal{P}^{fr}(\boldsymbol{0})$, and $\boldsymbol{j}_{st}(\boldsymbol{x}) \in \mathcal{P}^{vl}(\boldsymbol{0})$ hold. In addition, $\boldsymbol{j}_{eq}(\boldsymbol{x})$ and $\boldsymbol{f}_{st}(\boldsymbol{x})$ are characterized geometrically as

$$\boldsymbol{j}_{eq}(\boldsymbol{x}) := \mathcal{P}^{vl}(\boldsymbol{j}) \cap \mathcal{M}_x^{fr}(\boldsymbol{0}), \qquad \boldsymbol{f}_{st}(\boldsymbol{x}) := \mathcal{M}_x^{vl}(\boldsymbol{0}) \cap \mathcal{P}^{fr}(\boldsymbol{f}). \tag{115}$$

Furthermore, $\boldsymbol{j}_{eq}$ and $\boldsymbol{f}_{st}$ are also characterized variationally as the minimizers of dissipation functions:

$$\boldsymbol{j}_{eq}(\boldsymbol{x}) = \arg \min_{\boldsymbol{j}' \in \mathcal{P}^{vl}(\boldsymbol{j})} \Psi_x(\boldsymbol{j}'), \qquad \boldsymbol{f}_{st}(\boldsymbol{x}) = \arg \min_{\boldsymbol{f}'' \in \mathcal{P}^{fr}(\boldsymbol{f})} \Psi_x^*(\boldsymbol{f}''). \tag{116}$$

Proof The uniqueness of $\boldsymbol{j}_{eq}(\boldsymbol{x})$ and $\boldsymbol{f}_{st}(\boldsymbol{x})$ as intersections in Eq. 115 follows immediately from the property of the dual foliations. Because, for any $\boldsymbol{j}' \in \mathcal{P}^{vl}(\boldsymbol{j})$ and $\boldsymbol{f}'' \in \mathcal{P}^{fr}(\boldsymbol{f})$, $\langle \boldsymbol{j}' - \boldsymbol{j}_{eq}, \boldsymbol{f}_{eq} \rangle = 0$ and $\langle \boldsymbol{j}_{st}, \boldsymbol{f}'' - \boldsymbol{f}_{st} \rangle = 0$ hold, the generalized Pythagorean relations lead to

$$\mathcal{D}_x^{\mathcal{J}}[\boldsymbol{j}' \| \boldsymbol{0}] = \mathcal{D}_x^{\mathcal{J}}[\boldsymbol{j}' \| \boldsymbol{j}_{eq}] + \mathcal{D}_x^{\mathcal{J}}[\boldsymbol{j}_{eq} \| \boldsymbol{0}], \quad \mathcal{D}_x^{\mathcal{F}}[\boldsymbol{f}'' \| \boldsymbol{0}] = \mathcal{D}_x^{\mathcal{F}}[\boldsymbol{f}'' \| \boldsymbol{f}_{st}] + \mathcal{D}_x^{\mathcal{F}}[\boldsymbol{f}_{st} \| \boldsymbol{0}].$$

Because $\mathcal{D}_x^{\mathcal{J}}[j'\|\mathbf{0}] = \Psi_x(j')$ and $\mathcal{D}_x^{\mathcal{F}}[f''\|\mathbf{0}] = \Psi_x^*(f'')$ hold, the relations are reduced to

$$\Psi_x(j') = \mathcal{D}_x^{\mathcal{J}}[j'\|j_{eq}] + \Psi_x(j_{eq}), \quad \Psi_x^*(f'') = \mathcal{D}_x^{\mathcal{F}}[f''\|f_{st}] + \Psi_x^*(f_{st}). \quad (117)$$

Then Eq. 116 follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The decomposed flux j_{eq} and force f_{st} play a particularly important role in dynamics. From the definition, j_{eq} is the equilibrium flux, which induces the same instantaneous velocity $\dot{x}$ as j does, i.e., $\dot{x} = -\text{div}_{\mathbb{S}} j = -\text{div}_{\mathbb{S}} j_{eq}$. Thus, j_{eq} is the equilibrium flux mimicking the instantaneous dynamics induced by the nonequilibrium flux j. This equilibrium flux is uniquely determined owing to the information-geometric orthogonality of $\mathcal{P}^{vl}(j)$ and $\mathcal{M}_x^{fr}(\mathbf{0})$. Moreover, the decomposition $j = j_{eq} + (j - j_{eq})$ can be regarded as an information-geometric extension of the Helmholtz-Hodge-Kodaira decomposition in vector calculus and differential form, because $(j - j_{eq})$ is divergence free, i.e., $\text{div}_{\mathbb{S}}(j - j_{eq}) = 0$, and f_{eq} is a curl-free equilibrium force, i.e., $f_{eq} \in \mathcal{P}^{fr}(\mathbf{0}) = \text{Im}[\mathbb{S}^T] = \text{Ker}[\mathbb{V}^T]$.

On the contrary, by definition, $j_{st} \in \mathcal{P}^v(\mathbf{0})$ is the flux that makes the state x a steady state, i.e., $\dot{x} = 0$, and is also induced by the force in the same quotient set of force $\mathcal{P}^{fr}(f)$ as f. The decomposition $f = f_{st} + (f - f_{st})$ is also a HHK decomposition because $j_{st} \in \mathcal{P}^{vl}(\mathbf{0})$ is divergence free, i.e., $\text{div}_{\mathbb{S}} j_{st} = 0$, and $f - f_{st}$ is a curl-free equilibrium force, i.e., $f - f_{st} \in \mathcal{P}^{fr}(\mathbf{0}) = \text{Im}[\mathbb{S}^T] = \text{Ker}[\mathbb{V}^T]$. [72] These decompositions are used in the subsequent sections (Sect. 8 and Sect. 9).

7 Central affine manifold and Hilbert orthogonality

The dual foliation is an essential geometric object in information geometry. While less common than the dual foliation, the central affine manifold defined by a convex function also plays an integral role in information geometry [145].

Definition 32 (*Central affine manifolds in $\mathcal{J}_x$ and $\mathcal{F}_x$*) The central affine manifolds in $\mathcal{J}_x$ and $\mathcal{F}_x$ are defined as the level sets of $\Psi_x(j)$ and $\Psi_x^*(f)$, respectively[73]:

$$\mathcal{C}_x^{\Psi}(c) := \{j|\Psi_x(j) = c\} \subset \mathcal{J}_x, \quad \mathcal{C}_x^{\Psi^*}(c) := \{f|\Psi_x^*(f) = c\} \subset \mathcal{F}_x, \quad (118)$$

where $c \in \mathbb{R}_{\geq 0}$. For a given $j' \in \mathcal{J}_x$ or $f' \in \mathcal{F}_x$, the manifolds are also denoted as

$$\mathcal{C}_x^{\Psi}(j') := \{j|\Psi_x(j) = \Psi_x(j')\}, \quad \mathcal{C}_x^{\Psi^*}(f') := \{f|\Psi_x^*(f) = \Psi_x^*(f')\}. \quad (119)$$

---

[72] It should be noted that, in general, $j_{st} \neq j - j_{eq}, f_{eq} \neq f - f_{st}, j \neq j_{eq} + j_{st}$, and $f \neq f_{eq} + f_{st}$ hold due to the nonlinearity of Legendre transformation, except when the dissipation functions are quadratic under which the Legendre dual relation is reduced to the linear inner-product relation.

[73] While we introduce the central affine manifolds only on the edge space, they can be defined on the vertex space as well. The central affine manifolds on the vertex space of CRN become fundamental when we work on the isobaric processes in which the volume changes in conjunction with the reactions [85]. In this case, the volume is a global variable affecting all the reactions simultaneously.

Their Legendre transformations are also called (dual) central affine manifolds:

$$\mathcal{M}_x^{\Psi}(c) := \partial\Psi_x[\mathcal{C}_x^{\Psi}(c)] \subset \mathcal{F}_x, \qquad \mathcal{M}_x^{\Psi^*}(c) := \partial\Psi_x^*[\mathcal{C}_x^{\Psi^*}(c)] \subset \mathcal{J}_x. \qquad (120)$$

7.1 Pseudo-Hilbert-isosceles orthogonality and decomposition

By employing the central affine manifold, we can introduce another type of generalized orthogonality:

Definition 33 (*Pseudo-Hilbert-isosceles orthogonality* [78, 79, 81]) Pseudo-Hilbert-isosceles orthogonality between $j_S, j_A \in \mathcal{J}_x$ and between $f_S, f_A \in \mathcal{F}_x$ are defined as follows:

$$j_S \perp_H j_A \iff \Psi_x(j_S + j_A) = \Psi_x(j_S - j_A) \qquad (121)$$
$$f_S \perp_H f_A \iff \Psi_x^*(f_S + f_A) = \Psi_x^*(f_S - f_A). \qquad (122)$$

This orthogonality is motivated by the relation $\|j_S + j_A\|^2 = \|j_S - j_A\|^2$ satisfied by an orthogonal pair $j_S \perp j_A$ under a usual inner product structure and its induced norm $\|\cdot\|^2$.[74] By employing this orthogonality, we obtain pseudo-Hilbert isosceles decompositions of j and f as follows:

Lemma 3 (Positive decompositions of the bilinear pairing via pseudo-Hilbert-isosceles orthogonality [78, 79, 81, 83]) *For a given $j \in \mathcal{J}_x$ and any j' on the same central affine manifold as j, i.e., $j' \in \mathcal{C}_x^{\Psi}(j)$, we obtain the pseudo-Hilbert-isosceles orthogonal decomposition $j = j_S + j_A$:*

$$j_S := \frac{1}{2}(j + j'), \qquad\qquad j_A := \frac{1}{2}(j - j'), \qquad (123)$$

where $j_S \perp_H j_A$ and $j' = j_S - j_A$ hold. In addition, this decomposition induces a positive decomposition of the bilinear product $\langle j, f \rangle = \langle j_S, f \rangle + \langle j_A, f \rangle$ where

$$\langle j_S, f \rangle = \frac{1}{2}\mathcal{D}_x^{\mathcal{J},\mathcal{F}}[j'; -f] \geq 0, \qquad \langle j_A, f \rangle = \frac{1}{2}\mathcal{D}_x^{\mathcal{J},\mathcal{F}}[j'; f] \geq 0, \qquad (124)$$

hold. Similarly, for $f \in \mathcal{F}_x$ and $f'' \in \mathcal{C}_x^{\Psi^*}(f)$, a positive orthogonal decomposition $f = f_S + f_A$ is obtained by $f_S := \frac{1}{2}(f + f'')$ and $f_A := \frac{1}{2}(f - f'')$, which satisfy the associated relations:

$$\langle j, f_S \rangle = \frac{1}{2}\mathcal{D}_x^{\mathcal{J},\mathcal{F}}[j; -f''] \geq 0, \qquad \langle j, f_A \rangle = \frac{1}{2}\mathcal{D}_x^{\mathcal{J},\mathcal{F}}[j; f''] \geq 0. \qquad (125)$$

---

[74] If $\|\cdot\|^2$ is a squared norm that is not necessarily induced by an inner product, the orthogonality is called isosceles or James orthogonality [147]. Here, $\|\cdot\|^2$ is further replaced with the dissipation function, which does not satisfy some conditions required to be a norm.

These decompositions were introduced in [78] for rMJP and extended to CRN in [79, 81], whereas we pointed out its information-geometric aspect in [83]. The decomposition plays a role of characterizing the gradient-flow-like property of non-gradient flows.

8 Information-geometric properties of Equilibrium flow

In this section, we describe several properties of the equilibrium flow (Eq. 58) from the viewpoint of information geometry by employing the objects introduced in the previous sections. Such properties include the existence and uniqueness of the steady state (static property), convergence to the state (kinetic property), and the balance between information-geometric quantities associated with the steady state and convergence along the trajectory (the connection between static and kinetic properties). These properties are consistent with those that thermodynamic equilibrium systems should have. In addition, several results are extensions of the results obtained for FPE in the context of functional analysis, partial differential equations, and optimal transport.

8.1 Properties of equilibrium flow

The following property of the equilibrium state characterizes the static aspect of the equilibrium flow and is fundamentally ascribed to the dually flat structure of density and potential spaces[75]:

Proposition 6 (Equilibrium state and its geometric and variational characterizations) *The steady state of the equilibrium flow x_t (Eq. 58) starting from $x(0) = x_0$ is called the equilibrium state x_{eq}. For each x_0, the equilibrium state is identical to the intersection $x^\dagger = \mathcal{P}^{sc}(x_0) \cap \mathcal{M}^{eq}(\tilde{x})$, i.e., $x_{eq} = x^\dagger$, and thus uniquely exists for a given pair of the initial state x_0 and the parameter of equilibrium flow $\tilde{x}$. The equilibrium state x_{eq} is also characterized variationally as*

$$x_{eq} = \arg\min_{x \in \mathcal{P}^{sc}(x_0)} \mathcal{D}_\Phi^{\mathcal{X}}[x\|\tilde{x}] = \arg\min_{x_q \in \mathcal{M}^{eq}(\tilde{x})} \mathcal{D}_\Phi^{\mathcal{X}}[x_0\|x_q]. \quad (126)$$

Moreover, $\mathcal{M}^{eq}(\tilde{x}) = \mathcal{M}^{DB}$ holds.

Proof From Prop. 3, $x_{eq} \in \mathcal{M}^{DB}$, from which $f(x_{eq}) = 0$ follows. For the equilibrium force (Eq. 57), $\mathcal{M}^{DB} = \{x|f(x) = 0\} = \{x|\mathbb{S}^T(\partial\Phi[x] - \partial\Phi[\tilde{x}]) = 0\} = \mathcal{M}^{eq}(\tilde{x})$ holds. Thus, $x_{eq} \in \mathcal{M}^{eq}(\tilde{x})$. Because the initial state is x_0, $x_{eq} \in \mathcal{P}^{sc}(x_0)$. Thus, $x_{eq} = x^\dagger \in \mathcal{P}^{sc}(x_0) \cap \mathcal{M}^{eq}(\tilde{x})$. The first equality of Eq. 126 is obvious from the proof of the dual foliation (Lemma 1). The second equality is from the generalized Pythagorean relation (Eq. 107). □

The second property of the equilibrium flow is kinetic in nature and characterizes the Bregman divergence as the generalized driving potential, which ensures the con-

---

[75] Actually, this result is independent of the detail of the dissipation functions.

vergence of x_t to the equilibrium state. This property is attributed to the dually flat structure on the edge spaces.

Proposition 7 (Bregman divergence and Gibbs' H-Theorem) *For the trajectory of the equilibrium flow x_t (Eq. 58) starting from $x(0) = x_0$, the thermodynamic function $\mathcal{D}_\Phi^{\mathcal{X}}[x_t \| \tilde{x}]$ decreases, that is, $d\mathcal{D}_\Phi^{\mathcal{X}}[x_t \| \tilde{x}]/dt < 0$ except at $x_t \in \mathcal{M}^{eq}(\tilde{x})$ where $d\mathcal{D}_\Phi^{\mathcal{X}}[x_t \| \tilde{x}]/dt = 0$ holds. Thus, the equilibrium state $x_{eq} \in \mathcal{P}^{sc}(x_0) \cap \mathcal{M}^{eq}(\tilde{x})$ is locally and asymptotically stable.*

Proof By replacing $\mathcal{F}(x)$ in Prop. 3 with $\mathcal{D}_\Phi^{\mathcal{X}}[x \| \tilde{x}]$, we obtain $d\mathcal{D}_\Phi^{\mathcal{X}}[x_t \| \tilde{x}]/dt = -\left[\Psi_{x_t}^*(f(x_t)) + \Psi_{x_t}(j(x_t)) \right] \leq 0$ and the equality holds if and only if $f(x_t) = 0 (\Leftrightarrow x_t \in \mathcal{M}^{eq}(\tilde{x}))$. $\qquad\square$

Because $\mathcal{D}_\Phi^{\mathcal{X}}[x \| \tilde{x}]$ can be identified with the difference of total entropy between x and $\tilde{x}$ for thermodynamic systems such as CRN [85], $d\mathcal{D}_\Phi^{\mathcal{X}}[x_t \| \tilde{x}_{eq}]/dt \leq 0$ corresponds to the nondecreasing property of thermodynamic entropy, which is also referred as Gibbs' H-theorem.[76]

The third property provides a connection between the thermodynamic function and the dissipation function, which is immediately obtained from the De Giorgi's formulation of the generalized gradient flow (Eq. 52):

Proposition 8 (Balancing of thermodynamic function and dissipation function) *For the trajectory of the equilibrium flow x_t (Eq. 58) starting from $x(0) = x_0$, the following relation holds for the thermodynamic function and the dissipation function:*

$$
\mathcal{D}_\Phi^{\mathcal{X}}[x_0 \| \tilde{x}] - \mathcal{D}_\Phi^{\mathcal{X}}[x_t \| \tilde{x}] = \int_{t'=0}^{t} \left[\Psi_{x_{t'}}^*(f(x_{t'})) + \Psi_{x_{t'}}(j(x_{t'})) \right] dt' = \int_{t'=0}^{t} \dot{\Sigma}_{t'} dt',
\tag{127}
$$

In physics and chemistry, this relation means that the difference in the thermodynamic (potential) function between x_t and x_0 (the left-hand side), i.e., the change in total entropy, is equal to the integral of dissipation along x_t (the right-hand side), i.e., the entropy production, for equilibrium systems.

All these results indicate that the equilibrium flow and its properties mathematically abstract the properties of physical equilibrium systems. The equilibrium state x_{eq} is characterized algebraically by the unique intersection of $\mathcal{P}^{sc}(x_0)$ and $\mathcal{M}^{eq}(\tilde{x})$ and also variationally by Eq. 126. The convergence to x_{eq} is guaranteed by $d\mathcal{D}_\Phi^{\mathcal{X}}[x_t \| \tilde{x}]/dt \leq 0$. Furthermore, the entropy-dissipation balance relation (Eq. 127) itself defines the equilibrium system abstractly as the De Giorgi's formulation (Eq. 52) does.

8.2 Induced dually flat structure on tangent–cotangent spaces

The equilibrium state is characterized geometrically and variationally via the information-geometric structure on the vertex spaces $(\mathcal{X}, \mathcal{Y})$ as in Prop. 6. Similarly,

---

[76] We have multiple types of H-theorems. The most famous one is Boltzmann's H theorem, in which the H function is derived from the microscopic dynamics of a system. In Gibbs' H-theorem, H function is obtained by coarse-graining the microscopic system [44]
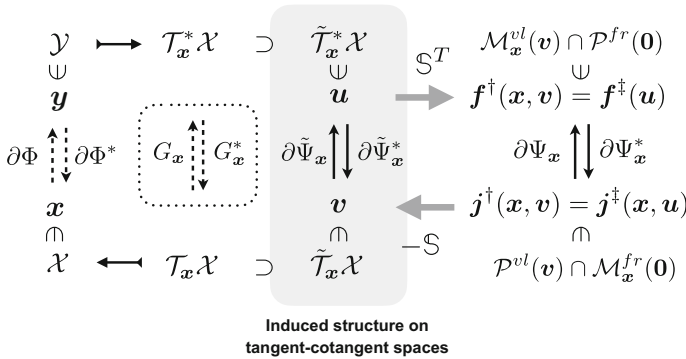
**Induced structure on
tangent-cotangent spaces**

Fig. 7 The induced dually flat structure on the restricted tangent and cotangent spaces from the dissipation functions on the edge spaces (gray region). The relationship is compared with the Riemannian metric on tangent and cotangent spaces via Fisher information matrices induced by the thermodynamic functions (dotted box)

the flux (kinetic law) of equilibrium systems (gradient systems) can be obtained variationally as the flux minimizing the dissipation function under the restriction of the continuity equation.

Lemma 4 (Equilibrium force as the minimizer of primal dissipation function) *For a given trajectory* $\{x_t\}$, *we define the trajectory of the flux* $\{j_t^\dagger\}$ *minimizing the primal dissipation:*

$$\{j_t^\dagger\} := \arg\min_{\{j_t\}} \int_0^t \Psi_{x_{t'}}[j_{t'}]dt', \quad s.t. \ \dot{x}_{t'} + \mathrm{div}_\mathbb{S} j_{t'} = 0 \ for \ all \ t' \in [0, t]. \quad (128)$$

Then, $j_t^\dagger$ *is generated by the equilibrium force,* $f_t^\dagger = \partial\Psi_x[j_t^\dagger] \in \mathcal{P}^{fr}(\mathbf{0})$. *Thus, the minimum primal dissipation flux that generates the given* $\{x_t\}$ *is the equilibrium flux.*

Proof Because the minimization of Eq. 128 can be conducted pointwise-manner for each $t' \in [0, t]$ and $\dot{x}_{t'} + \mathrm{div}_\mathbb{S} j_{t'} = 0 \iff j_{t'} \in \mathcal{P}^{vl}(\dot{x}_{t'})$, we have

$$j_{t'}^\dagger = j^\dagger(x_{t'}, \dot{x}_{t'}) = \arg\min_{j \in \mathcal{P}^{vl}(\dot{x}_{t'})} \Psi_{x_{t'}}[j] = \mathcal{P}^{vl}(\dot{x}_{t'}) \cap \mathcal{M}_{x_{t'}}^{fr}(\mathbf{0}), \quad (129)$$

where we used Eq. 115 and Eq. 116. Thus, from $j_{t'}^\dagger \in \mathcal{M}_{x_{t'}}^{fr}(\mathbf{0}) \iff f_{t'}^\dagger \in \mathcal{P}^{fr}(\mathbf{0})$, the minimum dissipation flux $\{j_t^\dagger\}$ is generated by the equilibrium force, $\{f_t^\dagger\} = \{f^\dagger(x_{t'}, \dot{x}_{t'})\} \in \mathcal{P}^{fr}(\mathbf{0})$ where $f^\dagger(x, \dot{x}) := \partial\Psi_x[j^\dagger(x, \dot{x})]$. $\square$

By exploiting this unique pairing between $\dot{x}_{t'}$ and $j_{t'}^\dagger$ or $\dot{x}_{t'}$ and $f_{t'}^\dagger$, we can obtain an induced dually flat structure on the restricted tangent and cotangent spaces of $\mathcal{X}$ and $\mathcal{Y}$ (Fig. 7), which can be regarded as an information-geometric extension of the Otto structure.

Theorem 2 (Induced dually flat structure on tangent and cotangent spaces) *Let $\tilde{\mathcal{T}}_x\mathcal{X} :=$ Im$\mathbb{S} \cong \mathcal{P}^{sc}(\mathbf{0}) \subset \mathcal{T}_x\mathcal{X}$ and $\tilde{\mathcal{T}}_x^*\mathcal{X} := \mathcal{T}_x^*\mathcal{X}/\mathrm{Ker}\mathbb{S}^T$ be tangent and cotangent spaces on $\mathcal{X}$ restricted by $\mathbb{S}$. On $\tilde{\mathcal{T}}_x\mathcal{X}$ and $\tilde{\mathcal{T}}_x^*\mathcal{X}$, we have the Legendre conjugate dissipation functions $\tilde{\Psi}_x : \tilde{\mathcal{T}}_x\mathcal{X} \to \mathbb{R}$ and $\tilde{\Psi}_x^* : \tilde{\mathcal{T}}_x^*\mathcal{X} \to \mathbb{R}$ induced by the dissipation functions on the edge spaces (Fig. 7).*

Proof By employing Eq. 129, for each $v \in \tilde{\mathcal{T}}_x\mathcal{X}$, we can uniquely determine $j^\dagger(x, v)$, $f^\dagger(x, v) \in \mathcal{P}^{fr}(\mathbf{0})$, and $u^\dagger(x, v) \in \tilde{\mathcal{T}}_x^*\mathcal{X}^{77}$. They satisfy

$$v = -\mathbb{S}j^\dagger(x, v), \quad j^\dagger(x, v) = \partial\Psi_x^*[f^\dagger(x, v)],$$
$$f^\dagger(x, v) = -\mathbb{S}^T u^\dagger(x, v). \tag{130}$$

Conversely, for a given $u \in \tilde{\mathcal{T}}_x^*\mathcal{X}$, we have $f^\ddagger(u)$, $j^\ddagger(x, u)$, and $v^\ddagger(x, u)$ as follows:

$$v^\ddagger(x, u) = -\mathbb{S}j^\ddagger(x, u), \quad j^\ddagger(x, u) = \partial\Psi_x^*[f^\ddagger(u)], \quad f^\ddagger(u) = -\mathbb{S}^T u. \tag{131}$$

Thus, for a pair of $(v, u)_x$ satisfying $u = u^\dagger(x, v)$, we have $v = v^\ddagger(x, u)$, $j^\dagger(x, v) = j^\ddagger(x, u)$, and $f^\dagger(x, v) = f^\ddagger(x, u)$. This pairing establishes a bijection between $\tilde{\mathcal{T}}_x\mathcal{X}$ and $\tilde{\mathcal{T}}_x^*\mathcal{X}$. Moreover, this bijection is realized by the Legendre transformations of the following induced dissipation functions on $\tilde{\mathcal{T}}_x\mathcal{X}$ and $\tilde{\mathcal{T}}_x^*\mathcal{X}$:

$$\tilde{\Psi}_x(v) := \Psi_x(j^\dagger(x, v)), \qquad \tilde{\Psi}_x^*(u) := \Psi_x^*(f^\ddagger(u)). \tag{132}$$

These functions are Legendre conjugate as follows:

$$\max_{u' \in \tilde{\mathcal{T}}_x^*\mathcal{X}} \left[\langle v, u' \rangle - \tilde{\Psi}_x^*(u') \right]$$

$$= \max_{\substack{u' \in \tilde{\mathcal{T}}_x^*\mathcal{X} \\ j \in \mathcal{P}^{vl}(v)}} \left[\langle -\mathbb{S}j, u' \rangle - \tilde{\Psi}_x^*(u') \right] = \max_{\substack{u' \in \tilde{\mathcal{T}}_x^*\mathcal{X} \\ j \in \mathcal{P}^{vl}(v)}} \left[\langle j, -\mathbb{S}^T u' \rangle - \Psi_x^*(-\mathbb{S}^T u') \right]$$

$$= \max_{\substack{f' \in \mathcal{P}^{fr}(\mathbf{0}) \\ j \in \mathcal{P}^{vl}(v)}} \left[\langle j, f' \rangle - \Psi_x^*(f') \right]$$

$$= \max_{\substack{f' \in \mathcal{P}^{fr}(\mathbf{0}) \\ j \in \mathcal{P}^{vl}(v)}} \left[\langle j^\dagger(x, v), f' \rangle - \Psi_x^*(f') + \langle (j - j^\dagger(x, v)), f' \rangle \right]$$

$$= \max_{f' \in \mathcal{P}^{fr}(\mathbf{0})} \left[\langle j^\dagger(x, v), f' \rangle - \Psi_x^*(f') \right] = \Psi_x(j^\dagger(x, v)) = \tilde{\Psi}_x(v),$$

where we used $\langle j - j^\dagger(x, v), f' \rangle = 0$ because $f' \in \mathcal{P}^{fr}(\mathbf{0}) = \mathrm{Im}\mathbb{S}^T$ and $(j - j^\dagger(x, v)) \in \mathrm{Ker}\mathbb{S}$. The inverse is also shown:

$$\max_{v' \in \tilde{\mathcal{T}}_x\mathcal{X}} \left[\langle v', u \rangle - \tilde{\Psi}_x(v') \right] = \max_{v' \in \tilde{\mathcal{T}}_x\mathcal{X}} \left[\langle -\mathbb{S}j^\dagger(x, v'), u \rangle - \Psi_x(j^\dagger(x, v')) \right]$$

---

[77] Because $f^\dagger(x, v) \in \mathcal{P}^{fr}(\mathbf{0}) = \mathrm{Im}[\mathbb{S}^T]$ and $\tilde{\mathcal{T}}_x^*\mathcal{X} := \mathcal{Y}/\mathrm{Ker}\mathbb{S}^T$, $u^\dagger(x, v)$ is uniquely determined.

$$= \max_{j^\dagger \in \mathcal{M}_x^{fr}(\mathbf{0})} \left[\langle j^\dagger, -\mathbb{S}^T u \rangle - \Psi_x(j^\dagger) \right] = \max_{j^\dagger \in \mathcal{M}_x^{fr}(\mathbf{0})} \left[\langle j^\dagger, f^\ddagger(u) \rangle - \Psi_x(j^\dagger) \right]$$

$$= \Psi_x^*(f^\ddagger(u)) = \tilde{\Psi}_x^*(u),$$

where we used the fact that $\{j^\dagger(x, v')\}_{v' \in \tilde{\mathcal{T}}_x \mathcal{X}} = \mathcal{M}_x^{fr}(\mathbf{0})$ in the second line. The pair $(v, u)_x$ is Legendre dual of these functions:

$$\partial_v \tilde{\Psi}_x(v) = \left[\frac{\partial j^\dagger(x, v)}{\partial v} \right]^T \frac{\partial \Psi_x(j)}{\partial j} \bigg|_{j = j^\dagger(x, v)} = \left[\frac{\partial j^\dagger(x, v)}{\partial v} \right]^T f^\dagger(x, v) \quad (133)$$

$$= - \left[\frac{\partial j^\dagger(x, v)}{\partial v} \right]^T \mathbb{S}^T u^\dagger(x, v) = u, \quad (134)$$

$$\partial_u \tilde{\Psi}_x^*(u) = \left[\frac{\partial f^\ddagger(u)}{\partial u} \right]^T \frac{\partial \Psi_x^*(f)}{\partial f} \bigg|_{f = f^\ddagger(u)} = \left[\frac{\partial f^\ddagger(u)}{\partial u} \right]^T j^\ddagger(x, u) \quad (135)$$

$$= -\mathbb{S} j^\ddagger(x, u) = v, \quad (136)$$

where we used $\left[\frac{\partial j^\dagger(x,v)}{\partial v} \right]^T \mathbb{S}^T = -I$ from $\frac{\partial}{\partial v}[v + \mathbb{S} j^\dagger(x, v)] = I + \mathbb{S} \frac{\partial j^\dagger(x,v)}{\partial v} = 0$ and $\frac{\partial f^\ddagger(u)}{\partial u} = -\mathbb{S}^T$. They are dissipation functions; strict convexity and 1-coercivity follow from those of the original dissipation functions. Also, we have

$$\text{Symmetry}: \quad \tilde{\Psi}_x(-v) = \Psi_x(j^\dagger(x, -v)) = \Psi_x(-j^\dagger(x, v)) = \tilde{\Psi}_x(v) \quad (137)$$

$$\text{Bounded by 0 at } \mathbf{0}: \quad \tilde{\Psi}_x(v = 0) = \Psi_x(j^\dagger(x, 0)) = \Psi_x(0) = 0. \quad (138)$$

$\square$

Using the induced dissipation functions, we define the Bregman divergence on $(\tilde{\mathcal{T}}_x \mathcal{X}, \tilde{\mathcal{T}}_x^* \mathcal{X})$, which is associated with the Bregman divergence on $(\mathcal{J}_x, \mathcal{F}_x)$:

$$\mathcal{D}_{\tilde{\Psi}_x}^{\mathcal{X}, \mathcal{Y}}[v \| u'] := \tilde{\Psi}_x(v) + \tilde{\Psi}_x^*(u') - \langle v, u' \rangle \quad (139)$$

$$= \Psi_x(j^\dagger) + \Psi_x^*(f'^\ddagger) - \langle j^\dagger, f'^\ddagger \rangle = \mathcal{D}_x^{\mathcal{J}, \mathcal{F}}[j^\dagger \| f'^\ddagger], \quad (140)$$

where $j^\dagger = j^\dagger(x, v)$ and $f'^\ddagger = f^\ddagger(x, u')$. Therefore, we have the induced dually flat structure on $(\tilde{\mathcal{T}}_x \mathcal{X}, \tilde{\mathcal{T}}_x^* \mathcal{X})$. This induced structure can be regarded as an extension to discrete manifolds of the Otto structure [65, 66]: the formal Riemannian structure induced by the L^2-Wasserstein distance. This is also related to Pistone's infinite-dimensional information geometry [72, 121].

8.3 Fisher information, natural gradient, mirror descent, evolutionary computation, and optimal transport

In information geometry, it is conventional to use the Fisher information matrices, i.e., the Hessian matrices $G_{\boldsymbol{x}}$ and $G_{\boldsymbol{y}}^*$ (Eq. 31) as the metric tensor (Fisher–Rao metric) on $(\mathcal{T}_{\boldsymbol{x}}\mathcal{X}, \mathcal{T}_{\boldsymbol{x}}^*\mathcal{X})$ or equivalently on $(\mathcal{T}_{\boldsymbol{y}}\mathcal{Y}, \mathcal{T}_{\boldsymbol{y}}^*\mathcal{Y})$ (Fig. 7). Gradient systems have been defined information-geometrically [25] as a Riemannian gradient flow using the Bregman divergence and the Fisher information matrix of $\Phi(\boldsymbol{x})$ as the gradient function and the metric tensor, respectively: $\dot{\boldsymbol{x}} = -G_{\boldsymbol{x}}^{-1}\partial\mathcal{D}_{\Phi}^{\mathcal{X}}[\boldsymbol{x}\|\tilde{\boldsymbol{x}}]$. Because both $G_{\boldsymbol{x}}$ and $\mathcal{D}_{\Phi}^{\mathcal{X}}$ are derived from $\Phi(\boldsymbol{x})$, this gradient flow becomes a geodesic in $\mathcal{Y}$ space: $\dot{\boldsymbol{y}} = -(\boldsymbol{y} - \tilde{\boldsymbol{y}})$. In natural gradient descent [67, 69, 148], the Fisher information matrix is used to find the steepest descent gradient of a function $\mathcal{F}(\boldsymbol{\theta})$ on a parameter space Θ as $\dot{\boldsymbol{\theta}} = -G_{\boldsymbol{\theta}}^{-1}\partial\mathcal{F}(\boldsymbol{\theta})$, where $G_{\boldsymbol{\theta}}$ is determined independently of $\mathcal{F}(\boldsymbol{\theta})$ by considering the underlying model parameter space. In optimization, the natural gradient is fundamental in information-geometric optimization algorithms, which contain various evolutionary optimization schemes [70]. In relation to machine learning, the mirror descent is identified with the natural gradient descent by a naive continuous limit [69, 149]. Furthermore, optimal transport has recently been employed to replace or integrate the Fisher-Rao metric with the Wasserstein metric [150, 151]. Because the Wasserstein metric can take the information of the base manifold into account, their integration may provide more amenable ways to accommodate various prior and structural information.

The doubly dual flat structure introduced in this work actually provides a solution to generalize those results and the associated problems. The base space $\mathcal{X}$ with the dually flat structure and the associated Fisher information matrix accommodates the conventional natural gradient. The graph or hypergraph structure endows the additional topological relation to the base space of $\mathcal{X}$. The dissipation functions on the edge spaces or their induced versions bestow a more flexible way than the Fisher-Rao metric to represent the loss of the potential function, i.e., the dissipation, at each point in the state space. Upon necessity, we may combine both of them (Fig. 7), for example, as $\dot{\boldsymbol{x}} = -G_{\boldsymbol{x}}^{-1}\partial\mathcal{F}^{(1)}(\boldsymbol{x}) - \mathrm{div}_{\mathbb{S}}\partial\Psi_{\boldsymbol{x}}^*[\mathrm{grad}_{\mathbb{S}}\partial\mathcal{F}^{(2)}(\boldsymbol{x})]$ where $\mathcal{F}^{(1)}(\boldsymbol{x})$ and $\mathcal{F}^{(2)}(\boldsymbol{x})$ could be different. This flexibility may contribute to the design of new algorithms for machine learning. Actually, this integrated representation is quite relevant to the filtering equations [152] in sequential inference where the first term, i.e., $-G_{\boldsymbol{x}}^{-1}\partial\mathcal{F}^{(1)}(\boldsymbol{x})$, can usually be associated with the update of posterior probability by observation and the second term, $-\mathrm{div}_{\mathbb{S}}\partial\Psi_{\boldsymbol{x}}^*[\mathrm{grad}_{\mathbb{S}}\partial\mathcal{F}^{(2)}(\boldsymbol{x})]$ can represent the prediction by the prior information on the dynamics. Our framework may provide a unified information-geometric perspective to various information-geometric analyses and extensions of filtering, e.g., projection-filters [153], information-geometric nonlinear filtering [154], and information geometric optimization [70]. Furthermore, the generalized gradient flow can be regarded as a continuous time limit of the mirror descent where the nonlinear Legendre duality between primal and dual spaces is preserved at the limit. This fact may be employed to design new gradient-based algorithms via the doubly dual flat structure.

9 Information-geometric properties of generalized nonequilibrium flow

In this section, we consider the nonequilibrium flow defined by Eq. 61, i.e.,

$$\dot{x} = -\mathrm{div}_{\mathbb{S}} \partial \Psi_x^* \left[\left[\mathrm{grad}_{\mathbb{S}} \partial \mathcal{D}_{\Phi}^{\mathcal{X}}[x \| \tilde{x}] \right] + f_{NE} \right], \tag{141}$$

with $f_{NE} \notin \mathrm{Im}\,\mathbb{S}^T$, and show how information geometry can be employed to analyze such dynamics. While we can obtain several properties of equilibrium flow independently of the detail of the thermodynamic function and the dissipation function, these functions should be related so as to obtain nice properties for the nonequilibrium flow. We will observe that the thermodynamic function and the dissipation function of LMA kinetics actually have such a relation.

9.1 Gradient-flow-like property and Lyapunov function of nonequilibrium flow

For the equilibrium flow (Eq. 58), the Bregman divergence $\mathcal{D}_{\Phi}^{\mathcal{X}}[x \| \tilde{x}]$ is a Lyapunov function. The Bregman divergence can still be a Lyapunov function even for the nonequilibrium flow (Eq. 141) under the following conditions:

Lemma 5 *Suppose that, for all $x \in \mathcal{X}$, the force $f(x) = \mathrm{grad}_{\mathbb{S}} \partial \mathcal{D}_{\Phi}^{\mathcal{X}}[x \| \tilde{x}] + f_{NE}$ is orthogonally decomposed as $f(x) = f_S(x) + f_A(x)$ where $f_S(x) := \mathrm{grad}_{\mathbb{S}} \partial_x \mathcal{D}_{\Phi}^{\mathcal{X}}[x \| \tilde{x}_{CB}]$ and $f_A(x) \in \mathcal{F}_x$ satisfy the pseudo-Hilbert-isosceles orthogonality $f_S(x) \perp_H f_A(x)$. Then $\frac{d}{dt} \mathcal{D}_{\Phi}^{\mathcal{X}}[x_t \| \tilde{x}_{CB}] \leq 0$ holds. In addition, $x_{CB} = \mathcal{P}^{sc}(x_0) \cap \mathcal{M}^{eq}(\tilde{x}_{CB})$ is the unique steady state of Eq. 141 with the initial state x_0 that attains $\frac{d}{dt} \mathcal{D}_{\Phi}^{\mathcal{X}}[x_t \| \tilde{x}_{CB}] = 0$. Thus, x_{CB} is locally and asymptotically stable.*[78]

Proof We can directly verify $\frac{d}{dt} \mathcal{D}_{\Phi}^{\mathcal{X}}[x_t \| \tilde{x}_{CB}] \leq 0$ as follows:

$$\frac{d}{dt} \mathcal{D}_{\Phi}^{\mathcal{X}}[x_t \| \tilde{x}_{CB}] = \langle \dot{x}, \partial_x \mathcal{D}_{\Phi}^{\mathcal{X}}[x_t \| \tilde{x}_{CB}] \rangle = -\langle \mathrm{div}_{\mathbb{S}} j(x), \partial_x \mathcal{D}_{\Phi}^{\mathcal{X}}[x_t \| \tilde{x}_{CB}] \rangle \tag{142}$$

$$= -\langle j(x), \mathrm{grad}_{\mathbb{S}} \partial_x \mathcal{D}_{\Phi}^{\mathcal{X}}[x_t \| \tilde{x}_{CB}] \rangle$$

$$= -\langle j(x), f_S(x) \rangle = -\frac{1}{2} \mathcal{D}_x^{\mathcal{J}, \mathcal{F}}[j(x) \| - f''(x)] \leq 0 \tag{143}$$

where we used Eq. 125 and $f''(x) := f_S(x) - f_A(x)$. The equality holds if and only if $f(x) = -f''(x)$, which means that

$$f(x) = -f''(x) \iff f_S(x) = 0 \iff \mathrm{grad}_{\mathbb{S}} \partial_x \mathcal{D}_{\Phi}^{\mathcal{X}}[x \| \tilde{x}_{CB}] = 0$$
$$\iff x \in \mathcal{M}^{eq}(\tilde{x}_{CB}). \tag{144}$$

Because $x_t \in \mathcal{P}^{sc}(x_0)$, $x_{CB} = \mathcal{P}^{sc}(x_0) \cap \mathcal{M}^{eq}(\tilde{x}_{CB})$ holds. □

Thus, if the pseudo-Hilbert-isosceles orthogonal decomposition exists, then the nonequilibrium flow behaves like the equilibrium flow.

---

[78] To have global stability, we have to consider the boundary of $\mathcal{X}$.

9.2 Complex-balanced state and pseudo-Hilbert-isosceles orthogonality

General conditions or situations under which the orthogonal decomposition in Lemma 5 exists are still an open problem. However, for CRN with LMA kinetics, the decomposition holds if a complex-balanced steady state exists.

Proposition 9 (Complex-balanced steady state and orthogonal decomposition for CRN with LMA kinetics [78, 79, 81, 83]) *Suppose that a complex balanced steady state exists, i.e., $\mathcal{M}^{CB} \neq \emptyset$ for CRN with LMA kinetics (Eq. 12). Using any $\tilde{x}_{CB} \in \mathcal{M}^{CB}$, consider a decomposition of the force $f_{MA}(x)$ as $f_{MA}(x) = f_S(x) + f_A$ where $f_S(x) = \mathrm{grad}_{\mathbb{S}} \partial_x \mathcal{D}_{\Phi}^{\mathcal{X}}[x \| \tilde{x}_{CB}]$, $f_A = \ln K + \mathbb{S}^T \ln \tilde{x}_{CB}$, and $\Phi(x)$ is as in Eq. 62. Then, for the dissipation functions in Eq. 68, the pseudo-Hilbert isosceles orthogonality $f_S(x) \perp_H f_A$ holds for all $x \in \mathcal{X}$. In addition, $\mathcal{M}^{eq}(\tilde{x}_{CB}) = \mathcal{M}^{CB}$ holds.*

Proof We can prove the orthogonality by direct computation. The orthogonality condition is

$$\Psi_x^*[f_S(x) + f_A] = \Psi_x^*[f_S(x) - f_A] \tag{145}$$

$$\Leftrightarrow \left\langle j_{MA}^+(x), \left(\frac{x}{\tilde{x}_{CB}}\right)^{-\mathbb{S}^T} - \mathbf{1}\right\rangle$$

$$+ \left\langle j_{MA}^-(x), \left(\frac{x}{\tilde{x}_{CB}}\right)^{\mathbb{S}^T} - \mathbf{1}\right\rangle = 0. \tag{146}$$

Consider the following equality:

$$\left\langle j_{MA}^{\pm}(x), \left(\frac{x}{\tilde{x}_{CB}}\right)^{\mp\mathbb{S}^T} - \mathbf{1}\right\rangle = \sum_{e=1}^{N_e} k_e^{\pm} \tilde{x}_{CB}^{\gamma_e^{\pm}} \left[\left(\frac{x}{\tilde{x}_{CB}}\right)^{\gamma_e^{\mp}} - \left(\frac{x}{\tilde{x}_{CB}}\right)^{\gamma_e^{\pm}}\right], \tag{147}$$

where $\gamma_e^{\pm} := \mathbb{\Gamma} b_e^{\pm}$. By using this, we have the following:

$$\text{Eq. } 146 = \sum_{e=1}^{N_e} \left(k_e^+ \tilde{x}_{CB}^{\gamma_e^+} - k_e^- \tilde{x}_{CB}^{\gamma_e^-}\right) \left[\left(\frac{x}{\tilde{x}_{CB}}\right)^{\gamma_e^-} - \left(\frac{x}{\tilde{x}_{CB}}\right)^{\gamma_e^+}\right] \tag{148}$$

$$= \left\langle (j_{MA}^+(\tilde{x}_{CB}) - j_{MA}^-(\tilde{x}_{CB})), \left[\left(\frac{x}{\tilde{x}_{CB}}\right)^{(\mathbb{\Gamma}\mathbb{B}^-)^T} - \left(\frac{x}{\tilde{x}_{CB}}\right)^{(\mathbb{\Gamma}\mathbb{B}^+)^T}\right]\right\rangle \tag{149}$$

$$= \left\langle j_{MA}(\tilde{x}_{CB}), \mathbb{B}_-^T \left(\frac{x}{\tilde{x}_{CB}}\right)^{\mathbb{\Gamma}^T} - \mathbb{B}_+^T \left(\frac{x}{\tilde{x}_{CB}}\right)^{\mathbb{\Gamma}^T}\right\rangle$$

$$= \left\langle \mathbb{B} j_{MA}(\tilde{x}_{CB}), \left(\frac{x}{\tilde{x}_{CB}}\right)^{\mathbb{\Gamma}^T}\right\rangle. \tag{150}$$

Thus, Eq. 145 holds for any $\boldsymbol{x} \in \mathcal{X}$ if $\mathbb{B} \boldsymbol{j}_{MA}(\tilde{\boldsymbol{x}}_{CB}) = \boldsymbol{0}$ holds.[79] $\mathcal{M}^{eq}(\tilde{\boldsymbol{x}}_{CB}) = \mathcal{M}^{CB}$ can be proved by obtaining the parametric representation of $\mathcal{M}^{CB}$ as $\mathcal{M}^{CB} = \{\boldsymbol{x} \in \mathcal{X} \mid \ln \boldsymbol{x} - \ln \tilde{\boldsymbol{x}}_{CB} \in \text{Ker} \mathbb{S}^{T}\}$ via solving $\boldsymbol{j}_{MA}(\boldsymbol{x}) = \boldsymbol{0}$.[80] This representation is identical to that of $\mathcal{M}^{eq}(\tilde{\boldsymbol{x}}_{CB})$ (Eq. 104). $\qquad\square$

Remark 10 (Algebraic structure of detailed balanced and complex balanced manifolds) We here mention about the underlying algebraic source of why $\mathcal{M}^{eq}(\tilde{\boldsymbol{x}}_{CB}) = \mathcal{M}^{CB}$ holds. First, we already showed that $\mathcal{M}^{DB} = \mathcal{M}^{eq}(\tilde{\boldsymbol{x}})$ holds generally if $\mathcal{M}^{DB} \neq \emptyset$. Under LMA kinetics (Eq. 12), the DB condition $\boldsymbol{j}_{MA}(\boldsymbol{x}) = \boldsymbol{0}$ is nothing but the binomial equations because $\boldsymbol{j}_{MA}^{\pm}(\boldsymbol{x})$ are vectors of monominals of $\boldsymbol{x}$. Owing to this, $\mathcal{M}^{DB}$ becomes a toric variety.[81] In contrast, the CB condition $\mathbb{B} \boldsymbol{j}_{MA}(\boldsymbol{x}_{CB}) = \boldsymbol{0}$ is a set of polynomial equations for LMA kinetics. Nonetheless, it was shown that $\mathcal{M}^{CB}$ is binomially generated and has the same structural matrix $\mathbb{S}^{T}$ as the equilibrium manifold [100]. Because of that, they become equivalent as manifolds.

Because rLDG (Eq. 3) is a subclass of CRN where $\mathbb{\Gamma} = I$ and thus $\mathbb{S} = \mathbb{B}$ holds, the complex-balanced condition is always satisfied for rLDG.

Corollary 1 (*rLDG is unconditionally complex-balanced [8]*) *All the steady states of rLDG are complex-balanced states, i.e., $\mathcal{M}^{ST} = \mathcal{M}^{CB}$ independently of the parameter values $\boldsymbol{k}^{\pm}$ of the flux (Eq. 3).*[82] *Thus, KL divergence (Eq. 64) always works as a Lyapunov function of rLDG.*[83]

The properties described in this corollary are well-known for rMJP and are usually obtained by using the Perron-Frobenius theorem for linear operators. The framework of the generalized flow enables us to extend them to the nonlinear regime.

9.3 Effective flux of the nonequilibrium flow by the primal information-geometric projection

In general, the nonequilibrium force or flux has redundant degrees of freedom in terms of generating a specific vector field or trajectory $\{\boldsymbol{x}_t\}$ on the density space. By using the extended HHK projective decomposition (Theorem 1), we can carve out the effective part of the flux for the trajectory $\{\boldsymbol{x}_t\}$. In addition, we can obtain an effective time-dependent equilibrium flux that mimics the trajectory $\{\boldsymbol{x}_t\}$:

Lemma 6 (Effective time-dependent equilibrium flux) *Suppose that $\boldsymbol{j}(\boldsymbol{x})$ is the flux of a generalized flow (Eq. 45), and define the corresponding effective equilibrium flux by $\boldsymbol{j}_{eq}(\boldsymbol{x}) = \partial \Psi_{\boldsymbol{x}}^{*}[\mathbb{S}^{T} \boldsymbol{u}_{eq}(\boldsymbol{x})]$ where $\boldsymbol{u}_{eq}(\boldsymbol{x}) := \partial \tilde{\Psi}_{\boldsymbol{x}}[-\mathbb{S} \boldsymbol{j}(\boldsymbol{x})]$. Then, $\boldsymbol{j}_{eq}(\boldsymbol{x})$ induces the same velocity as $\boldsymbol{j}(\boldsymbol{x})$ does.*[84] *Furthermore, for a given trajectory $\{\boldsymbol{x}_t\}$ of $\boldsymbol{j}(\boldsymbol{x})$, we*

---

[79] The transformation of Eq. 146 here is strongly dependent on the specific functional form of $\boldsymbol{j}_{MA}(\boldsymbol{x})$.

[80] We skip the derivation because it is involved. See the original derivation [100] or our rephrased version [84]

[81] The real variety generated by a toric ideal, i.e., a binomial and prime ideal [100].

[82] Such a situation is called unconditionally complex balanced.

[83] If Rank$[\mathbb{S}] = $ Rank$[\mathbb{B}]$, CRN is unconditionally complex-balanced. This condition is called the deficiency zero condition [8].

[84] $\boldsymbol{j}_{eq}(\boldsymbol{x})$ may not be equilibrium flux because $\boldsymbol{u}_{eq}(\boldsymbol{x})$ is not necessarily represented by the gradient of a certain function $\mathcal{F}(\boldsymbol{x})$ as $\boldsymbol{u}_{eq}(\boldsymbol{x}) = \partial_{\boldsymbol{x}} \mathcal{F}(\boldsymbol{x})$.

can construct a time-dependent equilibrium flux $\boldsymbol{j}_{eq}(t, \boldsymbol{x})$ that can generate the same $\{\boldsymbol{x}_t\}$ as follows

$$\boldsymbol{j}_{eq}(t, \boldsymbol{x}) := \partial \Psi_{\boldsymbol{x}}^* \left[\mathrm{grad}_{\mathbb{S}} \partial \mathcal{D}_{\Phi}^{\mathcal{X}}[\boldsymbol{x} \| \tilde{\boldsymbol{x}}_t] \right], \quad \tilde{\boldsymbol{x}}_t := \partial \Phi^*[\partial \Phi(\boldsymbol{x}_t) + \boldsymbol{u}_{eq}(\boldsymbol{x}_t)]. \quad (151)$$

Proof From Theorem 1, $\boldsymbol{j}(\boldsymbol{x})$ can be decomposed as $\boldsymbol{j}(\boldsymbol{x}) = \boldsymbol{j}_{eq}(\boldsymbol{x}) + (\boldsymbol{j}(\boldsymbol{x}) - \boldsymbol{j}_{eq}(\boldsymbol{x}))$. Because $\boldsymbol{f}_{eq}(\boldsymbol{x}) = \partial \Psi_{\boldsymbol{x}}^*[\boldsymbol{j}_{eq}(\boldsymbol{x})] \in \mathcal{P}^{fr}(\boldsymbol{0})$, there exists $\boldsymbol{u}_{eq}(\boldsymbol{x})$ satisfying $-\mathbb{S}^T \boldsymbol{u}_{eq}(\boldsymbol{x}) = \boldsymbol{f}_{eq}(\boldsymbol{x})$. By employing the duality introduced in Theorem 2, $\boldsymbol{u}_{eq}(\boldsymbol{x})$ can be represented as $\boldsymbol{u}_{eq}(\boldsymbol{x}) = \partial \tilde{\Psi}_{\boldsymbol{x}}[\boldsymbol{v}(\boldsymbol{x})]$ where $\boldsymbol{v}(\boldsymbol{x}) = -\mathbb{S}\boldsymbol{j}(\boldsymbol{x})$. Because $-\mathbb{S}\boldsymbol{j}_{eq}(\boldsymbol{x}) = \boldsymbol{v}(\boldsymbol{x})$, $\boldsymbol{j}_{eq}(\boldsymbol{x})$ generates the same dynamics or vector field as $\boldsymbol{j}(\boldsymbol{x})$ does. By solving $-\boldsymbol{u}_{eq}(\boldsymbol{x}_t) = \partial \Phi(\boldsymbol{x}_t) - \partial \Phi(\tilde{\boldsymbol{x}}_t)$, we have Eq. 151. $\qquad \square$

The effective time-dependent equilibrium flux is obtained more explicitly for CRN with LMA kinetics:

Corollary 2 *(Effective equilibrium force and flux of LMA kinetics) Consider the following quantities of CRN with LMA kinetics:*

Flux defined in Eq. 12: $\quad \boldsymbol{j}_{\mathrm{MA}}(\boldsymbol{x}; \boldsymbol{k}^{\pm})$ $\hfill (152)$

Force defined in Eq. 70: $\quad \boldsymbol{f}_{\mathrm{MA}}(\boldsymbol{x}; \boldsymbol{K})$ $\hfill (153)$

Thermodynamic function defined in Eq. 62: $\quad \Phi(\boldsymbol{x})$ $\hfill (154)$

Dissipation functions defined in Eq. 68 with Eq. 70: $\quad \Psi_{\boldsymbol{x}, \kappa}^*[\boldsymbol{f}] = \Psi_{\omega_{\mathrm{MA}}(\boldsymbol{x}; \kappa)}^*[\boldsymbol{f}],$ $\hfill (155)$

where $\boldsymbol{k}^{\pm} = \kappa \circ \boldsymbol{K}^{\pm 1/2}$ holds. For a trajectory $\{\boldsymbol{x}_t\}$ generated by $\boldsymbol{j}_{MA}(\boldsymbol{x}; \boldsymbol{k}^{\pm})$, the effective time-dependent equilibrium force $\boldsymbol{f}_{eq}(t, \boldsymbol{x})$ can be described as $\boldsymbol{f}_{MA}(\boldsymbol{x}; \boldsymbol{K}_{eq}(t))$ where $\boldsymbol{K}_{eq}(t)$ is determined by

$$\boldsymbol{K}_{eq}(t) := \exp \left[-\mathbb{S}^T \left(\boldsymbol{u}_{eq}(\boldsymbol{x}_t; \kappa) + \partial \Phi(\boldsymbol{x}_t) \right) \right],$$
$$\boldsymbol{u}_{eq}(\boldsymbol{x}_t; \kappa) := \partial \tilde{\Psi}_{\boldsymbol{x}_t, \kappa}[-\mathbb{S}\boldsymbol{j}_{MA}(\boldsymbol{x}_t)] \quad (156)$$

Thus, the effective time-dependent equilibrium flux $\boldsymbol{j}_{eq}(t, \boldsymbol{x})$ is represented as $\boldsymbol{j}_{eq}(t, \boldsymbol{x}) = \boldsymbol{j}_{MA}(\boldsymbol{x}; \boldsymbol{k}_{eq}^{\pm}(t))$ where $\boldsymbol{k}_{eq}^{\pm}(t) = \kappa \circ \boldsymbol{K}_{eq}^{\pm 1/2}(t)$.

This corollary means that the effective time-dependent flux of LMA kinetics is always obtained by a time-dependent modulation of the kinetic parameters $\boldsymbol{k}^{\pm}$. More specifically, the modulation of force part $\boldsymbol{K}_{eq}(t)$ is sufficient while the activity part κ is kept constant.[85]

Example 3 (Simplified Brusselator CRN [8, 104] (continued)) By using the Brusselator CRN (Ex. 1), we numerically obtained a nonequilibrium trajectory $\{\boldsymbol{x}_t\}$

---

[85] While this result may sound not so significant mathematically, for physics and chemistry, the result means that the effective flux is physically realizable and testable.
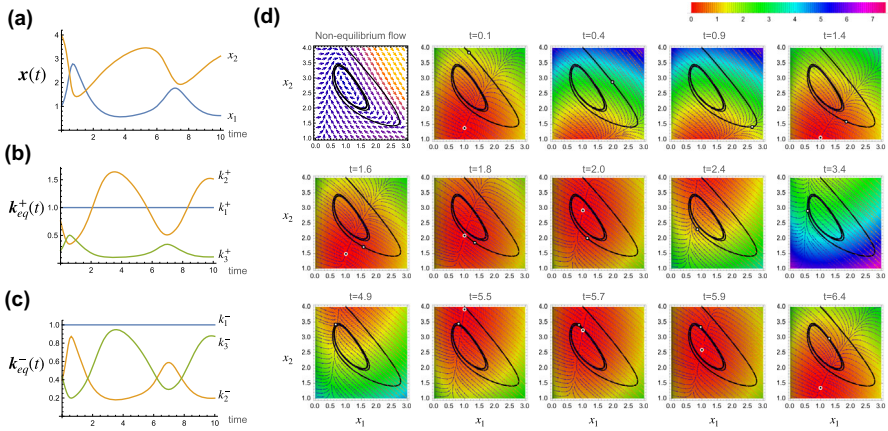
Fig. 8 A nonequilibrium trajectory of the Brusselator CRN (Ex. 1) and the associated time-dependent effective equilibrium flux $j_{eq}(t, x) = j_{MA}(x; k_{eq}^\pm(t))$. (a) A nonequilibrium trajectory $\{x_t\}$ of the Brusselator CRN for a set of the kinetic parameters, $k_1^+ = 1.0$, $k_1^- = 1.0$, $k_2^+ = 3.0$, $k_2^- = 0.1$, $k_3^+ = 1.0$, $k_3^- = 0.1$, and the initial state $(x_1(0), x_2(0)) = (1.0, 4.0)$. (b,c) The time-dependent kinetic parameter set $k_{eq}^\pm(t)$, which generates the time-dependent equilibrium flux $j_{eq}(t, x)$. (d) The top left panel shows the nonequilibrium trajectory $\{x_t\}$ (black curve) and the vector field $v(x) = -\mathbb{S}j(x)$ (arrows). The other panels show the time-dependent vector field $v_{eq}(t, x)$ induced by $j_{eq}(t, x)$: $v_{eq}(t, x) = -\mathbb{S}j_{eq}(t, x)$. The nonequilibrium trajectory $\{x_t\}$ is also depicted as a reference (the black curve). The color indicates the value of $\mathcal{D}_\Phi^\mathcal{X}[x \| \tilde{x}_t]$. In each panel, the white circle on the trajectory is x_t at which $j_{eq}(t, x)$ is computed, and the black circle with a white border corresponds to $\tilde{x}_t$ at which $\mathcal{D}_\Phi^\mathcal{X}[x \| \tilde{x}_t]$ is 0

(Fig. 8a, d top left panel). By using Cor. 2, we also computed the corresponding time-dependent kinetic parameter set $k_{eq}^\pm(t)$ that generates the time-dependent equilibrium flux $j_{eq}(t, x) = j_{MA}(t, x; k_{eq}^\pm(t))$ (Fig. 8b, c). Figure 8d shows the vector field $v_{eq}(t, x) = -\mathbb{S}j_{eq}(t, x)$ induced by the time-dependent equilibrium flux $j_{eq}(t, x)$ and the contours of $\mathcal{D}_\Phi^\mathcal{X}[x \| \tilde{x}_t]$ where $\tilde{x}_t$ follows from Eq. 151. From Fig. 8, we can see that the trajectory $\{x_t\}$ originally generated by the nonequilibrium flux $j(x)$ can be traced by the time-dependent equilibrium flux $j_{eq}(t, x)$ and also that $j_{eq}(t, x)$ can be physically realized by the modulation of the kinetic parameters $k_{eq}^\pm(t)$.

9.4 Characterization of the nonequilibrium flow by the dual information geometric projection

The nonequilibrium flow is redundant in terms of generating a specific vector field or trajectory $\{x_t\}$. Such redundancy is crucial to characterize the extent of nonequilibrium. One approach for the characterization is to investigate the cycle force or flux, which has been employed in the linear theory of dynamics on graphs and also in graph-theoretic approaches to nonequilibrium phenomena [89, 155–158]. To extract such cyclic components, we can use $\mathbb{V}^T = \text{curl}_\mathbb{V}$, its adjoint $\mathbb{V} = \text{curl}_\mathbb{V}^*$, the associated cycle subspaces $C^2(\mathbb{H})$ and $C_2(\mathbb{H})$, and also the generalized HKK decomposition (Theorem 1).
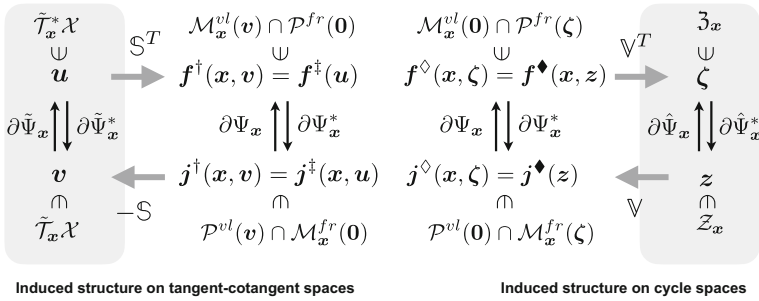
$$
\begin{array}{ccccc}
\tilde{\mathcal{T}}_x^* \mathcal{X} & \mathcal{M}_x^{vl}(v) \cap \mathcal{P}^{fr}(\mathbf{0}) & \mathcal{M}_x^{vl}(\mathbf{0}) \cap \mathcal{P}^{fr}(\zeta) & & \mathfrak{Z}_x \\
\mathbb{U} & \mathbb{S}^T \quad \mathbb{U} & \mathbb{U} & \mathbb{V}^T \quad \mathbb{U} \\
\boldsymbol{u} & \boldsymbol{f}^\dagger(x,v) = \boldsymbol{f}^\ddagger(u) & \boldsymbol{f}^\diamond(x,\zeta) = \boldsymbol{f}^\blacklozenge(x,z) & & \zeta \\
\partial \tilde{\Psi}_x \Big\Updownarrow \partial \tilde{\Psi}_x^* & \partial \Psi_x \Big\Updownarrow \partial \Psi_x^* & \partial \Psi_x \Big\Updownarrow \partial \Psi_x^* & & \partial \hat{\Psi}_x \Big\Updownarrow \partial \hat{\Psi}_x^* \\
\boldsymbol{v} & \boldsymbol{j}^\dagger(x,v) = \boldsymbol{j}^\ddagger(x,u) & \boldsymbol{j}^\diamond(x,\zeta) = \boldsymbol{j}^\blacklozenge(z) & & \boldsymbol{z} \\
\mathbb{\cap} & -\mathbb{S} \quad \mathbb{\cap} & \mathbb{\cap} & \mathbb{V} \quad \mathbb{\cap} \\
\tilde{\mathcal{T}}_x \mathcal{X} & \mathcal{P}^{vl}(v) \cap \mathcal{M}_x^{fr}(\mathbf{0}) & \mathcal{P}^{vl}(\mathbf{0}) \cap \mathcal{M}_x^{fr}(\zeta) & & \mathcal{Z}_x
\end{array}
$$

Induced structure on tangent-cotangent spaces   **Induced structure on cycle spaces**

Fig. 9 The induced dually flat structure on the cycle spaces (right gray region) in comparison with the induced structure on the restricted tangent and cotangent spaces (left gray region)

Definition 34 (*Cycle spaces*) The cycle spaces at $x \in \mathcal{X}$ are defined as $\mathcal{Z}_x = C_2(\mathbb{H}) = \mathrm{Ker}[\mathbb{V}]$ and $\mathfrak{Z}_x = C^2(\mathbb{H}) = \mathrm{Im}[\mathbb{V}^T]$.

For a given nonequilibrium force $\boldsymbol{f}(x)$, we can obtain its cycle component $\zeta = \mathrm{curl}_\mathbb{V} \boldsymbol{f}(x) = \mathbb{V}^T \boldsymbol{f}(x) \in \mathfrak{Z}_x$. ζ contains the information to categorize the force because $\mathcal{P}^{fr}(\boldsymbol{f}) = \mathcal{P}^{fr}(\zeta)$ is the quotient space of force by the equilibrium force. For each $\zeta \in \mathfrak{Z}_x$, we obtain the representative force $\boldsymbol{f}^\diamond(x,\zeta)$ via the following variational problem:

Lemma 7 (Steady (zero-velocity) force as the minimizer of the dual dissipation function) *For a given $\zeta \in \mathfrak{Z}_x$, we define the force $\boldsymbol{f}^\diamond(x,\zeta)$ minimizing the dual dissipation function:*

$$
\boldsymbol{f}^\diamond(x,\zeta) := \arg\min_{\boldsymbol{f}} \Psi_x^*[\boldsymbol{f}], \quad s.t. \ \mathrm{curl}_\mathbb{V} \boldsymbol{f} = \zeta. \tag{157}
$$

Then, $\boldsymbol{f}^\diamond(x,\zeta)$ is the steady (zero-velocity) force, i.e., $\boldsymbol{j}^\diamond = \partial \Psi_x^*[\boldsymbol{f}^\diamond] \in \mathcal{P}^{vl}(\mathbf{0})$.

Proof From Eq. 115 in Theorem 1, $\boldsymbol{f}^\diamond(x,\zeta) \in \mathcal{P}^{fr}(\zeta) \cap \mathcal{M}^{vl}(\mathbf{0})$. Thus, $\boldsymbol{j}^\diamond \in \mathcal{P}^{vl}(\mathbf{0})$. □

Among various forces $\boldsymbol{f}$ that has the same cyclic component ζ, the force $\boldsymbol{f}^\diamond(x,\zeta)$ is the one that induces no dynamics of x. Because any dynamics of x can be represented by the effective equilibrium flux as in Lemma 6, $\boldsymbol{f}^\diamond(x,\zeta)$ can be regarded as the force being purely relevant to the cycle.

Using $\boldsymbol{f}^\diamond(x,\zeta)$, we can establish the induced duality between $\mathcal{Z}_x$ and $\mathfrak{Z}_x$ spaces:

Theorem 3 (Induced dually flat structure on cycle spaces) *On the cycle spaces, $\mathcal{Z}_x$ and $\mathfrak{Z}_x$, we have the Legendre conjugate dissipation functions $\hat{\Psi}_x : \mathcal{Z}_x \to \mathbb{R}$ and $\hat{\Psi}_x^* : \mathfrak{Z}_x \to \mathbb{R}$ induced by the dissipation functions on the edge spaces (Fig. 9).*

Proof For each $\zeta \in \mathfrak{Z}_x$, we can uniquely determine $\boldsymbol{f}^\diamond(x,\zeta) \in \mathcal{F}_x$ and $\boldsymbol{j}^\diamond(x,\zeta) \in \mathcal{J}_x$ as

$$
\boldsymbol{f}^\diamond(x,\zeta) := \mathcal{P}^{fr}(\zeta) \cap \mathcal{M}_x^{vl}(\mathbf{0}) = \arg\min_{\boldsymbol{f} \in \mathcal{P}^{fr}(\zeta)} \Psi_x^*[\boldsymbol{f}], \tag{158}
$$

$$j^\diamond(x, \zeta) := \partial \Psi_x^*[f^\diamond(x, \zeta)] \in \mathcal{P}^{vl}(\mathbf{0}). \tag{159}$$

In addition, $z^\diamond(x, \zeta) \in \mathcal{Z}_x$ satisfying $\mathbb{V}z^\diamond(x, \zeta) = j^\diamond(x, \zeta)$ is also uniquely determined because $\mathbb{V} : \mathcal{Z}_x \to \mathcal{J}_x$, $j^\diamond(x, \zeta) \in \mathcal{P}^{vl}(\mathbf{0}) = \mathrm{Im}[\mathbb{V}]$ and $\mathrm{Ker}\mathbb{V} = \{\mathbf{0}\}$. For these quantities,

$$\zeta = \mathbb{V}^T f^\diamond(x, \zeta), \quad f^\diamond(x, \zeta) = \partial \Psi_x[j^\diamond(x, \zeta)], \quad j^\diamond(x, \zeta) = \mathbb{V}z^\diamond(x, \zeta), \tag{160}$$

hold. Conversely, for a given $z \in \mathcal{Z}_x$, we have $j^\blacklozenge(z)$, $f^\blacklozenge(x, z)$, and $\zeta^\blacklozenge(x, z)$ as follows:

$$\zeta^\blacklozenge(x, z) = \mathbb{V}^T f^\blacklozenge(x, z), \quad f^\blacklozenge(x, z) = \partial \Psi_x[j^\blacklozenge(z)], \quad j^\blacklozenge(z) = \mathbb{V}z. \tag{161}$$

For a pair of $(z, \zeta)_x$ that satisfies $z = z^\diamond(x, \zeta)$, then $\zeta = \zeta^\blacklozenge(x, z)$, $f^\diamond(x, \zeta) = f^\blacklozenge(x, z)$, and $j^\diamond(x, \zeta) = j^\blacklozenge(z)$ hold. This pairing establishes a bijection between $\mathcal{Z}_x$ and $\mathfrak{Z}_x$. This bijection is realized by the Legendre transformations of the following induced dissipation functions on $\mathcal{Z}_x$ and $\mathfrak{Z}_x$:

$$\hat{\Psi}_x(z) := \Psi_x(j^\blacklozenge(z)), \qquad\qquad \hat{\Psi}_x^*(\zeta) := \Psi_x^*(f^\diamond(x, \zeta)). \tag{162}$$

These functions are Legendre conjugate as follows:

$$\max_{z' \in \mathcal{Z}_x} \left[\langle z', \zeta \rangle - \hat{\Psi}_x(z') \right] = \max_{\substack{z' \in \mathcal{Z}_x \\ f \in \mathcal{P}^{fr}(\zeta)}} \left[\langle z', \mathbb{V}^T f \rangle - \hat{\Psi}_x(z') \right]$$

$$= \max_{\substack{z' \in \mathcal{Z}_x \\ f \in \mathcal{P}^{fr}(\zeta)}} \left[\langle \mathbb{V}z', f \rangle - \Psi_x(\mathbb{V}z') \right]$$

$$= \max_{\substack{j' \in \mathcal{P}^{vl}(\mathbf{0}) \\ f \in \mathcal{P}^{fr}(\zeta)}} \left[\langle j', f \rangle - \Psi_x(j') \right]$$

$$= \max_{\substack{j' \in \mathcal{P}^{vl}(\mathbf{0}) \\ f \in \mathcal{P}^{fr}(\zeta)}} \left[\langle j', f^\diamond(x, \zeta) \rangle - \Psi_x(j') + \langle j', (f - f^\diamond(x, \zeta)) \rangle \right]$$

$$= \max_{j' \in \mathcal{P}^{vl}(\mathbf{0})} \left[\langle j', f^\diamond(x, \zeta) \rangle - \Psi_x(j') \right] = \Psi_x^*(f^\diamond(x, \zeta))$$

$$= \hat{\Psi}_x^*(\zeta),$$

where we used $\langle j', (f - f^\diamond(x, \zeta)) \rangle = 0$ because $j' \in \mathcal{P}^{vl}(\mathbf{0}) = \mathrm{Im}\mathbb{V}$ and $f - f^\diamond(x, \zeta) \in \mathcal{P}^{fr}(\mathbf{0}) = \mathrm{Ker}\mathbb{V}^T$. The inverse is also shown:

$$\max_{\zeta' \in \mathfrak{Z}_x} \left[\langle z, \zeta' \rangle - \hat{\Psi}_x^*(\zeta') \right] = \max_{\zeta' \in \mathfrak{Z}_x} \left[\langle z, \mathbb{V}^T f^\diamond(x, \zeta') \rangle - \Psi_x^*(f^\diamond(x, \zeta')) \right]$$

$$= \max_{f^\diamond \in \mathcal{M}_x^{vl}(\mathbf{0})} \left[\langle \mathbb{V}z, f^\diamond \rangle - \Psi_x^*(f^\diamond) \right] = \max_{f^\diamond \in \mathcal{M}_x^{vl}(\mathbf{0})} \left[\langle j^\blacklozenge(z), f^\diamond \rangle - \Psi_x^*(f^\diamond) \right]$$

$$= \Psi_x(j^\blacklozenge(z)) = \hat{\Psi}_x(z)$$

The pair $(z, \zeta)_x$ is Legendre dual with respect to these functions:

$$\partial_\zeta \hat{\Psi}_x^*(\zeta) = \left[\frac{\partial f^\lozenge(x, \zeta)}{\partial \zeta}\right]^T \frac{\partial \Psi_x^*(f)}{\partial f}\bigg|_{f=f^\lozenge(x,\zeta)}$$

$$= \left[\frac{\partial f^\lozenge(x, \zeta)}{\partial \zeta}\right]^T j^\lozenge(x, \zeta) \tag{163}$$

$$= \left[\frac{\partial f^\lozenge(x, \zeta)}{\partial \zeta}\right]^T \mathbb{V} z^\lozenge(x, \zeta) = z, \tag{164}$$

$$\partial_z \hat{\Psi}_x(z) = \left[\frac{\partial j^\blacklozenge(z)}{\partial z}\right]^T \frac{\partial \Psi_x(j)}{\partial j}\bigg|_{j=j^\blacklozenge(z)} = \left[\frac{\partial j^\blacklozenge(z)}{\partial z}\right]^T f^\blacklozenge(x, z)$$

$$= \mathbb{V}^T f^\blacklozenge(x, z) = \zeta, \tag{165}$$

where we used $\left[\frac{\partial f^\lozenge(x,\zeta)}{\partial \zeta}\right]^T \mathbb{V} = I$ from $\frac{\partial}{\partial \zeta}[\zeta - \mathbb{V}^T f^\lozenge(x, \zeta)] = I - \mathbb{V}^T \frac{\partial f^\lozenge(x,\zeta)}{\partial \zeta} = 0$
and $\frac{\partial j^\blacklozenge(x,z)}{\partial z} = \mathbb{V}$. They are dissipation functions: strict convexity and 1-coercivity follow from those of the original dissipation functions. Also, we have

$$\text{Symmetry:} \hat{\Psi}_x(-z) = \Psi_x(j^\blacklozenge(-z)) = \Psi_x(-j^\blacklozenge(z)) = \hat{\Psi}_x(z) \tag{166}$$

$$\text{Bounded by 0 at } \mathbf{0}: \hat{\Psi}_x(z=0) = \Psi_x(j^\blacklozenge(0)) = \Psi_x(0) = 0. \tag{167}$$

$\square$

For a given force $f(x)$ of the generalized flow (Eq. 45), $f_{st}(x) = f^\lozenge(x, \mathbb{V}^T f(x))$ works as the effective cycle force for each $x \in \mathcal{X}$. Similarly to Cor. 2, we obtain the effective cycle force and the flux for LMA kinetics by parametric modulation:

Corollary 3 *(Effective cycle force and flux for LMA kinetics) Consider CRN with LMA kinetics as in Cor. 2. For each $x \in \mathcal{X}$, the effective cycle force $f_{st}(x)$ associated with $j_{MA}(x; k^\pm)$ can be described as $f_{st}(x) = f_{MA}(x; K_{st}(x))$ where $K_{st}(x)$ is determined by*

$$K_{st}(x) := \exp\left[f^\lozenge(x; \mathbb{V}^T f_{MA}(x; K)) - \mathbb{S}^T \partial \Phi(x)\right]. \tag{168}$$

Thus, the effective cycle flux $j_{st}(x)$ is represented as $j_{st}(x) = j_{MA}(x; k_{st}^\pm(x))$ where $k_{st}^\pm(x) = \kappa \circ K_{st}^{\pm 1/2}(x)$. For a given trajectory $\{x_t\}$, which is generated by a generalized flow, we have the effective time-dependent cycle flux $j_{st}(t, x)$ as $j_{st}(t, x) = j_{MA}(x; k_{st}^\pm(x_t))$. From the construction, this time-dependent cycle flux makes x_t a steady state for each t, i.e., $\mathbb{S} j_{st}(t, x_t) = \mathbf{0}$ holds for any t.
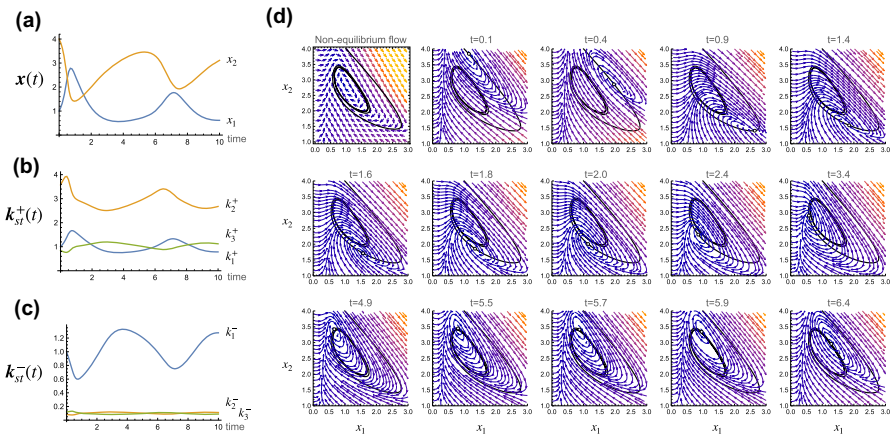
Fig. 10 A nonequilibrium trajectory of the Brusselator CRN (Ex. 1) and the associated time-dependent cycle flux $j_{st}(t, x) = j_{MA}(x; k_{st}^{\pm}(t))$. **a** A nonequilibrium trajectory $\{x_t\}$ of the Brusselator CRN obtained with the same parameter values in Fig. 8a. **b, c** The time-dependent kinetic parameter set $k_{st}^{\pm}(t)$, which generates the time-dependent cycle flux $j_{st}(t, x)$. **d** The top left panel shows the nonequilibrium trajectory $\{x_t\}$ (the black curve) and vector field $v(x) = -\mathbb{S}j(x)$ (arrows). The other panels show the time-dependent vector field $v_{st}(t, x)$ induced by $j_{st}(t, x)$: $v_{st}(t, x) = -\mathbb{S}j_{st}(t, x)$. The nonequilibrium trajectory $\{x_t\}$ is also depicted for a reference (the black curve). In each panel, the white circle on the trajectory is x_t at which $j_{st}(t, x)$ is computed

Example 4 (Simplified Brusselator CRN [8, 104] (continued)) For the nonequilibrium trajectory of the Brusselator CRN in Fig. 10a, we numerically obtained the effective cycle flux $j_{st}(t, x)$ and the corresponding time-dependent kinetic parameter set $k_{st}^{\pm}(t)$ (Fig. 10b, c). Figure 10d shows the vector field $v_{st}(t, x) = -\mathbb{S}j_{st}(t, x)$ induced by the effective cycle flux $j_{st}(t, x)$. From Fig. 10, we can see that any point on the trajectory $\{x_t\}$ originally generated by the nonequilibrium flux $j(x_t)$ can be kept steady with the time-dependent cycle flux $j_{st}(t, x)$ realized by the modulation of the kinetic parameter $k_{st}^{\pm}(t)$.

In modern nonequilibrium thermodynamics, it has been a great challenge to establish thermodynamic characterizations for nonequilibrium phenomena. To this end, the dissection of dynamics and the corresponding flux and force has been attempted [104, 159–162]. For a given trajectory $\{x_t\}$, the effective time-dependent equilibrium flux $j_{eq}(t, x)$ generates exactly the same trajectory, which dissects and mimics the dynamic aspect of the trajectory. On the other hand, the effective time-dependent cycle flux $j_{st}(t, x)$ makes each point on the trajectory steady, which can be recognized as the nonequilibrium aspect of the trajectory. Moreover, these two types of fluxes can be realized by appropriately modulating the kinetic parameter set $k^{\pm}$, which makes the dissected fluxes physically meaningful and accessible. More specifically, the modulation of the force part K of the kinetic parameter is sufficient for realization (Eq. 156 and Eq. 168), while the activity part κ is kept constant. In the case of CRN, the former is linked to the free energy difference between reactants and products of each reaction, and the latter is associated with the height of the energy barrier between them. This clear separation of different physical parameters in our framework is advantageous to

further investigate physical aspects of dynamics on graphs and hypergraphs. Thus, the dually flat structure on the edge space and the HHK decomposition provides a new and promising way to characterize the nonequilibrium flow.[86]

10 Summary and discussion

In this work, we have shown that the doubly dual flat structure of the vertex and edge spaces on graphs and hypergraphs provides the information-geometric basis for the dynamics on graphs and hypergraphs. Two notions of orthogonality, pseudo-Hilbert isosceles orthogonality and information-geometric orthogonality, have been introduced and shown to dissect the equilibrium and nonequilibrium aspects of the dynamics into the induced structures on the tangent and cotangent spaces and the cycle spaces. The doubly dual flat structure naturally connects the topological information of underlying discrete manifolds, i.e., graphs and hypergraphs, with the dynamics on them and thus endows more flexibility and representation power to the information-geometric modeling of dynamics. Furthermore, the generalized equilibrium and nonequilibrium flows, as well as the generalized flow, accommodate a sufficiently wide range of models, which include the reversible Markov jump processes on finite graphs and CRN with LMA kinetics (a class of PDS). These results could substantially extend the applicability of information geometry to dynamical problems.

10.1 Extension of other relations involving information measures

While we demonstrated that the generalized flow and the doubly dual flat structure can extend several results known for FPE and diffusion processes, we still have potentially relevant results and problems that could be explained and extended in our framework. For example, for FPE and diffusion processes, the Fisher information number $\mathbb{I}_F$ was extended to the relative Fisher information (also known as Hyvärinen divergence [120, 144]). The relative Fisher information of two trajectories $p_t^{(1)}(\mathbf{r})$ and $p_t^{(2)}(\mathbf{r})$ is known to satisfy information–theoretic relations such as the De Brujin identity [54] and its extensions [56, 62]. In addition, the logarithmic Sobolev inequality also constitutes a relationship between the Fisher information number and the KL divergence (or Shannon information) [63, 64]. It would be an important future problem to associate these results with the doubly dual flat structure.

Moreover, several relations potentially being related to De Giorgi's formulation (Eq. 52) have been known for mutual information in filtering and control theories. For example, Guo, Shamai, and Verdu found a relation between mutual information and the minimum mean square error (MMSE) in Gaussian channels [163]. Relations similar to these have also been reported by Mayer-Wolf and Zakai [164, 165]. Our framework may offer a unified perspective behind these different types of relations involving information measures.

---

[86] It should be noted that various definitions of housekeeping and excess EPRs have been proposed in the research field of nonequilibrium thermodynamics. The definition here is just one of them. It is an open problem how to define the notion of effective entropy for nonequilibrium dynamics.

10.2 Extensions of the doubly dual flat structure

There is also room for extensions of the doubly dual flat structure. While we consider only strictly convex thermodynamic functions and dissipation functions, the strict convexity is not necessarily required, at least for defining the generalized flow and the equilibrium and nonequilibrium flows[87]. Actually, in terms of thermodynamics, the thermodynamic function can be non-strictly convex when a phase transition of the system occurs [117]. The loss of bijectivity via the loss of the strict convexity can happen in complicated and degenerate statistical models [166]. Techniques from algebraic geometry could be employed to address such situations [167].

Moreover, the structure introduced for rLDG and CRN may be extended to irreversible cases, where some edges have only either forward or reverse jumps or reactions. For this purpose, we may take advantage of several results about the CB states obtained in CRN theory [8] where the reversibility is not necessarily assumed and those in stochastic thermodynamics for absolute irreversible processes [168].

While the nonequilibrium flow is general enough to cover at least all reversible CRN with LMA kinetics, the classes of nonlinear dynamics other than CRN are much wider in general. To further extend the range of models that can be covered, GENERIC (General Equation for Non-Equilibrium Reversible–Irreversible Coupling) would be a good candidate [169]. GENERIC is a theoretical framework to integrate dissipative dynamics (gradient flow dynamics) and conservative dynamics (Hamiltonian dynamics). The extension of the generalized flow to GENERIC has already been attempted but is still ongoing [77, 79]. One might also consider Hamiltonian-type dynamics, which differs from the GENERIC structure mentioned above. In the doubly dual flat structure, dual spaces are statically coupled by the Legendre duality. However, we could consider the coupling of two dynamics, each of which is defined on the primal and the dual spaces. Such coupling has been investigated in relation to accelerations of gradient flows [170], optimal control problems [171], and also mean field game problems [172]. It would be an interesting problem to formulate this dynamic coupling in relation to our results and also the results of GENERIC. The information geometry could offer new insights and techniques to achieve these missions.

10.3 Homological algebra and differential geometric formulations

From the viewpoint of the standard homological algebra, the doubly dual flat structure that we introduced is an extension of chain and cochain complexes with inner product structure. Because the homological algebra used here is an abstraction of the differential form, the doubly dual flat structure can also be viewed as an extension of the differential form and might be called *dually flat form*. It would be an interesting mission to characterize this stricture under a more rigorous mathematical formulation and to investigate if the Legendre duality can be consistently introduced for chains

---

[87] The strict convexity of the dissipation functions is assumed to work on the projections in the edge spaces where the bijection between $\mathcal{J}_x$ and $\mathcal{F}_x$ are important. In addition, $\mathcal{Y} = \mathbb{R}^{N_v}$ is assumed to make the induced dually flat structure well-defined for all given v. Only to define the generalized flow and equilibrium and nonequilibrium flow, injectivity of $\partial \Psi_x^*$ and $\partial \Phi$ is sufficient and $\mathcal{Y} = \mathbb{R}^{N_v}$ is not required.

and cochains higher than those of the edge space. From the viewpoint of differential geometry, the dual flat structure can be defined independently of the specific coordinate by the Hessian geometry [145]. While we stick to the standard basis of the graph or hypergraph on which the convex thermodynamic function is defined, we can formulate it more generally. It would be an important future work to clarify how the doubly dual flat structure can be formulated from a differential geometric perspective.

10.4 Consistency and Persistence

Finally, we would like to mention the problem of consistency and persistence. In this work, we presume that the flux $j(x)$ is consistent with $\mathbb{H}$ and that the trajectory is persistent. The explicit conditions when these are satisfied are still elusive. Actually, the condition for consistency is intricate, even for the separable cases. For an illustrative example, suppose that the ith molecule is involved as a reactant in the eth reaction of a CRN with LMA kinetics. For $x_i \to 0$, $j_e^+(x) \to 0$ holds. However, their Legendre dual diverges as $y_i \to -\infty$ and $f_e(x) \to -\infty$. The flux $j_e(x)$ stays finite because $\omega_e(x) \to 0$ holds. This example suggests that we have to consider a certain limit of relevant quantities to appropriately address the consistency condition.

The persistence of the nonequilibrium flow would be a much harder problem. While persistence has been approached for CB CRN with LMA kinetics using techniques from algebraic geometry [112], its connection to information geometry has yet to be clarified. Moreover, from an information-geometric viewpoint, the loss of persistence means a change in the support of probability or positive density, which effectively results in a change in the topology of the underlying graph or hypergraph. To resolve the problem, we may need a deeper understanding of the interrelationship among dynamics, information-geometric structure, and the underlying topology.

Data availability statement Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Declarations

Conflict of interest The authors have no conflict of interest, financial or otherwise.

A Symbols, Notations, and Abbreviations

Table 1 List of symbols and notations related to graph, hypergraph, and homological algebra

$\mathbb{G}, \mathbb{G}_{\boldsymbol{k}\pm}$	Finite graph and edge-weighted finite graph	Def. 1
$\mathbb{v}, \mathbb{e}$	Vertex and edges of graph $\mathbb{G}$	Def. 1
$\mathbb{B}$	Incidence matrix	Def. 1
$\mathbb{H}, \mathbb{H}_{\boldsymbol{k}\pm}$	Reversible CRN hypergraph and edge-weighted hypergraph	Def. 4
$\mathbb{X}, \hat{\mathbb{v}}, \mathbb{e}$	Vertex, hypervertex, and hyperedges of CRN hypergraph $\mathbb{H}$	Def. 4
$\mathbb{S}$	Hypergraph incidence matrix	Def. 4
$\mathbb{\Gamma}$	Hypervertex matrix	Def. 4
$C_p(\mathbb{G}), C^p(\mathbb{G})$	p-chain and p-cochain of graph with field $\mathbb{R}$	Sect. 3.1
$C_p(\mathbb{H}), C^p(\mathbb{H})$	p-chain and p-cochain of hypergraph with field $\mathbb{R}$	Def. 9
δ_p, δ^p	p-discrete differentials	Sect. 3.1, Def. 9
$\mathrm{grad}_{\mathbb{S}} := \delta^0 = \mathbb{S}^T$	Discrete gradient	Def. 10
$\mathrm{div}_{\mathbb{S}} = \mathrm{grad}_{\mathbb{S}}^* := \delta_1 = \mathbb{S}$	Discrete divergence (adjoint of discrete gradient)	Def. 10
$\mathrm{curl}_{\mathbb{V}} := \delta^1 = \mathbb{V}^T$	Discrete curl	Def. 10
$\mathrm{curl}_{\mathbb{V}}^* := \delta_2 = \mathbb{V}$	Adjoint of discrete curl	Def. 10

Table 2 List of symbols and notations related to the dynamics on graphs and hypergraphs

$\mathcal{X}, \mathcal{Y}$	Primal and dual vertex spaces (density and potential spaces)	Def. 11, Def. 13
$\boldsymbol{x}$	The density defined on vertices of graph or hypergraph	Def. 2, Def. 4
$\boldsymbol{y}$	The potential field defined on vertices of graph or hypergraph	Def. 13
$\mathcal{J}_{\boldsymbol{x}}, \mathcal{F}_{\boldsymbol{x}}$	Primal and dual edge spaces	Def. 17
$\boldsymbol{j}^+, \boldsymbol{j}^-$	The forward and reverse oneway fluxes of graph or hypergraph	Def. 2
$\boldsymbol{j}$	The total flux on edges of graph or hypergraph	Def. 2
$\boldsymbol{f}$	The force on edges of graph or hyergraph	Eq. 36
$\boldsymbol{j}_{\mathrm{MA}}^{\pm}, \boldsymbol{j}_{\mathrm{eMA}}^{\pm}$	Oneway fluxes of the normal and extended LMA kinetics	Eq. 12, Eq. 13
$\boldsymbol{j}_{\mathrm{FP}}$	Flux of the FPE	Eq. 18
$\boldsymbol{f}_{NE}$	Nonequilibrium force	Eq. 60
$\boldsymbol{f}_{\mathrm{MA}}(\boldsymbol{x}; \boldsymbol{K}), \omega_{\mathrm{MA}}(\boldsymbol{x}; \boldsymbol{\kappa})$	Force and frenetic activity of LMA kinetics	Eq. 70
$\mathcal{M}^{\mathrm{ST}}$	Steady state manifold	Eq. 47
$\mathcal{M}^{\mathrm{CB}}$	Complex-balanced manifold	Eq. 47
$\mathcal{M}^{\mathrm{DB}}$	Detailed-balanced manifold	Eq. 47
$\mathcal{L}_{\boldsymbol{\theta}}$	Weighted asymmetric graph Laplacian	Def. 3

Table 3 List of symbols and notations related to the information-geometric structure

Φ, Φ^*	Primal and dual thermodynamic functions	Def. 12, Def. 13
$\mathcal{D}_{\Phi}^{\mathcal{X}}[\boldsymbol{x}\|\boldsymbol{x}'], \mathcal{D}_{\Phi^*}^{\mathcal{Y}}[\boldsymbol{y}'\|\boldsymbol{y}]$	Bregman divergences on vertex spaces	Def. 14
$\mathcal{T}_{\boldsymbol{x}}\mathcal{X}, \mathcal{T}_{\boldsymbol{x}}^*\mathcal{X}$	Tangent and cotangent spaces of $\mathcal{X}$	Def. 15
$\mathcal{T}_{\boldsymbol{y}}\mathcal{Y}, \mathcal{T}_{\boldsymbol{y}}^*\mathcal{Y}$	Tangent and cotangent spaces of $\mathcal{Y}$	Def. 15
$G_{\boldsymbol{x}}, G_{\boldsymbol{y}}^*$	Hessian matrices on vertex spaces	Def. 15
$\Psi_{\boldsymbol{x}}, \Psi_{\boldsymbol{x}}^*$	Primal and dual dissipation functions	Def. 18
$\mathcal{D}_{\boldsymbol{x}}^{\mathcal{J},\mathcal{F}}[\boldsymbol{j};\boldsymbol{f}']$	Bregman divergence on the edge space	Eq. 41
$G_{\boldsymbol{x},\boldsymbol{j}}, G_{\boldsymbol{x},\boldsymbol{f}}^*$	Hessian matrices on the edge space	Eq. 42
$\Psi_{\boldsymbol{x}}^{q,*}(\boldsymbol{f})$	Quadratic dissipation function	Eq. 43
$\mathcal{P}^{sc}(\boldsymbol{x}_0)$	Stoichiometric subspace	Def. 26
$\mathcal{P}^{eq}(\tilde{\boldsymbol{y}})$	Equilibrium subspace	Def. 27
$\mathcal{P}^{vl}(\hat{\boldsymbol{j}})$	Iso-velocity subspace	Def. 28
$\mathcal{P}^{fr}(\boldsymbol{f}')$	Iso-force subspace	Def. 29
$\mathcal{M}^{sc}(\boldsymbol{y}_0), \mathcal{M}^{eq}(\tilde{\boldsymbol{x}})$	Stoichiometric and equilibrium manifolds	Def. 30
$\mathcal{M}_{\boldsymbol{x}}^{vl}(\hat{\boldsymbol{f}}), \mathcal{M}_{\boldsymbol{x}}^{fr}(\boldsymbol{j}')$	Iso-velocity and Iso-force manifolds	Def. 31
$\mathcal{C}_{\boldsymbol{x}}^{\Psi}(c), \mathcal{C}_{\boldsymbol{x}}^{\Psi^*}(c)$	Primal central affine manifolds	Def. 32
$\mathcal{M}_{\boldsymbol{x}}^{\Psi}(c), \mathcal{M}_{\boldsymbol{x}}^{\Psi^*}(c)$	Dual central affine manifolds	Def. 120
$\perp_H$	Pseudo-Hilbert-isosceles orthogonality	Def. 33
$\tilde{\mathcal{T}}_{\boldsymbol{x}}\mathcal{X}, \tilde{\mathcal{T}}_{\boldsymbol{x}}^*\mathcal{X}$	Induced tangent and cotangent spaces of $\mathcal{X}$	Theorem 2
$\tilde{\Psi}_{\boldsymbol{x}}(\boldsymbol{v}), \tilde{\Psi}_{\boldsymbol{x}}^*(\boldsymbol{u})$	Induced dissipation functions on tangent and cotangent spaces	Eq. 132
$\mathcal{Z}_{\boldsymbol{x}}, \mathfrak{Z}_{\boldsymbol{x}}$	Cycle spaces	Def. 34
$\hat{\Psi}_{\boldsymbol{x}}(\boldsymbol{z}), \hat{\Psi}_{\boldsymbol{x}}^*(\boldsymbol{\zeta})$	Induced dissipation functions on cycle spaces	Eq. 162

Table 4 List of abbreviations

CRN	Chemical reaction network
LMA	Law of mass action
(r)MJP	(Reversible) Markov jump process
(r)LDG	(Reversible) linear dynamics on graph
PDS	Polynomial dynamical systems
FPE	Fokker Planck Equation
CB	complex-balanced
DB	detail-balanced
LDB	local detailed balance
EPR	entropy production rate
HHK	Helmholtz-Hodge-Kodaira
MCMC	Markov chain Monte Carlo

References

1. Amari, S.-I.: Information Geometry and Its Applications. Springer, New York (2016)

2. Ay, N., Gibilisco, P., Matúš, F.: Information Geometry and Its Applications: On the Occasion of Shun-ichi Amari's 80th Birthday, IGAIA IV Liblice, Czech Republic, June 2016. Springer, New York (2018)

3. Risken, H., Frank, T.: The Fokker-Planck Equation: Methods of Solution And Applications. Springer Science & Business Media, New York (1996)

4. Horsthemke, W., Lefever, R.: Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology. Springer Science & Business Media, New York (2006)

5. Gardiner, C.: Stochastic Methods: A Handbook for the Natural and Social Sciences. Springer, Berlin (2010)

6. Murray, J.D.: In: Mathematical Biology: I. An Introduction. Springer Science & Business Media, New York (2007)

7. Beard, D.A., Qian, H.: Chemical Biophysics: Quantitative Analysis of Cellular Systems. Cambridge Texts in Biomedical Engineering. Cambridge University Press, Cambridge (2008). (**10.1017/CBO9780511803345**)

8. Feinberg, M.: Foundations of Chemical Reaction Network Theory. Springer, New York (2019)

9. Amari, S.: Differential Geometry in Statistical Inference. Institute of mathematical Statistics, Hayward, Calif (1987)

10. Ravishanker, N., Melnick, E.L., Tsai, C.-L.: Differential Geometry of Arma Models. J. Time Ser. Anal. **11**(3), 259–274 (1990). https://doi.org/10.1111/j.1467-9892.1990.tb00057.x

11. Tanaka, F., Komaki, F.: Asymptotic expansion of the risk difference of the Bayesian spectral density in the autoregressive moving average model. Sankhya A **73**(1), 162–184 (2011). https://doi.org/10.1007/s13171-011-0005-1

12. Amari, S.-I: Differential geometry of a parametric family of invertible linear systems-Riemannian metric, dual affine connections, and divergence. Math. Systems Theory **20**(1), 53–82 (1987). https://doi.org/10.1007/BF01692059

13. Amari, S.-I.: Information geometry on hierarchy of probability distributions. IEEE Trans. Inf. Theory **47**(5), 1701–1711 (2001). https://doi.org/10.1109/18.930911

14. Nakagawa, K., Kanaya, F.: On the converse theorem in statistical hypothesis testing for Markov chains. IEEE Trans. Inf. Theory **39**(2), 629–633 (1993). https://doi.org/10.1109/18.212294

15. Takeuchi, J., Barron, A.R.: Asymptotically minimax regret by Bayes mixtures. In: Proc. 1998 IEEE Int. Symp. Inf. Theory Cat No98CH36252, p. 318 (1998). https://doi.org/10.1109/ISIT.1998.708923

16. Nagaoka, H.: The Exponential Family of Markov Chains and Its Information. Geometry (2017). https://doi.org/10.48550/arXiv.1701.06119

17. Takeuchi, J., Kawabata, T.: Exponential Curvature of Markov Models. In: 2007 IEEE Int. Symp. Inf. Theory, pp. 2891–2895 (2007). https://doi.org/10.1109/ISIT.2007.4557657

18. Hayashi, M., Watanabe, S.: Information geometry approach to parameter estimation in Markov chains. Ann. Stat. **44**(4), 1495–1535 (2016). https://doi.org/10.1214/15-AOS1420

19. Wolfer, G., Watanabe, S.: Information Geometry of Reversible Markov Chains. Info. Geo. **4**(2), 393–433 (2021). https://doi.org/10.1007/s41884-021-00061-7

20. Pistone, G., Rogantin, M.P.: The algebra of reversible Markov chains. Ann. Inst. Stat. Math. **65**(2), 269–293 (2013). https://doi.org/10.1007/s10463-012-0368-7

21. Obata, T., Hara, H., Endo, K.: Differential geometry of nonequilibrium processes. Phys. Rev. A **45**(10), 6997–7001 (1992). https://doi.org/10.1103/PhysRevA.45.6997

22. Ohara, A.: Geometric study for the Legendre duality of generalized entropies and its application to the porous medium equation. Eur. Phys. J. B **70**(1), 15–28 (2009). https://doi.org/10.1140/epjb/e2009-00170-y

23. Ohara, A., Zhang, X.: Properties of Nonlinear Diffusion Equations on Networks and Their Geometric Aspects. In: Nielsen, F., Barbaresco, F. (eds.) Geom. Sci. Inf. Lecture Notes in Computer Science, pp. 736–743. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-80209-7_79

24. Nakamura, Y.: Completely integrable gradient systems on the manifolds of Gaussian and multinomial distributions. Japan J. Indust. Appl. Math. **10**(2), 179 (1993). https://doi.org/10.1007/BF03167571

25. Fujiwara, A., Amari, S.-I.: Gradient systems in view of information geometry. Physica D **80**(3), 317–327 (1995). https://doi.org/10.1016/0167-2789(94)00175-P

26. Felice, D., Ay, N.: Dynamical Systems Induced by Canonical Divergence in Dually Flat Manifolds. arXiv (2018). https://doi.org/10.48550/arXiv.1812.04461

27. Goto, S.-I., Wada, T.: Hessian–information geometric formulation of Hamiltonian systems and generalized Toda's dual transform. J. Phys. A: Math. Theor. **51**(32), 324001 (2018). https://doi.org/10.1088/1751-8121/aacbdf

28. Chirco, G., Malagò, L., Pistone, G.: Lagrangian and Hamiltonian Mechanics for Probabilities on the Statistical Manifold. Int. J. Geom. Methods Mod. Phys., 2250214 (2022) https://doi.org/10.1142/S0219887822502140 arxiv:2009.09431 [hep-th, stat]

29. Ihara, S.: Information Theory for Continuous Systems. World Scientific (1993)

30. Brigo, D., Hanzon, B., Gland, F.L.: Approximate nonlinear filtering by projection on exponential manifolds of densities. Bernoulli **5**(3), 495–534 (1999)

31. Newton, N.J.: Nonlinear Filtering and Information Geometry: A Hilbert Manifold Approach. In: Ay, N., Gibilisco, P., Matúš, F. (eds.) Inf. Geom. Its Appl. Springer Proceedings in Mathematics & Statistics, pp. 189–208. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-97798-0_7

32. Fleming, W.H., Mitter, S.K.: Optimal control and nonlinear filtering for nondegenerate diffusion processes. Stochastics **8**(1), 63–77 (1982). https://doi.org/10.1080/17442508208833228

33. Todorov, E.: Efficient computation of optimal actions. Proc. Natl. Acad. Sci. **106**(28), 11478–11483 (2009). https://doi.org/10.1073/pnas.0710743106

34. Theodorou, E.A., Todorov, E.: Relative entropy and free energy dualities: Connections to Path Integral and KL control. In: 2012 IEEE 51st IEEE Conf. Decis. Control CDC, pp. 1466–1473 (2012). https://doi.org/10.1109/CDC.2012.6426381

35. Jaynes, E.T.: Information Theory and Statistical Mechanics. Phys. Rev. **106**(4), 620–630 (1957). https://doi.org/10.1103/PhysRev.106.620

36. Frieden, B.R.: Science from Fisher Information: A Unification. Cambridge University Press, Cambridge (2004). https://doi.org/10.1017/CBO9780511616907

37. Sagawa, T.: Thermodynamics of Information Processing in Small Systems. Springer Science & Business Media (2012)

38. Kullback, S., Leibler, R.A.: On Information and Sufficiency. Ann. Math. Stat. **22**(1), 79–86 (1951). https://doi.org/10.1214/aoms/1177729694

39. Lebowitz, J.L., Bergmann, P.G.: Irreversible gibbsian ensembles. Ann. Phys. **1**(1), 1–23 (1957). https://doi.org/10.1016/0003-4916(57)90002-7

40. Shear, D.: An analog of the Boltzmann H-theorem (a Liapunov function) for systems of coupled chemical reactions. J. Theor. Biol. **16**(2), 212–228 (1967). https://doi.org/10.1016/0022-5193(67)90005-7

41. Horn, F., Jackson, R.: General mass action kinetics. Arch. Rational Mech. Anal. **47**(2), 81–116 (1972). https://doi.org/10.1007/BF00251225

42. Goh, B.S.: Global Stability in Many-Species Systems. Am. Nat. **111**(977), 135–143 (1977). https://doi.org/10.1086/283144

43. Figueiredo, A., Gléria, I.M., Rocha Filho, T.M.: Boundedness of solutions and Lyapunov functions in quasi-polynomial systems. Phys. Lett. A **268**(4), 335–341 (2000). https://doi.org/10.1016/S0375-9601(00)00175-4

44. Gibbs, J.W.: Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics. Cambridge Library Collection - Mathematics. Cambridge University Press, Cambridge (2010). https://doi.org/10.1017/CBO9780511686948

45. Waage, P., Gulberg, C.M.: Studies concerning affinity. J. Chem. Educ. **63**(12), 1044 (1986). https://doi.org/10.1021/ed063p1044

46. 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). https://doi.org/10.1016/j.chemphys.2016.03.026

47. Rao, R., Esposito, M.: Nonequilibrium Thermodynamics of Chemical Reaction Networks: Wisdom from Stochastic Thermodynamics. Phys. Rev. X **6**(4), 041064 (2016). https://doi.org/10.1103/PhysRevX.6.041064

48. Sughiyama, Y., Loutchko, D., Kamimura, A., Kobayashi, T.J.: Hessian geometric structure of chemical thermodynamic systems with stoichiometric constraints. Phys. Rev. Research **4**(3), 033065 (2022). https://doi.org/10.1103/PhysRevResearch.4.033065

49. Seifert, U.: Stochastic thermodynamics, fluctuation theorems and molecular machines. Rep. Prog. Phys. **75**(12), 126001 (2012). https://doi.org/10.1088/0034-4885/75/12/126001

50. Ito, S.: Stochastic Thermodynamic Interpretation of Information Geometry. Phys. Rev. Lett. **121**(3), 030605 (2018). https://doi.org/10.1103/PhysRevLett.121.030605

51. Kolchinsky, A., Wolpert, D.H.: Work, Entropy Production, and Thermodynamics of Information under Protocol Constraints. Phys. Rev. X **11**(4), 041024 (2021). https://doi.org/10.1103/PhysRevX.11.041024

52. Yoshimura, K., Ito, S.: Information geometric inequalities of chemical thermodynamics. Phys. Rev. Research **3**(1), 013175 (2021). https://doi.org/10.1103/PhysRevResearch.3.013175

53. Ohga, N., Ito, S.: Information-geometric Legendre duality in stochastic thermodynamics. ArXiv211211008 Cond-Mat (2021) arxiv:2112.11008 [cond-mat]

54. Stam, A.J.: Some inequalities satisfied by the quantities of information of Fisher and Shannon. Inf. Control **2**(2), 101–112 (1959). https://doi.org/10.1016/S0019-9958(59)90348-1

55. Plastino, A.R., Casas, M., Plastino, A.: Fisher's information, Kullback's measure, and H-theorems. Phys. Lett. A **246**(6), 498–504 (1998). https://doi.org/10.1016/S0375-9601(98)00567-2

56. Wibisono, A., Jog, V., Loh, P.-L.: Information and estimation in Fokker-Planck channels. In: 2017 IEEE Int. Symp. Inf. Theory ISIT, pp. 2673–2677 (2017). https://doi.org/10.1109/ISIT.2017.8007014

57. Fisher, R.A., Russell, E.J.: On the mathematical foundations of theoretical statistics. Philos. Trans. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character **222**(594–604), 309–368 (1922). https://doi.org/10.1098/rsta.1922.0009

58. Rao, B.R.: On an analogue of Cramér-Rao's inequality. Scand. Actuar. J. **1958**(1–2), 57–67 (1958). https://doi.org/10.1080/03461238.1958.10405982

59. Papaioannou, T., Ferentinos, K.: On Two Forms of Fisher's Measure of Information. Commun. Stat. - Theory Methods **34**(7), 1461–1470 (2005). https://doi.org/10.1081/STA-200063386

60. Kharazmi, O., Asadi, M.: On the time-dependent Fisher information of a density function. Braz. J. Probab. Stat. **32**(4), 795–814 (2018). https://doi.org/10.1214/17-BJPS366

61. Johnson, O.: Information Theory and the Central Limit Theorem. World Scientific (2004)

62. Yamano, T.: De Bruijn-type identity for systems with flux. Eur. Phys. J. B **86**(8), 363 (2013). https://doi.org/10.1140/epjb/e2013-40634-9

63. Gross, L.: Logarithmic Sobolev Inequalities. Am. J. Math. **97**(4), 1061–1083 (1975). https://doi.org/10.2307/2373688

64. Gross, L.: Hypercontractivity and logarithmic Sobolev inequalities for the Clifford-Dirichlet form. Duke Math. J. **42**(3), 383–396 (1975). https://doi.org/10.1215/S0012-7094-75-04237-4

65. Otto, F.: The Geometry of Dissipative Evolution Equations: The Porous Medium Equation. Commun. Partial Differ. Equ. **26**(1–2), 101–174 (2001). https://doi.org/10.1081/PDE-100002243

66. Villani, C.: Topics in Optimal Transportation. American Mathematical Soc, New York (2003)

67. Amari, S.-I: Natural Gradient Works Efficiently in Learning. Neural Comput. **10**(2), 251–276 (1998). https://doi.org/10.1162/089976698300017746

68. Beck, A., Teboulle, M.: Mirror descent and nonlinear projected subgradient methods for convex optimization. Oper. Res. Lett. **31**(3), 167–175 (2003). https://doi.org/10.1016/S0167-6377(02)00231-6

69. Raskutti, G., Mukherjee, S.: The Information Geometry of Mirror Descent. IEEE Trans. Inf. Theory **61**(3), 1451–1457 (2015). https://doi.org/10.1109/TIT.2015.2388583

70. Ollivier, Y., Arnold, L., Auger, A., Hansen, N.: Information-Geometric Optimization Algorithms: A Unifying Picture via Invariance Principles. J. Mach. Learn. Res. **18**(18), 1–65 (2017)

71. Hino, H., Akaho, S., Murata, N.: Geometry of EM and related iterative algorithms. Info. Geo. (2022). https://doi.org/10.1007/s41884-022-00080-y

72. Pistone, G.: Information Geometry of Smooth Densities on the Gaussian Space: Poincaré Inequalities. In: Nielsen, F. (ed.) Progress in Information Geometry: Theory and Applications. Signals and Communication Technology, pp. 1–17. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-65459-7_1

73. Grady, L.J., Polimeni, J.R.: Discrete Calculus: Applied Analysis on Graphs for Computational Science. Springer Science & Business Media, New York (2010)

74. Musielak, J.: Orlicz Spaces and Modular Spaces. Springer, New York (1983)

75. Mielke, A., Peletier, M.A., Renger, D.R.M.: On the Relation between Gradient Flows and the Large-Deviation Principle, with Applications to Markov Chains and Diffusion. Potential Anal **41**(4), 1293–1327 (2014). https://doi.org/10.1007/s11118-014-9418-5

76. Mielke, A., Patterson, R.I.A., Peletier, M.A., Michiel Renger, D.R.: Non-equilibrium Thermodynamical Principles for Chemical Reactions with Mass-Action Kinetics. SIAM J. Appl. Math. **77**(4), 1562–1585 (2017). https://doi.org/10.1137/16M1102240

77. Renger, D.R.M.: Gradient and GENERIC Systems in the Space of Fluxes. Applied to Reacting Particle Systems. Entropy **20**(8), 596 (2018). https://doi.org/10.3390/e20080596

78. Kaiser, M., Jack, R.L., Zimmer, J.: Canonical Structure and Orthogonality of Forces and Currents in Irreversible Markov Chains. J. Stat. Phys. **170**(6), 1019–1050 (2018). https://doi.org/10.1007/s10955-018-1986-0

79. Patterson, R.I.A., Renger, D.R.M., Sharma, U.: Variational structures beyond gradient flows: A macroscopic fluctuation-theory perspective. ArXiv210314384 Math-Ph (2021) arxiv:2103.14384 [math-ph]

80. Peletier, M.A., Rossi, R., Savaré, G., Tse, O.: Jump processes as generalized gradient flows. Calc. Var. **61**(1), 33 (2022). https://doi.org/10.1007/s00526-021-02130-2

81. Renger, D.R.M., Zimmer, J.: Orthogonality of fluxes in general nonlinear reaction networks. Discrete Contin. Dyn. Syst. - S **14**(1), 205 (2021). https://doi.org/10.3934/dcdss.2020346

82. Peletier, M.A., Schlichting, A.: Cosh gradient systems and tilting. Nonlinear Anal. **231**, 113094 (2023). https://doi.org/10.1016/j.na.2022.113094

83. Kobayashi, T.J., Loutchko, D., Kamimura, A., Sughiyama, Y.: Hessian geometry of nonequilibrium chemical reaction networks and entropy production decompositions. Phys. Rev. Research **4**(3), 033208 (2022). https://doi.org/10.1103/PhysRevResearch.4.033208

84. Kobayashi, T.J., Loutchko, D., Kamimura, A., Sughiyama, Y.: Kinetic derivation of the Hessian geometric structure in chemical reaction networks. Phys. Rev. Research **4**(3), 033066 (2022). https://doi.org/10.1103/PhysRevResearch.4.033066

85. Sughiyama, Y., Kamimura, A., Loutchko, D., Kobayashi, T.J.: Chemical thermodynamics for growing systems. Phys. Rev. Research **4**(3), 033191 (2022). https://doi.org/10.1103/PhysRevResearch.4.033191

86. Godsil, C., Royle, G.F.: Algebraic Graph Theory. Springer Science & Business Media, New York (2013)

87. Bretto, A.: Hypergraph Theory: An Introduction. Springer Science & Business Media, New York (2013)

88. Meyn, S., Tweedie, R.L.: Markov Chains and Stochastic Stability. Cambridge University Press, Cambridge (2009)

89. Schnakenberg, J.: Network theory of microscopic and macroscopic behavior of master equation systems. Rev. Mod. Phys. **48**(4), 571–585 (1976). https://doi.org/10.1103/RevModPhys.48.571

90. Craciun, G.: Polynomial Dynamical Systems, Reaction Networks, and Toric Differential Inclusions. arXiv (2019). https://doi.org/10.48550/arXiv.1901.02544

91. Biggs, N.: Algebraic Potential Theory on Graphs. Bull. Lond. Math. Soc. **29**(6), 641–682 (1997). https://doi.org/10.1112/S0024609397003305

92. Chung, F.R.K., Graham, F.C.: Spectral Graph Theory. American Mathematical Soc (1997)

93. Dörfler, F., Simpson-Porco, J.W., Bullo, F.: Electrical Networks and Algebraic Graph Theory: Models, Properties, and Applications. Proc. IEEE **106**(5), 977–1005 (2018). https://doi.org/10.1109/JPROC.2018.2821924

94. Saber, R.O., Murray, R.M.: Consensus protocols for networks of dynamic agents. In: Proc. 2003 Am. Control Conf. 2003, vol. 2, pp. 951–956 (2003). https://doi.org/10.1109/ACC.2003.1239709

95. Veerman, J.J.P., Lyons, R.: A Primer on Laplacian Dynamics in Directed Graphs. arXiv (2020). https://doi.org/10.48550/arXiv.2002.02605

96. Qian, H., Ge, H.: Stochastic Chemical Reaction Systems in Biology. Springer Nature, New York (2021)

97. Keener, J., Sneyd, J.: Mathematical Physiology: I: Cellular Physiology. Springer Science & Business Media, New York (2008)

98. Sottile, F.: Toric ideals, real toric varieties, and the algebraic moment map. arXiv:math/0212044 (2008) arxiv:math/0212044

99. Cox, D.A., Little, J.B., Schenck, H.K.: Toric Varieties. American Mathematical Soc, New York (2011)

100. Craciun, G., Dickenstein, A., Shiu, A., Sturmfels, B.: Toric dynamical systems. J. Symb. Comput. **44**(11), 1551–1565 (2009). https://doi.org/10.1016/j.jsc.2008.08.006

101. Rapallo, F.: Toric statistical models: Parametric and binomial representations. AISM **59**(4), 727–740 (2007). https://doi.org/10.1007/s10463-006-0079-z

102. Müller, S., Regensburger, G.: Generalized Mass Action Systems: Complex Balancing Equilibria and Sign Vectors of the Stoichiometric and Kinetic-Order Subspaces. SIAM J. Appl. Math. **72**(6), 1926–1947 (2012). https://doi.org/10.1137/110847056

103. Noor, E., Flamholz, A., Liebermeister, W., Bar-Even, A., Milo, R.: A note on the kinetics of enzyme action: A decomposition that highlights thermodynamic effects. FEBS Lett. **587**(17), 2772–2777 (2013). https://doi.org/10.1016/j.febslet.2013.07.028

104. Yoshimura, K., Kolchinsky, A., Dechant, A., Ito, S.: Housekeeping and excess entropy production for general nonlinear dynamics. Phys. Rev. Res. **5**(1), 013017 (2023). https://doi.org/10.1103/PhysRevResearch.5.013017

105. Lim, L.-H.: Hodge Laplacians on Graphs. SIAM Rev. **62**(3), 685–715 (2020). https://doi.org/10.1137/18M1223101

106. Sunada, T.: Topological Crystallography: With a View Towards Discrete Geometric Analysis. Springer Science & Business Media, New York (2012)

107. Desbrun, M., Hirani, A.N., Leok, M., Marsden, J.E.: Discrete Exterior Calculus. arXiv (2005). https://doi.org/10.48550/arXiv.math/0508341

108. Hirani, A.N.: Discrete Exterior Calculus. PhD thesis, California Institute of Technology (2003). https://doi.org/10.7907/ZHY8-V329

109. Knauer, U.: Algebraic Graph Theory: Morphisms. Monoids and Matrices. Walter de Gruyter, New York (2011)

110. Chen, W.-K.: Applied Graph Theory. Elsevier, Amsterdam (2012)

111. Craciun, G., Nazarov, F., Pantea, C.: Persistence and Permanence of Mass-Action and Power-Law Dynamical Systems. SIAM J. Appl. Math. **73**(1), 305–329 (2013). https://doi.org/10.1137/100812355

112. Craciun, G.: Toric Differential Inclusions and a Proof of the Global Attractor Conjecture (2016). https://doi.org/10.48550/arXiv.1501.02860

113. Bregman, L.M.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR Comput. Math. Math. Phys. **7**(3), 200–217 (1967). https://doi.org/10.1016/0041-5553(67)90040-7

114. Rockafellar, R.T.: Convex Analysis. Princeton University Press (1997)

115. Mitroi, F.-C., Niculescu, C.P.: An Extension of Young's Inequality. Abstr. Appl. Anal. **2011**, 162049 (2011). https://doi.org/10.1155/2011/162049

116. Nielsen, F.: On Geodesic Triangles with Right Angles in a Dually Flat Space. In: Nielsen, F. (ed.) Progress in Information Geometry: Theory and Applications. Signals and Communication Technology, pp. 153–190. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-65459-7_7

117. Callen, H.B., Callen, H.B.: Thermodynamics and an Introduction to Thermostatistics. Wiley, New York (1985)

118. Hiriart-Urruty, J.-B., Lemarechal, C.: Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods. Springer Science & Business Media, New York (1996)

119. Krasnosel'skij, M.A., Rutickij, J.B.: Convex Functions and Orlicz Spaces. Hindustan Publ, New York (1962)

120. Lods, B., Pistone, G.: Information Geometry Formalism for the Spatially Homogeneous Boltzmann Equation. Entropy **17**(6), 4323–4363 (2015). https://doi.org/10.3390/e17064323

121. Pistone, G.: Information Geometry of the Gaussian Space. In: Ay, N., Gibilisco, P., Matúš, F. (eds.) Inf. Geom. Its Appl. Springer Proceedings in Mathematics & Statistics, pp. 119–155. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-97798-0_5

122. Ambrosio, L., Gigli, N., Savare, G.: Gradient Flows: In Metric Spaces and in the Space of Probability Measures. Springer Science & Business Media, New York (2006)

123. Liero, M., Mielke, A., Savaré, G.: Optimal Transport in Competition with Reaction: The Hellinger-Kantorovich Distance and Geodesic Curves. SIAM J. Math. Anal. **48**(4), 2869–2911 (2016). https://doi.org/10.1137/15M1041420

124. Onsager, L.: Reciprocal Relations in Irreversible Processes. I. Phys. Rev. **37**(4), 405–426 (1931). https://doi.org/10.1103/PhysRev.37.405

125. Onsager, L.: Reciprocal Relations in Irreversible Processes. II. Phys. Rev. **38**(12), 2265–2279 (1931). https://doi.org/10.1103/PhysRev.38.2265

126. Machlup, S., Onsager, L.: Fluctuations and Irreversible Process. II. Systems with Kinetic Energy. Phys. Rev. **91**(6), 1512–1515 (1953). https://doi.org/10.1103/PhysRev.91.1512

127. Lisini, S.: Nonlinear diffusion equations with variable coefficients as gradient flows in Wasserstein spaces. ESAIM: COCV **15**(3), 712–740 (2009). https://doi.org/10.1051/cocv:2008044
128. Peletier, M.A.: Variational Modelling: Energies, Gradient Flows, and Large Deviations. arXiv (2014). https://doi.org/10.48550/arXiv.1402.1990
129. Truesdell, C.A., Truesdell, C., Noll, W., Antman, S., Noll, W.: The Non-Linear Field Theories of Mechanics. Springer Science & Business Media (2004)
130. Bergmann, P.G., Lebowitz, J.L.: New Approach to Nonequilibrium Processes. Phys. Rev. **99**(2), 578–587 (1955). https://doi.org/10.1103/PhysRev.99.578
131. Maes, C.: Local detailed balance. SciPost Phys. Lect. Notes, 032 (2021) https://doi.org/10.21468/SciPostPhysLectNotes.32
132. Maes, C.: Non-Dissipative Effects in Nonequilibrium Systems. Springer, New York (2017)
133. Patterson, R., Renger, M.: Large deviations of reaction fluxes. Math. Phys. Anal. Geom. **22**(3), 21 (2019). https://doi.org/10.1007/s11040-019-9318-4. arxiv:1802.02512 [math]
134. Wegscheider, R.: Über simultane Gleichgewichte und die Beziehungen zwischen Thermodynamik und Reaktionskinetik homogener Systeme. Z. Für Phys. Chem. **39U**(1), 257–303 (1902). https://doi.org/10.1515/zpch-1902-3919
135. Pistone, G., Sempi, C.: An Infinite-Dimensional Geometric Structure on the Space of all the Probability Measures Equivalent to a Given One. Ann. Stat. **23**(5), 1543–1561 (1995). https://doi.org/10.1214/aos/1176324311
136. Pistone, G.: Examples of the Application of Nonparametric Information Geometry to Statistical Physics. Entropy **15**(10), 4042–4065 (2013). https://doi.org/10.3390/e15104042
137. Shiraishi, N.: Optimal Thermodynamic Uncertainty Relation in Markov Jump Processes. J. Stat. Phys. **185**(3), 19 (2021). https://doi.org/10.1007/s10955-021-02829-8
138. Chow, S.-N., Huang, W., Li, Y., Zhou, H.: Fokker-Planck Equations for a Free Energy Functional or Markov Process on a Graph. Arch Rational Mech Anal **203**(3), 969–1008 (2012). https://doi.org/10.1007/s00205-011-0471-6
139. Maas, J.: Gradient flows of the entropy for finite Markov chains. J. Funct. Anal. **261**(8), 2250–2292 (2011). https://doi.org/10.1016/j.jfa.2011.06.009
140. Mielke, A.: Geodesic convexity of the relative entropy in reversible Markov chains. Calc. Var. **48**(1), 1–31 (2013). https://doi.org/10.1007/s00526-012-0538-8
141. Liero, M., Mielke, A.: Gradient structures and geodesic convexity for reaction–diffusion systems. Philos. Trans. R. Soc. Math. Phys. Eng. Sci. **371**(2005), 20120346 (2013). https://doi.org/10.1098/rsta.2012.0346
142. 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). https://doi.org/10.1103/PhysRevX.13.011013
143. Yamano, T.: Phase space gradient of dissipated work and information: A role of relative Fisher information. J. Math. Phys. **54**(11), 113301 (2013). https://doi.org/10.1063/1.4828855
144. Hyvärinen, A.: Estimation of Non-Normalized Statistical Models by Score Matching. J. Mach. Learn. Res. **6**, 695–709 (2005)
145. Shima, H.: The Geometry of Hessian Structures. World Scientific, Singapore (2007)
146. Wolsey, L.A., Nemhauser, G.L.: Integer and Combinatorial Optimization. Wiley, New York (1999)
147. Davis, H.F.: On Isosceles Orthogonality. Math. Mag. **32**(3), 129–131 (1959). https://doi.org/10.2307/3029494
148. Amari, S.-I.: Neural learning in structured parameter spaces: Natural Riemannian gradient. In: Proc. 9th Int. Conf. Neural Inf. Process. Syst. NIPS'96, pp. 127–133. MIT Press, Cambridge, MA, USA (1996)
149. Gunasekar, S., Woodworth, B., Srebro, N.: Mirrorless Mirror Descent: A Natural Derivation of Mirror Descent. In: Proc. 24th Int. Conf. Artif. Intell. Stat., pp. 2305–2313. PMLR (2021)
150. Li, W., Montúfar, G.: Natural gradient via optimal transport. Info. Geo. **1**(2), 181–214 (2018). https://doi.org/10.1007/s41884-018-0015-3
151. Amari, S.-I., Karakida, R., Oizumi, M.: Information geometry connecting Wasserstein distance and Kullback-Leibler divergence via the entropy-relaxed transportation problem. Info. Geo. **1**(1), 13–37 (2018). https://doi.org/10.1007/s41884-018-0002-8
152. Bain, A., Crisan, D.: Fundamentals of Stochastic Filtering. Springer Science & Business Media, New York (2008)

153. Brigo, D., Hanzon, B., LeGland, F.: A differential geometric approach to nonlinear filtering: The projection filter. IEEE Trans. Autom. Control **43**(2), 247–252 (1998). https://doi.org/10.1109/9.661075
154. Li, Y., Cheng, Y., Li, X., Wang, H., Hua, X., Qin, Y.: Information geometric approach for nonlinear filtering. In: 2017 36th Chin. Control Conf. CCC, pp. 1211–1216 (2017). https://doi.org/10.23919/ChiCC.2017.8027514
155. Hill, T.L.: Free Energy Transduction and Biochemical Cycle Kinetics. Courier Corporation, New York (2005)
156. Altaner, B., Grosskinsky, S., Herminghaus, S., Katthän, L., Timme, M., Vollmer, J.: Network representations of nonequilibrium steady states: Cycle decompositions, symmetries, and dominant paths. Phys. Rev. E **85**(4), 041133 (2012). https://doi.org/10.1103/PhysRevE.85.041133
157. Polettini, M., Esposito, M.: Irreversible thermodynamics of open chemical networks. I. Emergent cycles and broken conservation laws. J. Chem. Phys. **141**(2), 024117 (2014). https://doi.org/10.1063/1.4886396
158. Strang, A.: Applications of the Helmholtz-Hodge Decomposition to Networks and Random Processes. PhD thesis, Ann Arbor, United States (August 2020)
159. Oono, Y., Paniconi, M.: Steady State Thermodynamics. Prog. Theor. Phys. Suppl. **130**, 29–44 (1998). https://doi.org/10.1143/PTPS.130.29
160. 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). https://doi.org/10.1103/PhysRevLett.100.230602
161. Maes, C., Netočný, K.: A Nonequilibrium Extension of the Clausius Heat Theorem. J. Stat. Phys. **154**(1), 188–203 (2014). https://doi.org/10.1007/s10955-013-0822-9
162. Dechant, A., Sasa, S.-I., Ito, S.: Geometric decomposition of entropy production in out-of-equilibrium systems. ArXiv210912817 Cond-Mat (2022) arxiv:2109.12817 [cond-mat]
163. Guo, D., Shamai, S., Verdu, S.: Mutual information and minimum mean-square error in Gaussian channels. IEEE Trans. Inf. Theory **51**(4), 1261–1282 (2005). https://doi.org/10.1109/TIT.2005.844072
164. Mayer-Wolf, E., Zakai, M.: On a formula relating the Shannon information to the fisher information for the filtering problem. In: Korezlioglu, H., Mazziotto, G., Szpirglas, J. (eds.) Filter. Control Random Process. Lecture Notes in Control and Information Sciences, pp. 164–171. Springer, Berlin, Heidelberg (1984). https://doi.org/10.1007/BFb0006569
165. Mayer-Wolf, E., Zakai, M.: Some relations between mutual information and estimation error in Wiener space. Ann. Appl. Probab. **17**(3), 1102–1116 (2007). https://doi.org/10.1214/105051607000000131
166. Amari, S.-I., Park, H., Ozeki, T.: Singularies Affect Dynamics of Learning in Neuromanifolds. Neural Comput. **18**(5), 1007–1065 (2006). https://doi.org/10.1162/neco.2006.18.5.1007
167. Watanabe, S.: Algebraic Geometry and Statistical Learning Theory. Cambridge University Press, Cambridge (2009)
168. Murashita, Y., Funo, K., Ueda, M.: Nonequilibrium equalities in absolutely irreversible processes. Phys. Rev. E **90**(4), 042110 (2014). https://doi.org/10.1103/PhysRevE.90.042110
169. Öttinger, H.C.: Beyond Equilibrium Thermodynamics. Wiley, New York (2005)
170. Wang, Y., Li, W.: Accelerated Information Gradient Flow. J. Sci. Comput. **90**(1), 11 (2021). https://doi.org/10.1007/s10915-021-01709-3
171. Li, W., Liu, S., Osher, S.: Controlling conservation laws I: Entropy–entropy flux. J. Comput. Phys. **480**(C), (2023) https://doi.org/10.1016/j.jcp.2023.112019
172. Gao, Y., Li, W., Liu, J.-G.: Master Equations for Finite State Mean Field Games with Nonlinear Activations. arXiv (2022). https://doi.org/10.48550/arXiv.2212.05675