Gradient Structures from Classical to Quantum

Oliver Tse

Eindhoven University of Technology

Updated on March 18, 2025

Tutorial notes given at IPAM in Spring 2025

Abstract. *Gradient structures* provide a powerful framework for understanding the evolution of dynamical systems by linking them to *variational principles*. This lecture series explores the role of gradient flows in *classical* and *quantum* settings, illustrating how they unify concepts from thermodynamics, optimal transport, and quantum mechanics. We begin with classical gradient structures, discussing their applications in dissipative systems and their connections to transport costs and energetic structures. Moving to the quantum realm, we examine quantum analogs of gradient flows, exploring their role in open quantum systems, quantum entropy dissipation, and non-commutative transport. Bridging these perspectives, we hope to highlight the emerging mathematical structures that offer new insights into both classical and quantum dynamics.

Contents

1 Finite Dimensional Gradient Structures

1.1 Gradient structures on Euclidean space

We consider an evolution equation of the form

$$\dot{\mathbf{x}}(t) = -\nabla\mathcal{F}(\mathbf{x}(t)) \quad \text{for } t > 0, \qquad \mathbf{x}(0) = \mathbf{x}_0 \in \mathbb{R}^d, \tag{1.1}$$

eq:gf-finite

where $\mathcal{F}\colon \mathbb{R}^d \to \mathbb{R}$ is a sufficiently smooth function, henceforth called a driving functional, and $\mathbf{x}_0 \in \mathbb{R}^d$ is a given initial point.

In literature, the quantity $\dot{\mathbf{x}}$ is the *velocity* of the curve $\mathbf{x}$ and $-\nabla\mathcal{F}(\mathbf{x})$ is known as the *restoring force* of the driving functional $\mathcal{F}$ at $\mathbf{x}$. While (1.1) is just an equation, it possesses a more geometric interpretation. Namely, the curve $t \mapsto \mathbf{x}(t)$ is regarded as a *gradient flow* (or a *curve of steepest descent*). However, without a specification of its associated *gradient structure*, this statement is incomplete.

A gradient structure encompasses the following ingredients:

(1) A *state space* $\mathfrak{X}$, which in this case is $\mathbb{R}^d$;

(2) A *driving energy* $\mathcal{F} : \mathfrak{X} \to \mathbb{R} \cup \{+\infty\}$;

(3) A notion of *temporal derivative* $\dot{\mathbf{x}}$, a notion of *gradient* $\nabla\mathcal{F}$;

(4) A *kinetic relation* (or *force-to-velocity* map), which relates the restoring force $-\nabla\mathcal{F}$ and velocity $\dot{\mathbf{x}}$, encoding the instantaneous *frictional* properties of the system.

In the Euclidean case, specifying a gradient structure may initially seem redundant. However, as we will see later, a single gradient flow equation can have several (or infinitely many) gradient structures, each offering a *different* physical interpretation.

Variational formulation A valuable characteristic of gradient flow equations is that they can be given a variational form. Due to the chain rule, we obtain the *utmost* important property of a gradient flow:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t)) &= \mathrm{D}\mathcal{F}(\mathbf{x}(t))[\dot{\mathbf{x}}(t)] = \nabla\mathcal{F}(\mathbf{x}(t)) \cdot \dot{\mathbf{x}}(t) \\ &= -|\dot{\mathbf{x}}(t)|^2 = -|\nabla\mathcal{F}(\mathbf{x}(t))|^2 = -|\nabla\mathcal{F}(\mathbf{x}(t))||\dot{\mathbf{x}}(t)| \\ &= -\frac{1}{2}|\dot{\mathbf{x}}(t)|^2 - \frac{1}{2}|\nabla\mathcal{F}(\mathbf{x}(t))|^2 \le 0,\end{aligned}$$

where $\nabla\mathcal{F}(\mathbf{x}) = 0$ if and only if $\mathbf{x}$ is a stationary point of $\mathcal{F}$. In particular, $t \mapsto \mathcal{F}(\mathbf{x}(t))$ is a strictly decreasing function until $\mathbf{x}$ hits a stationary point. The final form on the right-hand side of the previous equation will play a fundamental role in the study of gradient flows and will provide us with the means to formulate (1.1) as a variational problem.

From the previous equalities, one deduces that a solution to (1.1), which satisfies the chain rule, can be characterized by the following scalar conditions for every $t > 0$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t)) = -\frac{1}{2}|\dot{\mathbf{x}}|^2(t) - \frac{1}{2}|\nabla\mathcal{F}(\mathbf{x}(t))|^2. \tag{1.2}$$

eq:gf-finite-edp

While (1.1) makes sense only in a Hilbert space, the formulation (1.2) is purely metric in nature and can be extended to more general spaces, provided we interpret $|\dot{\mathbf{x}}|$ as the *metric speed* of a curve $\mathbf{x}$ and $|\nabla\mathcal{F}(\mathbf{x})|$ as the *metric slope*.

Additionally, (1.2) gives us a way to formulate (1.1) as a variational problem. Integrating (1.2) over any interval $[s,t] \subset [0,\infty)$, we obtain the *energy-dissipation balance*

$$\mathscr{L}(\mathbf{x},[s,t]) \coloneqq \int_s^t \left\{ \frac{1}{2}|\dot{\mathbf{x}}|^2(r) + \frac{1}{2}|\nabla\mathcal{F}(\mathbf{x}(r))|^2 \right\} \mathrm{d}r + \mathcal{F}(\mathbf{x}(t)) - \mathcal{F}(\mathbf{x}(s)) = 0, \tag{1.3}$$

where $\mathscr{L}$ is called the *energy-dissipation functional*. Clearly, any curve $\mathbf{x}$ satisfying (1.1) will give $\mathscr{L}(\mathbf{x},[s,t]) = 0$ for every $[s,t] \subset [0,\infty)$ via (1.2). Conversely, if $\mathbf{x}$ satisfies (1.3) and the chain rule for $\mathcal{F}$ along the curve $\mathbf{x}$ holds, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t)) = \nabla\mathcal{F}(\mathbf{x}(t)) \cdot \dot{\mathbf{x}}(t) \qquad \text{for every } t > 0, \tag{1.4}$$

then $\mathscr{L}(\mathbf{x},[s,t])$ can be expressed as

$$\mathscr{L}(\mathbf{x},[s,t]) = \int_s^t \left\{ \frac{1}{2}|\dot{\mathbf{x}}|^2(r) + \frac{1}{2}|-\nabla\mathcal{F}(\mathbf{x}(r))|^2 - (-\nabla\mathcal{F}(\mathbf{x}(r))) \cdot \dot{\mathbf{x}}(r) \right\} \mathrm{d}r = 0.$$

Due to Young's inequality, the integrand in the identity above is nonnegative, and thus

$$\frac{1}{2}|\dot{\mathbf{x}}|^2(r) + \frac{1}{2}|-\nabla\mathcal{F}(\mathbf{x}(r))|^2 - (-\nabla\mathcal{F}(\mathbf{x}(r))) \cdot \dot{\mathbf{x}}(r) = 0 \quad \iff \quad \dot{\mathbf{x}}(r) = -\nabla\mathcal{F}(\mathbf{x}(r)),$$

for almost every $r \in [0,+\infty)$ (since $[s,t]$ was arbitrary). Since $t \mapsto \nabla\mathcal{F}(\mathbf{x}(t))$ is continuous, the equality holds true also for all $r \in [0,+\infty)$, i.e. $\mathbf{x}$ is a solution of (1.1).

More importantly, the observation that $\mathscr{L}(\mathbf{x},[s,t]) \geq 0$ for every curve $\mathbf{x}$ satisfying the chain rule (1.4) leads us to a variational characterization of gradient flows:

$$\begin{aligned} \mathbf{x} \text{ solves (1.1)} \quad &\iff \quad \mathscr{L}(\mathbf{x},[s,t]) = 0 \text{ for every } [s,t] \subset [0,+\infty) \\ &\iff \quad \mathbf{x} \text{ minimizes } \mathscr{L}(\,\cdot\,,[s,t]) \text{ for every } [s,t] \subset [0,+\infty). \end{aligned}$$

In particular, this means that the Cauchy problem (1.1) can be analyzed using tools from the Calculus of Variations, which brings us to an alternative definition of a solution to (1.1) as a gradient flow (or curve of maximal slope):

Definition 1.1 The gradient flow of a driving energy $\mathcal{F}\colon \mathbb{R}^d \to \mathbb{R}$ is a continuous curve $\mathbf{x} \in \mathcal{C}([0,+\infty);\mathbb{R}^d)$ characterized by the following properties:

(a) For every $\mathbf{x}_0 \in \mathbb{R}^d$, $\mathbf{x}(0) = \mathbf{x}_0$;

(b) The curve $\mathbf{x}$ satisfies the energy-dissipation balance (or is a *zero-cost flow*)

$$\mathscr{L}(\mathbf{x},[s,t]) = 0 \qquad \text{for every } [s,t] \subset [0,+\infty),$$

or equivalently, the *energy-dissipation principle*

$$\mathbf{x} \in \operatorname{argmin}\Big\{ \mathscr{L}(\mathbf{y},[s,t]) : \mathbf{y} \in \mathcal{C}([0,+\infty);\mathbb{R}^d) \Big\} \qquad \text{for every } [s,t] \subset [0,+\infty).$$

1.2 Gradient flows on nonlinear geometry

A more general perspective of (1.1) is necessary when the state space $\mathfrak{X}$ is, e.g., a Riemannian manifold, i.e., $\mathfrak{X} = (M, g)$ where M is a smooth manifold with a symmetric and positive definite family of *metric tensors* $g_{\mathbf{x}} : T_{\mathbf{x}}M \times T_{\mathbf{x}}M \to \mathbb{R}$, where $T_{\mathbf{x}}M$ is the tangent space of M at $\mathbf{x} \in M$. For a curve $\gamma\colon [a, b] \to M$, its length is given by

$$\mathrm{Length}_g(\gamma, [a, b]) \coloneqq \int_a^b \sqrt{g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t))}\,\mathrm{d}t.$$

Since $g_{\mathbf{x}}$ is symmetric and coercive for each $\mathbf{x} \in M$, it defines a symmetric and positive definite linear map $\mathbb{G}(\mathbf{x})\colon T_{\mathbf{x}}M \to T^*_{\mathbf{x}}M$ via $\langle \mathbb{G}(\mathbf{x})u, v\rangle \coloneqq g_{\mathbf{x}}(u, v)$, $u, v \in T_{\mathbf{x}}M$, where $T^*_{\mathbf{x}}M$ is the cotangent space of M at $\mathbf{x}$, and $\langle\cdot,\cdot\rangle$ denotes the duality pairing between the tangent space $T_{\mathbf{x}}M$ at $x \in M$ and its dual $T^*_{\mathbf{x}}M$.

Given a driving functional $\mathcal{F} : M \to \mathbb{R}$, the differential $\mathrm{D}\mathcal{F}(\mathbf{x})$ of $\mathcal{F}$ at $\mathbf{x}$ is defined by

$$\mathrm{D}\mathcal{F}(\mathbf{x})[\mathbf{v}] \coloneqq \lim_{t\to 0} \frac{1}{h}\Big\{\mathcal{F}(\gamma(t)) - \mathcal{F}(\gamma(0))\Big\}, \qquad \forall v \in T_{\mathbf{x}}M,$$

where γ is a *geodesic curve*—curve of minimal length—emitting from $\gamma(0) = \mathbf{x}$ with initial tangent vector $\dot\gamma(0) = \mathbf{v}$. In particular, $\mathrm{D}\mathcal{F}(\mathbf{x}) \in T^*_{\mathbf{x}}M$ is an element of the cotangent space $T^*_{\mathbf{x}}M$. The invertibility of $\mathbb{G}(\mathbf{x})$ then allows us to uniquely associate a cotangent vector with a tangent vector.

Definition 1.2 (Gradient) The *gradient* of a function $\mathcal{F} : M \to \mathbb{R}$ on a Riemannian manifold (M, g) is defined via

$$\nabla_g \mathcal{F}(\mathbf{x}) \coloneqq \mathbb{K}(\mathbf{x})\mathrm{D}\mathcal{F}(\mathbf{x}),$$

where $\mathbb{K}(\mathbf{x}) \coloneqq \mathbb{G}(\mathbf{x})^{-1}\colon T_{\mathbf{x}}M^* \to T_{\mathbf{x}}M$ is called the *Onsager operator*.

The gradient flow equation (1.1) in this setting then reads

$$\dot{\mathbf{x}}(t) = -\nabla_g \mathcal{F}(\mathbf{x}) = -\mathbb{K}(\mathbf{x}(t))\mathrm{D}\mathcal{F}(\mathbf{x}(t)) \in T_{\mathbf{x}(t)}M, \tag{1.5}$$

eq:gf-manifold

or equivalently,

$$0 = \mathbb{G}(\mathbf{x}(t))\dot{\mathbf{x}}(t) + \mathrm{D}\mathcal{F}(\mathbf{x}(t)) \in T^*_{\mathbf{x}(t)}M.$$

In the terminology of gradient structures, (1.5) is a kinetic relation.

As before, one applies the chain rule to obtain the gradient flow property

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t)) &= \mathrm{D}\mathcal{F}(\mathbf{x}(t))[\dot{\mathbf{x}}(t)] = g_{\mathbf{x}(t)}(\dot{\mathbf{x}}(t), \mathbb{K}(\mathbf{x}(t))\mathrm{D}\mathcal{F}(\mathbf{x}(t))) \\
&= -\langle \mathbb{G}(\mathbf{x}(t))\dot{\mathbf{x}}(t), \dot{\mathbf{x}}(t)\rangle = |\dot{\mathbf{x}}(t)|^2_{\mathbb{G}(\mathbf{x}(t))} \\
&= -\langle \mathrm{D}\mathcal{F}(\mathbf{x}(t)), \mathbb{K}(\mathbf{x}(t))\mathrm{D}\mathcal{F}(\mathbf{x}(t))\rangle = |\mathrm{D}\mathcal{F}(\mathbf{x}(t))|^2_{\mathbb{K}(\mathbf{x}(t))} \\
&= -\frac{1}{2}|\dot{\mathbf{x}}(t)|^2_{\mathbb{G}(\mathbf{x}(t))} - \frac{1}{2}|\mathrm{D}\mathcal{F}(\mathbf{x}(t))|^2_{\mathbb{K}(\mathbf{x}(t))} \le 0,
\end{aligned}$$

resulting in the single scalar equation

$$\frac{1}{2}|\dot{\mathbf{x}}(t)|^2_{\mathbb{G}(\mathbf{x}(t))} + \frac{1}{2}|-\mathrm{D}\mathcal{F}(\mathbf{x}(t))|^2_{\mathbb{K}(\mathbf{x}(t))} = \langle \dot{\mathbf{x}}(t), -\mathrm{D}\mathcal{F}(\mathbf{x}(t))\rangle, \tag{1.6}$$

eq:edp-manifold

which characterizes the solution to the gradient flow equation (1.5).

Remark 1.3 The symmetry and positive definiteness of $\mathbb{G}$ are fundamental from the point of view of thermodynamics, as was shown by Lars Onsager in his work on *"Reciprocal relations in irreversible processes,"* (1931), which won him the Nobel prize in 1963. His *"reciprocal relations"* were derived in the context of linearized irreversible thermodynamics and simply mean, in modern language, the symmetry relation $\mathbb{G}^\top = \mathbb{G}$.

An equation $\dot{\mathbf{x}}(t) = V(\mathbf{x}(t))$ may have multiple gradient structures or none at all. If it has at least one, then the question remains: *Which one do we take?*

Example 1.4 Let $M = \mathbb{R}^2$ and consider

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

Setting $\mathcal{F}(\mathbf{x}) = \frac{1}{2}(x_1^2 + x_2^2) + \frac{\alpha}{4}x_2^4$, we find $\nabla\mathcal{F}(\mathbf{x}) = (x_1, x_2 + \alpha x_2^3)^\top = -V(\mathbf{x})$. Hence, this evolution has a gradient structure with

$$\mathcal{F}(\mathbf{x}) = \frac{1}{2}(x_1^2 + x_2^2) + \frac{\alpha}{4}x_2^4, \qquad \mathbb{K}(\mathbf{x}) = I_d.$$

Alternatively, one coulda also consider the gradient structure

$$\widehat{\mathcal{F}}(\mathbf{x}) = \frac{1}{2}(x_1^2 + x_2^2), \qquad \widehat{\mathbb{K}}(\mathbf{x}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 + \alpha x_2^2 \end{bmatrix}.$$

Therefore, when looking at the ODE, it is unclear whether the nonlinear term $-\alpha x_2^3$, $\alpha > 0$, arises from a non-quadratic energy $\mathcal{F}$ or from a state-dependent friction law $\widehat{\mathbb{G}}$. This means that the choice of a gradient structure is thus a choice of modeling and usually contains *more* information than the ODE itself.

Looking at the scalar equation (1.6) and setting

$$\mathcal{R}(\mathbf{x}, \mathbf{s}) := \frac{1}{2}\langle \mathbf{s}, \mathbb{G}(\mathbf{x})\mathbf{s}\rangle, \qquad \mathcal{R}^*(\mathbf{x}, \mathbf{p}) := \frac{1}{2}\langle \xi, \mathbb{K}(\mathbf{x})\xi\rangle, \qquad (\mathbf{s}, \xi) \in T_{\mathbf{x}}M \times T_{\mathbf{x}}^*M,$$

we see that (1.6) may be expressed as

$$\mathcal{R}(\mathbf{x}, \dot{\mathbf{x}}) + \mathcal{R}^*(\mathbf{x}, -\mathrm{D}\mathcal{F}(\mathbf{x}(t))) = \langle \dot{\mathbf{x}}(t), -\mathrm{D}\mathcal{F}(\mathbf{x}(t))\rangle = -\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t)). \tag{1.7}$$

The function $\mathcal{R}$ ($\mathcal{R}^*$) is called the *(dual) dissipation potential*.

It is precisely this form that enables us to extend the concept of gradient flows from quadratic (dual) dissipation potentials to general convex functions. As we will see below, the gradient flow equation we obtain for general convex functions reads

$$\dot{\mathbf{x}}(t) = \mathrm{D}_\xi\mathcal{R}^*(\mathbf{x}(t), -\mathrm{D}\mathcal{F}(\mathbf{x}(t))), \tag{1.8}$$

or equivalently,

$$\mathrm{D}_{\mathbf{s}}\mathcal{R}(\mathbf{x}(t), \dot{\mathbf{x}}(t)) + \mathrm{D}\mathcal{F}(\mathbf{x}(t)) = 0.$$

Variational formulation As in the previous section, we would like to formulate the Cauchy problem (1.5) as a variational problem. To do so, we will need a generalization of Young's inequality for general convex functions.

Definition 1.5 (Legendre-Fenchel conjugate) Let $\Psi\colon \mathfrak{Z} \to \mathbb{R} \cup \{+\infty\}$ be a proper functional on a Banach space $\mathfrak{Z}$. The *Legendre-Fenchel conjugate* Ψ^* of Ψ is defined by

$$\mathfrak{Z}^* \ni \xi \mapsto \Psi^*(\xi) := \sup\Big\{ \langle \xi, \mathbf{s}\rangle - \Psi(\mathbf{s}) : \mathbf{s} \in \mathfrak{Z} \Big\}.$$

If Ψ is convex and lower semicontinuous, the pair (Ψ, Ψ^*) is called a *conjugate pair*.

The *subdifferential* $\partial\Psi$ of a convex function $\Psi : \mathfrak{Z} \to \mathbb{R} \cup \{+\infty\}$ is given by

$$\partial\Psi(\mathbf{s}) := \Big\{ \xi \in \mathfrak{X}^* : \Psi(\mathbf{r}) \geq \Psi(\mathbf{s}) + \langle \xi, \mathbf{r} - \mathbf{s}\rangle \quad \forall\, \mathbf{r} \in \mathfrak{Z} \Big\} \subset \mathfrak{Z}^*, \qquad \mathbf{s} \in \mathrm{dom}\Psi,$$

and $\partial\Psi(\mathbf{s}) = \emptyset$ otherwise. If Ψ is differentiable at $\mathbf{s}$, then $\partial\Psi(\mathbf{s}) = \{D\Psi(\mathbf{s})\}$. In fact, Ψ is differentiable at $\mathbf{s}$ if and only if $\partial\Psi(\mathbf{s})$ is a singleton.

Remark 1.6 (a) Ψ^* is convex and lower semicontinuous, even if Ψ is not.

(b) A trivial consequence of the definition of Ψ^* is the *Fenchel–Young inequality*

$$\Psi(\mathbf{s}) + \Psi^*(\xi) \geq \langle \xi, \mathbf{s}\rangle \qquad \forall\, (\mathbf{s}, \xi) \in \mathfrak{Z} \times \mathfrak{Z}^*. \tag{1.9}$$

eq:fenchel-young

(c) *Involution property:* If $\mathfrak{X}$ is reflexive and Ψ is convex and lower semicontinuous, then $\Psi^{**} = (\Psi^*)^* = \Psi$.

prop:variational-equivalence

Proposition 1.7. *Let $\mathfrak{Z}$ be a reflexive Banach space and (Ψ, Ψ^*) be a conjugate pair of proper, convex and lower semicontinuous functionals. Then, for every $(\mathbf{s}, \xi) \in \mathfrak{Z} \times \mathfrak{Z}^*$, the following statements are equivalent:*

(i) $\xi \in \partial\Psi(\mathbf{s})$ *(subdifferential inclusion in $\mathfrak{X}^*$)*

(ii) $\Psi(\mathbf{s}) + \Psi^*(\xi) = \langle \xi, \mathbf{s}\rangle$ *(optimality condition in $\mathbb{R}$)*

(iii) $\mathbf{s} \in \partial\Psi^*(\xi)$ *(subdifferential inclusion in $\mathfrak{X}$)*

Note that "$=$" in (ii) may be replaced with "$\leq$" instead since the Young–Fenchel inequality always gives "$\geq$".

Example 1.8 On a Hilbert space $\mathfrak{Z}$ with $\mathbb{G}\colon \mathfrak{Z} \to \mathfrak{Z}^*$ symmetric and positive definite,

$$\Psi(\mathbf{s}) = \frac{1}{2}\langle \mathbb{G}\mathbf{s}, \mathbf{s}\rangle \quad \Longleftrightarrow \quad \Psi^*(\xi) = \frac{1}{2}\langle \xi, \mathbb{K}\xi\rangle, \quad \mathbb{K} = (\mathbb{G})^{-1}.$$

Since $\Psi(\mathbf{s})$ and $\Psi^*(\xi)$ are differentiable with derivatives $\mathbb{G}\mathbf{s}$ and $\mathbb{K}\xi$ respectively, we have that $\partial\Psi(\mathbf{s}) = \{\mathbb{G}\mathbf{s}\}$ and $\partial\Psi^*(\xi) = \{\mathbb{K}\xi\}$.

If $p \in (1,\infty)$ and $\mathfrak{Z}$ is a Banach space, then

$$\Psi(\mathbf{s}) = \frac{1}{p}\|\mathbf{s}\|_{\mathfrak{Z}}^p \quad \Longleftrightarrow \quad \Psi^*(\xi) = \frac{1}{q}\|\xi\|_{\mathfrak{Z}^*}^q, \quad q = \frac{p}{p-1}.$$

Proposition 1.7 justifies the connection between (1.7) and (1.8) in the case when $\mathcal{R}$ and $\mathcal{R}^*$ are differentiable in its second argument. Upon integrating over an arbitrary intervals $[s,t] \subset [0,\infty]$, we obtain the energy-dissipation balance

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

which now holds for fairly general nonlinear manifolds. As in the previous section, and under the chain rule, we obtain the following characterization:

$$\mathbf{x} \text{ solves (1.1)} \quad \Longleftrightarrow \quad \mathbf{x} \text{ is a minimizer of the } \mathscr{L}(\,\cdot\,,[s,t]) \text{ for every } [s,t] \subset [0,\infty).$$

This formulation enables one to rigorously deduce the existence of (generalized) gradient flow solutions in two steps:

(1) $\mathscr{L}(\mathbf{x},[s,t]) \le 0$: Deduce the existence of curves via incremental minimization.

(2) $\mathscr{L}(\mathbf{x},[s,t]) \ge 0$: Prove a chain rule inequality for an admissible class of curves.

Remark 1.9 Consider a small time horizon $[0,\tau]$, $\tau \ll 1$. Making a change of variables $r = \tau s$, $s \in [0,1]$ and setting $\mathbf{x}^\tau(s) := \mathbf{x}(\tau s)$, the energy-dissipation functional can be expressed as

$$\mathscr{L}(\mathbf{x},[0,\tau]) = \int_0^1 \tau\mathcal{R}\left(\mathbf{x}^\tau(s),\frac{\dot{\mathbf{x}}^\tau(s)}{\tau}\right) + \tau\mathcal{R}^*(\mathbf{x}^\tau(s),-\mathrm{D}\mathcal{F}(\mathbf{x}^\tau(s)))\,\mathrm{d}s + \mathcal{F}(\mathbf{x}^\tau(1)) - \mathcal{F}(\mathbf{x}^\tau(0)).$$

Therefore, for $\tau \ll 1$, the second term under the temporal integral is small, which gives a motivation for the incremental minimization problem: For each $k \ge 0$, set

$$\mathbf{x}_{k+1} = \mathbf{y}_*(1), \quad \text{with} \quad \mathbf{y}_* \in \operatorname*{argmin}_{\mathbf{y}:\,\mathbf{y}(0)=\mathbf{x}_k} \left\{ \int_0^1 \tau\mathcal{R}\left(\mathbf{y}(s),\frac{\dot{\mathbf{y}}(s)}{\tau}\right)\mathrm{d}s + \mathcal{F}(\mathbf{y}(1)) - \mathcal{F}(\mathbf{x}_k) \right\}.$$

2 Gradient Structures in Spaces of Measures

In this section, we are concerned with the evolution of (probability) measures. We will (formally) investigate two classes of evolutions that can be expressed as gradient flows.

2.1 Diffusion processes

sec:diffusion

Consider the partial differential equation (also called the Fokker-Planck equation)

$$\partial_t \rho_t = \Delta \rho_t + \operatorname{div}(\rho_t \nabla V) \tag{2.1}$$

eq:pde

where $t \mapsto \rho_t$ is a curve in the space of probability measures $\mathcal{P}(\mathbb{R}^d)$ and $V\colon \mathbb{R}^d \to \mathbb{R}$ is a given (smooth) potential. In line with our previous discussions, we would like to consider gradient structures in the state space $\mathfrak{X} = \mathcal{P}(\mathbb{R}^d)$, and therewith a variational formulation for (2.1). It is not surprising that there may be many gradient structures to (2.1).

We are interested in a family of gradient structures originating from statistical mechanics. These structures are obtained via an upscaling process from the microscopic to the macroscopic level. In the situation of (2.1), we would like a structure that is derived from its *stochastic* counterpart given by the stochastic differential equation

$$dX_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \qquad X_0 \sim \rho_0, \tag{2.2}$$

eq:SDE

where $\rho_t = \operatorname{Law}(X_t)$ satisfies (2.1). Under certain Lipschitz and growth conditions on V, we know that (2.2) has a stationary measure given by $\pi \propto \exp(-V)\mathsf{Leb}_d$, where Leb_d is the Lebesgue measure on $\mathbb{R}^d$. Moreover, large-deviation theory tells us that the sequence of empirical laws

$$\rho_0^n := \frac{1}{n}\sum_{i=1}^n \delta_{X_0^i} \in \mathcal{P}(\mathbb{R}^d), \qquad X_0^i \sim \pi,$$

satisfies a large deviation principle with the *rate function* (Sanov's theorem) [8]

$$\mathcal{P}(\mathbb{R}^d) \ni \rho \mapsto \mathcal{F}(\rho) = \mathsf{Ent}(\rho|\pi) := \begin{cases} \displaystyle\int \log\left(\frac{\mathrm{d}\rho}{\mathrm{d}\pi}\right)\mathrm{d}\rho & \text{for } \rho \ll \pi, \\ +\infty & \text{otherwise,} \end{cases}$$

with Fréchet derivative $\mathrm{D}\mathcal{F}(\rho) = \log(\mathrm{d}\rho/\mathrm{d}\pi)$. In particular, large-deviation theory postulates a driving energy for the gradient structure that we are searching for, and we are now left to specify a (dual) dissipation potential. However, before we do that, let us

$\mathcal{P}(\mathbb{R}^d)$ as a nonlinear geometry The space of probability measures $\mathcal{P}(\mathbb{R}^d)$ can be considered as a subset of the space of finite measures $(\mathcal{M}^+(\mathbb{R}^d), \|\cdot\|_{\mathrm{TV}})$, where $\|\cdot\|_{\mathrm{TV}}$ is the total variation norm. Notice that since $\|\rho\|_{\mathrm{TV}} = 1$ for all $\rho \in \mathcal{P}(\mathbb{R}^d)$, it is not a convex subset of $\mathcal{M}^+(\mathbb{R}^d)$ but, still a submanifold with a differentiable structure as we show.

We begin by specifying the meaning of $\partial_t \rho_t$ for a curve $t \mapsto \rho_t \in \mathcal{P}(\mathbb{R}^d)$. For this purpose, we consider a point $\bar\rho \in \mathcal{P}(\mathbb{R}^d)$ and a smooth vector field $\xi \in \mathcal{C}_c^\infty([0,+\infty) \times \mathbb{R}^d; \mathbb{R}^d)$. Further, let $X\colon [0,+\infty)\times\mathbb{R}^d \to \mathbb{R}^d$ be a solution map to

$$\dot X(t,x) = \xi(t, X(t,x)), \qquad X(0,x) = x \in \mathbb{R}^d.$$

Defining the curve $t \mapsto \rho_t := (X(t,\cdot))_{\#}\bar{\rho} \in \mathcal{P}(\mathbb{R}^d)$, we see that for every $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^d)$,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int \varphi(x)\,\rho_t(\mathrm{d}x) &= \frac{\mathrm{d}}{\mathrm{d}t}\int \varphi(X(t,y))\,\bar{\rho}(\mathrm{d}y)\\ &= \int \nabla\varphi(X(t,y))\cdot\xi(t,X(t,y))\,\bar{\rho}(\mathrm{d}y) = \int \nabla\varphi(x)\cdot\xi(t,x)\,\rho_t(\mathrm{d}x).\end{aligned}$$

Integrating over arbitrary time intervals $[s,t] \subset [0,+\infty)$, we obtain

$$\int \varphi(x)\,\rho_t(\mathrm{d}x) - \int \varphi(x)\,\rho_s(\mathrm{d}x) = \int_s^t \int \nabla\varphi(x)\cdot\xi(r,x)\,\rho_r(\mathrm{d}x)\,\mathrm{d}r,$$

which is a weak formulation of the equation

$$\partial_t\rho_t + \operatorname{div}\big(\rho_t\,\xi(t,\cdot)\big) = 0, \qquad \rho_0 = \bar{\rho},$$

i.e., $t \mapsto \rho_t$ satisfies the continuity equation with the vector field $t \mapsto \xi(t,\cdot)$. For this reason, we can formally identify elements 'smooth' elements of $T_\rho\mathcal{P}(\mathbb{R}^d)$ with vector fields in $\mathcal{C}_c^\infty(\mathbb{R}^d;\mathbb{R}^d)$. A more precise description of this construction can be found in [2].

Now let $\mathcal{F}\colon \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ be a function of the form

$$\mathcal{F}(\rho) = \int \phi\left(\frac{\mathrm{d}\rho}{\mathrm{d}\mathsf{Leb}_d}\right)\mathrm{d}\mathsf{Leb}_d + \int V\,\mathrm{d}\rho = \text{internal energy} + \text{external energy},$$

where $\phi\colon [0,\infty) \to [0,+\infty)$ and $V\colon \mathbb{R}^d \to \mathbb{R}$ are smooth maps. Taking the curve $t \mapsto \rho_t$ above and using the change-of-variables formula

$$\frac{\mathrm{d}\rho}{\mathrm{d}\mathsf{Leb}_d}(x) = \frac{1}{\det(DX(t,X^{-1}(t,x)))}\frac{\mathrm{d}\bar{\rho}}{\mathrm{d}\mathsf{Leb}_d}(X^{-1}(t,x)), \qquad x \in \mathbb{R}^d,$$

we obtain, after sending $t \to 0$,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\rho_t)\Big|_{t=0} &= \int \left[\nabla\mathsf{p}_\phi\left(\frac{\mathrm{d}\bar{\rho}}{\mathrm{d}\mathsf{Leb}_d}(x)\right) + \frac{\mathrm{d}\bar{\rho}}{\mathrm{d}\mathsf{Leb}_d}(x)\nabla V(x)\right]\cdot\xi(0,x)\,\mathsf{Leb}_d(\mathrm{d}x)\\ &= -\int \nabla\left[\phi'\left(\frac{\mathrm{d}\bar{\rho}}{\mathrm{d}\mathsf{Leb}_d}\right) + V\right](x)\,\xi(0,x)\,\bar{\rho}(\mathrm{d}x)\\ &= \int \mathrm{D}\mathcal{F}(\bar{\rho})(x)\,\partial_t\rho_t\big|_{t=0}(\mathrm{d}x) = \mathrm{d}\mathcal{F}(\bar{\rho})\big[\partial_t\rho_t\big|_{t=0}\big],\end{aligned}$$

where $\mathsf{p}_\phi(r) := r\phi'(r) - \phi(r)$ is the so-called ϕ-*pressure* function.

Consequently, the steepest descent direction is obtained by choosing

$$\xi(0,\cdot) = \nabla\left[\phi'\left(\frac{\mathrm{d}\bar{\rho}}{\mathrm{d}\mathsf{Leb}_d}\right) + V\right] = \nabla\mathrm{D}\mathcal{F}(\bar{\rho}).$$

Variational formulation To this end, we express (2.1) in the form

$$\partial_t \rho_t + \operatorname{div} j_t = 0, \tag{CE}$$
$$j_t = -(\nabla \rho_t + \rho_t \nabla V) = -\rho_t \nabla \mathrm{D}\mathcal{F}(\rho_t). \tag{KR}$$

Taking the (dual) dissipation potential as

$$\mathcal{P}(\mathbb{R}^d) \times L^2(\mathbb{R}^d, \rho; \mathbb{R}^d) \ni (\rho, \xi) \longmapsto \mathcal{R}^*(\rho, \xi) = \frac{1}{2} \int |\xi(x)|^2 \rho(\mathrm{d}x),$$

we see that the kinetic relation (KR) can be expressed as

$$j_t = \mathrm{D}_\xi \mathcal{R}^*(\rho_t, -\nabla \mathrm{D}\mathcal{F}(\rho_t)) \iff \mathcal{R}(\rho_t, j_t) + \mathcal{R}^*(\rho_t, -\nabla \mathrm{D}\mathcal{F}(\rho_t)) = \langle -\mathrm{D}\mathcal{F}(\rho_t), j_t \rangle,$$

whenever the chain rule holds:

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{F}(\rho_t) = \langle \nabla \mathrm{D}\mathcal{F}(\rho_t), j_t \rangle \qquad \text{for almost every } t > 0. \tag{CR}$$

Following the same line of thought as in the previous chapter, we obtain the gradient structure $(\mathcal{P}(\mathbb{R}^d), \mathcal{F}, \mathcal{R}, \mathcal{R}^*)$ and the energy-dissipation balance

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

where the *density-flux pair* (ρ, j) satisfies the continuity equation (CE).

Remark 2.1 Several remarks are in order based on the observations above.

(i) The dissipation potential $\mathcal{R}\colon \mathcal{P}(\mathbb{R}^d) \times \mathcal{M}(\mathbb{R}^d; \mathbb{R}^d) \to [0, +\infty]$ is given by

$$\mathcal{R}(\rho, j) = \begin{cases} \displaystyle\int \left| \frac{\mathrm{d}j}{\mathrm{d}\rho}(x) \right|^2 \rho(\mathrm{d}x) & \text{for } j \ll \rho, \\ +\infty & \text{otherwise.} \end{cases}$$

Together with the continuity equation (CE), we recover the *Benamou-Brenier* representation of the Wasserstein distance, i.e.,

$$\mathbb{W}_2^2(\mu, \nu) = \inf_{(\rho, j)} \left\{ \int_0^1 \mathcal{R}(\rho_t, j_t) \, \mathrm{d}t : (\rho, j) \in (\mathrm{CE}),\ \rho_0 = \mu,\ \rho_1 = \nu \right\}.$$

Formally, one then obtains the gradient flow equation

$$\partial_t \rho_t = -\nabla_{\mathbb{W}_2} \mathcal{F}(\rho_t) = -\mathbb{K}(\rho_t) \mathrm{D}\mathcal{F}(\rho_t),$$

with Onsager operator

$$\mathbb{K}(\rho)\colon T_\rho^* \mathcal{P}(\mathbb{R}^d) \to T_\rho \mathcal{P}(\mathbb{R}^d); \qquad \varphi \mapsto -\operatorname{div}(\rho \nabla \varphi),$$

which was first observed by Otto in [17].

(ii) The energy-dissipation functional $\mathscr{L}$ turns out to be related to the rate function associated with the large deviation of the empirical laws

$$\mathsf{P}^n := \frac{1}{n}\sum_{i=1}^{n} \delta_{X^n}, \qquad X^n \sim \mathsf{R} := \mathrm{Law} X,$$

where X satisfies the SDE (2.2) with initial distribution $X_0 \sim \rho_0$. Indeed, recall that the large-deviation rate function is the relative entropy $\mathsf{Ent}(\cdot|\mathsf{R})$ on path space due to Sanov's theorem. Now let the density-flux pair (ρ, j) satisfy (CE), then [5]

$$\inf_j \mathscr{L}((\rho, j), [s, t]) = \inf_{\mathsf{Q}} \Big\{ \mathsf{Ent}(\mathsf{Q}|\mathsf{R}) \,:\, (X_t)_{\#}\mathsf{Q} = \rho_t \ \text{ for all } t \in [0, T] \Big\}.$$

(iii) Many different equations may be obtained by simply changing the driving energy. Indeed, choosing

$$\mathcal{F}(\rho) = \int U\left(\frac{\mathrm{d}\rho}{\mathrm{d}\mathsf{Leb}_d}\right) \mathrm{d}\mathsf{Leb}_d + \int V \,\mathrm{d}\rho + \frac{1}{2}\int K * \rho \,\mathrm{d}\rho,$$

we obtain a large family of nonlinear aggregation-diffusion equations. However, not all of them are justified by large-deviation theory (e.g., the gradient structure arising from the zero-range process for $U(r) = r^2$ has a different dissipation potential).

(iv) Alternatively, one could consider different dual dissipation potentials, such as

$$\mathcal{R}^*(\rho, \xi) = \frac{1}{2}\int |\xi(x)|^2 \,\vartheta(\rho)(\mathrm{d}x),$$

which leads to *aggregation-diffusion equations with nonlinear mobility* [4, 9] or the Hellinger-Kantorovich (or Wasserstein-Fisher-Rao) dual dissipation potential

$$\mathcal{R}^*(\rho, \varphi) = \frac{1}{2}\int |\nabla\varphi(x)|^2 \rho(\mathrm{d}x) + \frac{1}{2}\int |\varphi(x)|^2 \rho(\mathrm{d}x),$$

together with its associated continuity equation

$$\partial_t \rho_t + \mathrm{div}(\rho_t \nabla\varphi_t) = \rho_t \varphi_t,$$

giving rise to nonlinear *reaction-aggregation-diffusion equations* [10, 11].

(v) The *heat equation* is obtained when considering the internal energy $U(r) = -\log r$ and the dual dissipation potential

$$\mathcal{R}^*(\rho, \xi) = \frac{1}{2}\int (u(x))^2 \,|\xi(x)|^2 \,\mathsf{Leb}_d(\mathrm{d}x), \qquad \rho = u\,\mathsf{Leb}_d.$$

A rigorous gradient flow formulation for the heat equation is still of interest.

2.2 Jump processes

sec:jump

In many situations, one is interested in evolutions on discrete states. Consider an evolution on the set of probability vectors on a finite set Ω with cardinality $\#\Omega = n$, i.e.,

$$\mathcal{P}(\Omega) \coloneqq \left\{ \mathbf{p} : \Omega \to [0,1] \;\middle|\; p_x \geq 0,\ \sum\nolimits_{x\in\Omega} p_x = 1 \right\},$$

which will be our state space $\mathfrak{X}$. Notice that $\mathcal{P}(\Omega)$ may be identified with a subset of $\mathbb{R}^n$.

Here, one is interested in the evolution of the probability vector of the form

$$\dot{\mathbf{p}}(t) = Q^\top \mathbf{p}(t) \quad \text{for } t > 0, \qquad \mathbf{p}(0) = \bar{\mathbf{p}} \in \mathcal{P}(\Omega), \tag{2.3}$$

eq:gf-probability

where the *infinitesimal generator* (or transition-rate matrix) $Q = (q_{xy})$ satisfies

$$\sum\nolimits_{y\in\Omega} q_{xy} = 0 \quad \text{for } x \in \Omega, \qquad q_{xy} \geq 0 \ \text{ for } x \neq y,$$

ensuring that $\mathbf{p}(t) \in \mathcal{P}(\Omega)$ for all $t > 0$, and generates a random walk $(X_t)_{t\geq 0}$ on Ω. We further denote by $\mathsf{T}_t = \exp(t\,Q)$ the associated semigroup.

Is the solution (2.3) a gradient flow? If it is, what is it a gradient flow of?

The answer to the first question happens to be positive if the generator Q satisfies the so-called *detailed balance condition*

$$\boldsymbol{\vartheta}_{xy} \coloneqq q_{xy}\pi_x = q_{yx}\pi_y = \boldsymbol{\vartheta}_{yx} \tag{π-DB}$$

eq:db-condition

w.r.t. some probability vector $\boldsymbol{\pi} \in \mathcal{P}(\Omega)$, i.e., the edge measures $\boldsymbol{\vartheta}$ is symmetric. In this scenario, $\boldsymbol{\pi}$ is a *stationary state* for (2.3) in the sense that $\mathsf{T}_t^\top \boldsymbol{\pi} = \boldsymbol{\pi}$ or $Q^\top \boldsymbol{\pi} = 0$. In the theory of Markov processes, this property is equivalent to

(1) the microscopic time reversibility of the sample paths, i.e., if $\mathsf{R} = \operatorname{Law} X_\bullet$ on Ω with initial distribution $X_0 \sim \boldsymbol{\pi}$ and $Y_t^T = X_{T-t}$, then $\operatorname{Law}(Y_\bullet^T) = \mathsf{R}$ for every $T > 0$.

(2) T_t is self-adjoint w.r.t. the inner product $\langle f, g\rangle_\pi \coloneqq \sum_{x\in\Omega} f_x\, g_x\, \pi_x$, i.e.,

$$\langle \mathsf{T}_t f, g\rangle_\pi = \langle f, \mathsf{T}_t g\rangle_\pi.$$

Indeed, the sufficiency of the detailed balance condition (π-DB) holds since

$$\langle Qf, g\rangle_\pi = \sum_{x,y\in\Omega} \boldsymbol{\vartheta}_{xy} f_y g_x = \sum_{x,y\in\Omega} \boldsymbol{\vartheta}_{yx} f_x g_y = \sum_{x,y\in\Omega} \boldsymbol{\vartheta}_{xy} f_x g_y = \langle f, Qg\rangle_\pi,$$

we obtain $\langle \mathsf{T}_t f, g\rangle_\pi = \langle f, \mathsf{T}_t g\rangle_\pi$ by recursion. As for its necessity, we simply test with the functions $f = \mathbf{1}_{\bar{x}}$ and $g = \mathbf{1}_{\bar{y}}$ for every $\bar{x}, \bar{y} \in \Omega$. In the next section, we will use this formulation for quantum Markov semigroups.

To show that (2.3) is indeed a gradient flow equation for some gradient structure, we introduce a continuity equation

$$\dot{\mathbf{p}} + \overline{\mathsf{div}}\,\mathbf{j} = 0,$$

where $\overline{\mathsf{div}}$ is the discrete divergence operator given by

$$\overline{\mathsf{div}} : \mathcal{M}(\Omega \times \Omega) \to \mathcal{M}(\Omega); \qquad \overline{\mathsf{div}}\, \mathbf{j}(x) = \sum_{y \in \Omega} \big(\mathbf{j}_{xy} - \mathbf{j}_{yx}\big),$$

which is the dual to the discrete gradient defined by

$$\overline{\nabla} : B(\Omega) \to B(\Omega \times \Omega); \qquad \overline{\nabla}\varphi(x,y) = \varphi_y - \varphi_x.$$

For any $\varphi \in B(\Omega)$, we then find that

$$\begin{aligned}\varphi \cdot \dot{\mathbf{p}}(t) &= \sum_{x,y} q_{xy}\varphi_y \mathbf{p}_x = \sum_{x,y} \overline{\nabla}\varphi(x,y)\, q_{xy}\mathbf{p}_x = \frac{1}{2}\sum_{x,y} \overline{\nabla}\varphi(x,y)\big(q_{xy}\mathbf{p}_x - q_{yx}\mathbf{p}_y\big) \\ &= \sum_{x,y} \overline{\nabla}\varphi(x,y)\, \mathbf{j}_{xy} = -\varphi \cdot \overline{\mathsf{div}}\, \mathbf{j}, \qquad \text{with } \mathbf{j}_{xy} \coloneqq \frac{1}{2}\big(q_{xy}\mathbf{p}_x - q_{yx}\mathbf{p}_y\big).\end{aligned}$$

Assuming for the moment that $\mathbf{p} \ll \boldsymbol{\pi}$ with $\mathbf{p}_x = u_x \boldsymbol{\pi}_x$, we further obtain

$$\mathbf{j}_{xy} = \frac{1}{2}(u_x q_{xy}\boldsymbol{\pi}_x - u_y q_{yx}\boldsymbol{\pi}_y) = \frac{1}{2}(u_x - u_y) q_{xy}\boldsymbol{\pi}_x = -\frac{1}{2}\overline{\nabla} u(x,y)\, \boldsymbol{\vartheta}_{xy},$$

where the detailed balance condition (π-DB) was used. Consequently, (2.3) reads

$$\dot{\mathbf{p}} + \overline{\mathsf{div}}\, \mathbf{j} = 0, \tag{CE}$$

$$\mathbf{j}_{xy} = -\frac{1}{2}\overline{\nabla} u(x,y)\, \boldsymbol{\vartheta}_{xy}. \tag{KR}$$

As in the diffusion case in Section 2.1, a driving energy for the jump process may be derived via Sanov's theorem, yielding

$$\mathcal{F}(\mathbf{p}) = \mathsf{Ent}(\mathbf{p}|\boldsymbol{\pi}) \coloneqq \begin{cases} \displaystyle\sum_{x \in \Omega} \mathbf{p}_x \log \frac{\mathbf{p}_x}{\boldsymbol{\pi}_x} & \text{for } \mathbf{p} \ll \boldsymbol{\pi}, \\ +\infty & \text{otherwise.} \end{cases}$$

with Fréchet derivative $\mathrm{D}\mathcal{F}(\mathbf{p})(x) = \log(\mathbf{p}_x/\boldsymbol{\pi}_x) = \log u_x$.

However, in contrast to diffusion processes, a spatial gradient is absent, and we have to work with the discrete gradient $\overline{\nabla}$ instead. This situation allows for a lot of freedom in selecting a dual dissipation potential $\mathcal{R}^*$ for which (KR) can be expressed as $\mathbf{j} = D_\xi \mathcal{R}^*(\mathbf{p}, -\overline{\nabla}\mathrm{D}\mathcal{F}(\mathbf{p}))$. For simplicity, we consider dual dissipation potentials of the form

$$\mathcal{P}(\Omega) \times B(\Omega \times \Omega) \ni (\mathbf{p}, \xi) \longmapsto \mathcal{R}^*(\mathbf{p}, \xi) = \frac{1}{2}\sum_{x,y} \Psi^*(\xi_{xy})\, \mathfrak{m}(q_{xy}\mathbf{p}_x, q_{yx}\mathbf{p}_y),$$

where $\Psi^* \colon \mathbb{R} \to [0, +\infty)$ is symmetric, convex, and superlinear, and $\mathfrak{m}$ is a mean function. The corresponding dissipation potential is given by

$$\mathcal{R}(\mathbf{p}, \mathbf{j}) = \begin{cases} \displaystyle\frac{1}{2}\sum_{x,y} \Psi\left(\frac{2\mathbf{j}_{xy}}{\mathfrak{m}(q_{xy}\mathbf{p}_x, q_{yx}\mathbf{p}_y)}\right) \mathfrak{m}(q_{xy}\mathbf{p}_x, q_{yx}\mathbf{p}_y) & \text{for } \mathbf{j} \ll \boldsymbol{\sigma}, \\ +\infty & \text{otherwise,} \end{cases}$$

where (Ψ, Ψ^*) is a conjugate pair.

Remark 2.2 If $\mathbf{p} = u\boldsymbol{\pi}$, then the 1-homogeneity of the mean function $\mathfrak{m}$ yields

$$\mathcal{R}^*(\mathbf{p}, \xi) = \frac{1}{2} \sum_{x,y} \Psi^*(\xi_{xy})\, \mathfrak{m}(u_x, u_y)\, \boldsymbol{\vartheta}_{xy}.$$

Example 2.3 (a) Maas-Mielke-Chung et.al. [12, 14] chose a quadratic dual dissipation potential with $\Psi_2^*(r) = r^2/2$ and $\mathfrak{m}$ being the logarithmic mean, i.e., $\mathfrak{m}_{\log}(u, v) = \frac{u-v}{\log u - \log v}$, which gives

$$\begin{aligned} \mathrm{D}_\xi \mathcal{R}_2^*(-\overline{\nabla}\mathrm{D}\mathcal{F}(\mathbf{p}))(x,y) &= -\frac{1}{2}\overline{\nabla}\mathrm{D}\mathcal{F}(\mathbf{p})(x,y)\frac{u_x - u_y}{\log u_x - \log u_y}\,\boldsymbol{\vartheta}_{xy} \\ &= -\frac{1}{2}\overline{\nabla}u(x,y)\,\boldsymbol{\vartheta}_{xy} = \mathbf{j}_{xy}. \end{aligned}$$

The corresponding dissipation potential is

$$\mathcal{R}_2(\mathbf{p}, \mathbf{j}) = \frac{1}{2} \sum_{x,y} \left| \frac{\mathbf{j}_{xy}}{\mathfrak{m}_{\log}(q_{xy}\mathbf{p}_x, q_{yx}\mathbf{p}_y)} \right|^2 \mathfrak{m}_{\log}(q_{xy}\mathbf{p}_x, q_{yx}\mathbf{p}_y).$$

Together with the continuity equation (CE), one can define a nonlocal 2-Wasserstein distance for probability measures on Ω via the Benamou-Brenier representation

$$\mathfrak{W}_2^2(\mu, \nu) = \inf_{(\mathbf{p},\mathbf{j})} \left\{ \int_0^1 \mathcal{R}_2(\mathbf{p}(t), \mathbf{j}(t))\, \mathrm{d}t \,:\, (\mathbf{p}, \mathbf{j}) \in (\mathrm{CE}),\ \mathbf{p}(0) = \mu,\ \mathbf{p}(1) = \nu \right\}.$$

(b) Inspired by large deviations theory, [15] considers the dual dissipation potential $\Psi_{\cosh}^*(r) = 4(\cosh(r/2) - 1)$ with the geometric mean $\mathfrak{m}_{\mathrm{geo}}(u, v) = \sqrt{uv}$, giving

$$\begin{aligned} \mathrm{D}_\xi \mathcal{R}_{\cosh}^*(-\overline{\nabla}\mathrm{D}\mathcal{F}(\mathbf{p}))(x,y) &= \sinh(-\overline{\nabla}\mathrm{D}\mathcal{F}(\mathbf{p})(x,y)/2)\sqrt{u_x u_y}\,\boldsymbol{\vartheta}_{xy} \\ &= \frac{1}{2}\left(\sqrt{\frac{u_x}{u_y}} - \sqrt{\frac{u_y}{u_x}}\right)\sqrt{u_x u_y}\,\boldsymbol{\vartheta}_{xy} \\ &= -\frac{1}{2}\overline{\nabla}u(x,y)\boldsymbol{\vartheta}_{xy} = \mathbf{j}_{xy}. \end{aligned}$$

This is the so-called *cosh* structure and has a lot of nice properties (cf. [**?**]). Unfortunately, $\Psi_{\cosh}^*$ does not give rise to a distance $\mathcal{P}(\Omega)$. Nevertheless, there is a length structure associated with $\mathcal{R}_{\cosh}$, which one can use to rigorously study generalized gradient flows using a time-incremental approach [18].

The examples above show that different gradient structures may exist for jump processes even with the same driving energy. It is also interesting to notice the form of their associated Fisher information. For the quadratic dual dissipation potential, one has

$$\mathcal{R}_2^*(\rho, -\overline{\nabla}\mathrm{D}\mathcal{F}(\rho)) = \frac{1}{4}\sum_{x,y}(\overline{\nabla}\log u(x,y))(\overline{\nabla}u(x,y))\,\boldsymbol{\vartheta}_{xy} =: \mathcal{I}_2(\rho),$$

which resembles the continuous Fisher information

$$\frac{1}{4}\int |\nabla \log u(x)|^2\, u(x)\, \mathsf{Leb}_d(\mathrm{d}x).$$

On the other hand, the cosh dual dissipation potential gives

$$\mathcal{R}^*_{\cosh}(\rho, -\overline{\nabla}\mathrm{D}\mathcal{F}(\rho)) = \sum_{x,y} |\overline{\nabla}\sqrt{u}(x,y)|^2\, \boldsymbol{\vartheta}_{xy} =: \mathcal{I}_{\cosh}(\rho),$$

clearly resembling its continuous counterpart

$$\int |\nabla\sqrt{u}(x)|^2\, \mathsf{Leb}_d(\mathrm{d}x).$$

In fact, there is an explicit link between $\mathcal{I}_2$ and $\mathcal{I}_{\cosh}$. Indeed, simple computations yield

$$\begin{aligned}
\frac{1}{4}(\log v - \log u)(v-u) &= \frac{1}{2}(\sqrt{u}+\sqrt{v})(\log\sqrt{v} - \log\sqrt{u})(\sqrt{v}-\sqrt{u}) \\
&= \frac{1}{2}(\sqrt{v}+\sqrt{u})\int_0^1 \frac{1}{(1-\lambda)\sqrt{u}+\lambda\sqrt{v}}\,\mathrm{d}\lambda\, |\sqrt{v}-\sqrt{u}|^2 \\
&= \frac{\mathfrak{m}_{\mathrm{arith}}(\sqrt{v},\sqrt{u})}{\mathfrak{m}_{\log}(\sqrt{v},\sqrt{u})}|\sqrt{v}-\sqrt{u}|^2 \geq |\sqrt{v}-\sqrt{u}|^2,
\end{aligned}$$

where the last inequality follows from the property $\mathfrak{m}_{\log} \leq \mathfrak{m}_{\mathrm{arith}}$. Hence, $\mathcal{I}_2 \geq \mathcal{I}_{\cosh}$.

Summarizing, we obtain, as before, a variational formulation for (2.3) in the form

$$\mathscr{L}((\mathbf{p},\mathbf{j}),[s,t]) \coloneqq \int_s^t \mathcal{R}(\mathbf{p}(r),\mathbf{j}(r)) + \mathcal{R}^*(\mathbf{p}(r), -\overline{\nabla}\mathrm{D}\mathcal{F}(\mathbf{p}(r)))\,\mathrm{d}r + \mathcal{F}(\mathbf{p}(t)) - \mathcal{F}(\mathbf{p}(s)) = 0,$$

where the density-flux pair $(\mathbf{p},\mathbf{j})$ satisfies the continuity equation (CE).

Remark 2.4 While the zero-cost flow $\mathbf{p}$ may be the same for every gradient structure, the energy-dissipation functionals differ. Hence, the different gradient structures measure deviations from the zero-cost flow differently.

3 Gradient Structures for Non-commutative States

We have now arrived at the exciting part of this lecture series. As the title suggests, the state space $\mathfrak{X}$ we will consider contains elements that do not commute. The archetype for $\mathfrak{X}$ is the space of matrices $\mathsf{Mat}(\mathbb{F}, d)$ (most of the time, a subset thereof), where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.

In the following, we will discuss two such evolutions, one of which is classical and stems from a *projection* onto a special class of probability measures, while the other is an evolution of states for an open quantum system.

3.1 Semipositive definite matrices

Our first example comes from 'projecting' the Fokker-Planck equation (2.1) onto the class of Gaussian measures

$$\mathfrak{G} \coloneqq \left\{ G(\Sigma, m)\mathsf{Leb}_d \in \mathcal{P}(\mathbb{R}^d) : \Sigma \in \mathcal{S}(\mathbb{R}, d),\ m \in \mathbb{R}^d \right\} \subset \mathcal{P}(\mathbb{R}^d),$$

where $\Sigma \in \mathcal{S}(\mathbb{R}, d) \coloneqq \{A \in \mathsf{Mat}(\mathbb{R}, d) : A^\top = A,\ A \geq 0\}$ is the set of symmetric and positive definite matrices, and

$$G(\Sigma, m)(x) = \frac{1}{\sqrt{(2\sigma)^d \det(\Sigma)}} \exp\left(-\frac{1}{2} \langle x - m, \Sigma^{-1}(x - m) \rangle \right) \tag{3.1}$$

eq:gaussian-map

is the Gaussian density for a given covariance $\Sigma \in \mathcal{S}(\mathbb{R}, d)$ and mean $m \in \mathbb{R}^d$. We also denote $\mathcal{S}^+(\mathbb{R}, d) = \{A \in \mathcal{S}(\mathbb{R}, d) : A > 0\}$ and $\mathcal{G}^+$ accordingly. It is easy to see that $\mathcal{G}^+$ is a smooth finite-dimensional submanifold of dimension $(d+2)(d+1)/2 - 1$.

In the following, we suppose that V takes the form $V(x) = \frac{1}{2}\langle x - n, \Gamma(x - n)\rangle$. Then, the associated stationary measure is given by the Gaussian measure $\sigma = G(\Gamma, n)\mathsf{Leb}_d$.

Using the ansatz $\rho = G(\Sigma, m)$ in the Fokker-Plank equation (2.1) we obtain the system

$$\dot{\Sigma} = -\big(\Gamma^{-1}\Sigma + \Sigma\Gamma^{-1} - 2I\big), \tag{3.2}$$

eq:covariance

$$\dot{m} = -\Gamma^{-1}(m - n). \tag{3.3}$$

Since the system is decoupled and we are only interested in the non-commutative part, we will focus on the equation for the covariance Σ.

In this simple scenario, the solution of (3.2) is given explicitly by

$$\Sigma(t) = \Gamma + e^{-t\Gamma^{-1}}\big(\Sigma(0) - \Gamma\big)e^{-t\Gamma^{-1}}.$$

However, we are interested in whether (3.2) can be seen as a gradient flow. More precisely, we are looking for an energy $\mathcal{E}$ and a family of Onsager operators $\mathbb{K}(\Sigma)$ such that

$$\dot{\Sigma} = -\mathbb{K}(\Sigma)\mathrm{D}\mathcal{F}(\Sigma).$$

It turns out that, since the system for (Σ, m) comes from a 'projection', one can consider a gradient structure induced by the gradient structure for the Fokker-Planck equation (2.1). This statement is made precise by the following proposition from [13].

prop:restriction

Proposition 3.1. *Let a gradient system $(\mathfrak{X}, \mathcal{F}, \mathbb{K})$ be given in terms of a family Onsager operators $\mathbb{K}(\rho) : T_\rho^*\mathfrak{X} \to T_\rho\mathfrak{X}$. Let $j : \mathfrak{Z} \to \mathfrak{X}$ be an embedding such that $j(\mathfrak{Z}) \subset \mathfrak{X}$ is a smooth submanifold that is invariant under the gradient flow evolution $\dot\rho = -\mathbb{K}(\rho)\mathrm{D}\mathcal{F}(\rho)$.*

Then, the restricted flow on $\mathfrak{Z}$ is again a gradient system given by the reduced gradient system $(\mathfrak{Z}, \mathcal{F}|_{\mathfrak{Z}}, \mathbb{K}|_{\mathfrak{Z}})$, where the reduced Onsager operator is given by

$$\forall\, w \in \mathfrak{Z},\ \forall\, \eta \in T_w^*\mathfrak{Z} : \quad \big\langle \eta, \mathbb{K}|_{\mathfrak{Z}}(w)\eta\big\rangle_{T_w\mathfrak{Z}} = \inf\Big\{ \big\langle \xi, \mathbb{K}(j(w))\xi\big\rangle_{T_{j(w)}\mathfrak{Z}} \ :\ \mathrm{D}j(w)^*\xi = \eta \Big\},$$

Let us consider the gradient structure for the Fokker-Planck equation (2.1) with

$$\mathfrak{X} = \mathcal{P}(\mathbb{R}^d), \qquad \mathcal{F} = \mathsf{Ent}(\cdot|G(\mathit{\Gamma}, n)), \qquad \mathbb{K}(\rho)[\varphi] = -\mathrm{div}(\rho\nabla\varphi).$$

It turns out that $j(\mathfrak{Z})$ with $\mathfrak{Z} = \mathcal{S}^+ \times \mathbb{R}^d$ and the embedding $j = G\colon \mathfrak{Z} \to \mathcal{P}(\mathbb{R}^d)$ is invariant under the evolution. Hence, Proposition 3.1 applies, and after some simple but tedious computations, we obtain the induced energy and Onsager operator given by

$$\mathcal{F}|_{\mathfrak{Z}}(\Sigma, m) = \frac{1}{2}\langle m - n, \mathit{\Gamma}^{-1}(m-n)\rangle + \underbrace{\frac{1}{2}\operatorname{tr}[\mathit{\Gamma}^{-1}\Sigma - I] - \frac{1}{2}\log\det(\mathit{\Gamma}^{-1}\Sigma)}_{=:\mathcal{E}(\Sigma)},$$

$$\mathbb{K}|_{\mathfrak{Z}}(\Sigma, m)[\Lambda] = \begin{bmatrix} 2\big(\Lambda_\Sigma\Sigma + \Sigma\Lambda_\Sigma\big) & 0 & 0 \\ 0 & \Lambda_m & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since we are only interested in the evolution for Σ, with an abuse of notation, we set

$$\mathbb{K}(\Sigma)[\Lambda] \coloneqq 2\big(\Lambda\Sigma + \Sigma\Lambda\big), \qquad \Lambda \in \mathsf{Mat}(\mathbb{R}, d).$$

We then check that $\mathrm{D}\mathcal{E}(\Sigma) = (\mathit{\Gamma}^{-1} - \Sigma^{-1})/2$, which yields

$$-\mathbb{K}(\Sigma)[\mathrm{D}\mathcal{E}(\Sigma)] = (\mathit{\Gamma}^{-1} - \Sigma^{-1})\Sigma + \Sigma(\mathit{\Gamma}^{-1} - \Sigma^{-1}) = \mathit{\Gamma}^{-1}\Sigma + \Sigma\mathit{\Gamma}^{-1} - 2I = \dot\Sigma.$$

Remark 3.2 (1) The dual dissipation potential $\mathcal{R}$ associated to $\mathbb{K}$ is given by

$$\mathcal{R}^*(\Sigma, \Lambda) = 2\operatorname{tr}[\Lambda\Sigma\,\Lambda], \qquad \mathrm{D}_\Lambda\mathcal{R}(\Sigma, \Lambda) = \mathbb{K}(\Sigma)[\Lambda].$$

(2) The dynamical transport cost associated with this Onsager $\mathbb{K}$ operator is given by

$$\begin{aligned}\mathcal{D}^2(\Sigma_0, \Sigma_1) &\coloneqq \inf_{(\Sigma,\Lambda)} \left\{ \int_0^1 \mathcal{R}^*(\Sigma(r), \Lambda(t))\,\mathrm{d}t \ :\ \dot\Sigma = \mathbb{K}(\Sigma)[\Lambda],\ \Sigma(0) = \Sigma_0,\ \Sigma(1) = \Sigma_1 \right\} \\ &= \inf_{\Sigma} \left\{ \int_0^1 \mathcal{R}(\Sigma(r), \dot\Sigma(r))\,\mathrm{d}r \ :\ \Sigma(0) = \Sigma_0,\ \Sigma(1) = \Sigma_1 \right\},\end{aligned}$$

which coincides with the distance

$$\mathcal{B}^2(\Sigma_0, \Sigma_1) = \operatorname{tr}\Big[\Sigma_0 + \Sigma_1 - 2(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2})^{1/2}\Big],$$

which is the *Bures distance* in quantum information theory or the *Bures-Wasserstein* in the statistical learning community). Whenever Σ_0 and Σ_1 commute, then $\mathcal{B}$ reduces to the Hellinger distance.

To conclude, we deduced that the evolution for Σ (3.2) is a gradient flow equation with the quadratic gradient structure $(\mathfrak{Z}, \mathbb{K}, \mathcal{E})$, from which a variational formulation follows.

3.2 Density matrices

Studying open quantum systems can be quite challenging. Fortunately, approximations can be performed on these systems to derive equations describing the evolution of quantum states, which for finite-dimensional systems can be identified with *density matrices*

$$\rho \in \mathcal{S}(\mathbb{C}, n) \coloneqq \Big\{ \nu \in \mathsf{Mat}(\mathbb{C}, n) : \nu \geq 0,\ \nu = \nu^\dagger,\ \mathrm{tr}[\nu] = 1 \Big\}.$$

In the following, we set $\mathcal{S}^+(\mathbb{C}, n) = \{\nu \in \mathcal{S}(\mathbb{C}, n) : \nu > 0\}$.

Here, we follow the construction of gradient structures for quantum evolutions by [3]. Similar works have been done in [6, 7, 16].

We begin our discussion by defining a *quantum Markov semigroup*.

Definition 3.3 (Quantum Markov semigroup) A quantum Markov semigroup on $\mathsf{Mat}(\mathbb{C}, n)$ is a $\mathcal{C}_0$-semigroup of operators $(\mathsf{T}_t)_{t\geq 0}$ that are

(i) *unital*: $\mathsf{T}_t(\mathrm{I}) = \mathrm{I}$, and

(ii) *completely positive*: For any $k \geq 0$, $\mathsf{T}_t \otimes \mathsf{id}_{\mathsf{Mat}(\mathbb{C},k)}$ is positivity preserving, i.e.,

$$\mathsf{T}_t \otimes \mathsf{id}_{\mathsf{Mat}(\mathbb{C},k)}(A) \geq 0 \quad \text{for every } A \geq 0.$$

In particular, $(\mathsf{T}_t A)^\dagger = \mathsf{T}_t A^\dagger$ for all $A \in \mathsf{Mat}(\mathbb{C}, n)$, i.e., T_t is *real* for all $t \geq 0$.

Its adjoint T_t^* with respect to the Hilbert-Schmidt inner product $\langle A, B\rangle = \mathrm{tr}[A^\dagger B]$ is completely positive and trace preserving, i.e., $\mathrm{tr}[\mathsf{T}_t^*(A)] = \mathrm{tr}[A]$ for all $A \in \mathsf{Mat}(\mathbb{C}, n)$. In other words, T_t^* is a quantum channel for every $t \geq 0$.

As in the classical case, we would like to end up with a semigroup T_t for which the curve $t \mapsto \rho_t \coloneqq \mathsf{T}_t^* \rho_0$ is a gradient flow. A sufficient condition for this is a quantum version of the detailed balance condition. Interestingly, there are several non-equivalent notions of detailed balance in the context of quantum evolutions, as described below.

Detailed balance Recall that the detailed balance condition (π-DB) in the classical case is equivalent to the corresponding semigroup being self-adjoint w.r.t. the π-weighted inner product $\langle \cdot, \cdot \rangle_\pi$. In the quantum case, there are many candidates for such a weighted inner product. For example, given $\sigma \in \mathcal{S}^+(\mathbb{C}, n)$ and $s \in [0, 1]$, we may define

$$\langle A, B\rangle_s \coloneqq \mathrm{tr}[A^\dagger \sigma^s B \sigma^{1-s}], \qquad A, B \in \mathsf{Mat}(\mathbb{C}, n).$$

By the cyclicity property of the trace, we find that $\langle A, A\rangle_s = \mathrm{tr}[|\sigma^{s/2} A \sigma^{(1-s)/2}|^2] \geq 0$. Therefore, $\langle \cdot, \cdot \rangle_s$ is a positive definite sesquilinear form.

Example 3.4 Given $\sigma \in \mathcal{S}^+(\mathbb{C}, n)$ and $s \in [0, 1]$, we may define

(a) the s-inner products

$$\langle A, B\rangle_s \coloneqq \mathrm{tr}[A^\dagger \sigma^s B \sigma^{1-s}], \qquad A, B \in \mathsf{Mat}(\mathbb{C}, n).$$

The inner products with $s = 0$ and $s = 1/2$ have their own names. Namely, the Gelfand-Naimark-Segal (GNS) inner product $\langle \cdot, \cdot \rangle_{GNS}$ and Kubo-Martin-Schwinger (KMS) inner product $\langle \cdot, \cdot \rangle_{KMS}$, respectively.

(b) Another inner product is the Bogoliubov-Kubo-Mori inner product given by

$$\langle A, B\rangle_{BKM} \coloneqq \int_0^1 \langle A, B\rangle_s \, \mathrm{d}s.$$

This inner product plays a special role in our current discussion.

Fortunately, the following interesting result simplifies things.

lem:self-adjoint

Lemma 3.5. *Let* T *be self-adjoint w.r.t. the* $\langle\cdot,\cdot\rangle_s$ *inner product for some* $s \in [0,1]\backslash\{\frac{1}{2}\}$. *Then,* T *commutes with the* σ*-modular operator*

$$\Delta_\sigma A = \sigma A \sigma^{-1}, \qquad A \in \mathsf{Mat}(\mathbb{C}, n),$$

and is self-adjoint w.r.t. $\langle\cdot,\cdot\rangle_s$ *for all* $s \in [0,1]$, *including* $s = \frac{1}{2}$.
In particular, T *is also self-adjoint w.r.t.* $\langle\cdot,\cdot\rangle_{BKM}$.

Remark 3.6 (i) Examples exists for which T is self-adjoint w.r.t. $\langle\cdot,\cdot\rangle_{KMS}$ or $\langle\cdot,\cdot\rangle_{BKM}$ but not w.r.t. $\langle\cdot,\cdot\rangle_{GNS}$ [**?**].

(ii) Notice that the s-inner product may be expressed as

$$\langle A, B\rangle_s = \mathrm{tr}[A^\dagger \sigma^s B \sigma^{-s} \sigma] = \mathrm{tr}[A^\dagger \Delta_\sigma^s B \sigma].$$

Therefore, one can consider for any $f\colon (0,+\infty) \to (0,+\infty)$ the f-inner product

$$\langle A, B\rangle_f \coloneqq \mathrm{tr}[A^\dagger f(\Delta_\sigma) B \sigma].$$

In this notation, the BKM-inner product is given by $f(r) = \int_0^1 r^s \, \mathrm{d}s$.

In view of Lemma 3.5, one can define the quantum detailed balance as follows:

Definition 3.7 Let $\sigma \in \mathcal{S}(\mathbb{C}, n)$. A quantum Markov semigroup $(\mathsf{T}_t)_{t\geq 0}$ is said to satisy the *quantum detailed balance condition* w.r.t. σ if for every $t > 0$,

$$\langle \mathsf{T}_t A, B\rangle_{GNS} = \langle A, \mathsf{T}_t B\rangle_{GNS} \qquad \forall A, B \in \mathsf{Mat}(\mathbb{C}, n). \qquad (\sigma\text{-DBC})$$

eq:db-condition-quantum

In this case, we say that $(\mathsf{T}_t)_{t\geq 0}$ satisfies (σ-DBC).

Remark 3.8 Let $(\mathsf{T}_t)_{t\geq 0}$ be a quantum Markov semigroup satisfying (σ-DBC). Using the fact that T_t is unital for all $t > 0$, we have that

$$\begin{aligned}\langle \mathsf{T}_t^* \sigma, A\rangle &= \langle \sigma, \mathsf{T}_t A\rangle = \mathrm{tr}[\sigma \mathsf{T}_t A] = \mathrm{tr}[\sigma^{1-s}\,\mathrm{I}\,\sigma^s \mathsf{T}_t A] = \langle \mathrm{I}, \mathsf{T}_t A\rangle_s \\ &= \langle \mathsf{T}_t(\mathrm{I}), A\rangle_s = \langle \mathrm{I}, A\rangle_s = \langle \sigma, A\rangle \qquad \text{for all } A \in \mathsf{Mat}(\mathbb{C}, n).\end{aligned}$$

Hence, σ is stationary for the semigroup $(\mathsf{T}_t)_{t\geq 0}$.

For quantum Markov semigroups, we have a nice representation of its generator due to Alicki 1976 as given by the following result.

thm:alicki

Theorem 3.9. *Let* $(\mathsf{T}_t)_{t\geq 0}$ *satisfy* (σ-DBC) *with* $\sigma \in \mathcal{S}^+(\mathbb{C}, n)$. *Then, its generator reads*

$$\mathsf{L} = \sum_{j\in J} e^{-\omega_j/2}\mathsf{L}_j, \qquad \mathsf{L}_j A = V_j^\dagger [A, V_j] + [V_j^\dagger, A]V_j,$$

where J *is a finite index set, the pair* $(\omega_j, V_j) \in \mathbb{R} \times \mathsf{Mat}(\mathbb{C}, n)$ *satisfy*

$$\{V_j\}_{j\in J} = \{V_j^\dagger\}_{j\in J} \qquad \text{and} \qquad \Delta_\sigma V_j = e^{-\omega_j} V_j, \qquad \text{for all } j \in J,$$

where Δ_σ *is the* σ*-modular operator. In other words,* $(e^{-\omega_j}, V_j)_{j\in J}$ *are eigenpairs of* Δ_σ. *Moreover, we have that* $(e^{\omega_j}, V_j^\dagger)_{j\in J}$ *are also eigenpairs.*

In view of Theorem 3.9, the evolution corresponding to $(\mathsf{T}_t)_{t\geq 0}$ is given by

eq:lindblad
$$\partial_t \rho = \mathsf{L}^* \rho, \qquad \rho_0 = \bar{\rho} \in \mathcal{S}(\mathbb{C}, n), \tag{3.4}$$

which is known as the *Lindblad equation* [**?**]. Recall that L^* is the adjoint of the generator L w.r.t. the Hilbert-Schmidt inner product.

As in the previous section, it is unclear how the Lindblad equation (3.4) can be interpreted as a gradient flow equation. In addition to the fact that we are dealing with superoperators—which is not inherently a problem since we can identify elements of $\mathcal{S}(\mathbb{C}, n)$ with vectors in $\mathbb{C}^{n^2}$—we also have the issue of non-commutativity. This makes identifying an appropriate driving energy $\mathcal{F}$, dissipation potential pair $(\mathcal{R}, \mathcal{R}^*)$, and continuity equation for the evolution (3.4) less obvious.

Even in the quantum setting, the relative entropy Ent plays an important role as a driving energy. However, due to noncommutativity, there may be many versions of the relative entropy. For example, there is the *Umegaki* relative entropy

$$\mathsf{Ent}_{\mathrm{U}}(\rho|\sigma) = \mathrm{tr}[\rho(\log \rho - \log \sigma)],$$

and the *Belavkin-Staszewski* relative entropy

$$\mathsf{Ent}_{\mathrm{BS}}(\rho|\sigma) = \mathrm{tr}[\rho \log(\rho^{1/2}\sigma^{-1}\rho^{1/2})] = \mathrm{tr}[\sigma(\sigma^{-1/2}\rho\sigma^{-1/2}) \log(\sigma^{-1/2}\rho\sigma^{-1/2})],$$

where they satisfy $\mathsf{Ent}_{\mathrm{BS}} \geq \mathsf{Ent}_{\mathrm{U}}$ and are equivalent whenever ρ and σ commute as they both result in the classical formula for Ent.

As in Section 2.2, we consider an appropriate continuity equation for the Lindblad equation (3.4). For any $A \in \mathsf{Mat}(\mathbb{C}, n)$, we define the partial derivative operators

$$\partial_j A \coloneqq [V_j, A], \qquad \partial_j^\dagger A = [V_j^\dagger, A], \qquad j \in J.$$

The associated gradient and divergence are given by

$$\overline{\nabla} A = (\partial_1 A, \ldots, \partial_{\#J} A)^\top, \qquad \overline{\mathsf{div}}\, E = -\sum_{j\in J} \partial_j^\dagger E_j.$$

The continuity equation then reads

$$\partial_t \rho + \overline{\mathsf{div}}\, E = 0.$$

It is easy to see that the flux associated with the Lindblad equation is given by

$$E_j = e^{\omega_j/2} \rho V_j - e^{-\omega_j/2} V_j \rho.$$

Indeed, for any $A \in \mathsf{Mat}(\mathbb{C}, n)$, we have that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \operatorname{tr}[\rho A] = \operatorname{tr}[\rho_t \mathsf{L} A] &= \sum_{j \in J} e^{-\omega_j/2} \operatorname{tr}[\rho \mathsf{L}_j A] \\
&= \sum_{j \in J} e^{-\omega_j/2} \operatorname{tr}\Big[\rho\big(V_j^\dagger [A, V_j] + [V_j^\dagger, A] V_j\big)\Big] \\
&= \sum_{j \in J} \operatorname{tr}\Big[\rho\big(e^{-\omega_j/2} V_j^\dagger [A, V_j] + e^{\omega_j/2} [V_j, A] V_j^\dagger\big)\Big] \\
&= \sum_{j \in J} \operatorname{tr}\Big[\big(e^{\omega_j/2} V_j^\dagger \rho - e^{-\omega_j/2} \rho V_j^\dagger\big) \partial_j A\Big] = \sum_{j \in J} \langle E_j, \partial_j A \rangle.
\end{aligned}$$

prop:q-space-chainrule

Proposition 3.10 ([?])**.** *For all $\rho \in \mathcal{S}^+(\mathbb{C}, n)$ and $j \in J$, we have*

$$E_j = -\widehat{\rho}_j \odot \partial_j (\log \rho - \log \sigma),$$

where

$$\widehat{\rho}_j \odot A = \int_0^1 \big(e^{\omega_j/2} \rho\big)^{1-s} A \big(e^{-\omega_j/2} \rho\big)^s \,\mathrm{d}s, \qquad j \in J,\ A \in \mathsf{Mat}(\mathbb{C}, n),$$

is self-adjoint with $(\widehat{\rho}_j \odot \cdot)^* = \widehat{\rho}_j \odot \cdot$ for every $j \in J$.

Due to Proposition 3.10, we deduce that

$$E_j = -\widehat{\rho}_j \odot \partial_j \mathrm{D}\mathcal{F}(\rho),$$

where $\mathcal{F} = \mathsf{Ent}_{\mathrm{U}}(\cdot|\sigma)$ is the Umegaki relative entropy. The Onsager operator is then

$$\mathbb{K}(\rho)[A] \coloneqq \sum_{j \in J} \partial_j^\dagger \big(\widehat{\rho}_j \odot \partial_j A\big), \qquad A \in \mathsf{Mat}(\mathbb{C}, n),$$

with its corresponding dual dissipation potential

$$\mathcal{R}^*(\rho, \Lambda) = \frac{1}{2} \langle \Lambda, \Lambda \rangle_{\widehat{\rho}} \coloneqq \frac{1}{2} \sum_{j \in J} \langle \Lambda_j, \widehat{\rho}_j \odot \Lambda_j \rangle, \qquad \Lambda \in [\mathsf{Mat}(\mathbb{C}, n)]^J.$$

All in all, we can express the Lindblad equation (3.4) in the form

$$\begin{aligned}
\partial_t \rho + \overline{\mathsf{div}}\, E &= 0, \\
E &= \mathrm{D}_\Lambda \mathcal{R}^*(\rho, -\overline{\nabla} \mathrm{D}\mathcal{F}(\rho)),
\end{aligned}$$

which is the form of a gradient flow with gradient structure $(\mathcal{S}(\mathbb{C}, n), \mathcal{F}, \mathcal{R}^*)$.

References

[1] S. Adams et al., From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage, Comm. Math. Phys. **307** (2011), no. 3, 791–815.

[2] 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.

[3] E. A. Carlen and J. Maas, Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance, J. Funct. Anal. **273** (2017), no. 5, 1810–1869.

[4] J. A. Carrillo de la Plata et al., Nonlinear mobility continuity equations and generalized displacement convexity, J. Funct. Anal. **258** (2010), no. 4, 1273–1309.

[5] P. Cattiaux and C. Léonard, Large deviations and Nelson processes, Forum Math. **7** (1995), no. 1, 95–115.

[6] Y. Chen et al., An efficient algorithm for matrix-valued and vector-valued optimal mass transport, J. Sci. Comput. **77** (2018), no. 1, 79–100.

[7] Y. Chen et al., On the matrix Monge-Kantorovich problem, European J. Appl. Math. **31** (2020), no. 4, 574–600.

[8] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, second edition, Applications of Mathematics (New York), 38, Springer, New York, 1998.

[9] S. Fagioli and O. Tse, On gradient flow and entropy solutions for nonlocal transport equations with nonlinear mobility, Nonlinear Anal. **221** (2022), Paper No. 112904, 35 pp.

[10] S. K. Kondratyev, L. Monsaingeon and D. A. Vorotnikov, A new optimal transport distance on the space of finite Radon measures, Adv. Differential Equations **21** (2016), no. 11-12, 1117–1164.

[11] M. Liero, A. Mielke and G. Savaré, Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures, Invent. Math. **211** (2018), no. 3, 969–1117.

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

[13] J. Maas and A. Mielke, Modeling of chemical reaction systems with detailed balance using gradient structures, J. Stat. Phys. **181** (2020), no. 6, 2257–2303.

[14] A. Mielke, Geodesic convexity of the relative entropy in reversible Markov chains, Calc. Var. Partial Differential Equations **48** (2013), no. 1-2, 1–31.

[15] 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.

[16] M. Mittnenzweig and A. Mielke, An entropic gradient structure for Lindblad equations and couplings of quantum systems to macroscopic models, J. Stat. Phys. **167** (2017), no. 2, 205–233.

[17] F. Otto, The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174.

[18] M. A. Peletier et al., Jump processes as generalized gradient flows, Calc. Var. Partial Differential Equations **61** (2022), no. 1, Paper No. 33, 85 pp.