Generalized Gradient Flows for Jump Processes

Oliver Tse

Eindhoven University of Technology

Updated on September 4, 2026
Lecture notes for a school on Gradient Flows at ICTS

Abstract. These notes accompany a short course on *generalized gradient structures* for Markov jump processes. Starting from reversible Markov chains on discrete state spaces, we see that one equation may carry many gradient structures, or none, and we make the *cosh*-type structure rigorous for jump processes on Polish spaces. Which structure to take is settled by a selection principle: the large-deviation rate functional of the underlying particle system is canonical, and splitting it into its symmetric and antisymmetric parts under time reversal produces a dissipation and a free energy—the reason jump processes come out *cosh*.

Contents

1 Gradient structures for discrete Markov chains

ch:discrete

Goal of Lecture 1

Starting from the classical Euclidean picture, we explain what a *(generalized) gradient structure* is, and we see that one equation may carry many such structures, or none at all. Onsager's reciprocal relations explain why any of them is a convex duality between conjugate dissipation potentials—though not which pair to take. We then show that a reversible continuous-time Markov chain on a finite state space is a gradient flow of the relative entropy. We meet the discrete continuity equation, the *graph gradient/graph divergence* duality, and the freedom in choosing a *dissipation potential*—in particular the quadratic (Maas–Mielke–Chow–Huang–Zhou) and the *cosh* structures.

1.1 Warm-up: gradient flows on Euclidean space

sec:euclidean

Consider the ODE

$$\dot{x}(t) = -\nabla F(x(t)), \qquad x(0) = x_0 \in \mathbb{R}^d, \tag{1.1}$$

eq:gf-euclid

with $F\colon \mathbb{R}^d \to \mathbb{R}$ smooth. Calling (1.1) a *gradient flow* is, strictly speaking, an incomplete statement: it becomes meaningful only once we fix a *gradient structure*, consisting of

(i) a state space $\mathcal{X}$ (here $\mathbb{R}^d$);
(ii) a driving energy $F\colon \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$;
(iii) notions of velocity $\dot{x}$ and of gradient ∇F;
(iv) a *kinetic relation* linking the restoring force $-\nabla F(x)$ to the velocity $\dot{x}$.

The same equation can carry several, or even infinitely many, gradient structures, each with possibly different physical or probabilistic meaning.

The kinetic relation. For the moment, we consider (iv) in the linear form

$$\dot{x}(t) = -K(x(t))\,\nabla F(x(t)), \tag{GF}$$

eq:kinetic-linear

where the *mobility*, or *Onsager operator* $K(x)$ is symmetric and positive definite. Its inverse $G(x) := K(x)^{-1}$ is then a metric tensor, and $K = \mathrm{Id}$ returns (1.1). The pair carries with it two quadratic forms, the *dissipation potential* and its *dual*:

$$\mathcal{R}(x,v) := \tfrac{1}{2}\langle v, G(x)v\rangle = \tfrac{1}{2}|v|_G^2, \qquad \mathcal{R}^*(x,\xi) := \tfrac{1}{2}\langle \xi, K(x)\xi\rangle = \tfrac{1}{2}|\xi|_K^2,$$

where the first measures a velocity v, while the second a force ξ.

Remark 1.1 (Onsager reciprocity) The symmetry and positive definiteness of G are not conveniences but physics. They are Onsager's *reciprocal relations* from "Reciprocal relations in irreversible processes" (1931), which earned him the 1963 Nobel Prize. In modern language, they are precisely the statement $G^\top = G$. The symmetry of the Onsager operator is the macroscopic shadow of *microscopic reversibility* of the underlying dynamics—a first hint that reversibility and gradient structure belong together, which Lecture 3 will make precise.

Variational reformulation. The defining characteristic that makes a gradient flow a gradient flow is that it dissipates a driving energy along the flow, i.e., along a solution of (GF) the *chain rule* gives

$$\frac{\mathrm{d}}{\mathrm{d}t}F(x) = \langle \nabla F(x), \dot{x}\rangle = -\langle \nabla F(x), K(x)\nabla F(x)\rangle \leq 0.$$

Yet this identity provides a stronger characterization of gradient flows. Indeed, from the Cauchy–Schwarz and Young inequalities, we obtain

$$\langle \xi, v\rangle \leq \|v\|_G \|\xi\|_K \leq \tfrac{1}{2}\|v\|_G^2 + \tfrac{1}{2}\|\xi\|_K^2 = \mathcal{R}(x,v) + \mathcal{R}^*(x,\xi)$$

for any velocity v and force ξ, with equality if and only if $v = K(x)\xi$. Therefore, taking $\xi = -\nabla F(x)$ and $v = \dot{x}$, this allows us to fully characterize the *vector* equation (GF) with a single *scalar* one:

$$\mathcal{R}(x,\dot{x}) + \mathcal{R}^*\big(x, -\nabla F(x)\big) = \langle \dot{x}, -\nabla F(x)\rangle = -\frac{\mathrm{d}}{\mathrm{d}t}F(x(t)). \tag{1.2}$$

eq:onsager-edp

Integrating over $[s,t]$ yields the so-called *energy–dissipation balance* (EDB)

$$\mathscr{L}(x,\dot{x};[s,t]) := \int_s^t \Big(\mathcal{R}(x,\dot{x}) + \mathcal{R}^*\big(x,-\nabla F(x)\big)\Big)\,\mathrm{d}r + F(x(t)) - F(x(s)) = 0. \tag{EDB}$$

eq:edb-euclid

Since the functional $\mathscr{L}$ is always nonnegative due to the chain rule and Young's inequality, the equality $\mathscr{L} = 0$ on every interval $[s,t]$ *characterizes* solutions of (GF). With $K = \mathrm{Id}$ the two potentials are the familiar $\frac{1}{2}|\dot{x}|^2$ and $\frac{1}{2}|\nabla F(x)|^2$.

Observe that (EDB) is purely *variational* and no longer refers to a Hilbert structure: it needs only a notion of *metric speed* $|\dot{x}|$ and of *metric slope* $|\nabla F|$, and therefore extends to general metric spaces and to spaces of measures.

Some examples. Let us provide concrete illustrations. Recall that a gradient structure consists of a state space, a driving energy, notions of velocity and gradient, and a kinetic relation, but the equation alone determines none of these objects. The point of the example below is that the vector field alone does not determine the gradient structure.

ex:two-structures

Example 1.2 (One equation, two structures) On $\mathcal{X} = \mathbb{R}^2$ consider

$$\dot{x}(t) = V(x(t)), \qquad V(x) = -\big(x_1,\ x_2 + \alpha x_2^3\big)^\top, \qquad \alpha > 0.$$

Setting $F(x) = \frac{1}{2}(x_1^2 + x_2^2) + \frac{\alpha}{4}x_2^4$ gives $\nabla F = -V$, so this solves (GF) with

$$F(x) = \tfrac{1}{2}(x_1^2 + x_2^2) + \tfrac{\alpha}{4}x_2^4, \qquad K(x) = \mathrm{Id}.$$

But it is *equally* a gradient flow with

$$\widetilde{F}(x) = \tfrac{1}{2}(x_1^2 + x_2^2), \qquad \widetilde{K}(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 + \alpha x_2^2 \end{pmatrix},$$

since $\widetilde{K}(x)\nabla\widetilde{F}(x) = (x_1, x_2 + \alpha x_2^3)^\top = -V(x)$.

Therefore, just looking at the ODE, one cannot tell whether the nonlinearity αx_2^3 comes from a non-quadratic *energy* or from a state-dependent *friction law*. The choice of a gradient structure is a modeling choice, and it carries strictly more information than the equation itself.

ex:damped

Example 1.3 (An equation with no gradient structure) The damped harmonic oscillator

$$\dot{x}(t) = V(x(t)), \qquad V(x) = \big(x_2,\ -x_1 - \gamma x_2\big)^\top, \qquad \gamma > 0,$$

possesses no gradient structure at all. It does carry a more general structure—the so-called GENERIC structure, combining a conservative (Hamiltonian) part with a dissipative one—which remains an active field of research. Dissipating energy is thus necessary but not sufficient: the oscillator's energy decays, yet the motion is not steepest descent for any choice of friction.

The moral of the story here is that an evolution $\dot{x} = V(x)$ may carry many gradient structures, or none. Either way, which one to take, if a gradient structure exists, is a question the equation alone cannot answer. We will discuss this in detail in Lecture 3.

1.2 Beyond quadratic dissipation potentials

sec:generalized

So far, we've only considered quadratic dissipation potentials. However, nothing in EDB actually requires this. Relaxing it is exactly what jump dynamics will force us to do in §3. To exploit that freedom, and to reach the (generalized) gradient structures for jump dynamics, we will need to replace the quadratic Young inequality by its convex-duality generalization, namely the *Fenchel–Young inequality*: for any conjugate pair (ψ, ψ^*) of proper, convex, lower-semicontinuous functions, one has $\psi(s) + \psi^*(\xi) \geq \langle \xi, s\rangle$, with the three equivalent characterizations of the equality case

$$\xi \in \partial\psi(s) \quad\Longleftrightarrow\quad \psi(s) + \psi^*(\xi) = \langle \xi, s\rangle \quad\Longleftrightarrow\quad s \in \partial\psi^*(\xi).$$

The generalized step

Replace the two quadratic terms in (1.2) by a pair of convex *dissipation potentials* $\mathcal{R}(x, \dot{x})$ and its Legendre dual $\mathcal{R}^*(x, \xi)$. The evolution equation becomes

$$\dot{x}(t) = D_\xi \mathcal{R}^*\big(x(t), -\nabla F(x(t))\big),$$

and, upon integration, the EDB becomes

$$\mathscr{L}(x, \dot{x}; [s,t]) := \int_s^t \big[\mathcal{R}(x, \dot{x}) + \mathcal{R}^*(x, -\nabla F(x))\big]\, \mathrm{d}r + F(x(t)) - F(x(s)) = 0,$$

which makes sense whenever $\mathcal{R}(x, \cdot)$ is convex with $\mathcal{R}(x, 0) = 0$. A *generalized gradient system* is the triple $(\mathcal{X}, F, \mathcal{R})$ (equivalently $(\mathcal{X}, F, \mathcal{R}^*)$).

When $\mathcal{R}(x, v) = \frac{1}{2}|v|^2$ we recover the classical case.

1.3 Reversible Markov chains as gradient flows

sec:markov

Let V be a finite set of $n \in \mathbb{N}$ *vertices*, and $E := V \times V$ the set of *edges*. The state space is the simplex of probability vectors

$$\mathcal{P}(V) := \Big\{ \rho \colon V \to [0,1] \;\Big|\; \rho_x \geq 0,\ \textstyle\sum_x \rho_x = 1 \Big\} \subset \mathbb{R}^n.$$

A continuous-time Markov chain $(X_t)_{t \geq 0}$ has a *jump kernel* $\kappa = (\kappa_{xy})$, the rate of jumps from x to y, and generator

$$(Q\varphi)(x) = \sum\nolimits_y \big[\varphi(y) - \varphi(x)\big]\kappa_{xy}, \qquad \varphi \in \mathcal{B}(V).$$

Its law $\rho(t)$ evolves by the *forward Kolmogorov equation*

$$\dot\rho(t) = Q^*\rho(t), \qquad (Q^*\rho)_x = \sum\nolimits_y \big[\rho_y \kappa_{yx} - \rho_x \kappa_{xy}\big], \qquad \rho(0) = \bar\rho \in \mathcal{P}(V). \tag{1.3}$$

eq:fke-discrete

def:db-discrete

Definition 1.4 (Detailed balance) κ satisfies *detailed balance* with respect to $\pi \in \mathcal{P}(V)$ if the *edge measure*

$$\vartheta_{xy} := \pi_x \kappa_{xy} = \pi_y \kappa_{yx} = \vartheta_{yx} \qquad \text{is symmetric.} \tag{DBC}$$

eq:db-discrete

Then π is stationary, $Q^*\pi = 0$.

Detailed balance is equivalent to time-reversibility of the chain started from π, and to self-adjointness of the semigroup $T_t = e^{tQ}$ in $L^2(\pi)$, $\langle f, g\rangle_\pi := \sum_x f_x g_x \pi_x$. Indeed,

$$\langle Qf, g\rangle_\pi = \sum_{x,y} \vartheta_{xy} f_y g_x = \sum_{x,y} \vartheta_{yx} f_x g_y = \langle f, Qg\rangle_\pi,$$

using symmetry of ϑ; reversibility follows by iterating.

1.3.1 The discrete continuity equation

Introduce the *graph gradient* and *graph divergence*

$$\begin{aligned} \nabla &\colon \mathcal{B}(V) \to \mathcal{B}(E), & (\nabla\varphi)(x,y) &= \varphi(y) - \varphi(x), \\ \operatorname{div} &\colon \mathcal{M}(E) \to \mathcal{M}(V), & (\operatorname{div} j)_x &= \sum\nolimits_y \big(j_{xy} - j_{yx}\big), \end{aligned} \tag{1.4}$$

eq:graph-grad

which are negative adjoints: $\sum_{x,y} \nabla\varphi(x,y)\, j_{xy} = -\sum_x \varphi(x)(\operatorname{div} j)_x$. Defining the *net flux* $j_{xy} := \frac{1}{2}\big(\rho_x\kappa_{xy} - \rho_y\kappa_{yx}\big)$—the half because each edge is counted in both orientations—a direct computation shows, for any test function φ,

$$\sum_x \varphi(x)\,\dot\rho_x = \sum_{x,y} \rho_x \kappa_{xy}\, \nabla\varphi(x,y) = \sum_{x,y} \nabla\varphi(x,y)\, j_{xy} = -\sum_x \varphi(x)(\operatorname{div} j)_x,$$

the middle equality because $\nabla\varphi$ is antisymmetric. Hence (1.3) is equivalent to the *continuity equation*

$$\dot\rho + \operatorname{div} j = 0, \qquad j_{xy} = \tfrac{1}{2}\big(\rho_x \kappa_{xy} - \rho_y \kappa_{yx}\big). \tag{CE}$$

Writing $\rho = u\pi$ (i.e. $\rho_x = u(x)\pi_x$) and using detailed balance,

$$j_{xy} = \frac{1}{2}\big(u(x)\pi_x\kappa_{xy} - u(y)\pi_y\kappa_{yx}\big) = \frac{1}{2}(u(x)-u(y))\,\vartheta_{xy} = -\frac{1}{2}\,\nabla u(x,y)\,\vartheta_{xy}, \tag{1.5}$$

eq:kr-discrete

i.e., the flux is proportional to the gradient of the density along edges and is the precursor to the *kinetic relation*. Next, we introduce the energy and dissipation potentials.

1.3.2 The driving energy and the dissipation potential

By Sanov's theorem, the natural driving energy for the empirical measure of N independent copies of the chain is the *relative entropy*

$$\mathcal{F}(\rho) = \mathsf{Ent}_\phi(\rho|\pi) := \begin{cases} \sum_x \phi\Big(\dfrac{\rho_x}{\pi_x}\Big)\pi_x = \sum_x \rho_x \log\dfrac{\rho_x}{\pi_x}, & \rho \ll \pi, \\ +\infty, & \text{otherwise,} \end{cases}$$

with $\phi(s) = s\log s - s + 1$, whose Fréchet derivative is $D\mathcal{F}(\rho) = \log u$, $u = \mathrm{d}\rho/\mathrm{d}\pi$. More importantly, in the discrete setting there is *no underlying spatial gradient*: we must use the graph gradient ∇ in (1.4). This leaves a lot of freedom in the choice of dissipation potential. Writing $\rho = u\pi$, we consider only dual dissipation potentials of the form

$$\mathcal{R}^*(\rho,\xi) = \frac{1}{2}\sum_{x,y} \psi^*\big(\xi(x,y)\big)\,\mathsf{m}\big(u(x),u(y)\big)\,\vartheta_{xy},$$

where $\psi^*\colon \mathbb{R} \to [0,\infty)$ is even, convex, and superlinear with $\psi^*(0) = 0$, and the *flux multiplier* m is a symmetric, concave, 1-homogeneous *mean*, i.e. $\mathsf{m}(u,u) = u$.

Exercise 1.1 Show that the corresponding primal potential $\mathcal{R}$, with (ψ,ψ^*) a conjugate pair and $2j_{xy} = w(x,y)\vartheta_{xy}$, is given by

$$\mathcal{R}(\rho,j) = \frac{1}{2}\sum_{x,y} \Upsilon\big(u(x),u(y),w(x,y)\big)\,\vartheta_{xy},$$

where

$$\Upsilon(u,v,w) := \begin{cases} \mathsf{m}(u,v)\,\psi\Big(\dfrac{w}{\mathsf{m}(u,v)}\Big) & \text{if } \mathsf{m}(u,v) > 0, \\ 0 & \text{if } w = 0, \\ +\infty & \text{if } w \neq 0 \text{ and } \mathsf{m}(u,v) = 0, \end{cases} \tag{1.6}$$

eq:upsilon

is 1-homogeneous, jointly convex and lower semicontinuous.

The $(\mathcal{F}, \mathcal{R}, \mathcal{R}^*)$ triple for a Markov chain

The chain (1.3) is the generalized gradient flow of

$$\dot\rho + \operatorname{div} j = 0, \qquad j = D_\xi \mathcal{R}^*\big(\rho, -\nabla D\mathcal{F}(\rho)\big),$$

that is, the flux is produced by the dual dissipation potential acting on the thermodynamic force $-\nabla D\mathcal{F}(\rho) = -\nabla \log u$. Solutions are exactly the null-minimizers of the EDB functional

$$\mathscr{L}(\rho, j; [0,T]) = \int_0^T \Big[\mathcal{R}(\rho_t, j_t) + \mathcal{R}^*\big(\rho_t, -\nabla \log u_t\big)\Big]\,\mathrm{d}t + \mathcal{F}(\rho_T) - \mathcal{F}(\rho_0) \geq 0.$$

1.3.3 Two canonical choices

ex:quadratic

Example 1.5 (Quadratic (Maas–Mielke–Chow–Huang–Zhou)) Take $\psi^*(\xi) = \frac{1}{2}\xi^2$ and $\mathsf{m} = \mathsf{m}_{\log}$ the *logarithmic mean* $\mathsf{m}_{\log}(u,v) = \frac{u-v}{\log u - \log v}$. Then

$$j_{xy} = D_\xi \mathcal{R}^*\big(\rho, -\nabla \log u\big)_{xy} = \tfrac{1}{2}\big(-\nabla \log u(x,y)\big)\,\mathsf{m}_{\log}(u(x),u(y))\,\vartheta_{xy} = \tfrac{1}{2}\big(u(x)-u(y)\big)\,\vartheta_{xy},$$

the factor $\frac{1}{2}$ coming from the one in $\mathcal{R}^*$, and $\big(-\nabla \log u(x,y)\big)\mathsf{m}_{\log}(u(x),u(y)) = u(x)-u(y)$ cancelling the logarithmic mean against its own denominator. This is exactly (1.5). This is the structure of Maas [8], Mielke [9], and Chow–Huang–Li–Zhou [2]; it makes $\mathcal{P}(V)$ a Riemannian-like metric space (a discrete Wasserstein space) and underlies the Erbar–Maas notion of discrete Ricci curvature [4].

ex:cosh

Example 1.6 (Cosh/large-deviation structure) Take

$$\psi^*(\xi) = 4\big(\cosh(\xi/2) - 1\big), \qquad \mathsf{m}(a,b) = \sqrt{ab},$$

so that, with $\rho = u\pi$, the edge weight is the geometric mean $\sqrt{u(x)u(y)}\,\vartheta_{xy}$. Since $(\psi^*)'(\xi) = 2\sinh(\xi/2)$, the flux relation reads

$$j_{xy} = \tfrac{1}{2}(\psi^*)'\big(-\nabla \log u(x,y)\big)\sqrt{u(x)u(y)}\,\vartheta_{xy} = \tfrac{1}{2}\Big[2\sinh\Big(\tfrac{1}{2}\log\tfrac{u(x)}{u(y)}\Big)\Big]\sqrt{u(x)u(y)}\,\vartheta_{xy}.$$

Using $2\sinh(\frac{1}{2}\log\frac{u(x)}{u(y)}) = \sqrt{u(x)/u(y)} - \sqrt{u(y)/u(x)}$, the bracket against $\sqrt{u(x)u(y)}$ is $u(x)-u(y)$, so $j_{xy} = \frac{1}{2}(u(x)-u(y))\,\vartheta_{xy}$, again recovering (1.5). The primal potential is

$$\psi(s) = 2s\log\left(\frac{s+\sqrt{s^2+4}}{2}\right) - 2\sqrt{s^2+4} + 4.$$

This structure is *not* p-homogeneous, hence *not* metric, but it is distinguished because it arises as the large-deviation rate functional of the empirical fluxes of the underlying chain [12]. It is this non-metric, large-deviation–induced structure that the rigorous theory of Lecture 2 is built to handle.

1.4 The Fisher information and dissipation along solutions

Setting $s = 0$ and differentiating the EDB, a solution dissipates energy at rate

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\rho_t) = -\mathcal{R}(\rho_t, j_t) - \mathcal{D}(\rho_t) \le 0, \qquad \mathcal{D}(\rho) := \mathcal{R}^*\big(\rho, -\nabla \log u\big),$$

where $\mathcal{D}$ is the (generalized) *Fisher information*. For the two canonical choices,

$$\mathcal{D}_{\mathrm{quad}}(\rho) = \frac{1}{4}\sum_{x,y} \big(\log u(x) - \log u(y)\big)\big(u(x) - u(y)\big)\,\vartheta_{xy},$$
$$\mathcal{D}_{\cosh}(\rho) = \sum_{x,y} \big(\sqrt{u(x)} - \sqrt{u(y)}\big)^2\,\vartheta_{xy}.$$

Both are nonnegative, vanish only at $u \equiv \text{const}$ (i.e., $\rho = \pi$), and serve as Lyapunov dissipation rates driving convergence to equilibrium.

2 Generalized gradient structures for jump processes

ch:jump

Goal of Lecture 2

We lift the discrete picture of Lecture 1 to jump processes on a general *Polish* state space V, where the set of vertices may be finite, countable, or uncountable. We make the cosh-type structure rigorous: the continuity equation on $\mathcal{M}^+(V)$, the dissipation potential $\mathcal{R}$, the chain rule, the Fisher information $\mathcal{D}$, and the $(\mathcal{F}, \mathcal{R}, \mathcal{R}^*)$ energy–dissipation balance that *defines* solutions. We sketch existence via minimizing movements. This is the framework of Peletier–Rossi–Savaré–Tse [14].

Think of a jump process as moving from a *vertex* x to a *vertex* y along an *edge* (x, y) of a *graph*. Now V is a *Polish space*—separable and completely metrizable—carrying its Borel σ-algebra $\mathcal{B}$, and $E := V \times V$. The law of the process is a time-dependent measure $t \mapsto \rho_t \in \mathcal{M}^+(V)$ solving the *forward Kolmogorov equation*

$$\partial_t \rho_t = Q^* \rho_t, \qquad (Q^*\rho)(\mathrm{d}x) = \int_V \rho(\mathrm{d}y)\, \kappa(y, \mathrm{d}x) - \rho(\mathrm{d}x) \int_V \kappa(x, \mathrm{d}y), \tag{2.1}$$

eq:fke-jump

where Q^* is the dual of the generator

$$(Q\varphi)(x) = \int_V [\varphi(y) - \varphi(x)]\, \kappa(x, \mathrm{d}y), \qquad \varphi \in \mathcal{B}_b(V).$$

The *jump kernel* $\kappa(x, \cdot) \in \mathcal{M}^+(V)$ is the infinitesimal rate of jumps from $x \in V$.

ass:Vpik

Assumption 2.1 V is Polish with Borel σ-algebra $\mathcal{B}$, the reference measure is $\pi \in \mathcal{M}^+(V)$, and $(\kappa(x, \cdot))_{x \in V}$ is a kernel satisfying $\sup_x \kappa(x, V) < +\infty$ and the *detailed-balance condition* w.r.t. π, i.e.,

$$\int_A \kappa(x, B)\, \pi(\mathrm{d}x) = \int_B \kappa(y, A)\, \pi(\mathrm{d}y) \qquad \text{for all } A, B \in \mathcal{B}. \tag{DBC}$$

eq:db-jump

Detailed balance (DBC) says exactly that the *edge measure*

$$\vartheta(\mathrm{d}x\, \mathrm{d}y) := \pi(\mathrm{d}x)\, \kappa(x, \mathrm{d}y) \in \mathcal{M}^+(E) \qquad \text{is symmetric.}$$

This is the continuum analog of the discrete condition in Lecture 1.

Graph gradient and divergence. These carry over directly from §1.3:

$$(\nabla\varphi)(x, y) := \varphi(y) - \varphi(x), \qquad \varphi \in \mathcal{B}_b(V)$$
$$(\operatorname{div} j)(\mathrm{d}x) := \int_{y \in V} \big[j(\mathrm{d}x, \mathrm{d}y) - j(\mathrm{d}y, \mathrm{d}x) \big], \qquad j \in \mathcal{M}(E).$$

As before, they satisfy the integration-by-parts formula $\langle \nabla\varphi, j \rangle = -\langle \varphi, \operatorname{div} j \rangle$.

Setwise convergence. The 'correct' notion of convergence for sequences of measures in this setting is that of *setwise convergence*. It sits strictly between convergence in total variation and narrow (or weak) convergence. Every limit below is taken in this sense.

def:setwise

Definition 2.2 (Setwise convergence) A sequence $(\mu_n)_n \subset \mathcal{M}(V)$ converges *setwise* to $\mu \in \mathcal{M}(V)$ if

$$\mu_n(B) \longrightarrow \mu(B) \qquad \text{for every } B \in \mathcal{B}.$$

The same definition on $\mathcal{M}(E)$ is used for fluxes.

Proposition 2.3. *For sequences $(\mu_n)_n \subset \mathcal{M}(V)$ and $\mu \in \mathcal{M}(V)$, the following characterizations of setwise convergence are equivalent:*

(i) $\mu_n \to \mu$ *setwise in* $\mathcal{M}(V)$.

(ii) (Convergence in duality) $\langle \varphi, \mu_n \rangle \to \langle \varphi, \mu \rangle$ *for every* $\varphi \in \mathcal{B}_b(V)$.

(iii) (Weak L^1-convergence of densities) There exists a common dominating measure $\gamma \in \mathcal{M}(V)$ for $(\mu_n)_n$ and μ, i.e., $\mu_n \ll \gamma$, $\mu \ll \gamma$, such that

$$\frac{\mathrm{d}\mu_n}{\mathrm{d}\gamma} \rightharpoonup \frac{\mathrm{d}\mu}{\mathrm{d}\gamma} \quad \text{weakly in } L^1(V, \gamma).$$

2.1 The three ingredients of the gradient structure

2.1.1 Driving energy

ass:Ephi

Assumption 2.4 ($\mathcal{F}\phi$) The driving functional is the ϕ-relative entropy

$$\mathcal{F}(\rho) = \mathsf{Ent}_\phi(\rho|\pi) := \begin{cases} \displaystyle\int_V \phi\Big(\frac{\mathrm{d}\rho}{\mathrm{d}\pi}\Big)\,\mathrm{d}\pi, & \rho \ll \pi, \\ +\infty, & \text{otherwise,} \end{cases}$$

where $\phi \in C([0,\infty)) \cap C^1((0,\infty))$ is convex with superlinear growth and $\min \phi = 0$. The central example is the Boltzmann–Shannon density $\phi_{\mathsf{B}}(s) = s \log s - s + 1$.

$\mathcal{F}$ is lower semicontinuous for setwise convergence, and its sublevels are setwise relatively compact—the compactness backbone of the existence theory.

2.1.2 The continuity equation on $\mathcal{M}^+(V)$

def:ce-jump

Definition 2.5 (Continuity equation) Let $I = [a, b]$. A pair (ρ, j) belongs to $\mathsf{CE}(I)$ if

(i) $(\rho_t)_{t\in I} \in \mathcal{M}^+(V)$ is a family of time-dependent measures,

(ii) $(j_t)_{t\in I} \subset \mathcal{M}(E)$ is a measurable family with $\int_I |j_t|(E)\,\mathrm{d}t < +\infty$, and

(iii) the continuity equation $\partial_t \rho_t + \operatorname{div} j_t = 0$ is satisfied in the following sense:

$$\int_V \varphi\,\mathrm{d}\rho_{t_2} - \int_V \varphi\,\mathrm{d}\rho_{t_1} = \int_{t_1}^{t_2}\!\!\int_E \nabla\varphi\,\mathrm{d}j_t\,\mathrm{d}t \qquad \forall\,\varphi \in \mathcal{B}_b(V),\ [t_1,t_2] \subseteq I, \qquad (\mathsf{CE})$$

We write $\mathsf{CE}(I;\rho_a)/\mathsf{CE}(I;\rho_a,\rho_b)$ when the endpoints are prescribed.

lem:CE-implications

Lemma 2.6. *If $(\rho, j) \in \mathsf{CE}(I)$, then there is a common dominating measure $\gamma \in \mathcal{M}^+(V)$ for $(\rho_t)_{t\in I}$ (i.e., $\rho_t \ll \gamma$ for all $t \in I$), and an absolutely continuous map $u : I \to L^1(V,\gamma)$ such that $\rho_t = u_t\gamma \ll \gamma$ for every $t \in I$.*

In particular, $t \mapsto \rho_t$ is continuous w.r.t. the total variation distance, and taking $\varphi \equiv 1$ shows that the total mass $\rho_t(V)$ is conserved in time.

2.1.3 The dissipation potential

ass:Rstaralpha

Assumption 2.7 The dual dissipation density $\psi^*\colon \mathbb{R} \to [0,\infty)$ is even, convex, differentiable, with $\psi^*(0) = 0$ and superlinear growth. The *flux density* $\mathsf{m}\colon [0,+\infty)^2 \to [0,+\infty)$ is continuous, symmetric, concave and 1-homogeneous, and is a *mean*, $\mathsf{m}(u,u) = u$.

Concavity and 1-homogeneity make m superadditive, so for $u \le v$ we get $\mathsf{m}(u,v) \ge \mathsf{m}(u,u) + \mathsf{m}(0,v-u) \ge u$: the mean dominates $\min(u,v)$, and in particular $\mathsf{m}(u,v) > 0$ whenever $uv > 0$. Along a curve $\rho = u\pi$, the *dual dissipation potential* is

$$\mathcal{R}^*(\rho,\xi) = \frac{1}{2}\int_E \psi^*\big(\xi(x,y)\big)\,\mathsf{m}\big(u(x),u(y)\big)\,\vartheta(\mathrm{d}x\,\mathrm{d}y), \qquad \xi \in \mathcal{B}_b(E).$$

Its Legendre dual, the *primal* potential, has the perspective-function form: writing $2j = w\,\vartheta$ (so w is a density of the flux against the symmetric edge measure),

$$\mathcal{R}(\rho,j) = \frac{1}{2}\int_E \Upsilon\big(u(x),u(y),w(x,y)\big)\,\vartheta(\mathrm{d}x\,\mathrm{d}y),$$

where Υ, defined as in (1.6), is jointly convex and lower semicontinuous.

The cosh structure on $\mathcal{M}^+(V)$

The distinguished choice, induced by the large deviations of the underlying particle system (Lecture 3), is

$$\psi^*(\xi) = 4\big(\cosh(\xi/2) - 1\big), \qquad \mathsf{m}(u,v) = \sqrt{uv},$$

for which the formal gradient-flow equation $\partial_t\rho = -\operatorname{div} D_\xi\mathcal{R}^*\big(\rho, -\nabla\phi'(u)\big)$ reduces, with $\phi'(u) = \log u$, to the linear forward Kolmogorov equation (2.1)

$$\partial_t u_t(x) = \int_V \big(u_t(y) - u_t(x)\big)\,\kappa(x,\mathrm{d}y).$$

Remark 2.8 (Linear compatibility and the general equation) For a triple (ψ^*,ϕ,m), equation $\partial_t\rho = -\operatorname{div} D_\xi\mathcal{R}^*(\rho,-\nabla\phi'(u))$ becomes the integro-differential equation

$$\partial_t u_t(x) = \int_V F\big(u_t(x),u_t(y)\big)\kappa(x,\mathrm{d}y)$$

with $F(u,v) = (\psi^*)'\big(\phi'(v) - \phi'(u)\big)\mathsf{m}(u,v)$. It coincides with the *linear* equation (2.1) precisely when the *linear compatibility condition* $F(u,v) = v - u$ holds for all $u,v > 0$. Both the *cosh* choice and the quadratic choice $\psi^*(\xi) = \frac{1}{2}\xi^2$, $\mathsf{m}(u,v) = \frac{u-v}{\log u - \log v}$ satisfy it.

2.2 The Fisher information

We now define the *Fisher information*. For this, we will need the following objects.

def:fisher-integrands

Definition 2.9 (Force and Fisher integrands) Following [14], set $\phi'(0) := \lim_{r\downarrow 0}\phi'(r) \in [-\infty, +\infty)$ and define the *discrete force*

$$A_\phi(u,v) := \begin{cases} \phi'(v) - \phi'(u), & (u,v) \in \mathbb{R}^2_+ \setminus \{(0,0)\}, \\ 0, & u = v = 0. \end{cases}$$

For the Boltzmann density, this is $A_\phi(u,v) = \log(v/u)$, with values $\pm\infty$ on the axes. Set

$$D^-_\phi(u,v) := \begin{cases} \psi^*\big(A_\phi(u,v)\big)\mathsf{m}(u,v), & \mathsf{m}(u,v) > 0, \\ 0, & \text{otherwise}, \end{cases}$$

$$D^+_\phi(u,v) := \begin{cases} \psi^*\big(A_\phi(u,v)\big)\mathsf{m}(u,v), & \mathsf{m}(u,v) > 0, \\ 0, & u = v = 0, \\ +\infty, & \mathsf{m}(u,v) = 0,\ u \neq v. \end{cases}$$

The *Fisher information* is then defined as

$$\mathrm{dom}(\mathcal{F}) \ni \rho \mapsto \mathcal{D}(\rho) = \frac{1}{2}\int_E D_\phi\big(u(x), u(y)\big)\,\vartheta(\mathrm{d}x\,\mathrm{d}y), \qquad \rho = u\pi,$$

where D_ϕ is the lower-semicontinuous envelope of D^+_ϕ on $\mathbb{R}^2_+$, i.e.,

$$\begin{aligned} D_\phi(u,v) &:= \sup\Big\{G(u,v) \,:\, G : \mathbb{R}^2_+ \to [0,+\infty] \ \text{ lower semicontinuous},\ G \le D^+_\phi\Big\} \\ &= \inf\Big\{\liminf_{n\to\infty} D^+_\phi(u_n, v_n) : u_n \to u,\ v_n \to v\Big\}. \end{aligned}$$

The integrands $D^\pm_\phi$ and D_ϕ differ only on the degenerate set $\{\mathsf{m} = 0,\ u \neq v\}$ and always satisfy $D^-_\phi \le D_\phi \le D^+_\phi$.

prop:fisher-lsc

Proposition 2.10 (Lower semicontinuity of $\mathcal{D}$)**.** *Assume either π is purely atomic or that D_ϕ is convex on $\mathbb{R}^2_+$. Then $\mathcal{D}$ is lower semicontinuous w.r.t. setwise convergence (Definition 2.2), i.e., for all $(\mu_n)_n \subset \mathrm{dom}(\mathcal{F})$, $\mu \in \mathrm{dom}(\mathcal{F})$, one has that*

$$\mu_n \to \mu \ \text{ setwise in } \mathcal{M}^+(V) \quad\Longrightarrow\quad \mathcal{D}(\mu) \le \liminf_{n\to\infty} \mathcal{D}(\mu_n).$$

ex:fisher

Example 2.11 For the linear equation with $\phi = \phi_{\mathsf{B}}$, so $A_\phi(u,v) = \log(v/u)$, we find

(1) if $\psi^*(s) = \frac{1}{2}s^2$ and $\mathsf{m}(u,v) = \frac{u-v}{\log u - \log v}$, then

$$D_\phi(u,v) = \begin{cases} \frac{1}{2}(\log u - \log v)(u-v) & \text{for } uv > 0, \\ 0 & \text{for } u = v = 0, \\ +\infty & \text{for } uv = 0,\ u \neq v; \end{cases}$$

(2) if $\psi^*(s) = 4\big(\cosh\frac{s}{2} - 1\big)$ and $\mathsf{m}(u,v) = \sqrt{uv}$, then

$$D_\phi(u,v) = 2\big(\sqrt{u} - \sqrt{v}\big)^2 \qquad \forall\,(u,v) \in [0,+\infty)\times[0,+\infty).$$

Both examples give convex D_ϕ on $\mathbb{R}^2_+$, so Proposition 2.10 applies and $\mathcal{D}$ is setwise lower semicontinuous in either case. The difference between them occurs on $\{uv = 0,\ u \neq v\}$. We note that finiteness of the Fisher information $\mathcal{D}$ is trivial for the *cosh* example, but not for the quadratic case. On the other hand, if we know that the $\mathcal{D}$ is finite for the quadratic case, then we know that $u(x)u(y) > 0$ for ϑ-a.e. $(x,y) \in E$, and hence, $u(x) > 0$ for every π-a.e. $x \in V$, while no such property can be claimed for (2).

2.3 Admissible curves and the chain rule

The heart of the rigorous theory is a *chain rule*, which we show to hold for the following class of curves.

def:admissible

Definition 2.12 (Admissible curves) A pair $(\rho, j) \in \mathsf{CE}(I)$ is *admissible* if

$$\sup_{t\in I} \mathcal{F}(\rho_t) < +\infty, \qquad \int_I \big[\mathcal{R}(\rho_t, j_t) + \mathcal{D}(\rho_t)\big]\,\mathrm{d}t < +\infty.$$

In this case, we write $(\rho, j) \in \mathscr{A}(I)$.

The three terms in the definition are exactly the terms of the balance (EDB), so $\mathscr{A}(I)$ is simply the class on which (EDB) is a statement about finite quantities.

A few remarks are in order:

Remark 2.13 (1) The finiteness of $\mathcal{F}(\rho_t)$ gives $\rho_t = u_t\pi$ with $u_t \in L^1(V,\pi)$. Hence, $(\rho, j) \in \mathscr{A}(I) \subset \mathsf{CE}(I)$ implies $\rho_t = u_t\pi$ with $u_t \in L^1(V,\pi)$ for all $t \in I$.

(2) The finiteness of $\mathcal{R}(\rho_t, j_t)$ forces j_t to vanish on $\{\mathsf{m} = 0\}$ for almost every $t \in I$, since $\Upsilon(u,v,w) = +\infty$ for $w \neq 0$ there. On $\{\mathsf{m} > 0\}$, we have that $D^\pm_\phi = D_\phi$.

Under finite action, we obtain

Proposition 2.14. *Let $(\rho, j) \in \mathsf{CE}(I)$ with finite action $\int_I \mathcal{R}(\rho_t, j_t)\,\mathrm{d}t < +\infty$. If $\rho_t \ll \pi$ for some $t \in I$, then there exists*

(1) an absolutely continuous and a.e. differentiable map $u : I \to L^1(V,\pi)$ with $\rho_t = u_t\pi$,

(2) a map $w \in L^1(I\times E, \lambda\otimes\vartheta)$ such that $2j_t = w_t\vartheta$ for a.e. $t \in I$, and

(3) the equation

$$\partial_t u_t = -\frac{1}{2}\int_V [w_t(x,y) - w_t(y,x)]\kappa(x,\mathrm{d}y)$$

is satisfied for almost every $t \in I$.

We conclude this subsection with the chain rule for admissible pairs $(\rho, j) \in \mathscr{A}(I)$.

thm:chainrule

Theorem 2.15 (Chain rule [14]). *Let $(\rho, j) \in \mathscr{A}(I)$ and write $\rho_t = u_t\pi$. Then $t \mapsto \mathcal{F}(\rho_t)$ is absolutely continuous on I and, for all $s \le t$ in I,*

$$\mathcal{F}(\rho_t) - \mathcal{F}(\rho_s) = \int_s^t\int_E \nabla\phi'(u_r)\,\mathrm{d}j_r\,\mathrm{d}r \geq -\int_s^t \mathcal{R}(\rho_r, j_r) + \mathcal{D}(\rho_r)\,\mathrm{d}r. \qquad \text{(CR)}$$

eq:chainrule

2.4 Definition of solutions: the $(\mathcal{F},\mathcal{R},\mathcal{R}^*)$ balance

def:edb-jump

Definition 2.16 ($(\mathcal{F},\mathcal{R},\mathcal{R}^*)$ energy–dissipation balance) A pair $(\rho,j)\in\mathsf{CE}(0,T)$ is a *solution* of the $(\mathcal{F},\mathcal{R},\mathcal{R}^*)$ system if for every $0\le s\le t\le T$,

$$\mathscr{L}(\rho,j;[s,t]) := \int_s^t \Big[\mathcal{R}(\rho_r,j_r)+\mathcal{D}(\rho_r)\Big]\,\mathrm{d}r + \mathcal{F}(\rho_t)-\mathcal{F}(\rho_s) = 0. \tag{EDB}$$

eq:edb-jump

EDB = EDI under the chain rule

The chain rule (CR) and Young's inequality $\mathcal{R}(\rho,j)+\mathcal{R}^*(\rho,-\nabla\phi'(u))+\langle\nabla\phi'(u),j\rangle\ge 0$ yields, for every $(\rho,j)\in\mathsf{CE}(0,T)$, the *chain-rule (lower) bound*

$$\mathscr{L}(\rho,j;[s,t])\ge 0.$$

Hence, the balance (EDB) is equivalent to the energy–dissipation *inequality* (EDI)

$$\mathscr{L}(\rho,j;[s,t])\le 0. \tag{EDI}$$

eq:edi

thm:eqform

Theorem 2.17 (Equation form of solutions [14]). *A curve (ρ_t) with $\mathcal{F}(\rho_0)<+\infty$ solves the $(\mathcal{F},\mathcal{R},\mathcal{R}^*)$ system iff*

(i) $\rho_t = u_t\pi$ *with* $t\mapsto u_t$ *absolutely continuous in* $L^1(V,\pi)$,

(ii) the equality case $D_\phi = D_\phi^-$ *holds* $\lambda\otimes\vartheta$*-a.e., and*

(iii) the flux takes the form $2j_t(\mathrm{d}x\,\mathrm{d}y) = -F_0(u_t(x),u_t(y))\,\vartheta(\mathrm{d}x\,\mathrm{d}y)$ *such that*

$$\partial_t u_t(x) = \int_V F_0\big(u_t(x),u_t(y)\big)\,\kappa(x,\mathrm{d}y)\quad(\lambda\otimes\pi)\text{-a.e.},$$

where $F_0(u,v) = (\psi^*)'\big(A_\phi(u,v)\big)\mathsf{m}(u,v)$ *is the skew-symmetric flux map.*

thm:unique

Theorem 2.18 (Uniqueness [14]). *If the Fisher information $\mathcal{D}$ is convex and ϕ is strictly convex, then solutions of (EDB) are uniquely determined by $\rho_0\in\mathrm{dom}(\mathcal{F})$.*

2.5 Existence via minimizing movements

sec:mm

Solutions in the sense of Definition 2.16 are produced by the *Minimizing-Movement* (or JKO) scheme. Fix a time step $\tau>0$ and let $\rho_\tau^0=\rho_0\in\mathrm{dom}(\mathcal{F})$. We then recursively set

$$\rho_\tau^n \in \operatorname*{arg\,min}_{\rho\in\mathcal{M}^+(V)}\Big\{\mathcal{W}_\tau\big(\rho_\tau^{n-1},\rho\big)+\mathcal{F}(\rho)\Big\}, \tag{2.2}$$

eq:jko

where the *Dynamical-Variational Transport* cost $\mathcal{W}_\tau\colon\mathcal{M}^+(V)^2\to[0,+\infty)$ is

$$\mathcal{W}_\tau(\mu,\nu) := \inf\Big\{\int_0^\tau \mathcal{R}(\rho_r,j_r)\,\mathrm{d}r : (\rho,j)\in\mathsf{CE}(0,\tau;\mu,\nu)\Big\}.$$

thm:mm

Theorem 2.19 (Convergence of the scheme [14]). *Let $\mathcal{F}(\rho_0)<+\infty$. Under Assumptions 2.1, 2.4, 2.7, the piecewise-constant interpolants of (2.2) admit, as $\tau\to 0$, a (sub)sequential limit $(\rho,j)\in\mathsf{CE}(0,T)$ satisfying the energy–dissipation inequality (EDI). In particular $(\rho,j)\in\mathscr{A}(0,T)$, which with the chain rule (CR) makes it a solution of the $(\mathcal{F},\mathcal{R},\mathcal{R}^*)$ system in the sense of Definition 2.16.*

3 Where gradient structures come from

ch:origins

Goal of Lecture 3

Lecture 1 exhibited *a* gradient structure for a reversible Markov chain and remarked, without justification, that the cosh structure is "singled out by large deviations". This lecture makes good on that promise. Its message is the *selection principle*: for a stochastic particle system the large-deviation rate functional is canonical; three classical theorems—Sanov, Girsanov, and disintegration—turn it into an explicit action on curves of measures and fluxes; and splitting that action into its symmetric and antisymmetric parts under *time reversal* produces a gradient structure, with the symmetric part the dissipation and the antisymmetric part the change in free energy. Detailed balance is exactly the condition making this work. Diffusions come out quadratic, jump processes come out cosh.

3.1 The selection principle: large deviations

sec:selection

Lecture 1 left a question open. The reversible Markov chain admits at least two gradient structures—the quadratic one of Example 1.5 and the cosh one of Example 1.6—and both generate the same forward Kolmogorov equation. Which should one choose? Answering "whichever is convenient" is unsatisfying, because different structures predict different behavior away from equilibrium and behave differently under limits. We want a principled answer, and probability gives us one.

The key point is that the macroscopic equation isn't fundamental. We can derive it from the law of large numbers for a stochastic particle system. This system captures not only the average behavior but also the *probability of deviations* from that average. This information is canonical—free of modeling choices—and also reveals a gradient structure.

The derivation has three steps, each of which is a classical theorem. We state it for $N \in \mathbb{N}$ independent copies of a Markov jump process with jump kernel κ on a Polish space V, writing $E := V \times V$ as in Lecture 1, started from measure $\bar{\rho}$. Let $\mathsf{P}_{\bar{\rho}}$ denote the law of a single 'particle', which is a probability measure on the space of *cádlág* paths $D([0,T];V)$, i.e., paths $\omega : [0,T] \to V$ that are right-continuous with left limits, equipped with the Skorohod J_1 topology. Since V is Polish, so is $D([0,T];V)$, w.r.t. the J_1 topology.

Step 1: Sanov, on path space. Consider the *empirical process* on path space

$$\mathsf{P}^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{X^i_\bullet} \;\in\; \mathcal{P}\big(D([0,T];V)\big).$$

This is a random measure not on states but on whole trajectories. Since the trajectories are independent, Sanov's theorem applies, and we obtain a large-deviation principle with speed N whose rate function is a relative entropy on path space given by

$$\mathrm{Prob}\big(\mathsf{P}^N \approx \mathsf{Q}\big) \asymp e^{-N\,\mathcal{I}_{LD}(\mathsf{Q})}, \qquad \mathcal{I}_{LD}(\mathsf{Q}) = \mathsf{Ent}(\mathsf{Q}|\mathsf{P}_{\bar{\rho}}). \tag{3.1}$$

eq:sanov-path

This is the full probabilistic input, and it offers no modeling freedom. Note that it vanishes exactly at $\mathsf{Q} = \mathsf{P}_{\bar{\rho}}$, the law of the typical trajectory.

Step 2: Girsanov for jump processes. Formula (3.1) is not useful in its current form because it represents an entropy between measures on an infinite-dimensional path space. The relevant tool, analogous to Girsanov's theorem for diffusions but for jump processes, precisely determines which path measures Q have finite entropy relative to $\mathsf{P}_{\bar\rho}$, and calculates their density.

Consider the law $\mathsf{Q}^{\hat\kappa}$ of the process which starts from ρ_0 and jumps with a *tilted*, possibly time-dependent kernel

$$\hat\kappa_t(x,\mathrm{d}y) = r_t(x,y)\,\kappa(x,\mathrm{d}y), \tag{3.2}$$

eq:tilted-kernel

so that the density r_t records, at each state and each target, the factor by which *observed* rates depart from the *expected* ones. Then $\mathsf{Q}^{\hat\kappa} \ll \mathsf{P}_{\bar\rho}$, with density given by the stochastic (Doléans-Dade) exponential: for a càdlàg path $\omega = (\omega_t)_{t\in[0,T]} \in D([0,T];V)$,

$$\frac{\mathrm{d}\mathsf{Q}^{\hat\kappa}}{\mathrm{d}\mathsf{P}_{\bar\rho}}(\omega) = \frac{\mathrm{d}\rho_0}{\mathrm{d}\bar\rho}(\omega_0)\,\exp\!\Big(\int_0^T\!\!\int_V \big(1-r_t(\omega_{t^-},y)\big)\kappa(\omega_{t^-},\mathrm{d}y)\,\mathrm{d}t\Big)\prod_{\substack{t\le T:\\ \omega_{t^-}\neq\omega_t}} r_t(\omega_{t^-},\omega_t)\,, \tag{3.3}$$

eq:girsanov-dd

one factor r_t per *observed* jump, times an exponential correction accounting for the jumps that were *not* observed. For each càdlàg path ω we denote by

$$\mathsf{N}_\omega := \sum_{t\le T:\ \omega_{t^-}\neq\omega_t} \delta_{(t,\,\omega_t)} \in \mathcal{M}\big((0,T]\times V\big),$$

its associated *jump measure*, which records when the path jumps and where it lands. Taking the logarithm of (3.3) turns the product into an integral against N_ω, and the resulting identity splits into an initial term and a dynamic one:

$$\log\frac{\mathrm{d}\mathsf{Q}^{\hat\kappa}}{\mathrm{d}\mathsf{P}_{\bar\rho}}(\omega) = \underbrace{\log\frac{\mathrm{d}\rho_0}{\mathrm{d}\bar\rho}(\omega_0)}_{\text{initial data}} + \underbrace{\int_0^T\!\!\int_V \log r_t(\omega_{t^-},y)\,\mathsf{N}_\omega(\mathrm{d}t\,\mathrm{d}y) - \int_0^T\!\!\int_V \big(r_t(\omega_{t^-},y)-1\big)\,\kappa(\omega_{t^-},\mathrm{d}y)\,\mathrm{d}t}_{\text{dynamic part}}. \tag{3.4}$$

eq:girsanov-jump

The dynamic part has a transparent structure: the first integral pays $\log r_t$ for each jump the path *actually makes*, while the second accounts for the jumps that were expected but *not observed*. The rigorous statement follows from the *martingale problem*. We note that the jump measure N_ω has compensator $\hat\kappa_t(\omega_t,\mathrm{d}y)\,\mathrm{d}t$ under $\mathsf{Q}^{\hat\kappa}$, i.e., for every nonnegative predictable function f, we have that

$$\mathsf{E}_{\mathsf{Q}^{\hat\kappa}}\Big[\int_0^T\!\!\int_V f(t,\omega_{t^-},y)\,\mathsf{N}_\omega(\mathrm{d}t\,\mathrm{d}y)\Big] = \mathsf{E}_{\mathsf{Q}^{\hat\kappa}}\Big[\int_0^T\!\!\int_V f(t,\omega_{t^-},y)\,\hat\kappa_t(\omega_{t^-},\mathrm{d}y)\,\mathrm{d}t\Big]. \tag{3.5}$$

eq:compensator

Now write $\rho_t := (e_t)_\#\mathsf{Q}^{\hat\kappa}$ for the time marginal along the map $e_t(\omega)=\omega_t$, and set

$$\mathsf{Ent}\big(\hat\kappa_t(x,\cdot)\big|\kappa(x,\cdot)\big) := \int_V \phi\big(r_t(x,y)\big)\kappa(x,\mathrm{d}y), \qquad \phi(s) = s\log s - s + 1,$$

By applying (3.5) with $f = \log r_t$, the jump-measure integral in (3.4) turns into a time integral against $\hat\kappa_t = r_t\,\kappa$. Together with the fact that

$$\mathsf{E}_{\mathsf{Q}^{\hat\kappa}}\left[\int_0^T g(t,\omega_{t^-})\,\mathrm{d}t\right] = \int_0^T\!\!\int_V g(t,x)\,\rho(\mathrm{d}x),$$

for every predictable function g, the dynamic part in (3.4) combine to a single ϕ-integral,

$$\int_0^T\!\!\int_E \phi(r_t(x,y))\,\kappa(x,\mathrm{d}y)\,\rho_t(\mathrm{d}x)\,\mathrm{d}t = \int_0^T\!\!\int_V \mathsf{Ent}\big(\hat\kappa_t(x,\cdot)\big|\kappa(x,\cdot)\big)\rho_t(\mathrm{d}x)\,\mathrm{d}t, \tag{3.6}$$

eq:dynamic-average

and the initial term of (3.4) gives $\mathsf{Ent}(\rho_0|\bar\rho)$. Since both terms are nonnegative,

$$\mathsf{Ent}\big(\mathsf{Q}^{\hat\kappa}\big|\mathsf{P}_{\bar\rho}\big) < \infty \quad\Longleftrightarrow\quad \begin{cases} \mathsf{Ent}(\rho_0|\bar\rho) < +\infty \quad \text{and} \\ \displaystyle\int_0^T\!\!\int_V \mathsf{Ent}\big(\hat\kappa_t(x,\cdot)\big|\kappa(x,\cdot)\big)\,\rho_t(\mathrm{d}x)\,\mathrm{d}t < +\infty. \end{cases}$$

Both conditions hold, for instance, when the initial entropy is finite, the tilt is bounded, $r_t \le C$, and the total rates are bounded, $\sup_x \kappa(x,V) < +\infty$—these are the standing assumptions of Lecture 2. Note the two regimes of ϕ: near $s=1$ one has $\phi(s) \approx \frac12(s-1)^2$, so mildly tilted rates cost quadratically, while $\phi(s) \sim s\log s$ as $s \to +\infty$, so large tilts pay a Boltzmann price—a first shadow of exponential-type dissipation.

The direction that matters for us is the converse, established in great generality by Conforti and Léonard [3]: if $\mathsf{Ent}(\mathsf{Q}|\mathsf{P}_{\bar\rho}) < \infty$, then Q is itself the law of a jump process with the *same* deterministic drift and a *predictable* jump intensity kernel $\hat\kappa_t \ll \kappa$. In other words, every finite-entropy path measure is of the tilted form (3.3). Finiteness of the entropy is thus not merely an integrability condition, but also a structural statement. It says that an atypical trajectory of the particle system is simply the same process with tilted rates, and $\hat\kappa_t$ is precisely the observed jump kernel.

rem:markovianization

Remark 3.1 (From predictable to Markov) A priori, the intensity produced by the converse is predictable, so Q need not be Markov. However, if the reference process $\mathsf{P}_{\bar\rho}$ is Markov, then we can assume that the tilt is Markovian (see [3]).

Step 3: From kernels to fluxes. Collecting the averages of Step 2, we obtain

$$\mathsf{Ent}\big(\mathsf{Q}\big|\mathsf{P}_{\bar\rho}\big) = \underbrace{\mathsf{Ent}(\rho_0|\bar\rho)}_{\text{initial data}} + \underbrace{\int_0^T\!\!\int_V \mathsf{Ent}\big(\hat\kappa_t(x,\cdot)\big|\kappa(x,\cdot)\big)\,\rho_t(\mathrm{d}x)\,\mathrm{d}t}_{\text{dynamic cost}}, \tag{3.7}$$

eq:disintegrate-entropy

Next, we eliminate the kernel in favor of the mass it transports. Define the *flux* and the *expected flux* at a state ρ by

$$j_t := \rho_t \otimes \hat\kappa_t, \qquad (\rho\otimes\kappa)(\mathrm{d}x\,\mathrm{d}y) := \rho(\mathrm{d}x)\,\kappa(x,\mathrm{d}y),$$

both measures on E. Relative entropy disintegrates over the first variable, so that

$$\mathsf{Ent}\big(j_t\big|\rho_t\otimes\kappa\big) = \int_V \mathsf{Ent}\big(\hat\kappa_t(x,\cdot)\big|\kappa(x,\cdot)\big)\,\rho_t(\mathrm{d}x),$$

which is precisely the inner integrand of the dynamic cost in (3.7). Fixing the initial datum $\rho_0 = \bar\rho$ so that the initial term vanishes, (3.1) becomes an action on curves of measures and fluxes with Lagrangian L, i.e.,

$$\mathcal{I}_{LD}(\rho, j) = \int_0^T L(\rho_t, j_t)\,\mathrm{d}t, \qquad L(\rho, j) = \mathsf{Ent}\big(j \big| \rho \otimes \kappa\big), \tag{3.8}$$

eq:LD-action

finite only along pairs (ρ, j) satisfying the continuity equation

$$\dot\rho + \operatorname{div} j = 0, \qquad (\operatorname{div} j)(\mathrm{d}x) := \int_{y \in V} \big[j(\mathrm{d}x, \mathrm{d}y) - j(\mathrm{d}y, \mathrm{d}x)\big]. \tag{CE}$$

eq:ce-ld

The presentation here follows [13, App. A], where it is carried out for the Kac process.

rem:static-sanov

Remark 3.2 (The free energy is a rate function too) Lecture 1 justified the relative entropy $\mathsf{Ent}(\cdot|\pi)$ as driving energy by appealing to Sanov's theorem for $N \in \mathbb{N}$ i.i.d. samples $Y^1, \dots, Y^N$ from the invariant measure π. That is the *static* counterpart of (3.1): writing $\mathsf{S}^N := \frac{1}{N}\sum_{i=1}^N \delta_{Y^i} \in \mathcal{P}(V)$ for the empirical measure of such samples—a measure on *states*, where P^N is a measure on *paths*—one has

$$\mathrm{Prob}\big(\mathsf{S}^N \approx \rho\big) \asymp e^{-N\mathsf{Ent}(\rho|\pi)}.$$

So *both* ingredients of the gradient structure are large-deviation rate functions of the same particle system—the free energy from its equilibrium statistics, the dissipation from its path statistics. The splitting discussed in §3.2 below is precisely the statement that the dynamic one contains the static one.

3.2 Time reversal performs the splitting

sec:reversal

Say that a gradient structure $(\mathcal{F}, \mathcal{R}, \mathcal{R}^*)$ is *induced by* the particle system if the large-deviation action (3.8) is, up to a positive constant, the energy–dissipation functional (EDB) of Lecture 2:

$$\begin{aligned} \mathcal{I}_{LD}(\rho, j) &= c\,\mathscr{L}(\rho, j; [0,T]), \\ \mathscr{L}(\rho, j; [0,T]) &= \int_0^T \big[\mathcal{R}(\rho_t, j_t) + \mathcal{D}(\rho_t)\big]\,\mathrm{d}t + \mathcal{F}(\rho_T) - \mathcal{F}(\rho_0). \end{aligned}$$

Such an identification gives two things at once. Nonnegativity of $\mathscr{L}$ needs no proof, since $\mathcal{I}_{LD}$ is a relative entropy; and the null-minimizers of $\mathscr{L}$, which is how Lecture 2 defines solutions, are exactly the typical trajectories of the particle system. The gradient structure is then not postulated but *derived*.

It remains to say why a rate functional should split in this way at all. The argument below is formal but complete, and produces the two halves of $\mathscr{L}$ with $c = \frac{1}{2}$.

The time-reversal functional. Let $\mathsf{s}(x,y) := (y,x)$ denote the swap on E and set

$$j^\dagger := \mathsf{s}_\# j,$$

the flux of the *time-reversed* observation: if j records mass moving from x to y, then $j^\dagger$ records the same mass moving from y to x. The divergence in (CE) antisymmetrizes, so swapping flips its sign,

$$\operatorname{div} j^\dagger = -\operatorname{div} j,$$

and consequently, if (ρ, j) solves (CE) on $[0,T]$, then so does the genuinely reversed curve $(\hat\rho_t, \hat j_t) := (\rho_{T-t}, j^\dagger_{T-t})$:

$$\partial_t \hat\rho_t = -(\partial_t \rho)(T-t) = \operatorname{div} j_{T-t} = -\operatorname{div} \hat j_t.$$

Reversal thus maps admissible pairs to admissible pairs, and it is an involution. Since the action (3.8) integrates a Lagrangian evaluated pointwise in time, the substitution $t \mapsto T-t$ shows that the price the particle system charges for the reversed observation is

$$\mathcal{I}_{LD}(\hat\rho, \hat j) = \int_0^T L\big(\rho_{T-t}, j^\dagger_{T-t}\big)\,\mathrm{d}t = \int_0^T L\big(\rho_t, j^\dagger_t\big)\,\mathrm{d}t =: \mathcal{I}^\dagger_{LD}(\rho, j).$$

This is the *time-reversal functional*. It is defined on the same class of pairs as $\mathcal{I}_{LD}$, and $\mathcal{I}^{\dagger\dagger}_{LD} = \mathcal{I}_{LD}$. Now split $\mathcal{I}_{LD}$ into its symmetric and antisymmetric parts under $\dagger$:

$$\mathcal{I}_{LD} = \underbrace{\tfrac12\big(\mathcal{I}_{LD} + \mathcal{I}^\dagger_{LD}\big)}_{\text{symmetric}} + \underbrace{\tfrac12\big(\mathcal{I}_{LD} - \mathcal{I}^\dagger_{LD}\big)}_{\text{antisymmetric}}.$$

For a process satisfying detailed balance with respect to π, these two pieces are precisely the two halves of an energy–dissipation balance, i.e., formally,

$$\begin{aligned} \tfrac12\big(\mathcal{I}_{LD} + \mathcal{I}^\dagger_{LD}\big) &= \tfrac12 \int_0^T \big[\mathcal{R}(\rho_t, j_t) + \mathcal{D}(\rho_t)\big]\,\mathrm{d}t, \\ \tfrac12\big(\mathcal{I}_{LD} - \mathcal{I}^\dagger_{LD}\big) &= \tfrac12\big(\mathcal{F}(\rho_T) - \mathcal{F}(\rho_0)\big). \end{aligned} \tag{3.9}$$

eq:antisym-part

The computation. Both identities come out of one change of variables. Take the reference measure of §3.1 to be the invariant one, $\bar\rho = \pi$, and assume the detailed-balance condition (DBC) with respect to it. The *edge measure*

$$\vartheta := \pi \otimes \kappa, \qquad \vartheta(\mathrm{d}x\,\mathrm{d}y) = \pi(\mathrm{d}x)\,\kappa(x, \mathrm{d}y)$$

on E is then symmetric, $\mathsf{s}_\#\vartheta = \vartheta$, and we write $\rho_t = u_t \pi$ and $j_t = w_t\,\vartheta$.

The second is legitimate wherever $L(\rho_t, j_t) < +\infty$, since then $j_t \ll \rho_t \otimes \kappa$, and the expected flux of Step 3 is

$$(\rho \otimes \kappa)(\mathrm{d}x\,\mathrm{d}y) = u(x)\,\vartheta(\mathrm{d}x\,\mathrm{d}y),$$

i.e., its density against ϑ is the density of ρ at the *tail* of the edge. Since $j_t = \rho_t \otimes \hat{\kappa}_t$ with $\hat{\kappa}_t = r_t \kappa$ by (3.2), $w_t(x,y) = u_t(x)\, r_t(x,y)$ is exactly the tilt, and the Lagrangian below is the dynamic cost (3.6) rewritten w.r.t. ϑ. So, with $\phi(s) = s \log s - s + 1$, we obtain

$$\begin{aligned} L(\rho, j) = \mathsf{Ent}\big(j \,\big|\, \rho \otimes \kappa\big) &= \int_E \phi\Big(\frac{w(x,y)}{u(x)}\Big) u(x)\, \vartheta(\mathrm{d}x\, \mathrm{d}y) \\ &= \int_E \Big[w(x,y) \log \frac{w(x,y)}{u(x)} - w(x,y) + u(x) \Big] \vartheta(\mathrm{d}x\, \mathrm{d}y). \end{aligned}$$

Because ϑ is symmetric, $j^\dagger = \mathsf{s}_\#(w\vartheta) = (w \circ \mathsf{s})\, \vartheta$ has density $w(y,x)$ at (x,y). Relabelling $(x,y) \leftrightarrow (y,x)$ under ϑ then gives

$$L(\rho, j^\dagger) = \int_E \Big[w(x,y) \log \frac{w(x,y)}{u(y)} - w(x,y) + u(y) \Big] \vartheta(\mathrm{d}x\, \mathrm{d}y).$$

The two Lagrangians differ only in whether the observed flux along an edge is priced against the density at the *tail* x or at the *head* y. This asymmetry is the whole content of the splitting, and both follow by subtracting and adding.

Antisymmetric part. Subtracting, the $w \log w$ and $-w$ terms cancel and

$$\frac{1}{2}\big(L(\rho, j) - L(\rho, j^\dagger)\big) = \frac{1}{2} \int_E \Big[w(x,y) \log \frac{u(y)}{u(x)} + u(x) - u(y) \Big] \vartheta(\mathrm{d}x\, \mathrm{d}y) = \frac{1}{2} \langle \nabla \log u, j \rangle,$$

since $\int_E [u(x) - u(y)] \vartheta = 0$ by the symmetry of ϑ. What is left is exactly half the integrand of the chain rule (CR) for the free energy $\mathcal{F} = \mathsf{Ent}(\cdot \,|\, \pi)$, whose density ϕ has $\phi'(u) = \log u$. Integrating in time gives (3.9) and shows that the antisymmetric part sees only the endpoints.

Symmetric part. Adding instead, and using $\frac{1}{2}\big(\log u(x) + \log u(y)\big) = \log \sqrt{u(x)u(y)}$,

$$\frac{1}{2}\big(L(\rho, j) + L(\rho, j^\dagger)\big) = \int_E \Big[w(x,y) \log \frac{w(x,y)}{\sqrt{u(x)u(y)}} - w(x,y) + \frac{u(x) + u(y)}{2} \Big] \vartheta(\mathrm{d}x\, \mathrm{d}y).$$

This is where the geometric mean $\mathsf{m}(u,v) = \sqrt{uv}$ comes from and is therefore not a modeling choice but the output of symmetrizing. Using the shorthand $\mathsf{m} := \sqrt{u(x)u(y)}$ and completing the square in $\sqrt{u}$,

$$\frac{1}{2}\big(L(\rho, j) + L(\rho, j^\dagger)\big) = \int_E \mathsf{m}\, \phi\Big(\frac{w}{\mathsf{m}}\Big) \vartheta \;+\; \frac{1}{2} \int_E \Big(\sqrt{u(x)} - \sqrt{u(y)} \Big)^2 \vartheta,$$

where we use the fact that $\mathsf{m}\, \phi(w/\mathsf{m}) = w \log(w/\mathsf{m}) - w + \mathsf{m}$ for the first term, and for the second, $\frac{1}{2}(\sqrt{u(x)} - \sqrt{u(y)})^2 = \frac{1}{2}(u(x) + u(y)) - \mathsf{m}$. These two terms are the two dissipation terms. The second is half the cosh Fisher information of Example 2.11, whose integrand is $2(\sqrt{u} - \sqrt{v})^2$.

References

[1] L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second edition, Lectures in Mathematics ETH Zürich, Birkhäuser, Basel, 2008.

[2] S.-N. Chow, W. Huang, Y. Li and H. Zhou, Fokker–Planck equations for a free energy functional or Markov process on a graph, Arch. Ration. Mech. Anal. **203** (2012), no. 3, 969–1008.

[3] G. Conforti and C. Léonard, Time reversal of Markov processes with jumps under a finite entropy condition, Stochastic Process. Appl. **144** (2022), 85–124.

[4] M. Erbar and J. Maas, Ricci curvature of finite Markov chains via convexity of the entropy, Arch. Ration. Mech. Anal. **206** (2012), no. 3, 997–1038.

[5] N. Fournier and S. Méléard, A microscopic probabilistic description of a locally regulated population and macroscopic approximations, Ann. Appl. Probab. **14** (2004), no. 4, 1880–1919.

[6] J. Hoeksema and O. Tse, Generalized gradient structures for measure-valued population dynamics and their large-population limit, Calc. Var. Partial Differential Equations **62** (2023), no. 5, Paper No. 158.

[7] A. Hraivoronska and O. Tse, Diffusive limit of random walks on tessellations via generalized gradient flows, SIAM J. Math. Anal. **55** (2023), no. 4, 2948–2995.

[8] J. Maas, Gradient flows of the entropy for finite Markov chains, J. Funct. Anal. **261** (2011), no. 8, 2250–2292.

[9] A. Mielke, A gradient structure for reaction–diffusion systems and for energy-drift-diffusion systems, Nonlinearity **24** (2011), no. 4, 1329–1346.

[10] A. Mielke, On evolutionary Γ-convergence for gradient systems, in *Macroscopic and Large Scale Phenomena*, Lect. Notes Appl. Math. Mech. **3**, Springer, 2016, 187–249.

[11] A. Mielke, A. Montefusco and M. A. Peletier, Exploring families of energy-dissipation landscapes via tilting: three types of EDP convergence, Contin. Mech. Thermodyn. **33** (2021), no. 3, 611–637.

[12] A. Mielke, M. A. Peletier and D. R. M. Renger, On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion, Potential Anal. **41** (2014), no. 4, 1293–1327.

[13] A. Montefusco, U. Sharma and O. Tse, Fourier–Cattaneo equation: stochastic origin, variational formulation, and asymptotic limits, Commun. Math. Sci. **24** (2026), no. 1, 1–42.

[14] M. A. Peletier, R. Rossi, G. Savaré and O. Tse, Jump processes as generalized gradient flows, Calc. Var. Partial Differential Equations **61** (2022), no. 1, Paper No. 33, 85 pp.

[15] E. Sandier and S. Serfaty, Gamma-convergence of gradient flows with applications to Ginzburg–Landau, Comm. Pure Appl. Math. **57** (2004), no. 12, 1627–1672.