Bocconi

Bachelor of Science
in Mathematical and Computing Sciences
for Artificial Intelligence

Dissipative-Hamiltonian Evolution within the GENERIC paradigm

Bachelor of Science thesis by:
Luca Martucci

Advisor:
Prof. **Giuseppe Savaré**

Academic Year 2024/2025

Università Commerciale
Luigi Bocconi

Dissipative-Hamiltonian evolution within the GENERIC paradigm

Luca Martucci

To my family and friends,
for their love and support.

Contents

Introduction

The GENERIC (General Equation for the Non-Equilibrium Reversible-Irreversible Coupling) formalism was first proposed, by Hans Christian Öttinger and Miroslav Grmela [GÖ97], as a general framework for expressing conservation laws of the total mass, momentum, and energy of some physical system. The bulk of the idea is that the dynamics of the system is driven by a time-evolution operator to be formally expressed as follows:

$$\partial_t = \mathrm{K}\nabla S + \mathrm{L}\nabla E,$$

for suitably chosen entropy S and energy E functionals, fully determining the mechanical and thermal properties of the evolution. In particular, the operator L, applied to the energetic derivative ∇E, encodes the proper conservative features of the system's behavior, while K, applied to the entropic derivative ∇S, characterizes the dissipative factor of any thermodynamically-consistent description. A sensible dissipative-conservative coupling can only be attained imposing non-interaction conditions:

$$\mathrm{L}\nabla S = 0 = \mathrm{K}\nabla E,$$

ensuring that the dissipative drift does not affect the energy and, contiguously, the conservation of energy still allows the (negative) entropy to dissipate. The crucial modeling assumption is an unambiguous choice of the state variables $u : \Omega \longrightarrow \mathbb{R}$, describing some state space Ω. Mathematically, this amounts to requiring the formal differential equation

$$\partial_t u = \mathrm{K}(u)\nabla S(u) + \mathrm{L}(u)\nabla E(u),$$

governing the evolution, to have a form that is independent of the coordinates chosen to chart the geometry of the set of state variables.

In Chapter I, we present the theory of gradient flows, namely, the most basic system of dissipative evolution, in the Euclidean space: $\partial_t u = -\nabla S(u)$, and show its convergence for a specific class of well-behaved entropic potentials, under the time-incremental variational scheme of Minimizing Movements, first proposed by De Giorgi [De 93], and thoroughly exploited to this end in the modern treatment of grandient flows in the most general setting of more abstract metric spaces, as developed by [AGS08]. Leveraging our characterization of solutions for the gradient flow equation in the Euclidean setting, we then illustrate its coordinate-unambiguous and naturally Riemannian geometric version, supplying us with the analytical form of the operator K. In this context, the Energy-Dissipation Principle now stands at the heart of the theory.

For the conservative dynamics, in Chapter II, we instead refer to the more ancient theory of Classical Mechanics and, in particular, its Hamiltonian formalism. Instead of a principle of dissipation of energy, we now have, as objective, Energy Conservation, which we discuss how to reformulate both in terms of the, now symplectic, differential geometric theory and approximation methods. Again, we get a clear expression for the operator L in terms of the geometric features of the system, and thus of the evolution it generates, which can now be properly called under the historical name of Hamilton's equations:

$$\dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q}$$

for ever-coupled state variables $u = (p, q) \in \Omega^2$.

Hopefully, by this time, the authentic symmetry between the geometric theory for gradient flows and the symplectic formulation of Hamiltonian systems will be manifestly evident, providing a natural motivation for coupling the two in a unified paradigm. While in the former we concentrate on a framework where coordinate transformations isometrically preserve the structure of the dissipative evolution, in the latter we hint at the rich and well-established theory of canonical transformations, those that keep intact Hamilton's equations of motion.

Finally, in Chapter III, we introduce the GENERIC formalism, after briefly discussing pioneering proposals for such a gradient-Hamiltonian coupling in the field of the thermodynamics, following the rules of which, dissipative systems are irreversible in nature while those that conserve energy are ruled by a more classical reversibility; these two dynamics respectively

corresponding to the second and first principle of thermodynamics, which we recover from the non-degeneracy axioms. The extension of classical discretization methods for Hamiltonian systems to the study of GENERIC is then discussed in Chapter IV, together with novel approaches, inspired by Minimizing Movements [JST22] and exploiting the approximation power of neural networks [ZSE22], along with relevant examples for each of the three.
The appendices contain the necessary background in differential geometry, convex and functional analysis. There, we also illustrate how these tools were developed and exploited before GENERIC, in the study of mechanics, both by Lagrange and Hamilton, and more recently to provide a mathematical explanation for the miraculous efficiency of neural networks.

CHAPTER I

Gradient flows

1. Gradient flows in the Euclidean space

Newton's equations describe the dynamics of a system under inertia and forces, governed by the Principle of Least Action:

$$m\ddot{x} = -\nabla\varphi(x).$$

The system can be severely damped: $m\ddot{x} + \gamma\dot{x} = -\nabla\varphi(x)$, for example by some friction or resistance, so that the inertial term becomes negligible and the equations of motion reduce to

$$\dot{x} = -\frac{1}{\gamma}\nabla\varphi(x), \tag{GF}$$

and, since γ is a constant, we are presented with a rather simple class of ordinary differential equations for trajectories:

$$t \in [0, +\infty) \longmapsto u_t \in \mathbb{R}^d,$$

moving towards decreasing the *potential* $\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}$,

$$\dot{u}_t = -\nabla\varphi(u_t), \quad \lim_{t\downarrow 0} u_t = u_0 \in \mathbb{R}$$

and these are usually referred to as *Gradient flow equations*[‡].

Example. In dimension $d = 2$, a gradient flow equation is simply a system

$$\dot{x} = -\frac{\partial}{\partial x}\varphi(x, y), \quad \dot{y} = -\frac{\partial}{\partial y}\varphi(x, y)$$

and, supposing $\varphi \in \mathbf{C}^2(\mathbb{R}^d)$ has an isolated critical point at (x_*, y_*):

- if it is a non-degenerate local minimum of φ, then (x_*, y_*) is a stable node for the system;

[‡]See [DS10] for an overview over modern formulations of the problem.

- if it is a non-degenerate local maximum of φ, then (x_*, y_*) is an unstable node for the system;
- if it is a saddle point of φ, then (x_*, y_*) is a saddle node for the system.

Remark (Curl-free characterization of gradient flows). An arbitrary two-dimensional system of ODEs

$$\begin{cases} \dot{x} = f(x,y) \\ \dot{y} = g(x,y) \end{cases} \qquad f, g \in \mathbf{C}^1(\Omega)$$

for a simply connected domain $\Omega \subseteq \mathbb{R}^2$, is a gradient flow equation if and only if

$$\frac{\partial f}{\partial y} = \frac{\partial g}{\partial x}. \tag{GFC}$$

The necessary condition follows from the Clairaut–Schwarz–Young integrability condition of commuting partial derivatives; for the sufficient condition, define the potential

$$\varphi(x,y) = -\int_0^y g(0,s)\,\mathrm{d}s - \int_0^x f(s,y)\,\mathrm{d}s$$

in integral form[§], so that the claim follows from the Fundamental Theorem of Calculus:

$$\begin{aligned} \frac{\partial \varphi}{\partial y} &= -g(0,y) - \int_0^x \frac{\partial f}{\partial y}(s,y)\,ds \\ &\overset{\text{(GFC)}}{=} -g(0,y) - \int_0^x \frac{\partial g}{\partial x}(s,y)\,ds \\ &= -g(0,y) - [g(x,y) - g(0,y)] = -g(x,y) \end{aligned}$$

and, similarly, $\partial\varphi/\partial x = -f(x,y)$. This argument directly generalizes to dimension $d = 3$, and the necessary condition is trivial since every gradient is always curl-free: $\nabla \times \nabla\varphi = 0$.

Definition 1.1 ([DS10]). Given $\varphi \in \mathbf{C}(\mathbb{R}^d)$, consider the time-evolution:

$$\dot{u}_t = -\nabla\varphi(u_t), \quad \lim_{t\downarrow 0} u_t = u_0 \in \mathbb{R}, \tag{$\mathfrak{G}$}$$

then the family of maps $G_t : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, $t \in [0, +\infty)$, such that for any $u_0 \in \mathbb{R}^d$, $G_0(u_0) \equiv u_0$ and the curve $u_t \equiv G_t(u_0)$ is the only $\mathbf{C}^1$ solution of ($\mathfrak{G}$), is a *Gradient flow* of φ.

[§]Here we use simple connectedness to justify that the choice of the integration path does not matter. As it is clear from Appendix C, (GFC) sits profoundly at the intersection of topology and geometry, and is an instance of the Poincaré lemma.

Energy identity. Trajectories solving ($\mathfrak{G}$) fulfill:

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi(u_t) \leqslant 0,$$

with equality for critical points of φ, which are thus fixed points of the dynamics.

Proof. By the chain rule, $\dot{\varphi}(u_t) = \langle \nabla\varphi, \dot{u}_t \rangle \overset{(\mathfrak{G})}{=} -|\dot{u}_t|^2 \leqslant 0$. ∎

1.1 Λ-convexity

If we try to impose the hypotheses of Peano-Picard's existence and uniqueness theorem on this class of differential equations, that is, $-\nabla\varphi$ is Lipschitz for some fixed constant λ, then we obtain the following *Hessian inequality*:

$$\nabla^2\varphi(x) \geqslant \lambda \mathrm{Id}_d \quad \text{i.e.} \quad \langle \nabla^2\varphi(x)\xi, \xi \rangle \geqslant \lambda|\xi|^2 \quad \forall \xi \in \mathbb{R}^d. \tag{HI}$$

Definition 1.2. A function $\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}$ is *λ-convex* if it satisfies (HI) with constant λ or, equivalently, the *Sub-gradient inequality*:

$$\langle \nabla\varphi(u), v - u \rangle \leqslant \varphi(v) - \varphi(u) - \frac{\lambda}{2}|u - v|^2. \tag{SGI}$$

We will assume $\lambda \geqslant 0$ throughout, as the case $\lambda = 0$ corresponds to a convex function and, for λ positive, our definition implies that the potential grows faster than quadratically. A negative coefficient would in fact allow for saddle points and "valleys" that impede the convergence of the flow towards global minima. Our intuition that λ-convex functions provide a nice class of potentials for ($\mathfrak{G}$), is confirmed by a powerful characterization of the solutions to the gradient flow they generate.

Evolution variation inequality. If $\varphi \in \mathbf{C}^2(\mathbb{R}^d)$ is λ-convex, u_t is a solution to ($\mathfrak{G}$) if and only if

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}|u_t - v|^2 + \frac{\lambda}{2}|u_t - v|^2 \leqslant \varphi(v) - \varphi(u_t) \tag{EVI$_\lambda$}$$

for all $v \in \mathbb{R}^d$.

Proof. If u_t is a solution to ($\mathfrak{G}$), then

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|u_t - v|^2 = \langle \dot{u}_t, u_t - v\rangle \overset{(\mathfrak{G})}{=} \langle \nabla\varphi(u_t), v - u_t\rangle \overset{\text{(SGI)}}{\leqslant} \varphi(v) - \varphi(u_t) - \frac{\lambda}{2}|u_t - v|^2.$$

Conversely, we can rewrite (EVI$_\lambda$) as the differential inequality

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(v) \leqslant \varphi(v) - \varphi(u_t) - \lambda\Phi(v) \tag{DEVI$_\lambda$}$$

and, by the chain rule on the distance function $\Phi(v) \equiv |u_t - v|^2/2$, we obtain:

$$\langle \dot{u}_t, u_t - v\rangle \leqslant \varphi(v) - \varphi(u_t) - \frac{\lambda}{2}|u_t - v|^2,$$

then, choosing $v \equiv u_t + \varepsilon x$ for $\varepsilon > 0, x \in \mathbb{R}^d$ and dividing by ε:

$$-\langle \dot{u}_t, x\rangle \leqslant \frac{1}{\varepsilon}(\varphi(v) - \varphi(u_t)) - \frac{\lambda\varepsilon}{2}|x|^2$$

and as $\varepsilon \downarrow 0$: $-\langle \dot{u}_t, x\rangle \leqslant \langle \nabla\varphi(u_t), \xi\rangle$ for all $x \in \mathbb{R}^d$, i.e. $-\dot{u}_t = \nabla\varphi(u_t)$. ∎

If we picture our flow as measuring the "dissipation" of a quantity φ over time, (EVI$_\lambda$) can be interpreted as an exponential bound, at around any point v, to the "decay of the flow" towards minimizing the potential. Recognizing the Leibniz rule, it can indeed be written as:

$$e^{-\lambda t}\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{\lambda t}\frac{1}{2}|u_t - v|^2\right) \leqslant \varphi(v) - \varphi(u_t)$$

and moreover, if v is another solution, because

$$\frac{\mathrm{d}}{\mathrm{d}t}|u_t - v_t|^2 = 2\langle \dot{u}_t - \dot{v}_t, u_t - v_t\rangle \overset{(\mathfrak{G})}{=} -2\langle \nabla u_t - \nabla v_t, u_t - v_t\rangle \overset{\text{(SGI)}}{\leqslant} -2\lambda|u_t - v_t|,$$

we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{\lambda t}|u_t - v_t|\right) = \lambda e^{\lambda t}|u_t - v_t| + e^{\lambda t}\frac{\mathrm{d}}{\mathrm{d}t}|u_t - v_t| \leqslant \lambda e^{\lambda t}|u_t - v_t| - \lambda e^{\lambda t}|u_t - v_t| = 0,$$

bounding the rate at which the two trajectories "collide".

1.2 Time-incremental minimization scheme

When presented with whatever system, if it is a gradient flow and we recognize it, the potential may be arbitrarily intricate and, in order to develop a theory for the existence and

uniqueness of solutions for ($\mathfrak{G}$), it may be useful to approach the problem by constructing discrete approximations of the solution.
Fixing a time step τ and a uniform partition

$$\mathscr{P}_\tau = \{0 = t_0 < t_1 < \cdots < t_n < \dots \tau\}, \quad t_n = n\tau,$$

of $[0, +\infty)$, we look for the sequence $(u_n)_{n\in\mathbb{N}}$ providing an approximation for the solution $(u(t_n))_{n\in\mathbb{N}}$. The most immediate iterative approach is the *explicit Euler scheme*:

$$u_{n+1} \equiv u_n - \tau\nabla\varphi(u_n),$$

which is the iterative update defining the most basic *Gradient Descent method*. This algorithm works "the fastest" [‡] precisely for λ-convex functions.
Another classical approximation scheme is the *implicit Euler method*:

$$\frac{u_n - u_{n-1}}{\tau} + \nabla\varphi(u_n) = 0,$$

which can be equivalently reformulated as the optimization problem for the functional

$$\Xi(u) \equiv \varphi(u) + \frac{1}{\tau}\Phi(u_{n-1}),$$

because:

$$0 = \nabla\varphi(u) + \frac{1}{\tau}\nabla\Phi(u_{n-1}) = \nabla\varphi(u) + \frac{1}{\tau}(u - u_{n-1})$$

hence,

$$u_n \in \operatorname*{arg\,min}_{u\in\mathbb{R}^d}\left(\varphi(u) + \frac{1}{\tau}|u - u_{n-1}|^2\right) \tag{MM}$$

ensuring stability and convergence as we refine the partition: $\tau \downarrow 0$, without even requiring a uniform Lipschitz condition on $\nabla\varphi$, differently from the explicit method. This idea of conveniently reformulating the implicit discretization of a dynamic evolution equation, that involve hard, or rather impossible, inversions of vector-valued maps, by means of some time-incremental variational minimization was first proposed by Ennio De Giorgi [De 93], which he called *Minimizing Movements* and presented as a unified framework to solve a broad class of evolutionary problems on a generic space $\mathbf{X}$: the objective is to find discrete solutions $(u_n)_n$

[‡] See [BV04, Sect. 9.3] for the "minutiæ" of the convergence theory.

to the recursive minimization objective

$$u_n \in \arg\min_{x \in \mathbf{X}} \mathsf{F}_\tau(n, u_{n-1}, x), \quad n = 1, \dots, N\tau$$

for some given initial datum u_0 and suitable family of functionals $\mathsf{F}_\tau : \mathbb{N} \times \mathbf{X} \times \mathbf{X} \longrightarrow \mathbb{R}$.

Definition 1.3 ([De 93]). We say that the curve $u : \mathbb{R} \longrightarrow \mathbf{X}$ is a *Minimizing Movement* associated to F on $\mathbf{X}$, and write $u \in \mathsf{MM}(\mathsf{F}, \mathbf{X})$, if there exist piecewise continuous approximating curves $w_\tau : \mathbb{N} \longrightarrow \mathbf{X}$ such that, for any $t \in \mathbb{R}$:

$$\lim_{\tau \downarrow 0} w_\tau(\lfloor t/\tau \rfloor) = u(t)$$

and for any $\tau \in (0, +\infty), n \in \mathbb{N}$:

$$\mathsf{F}_\tau(n, w_\tau(n-1), w_\tau(n)) = \min_{x \in \mathbf{X}} \mathsf{F}_\tau(n, w_\tau(n-1), x).$$

So, Minimizing Movements are limits, as $\tau \downarrow 0$, of the piecewise-constant interpolations of u_n.

Example. In the Euclidean space $\mathbf{X} = \mathbb{R}^d$, in the case of gradient flows of a time-dependent potential $\varphi \in \mathrm{Lip}(\mathbf{X}) \cap \mathbf{C}^2(\mathbf{X})$, the related Minimizing Movement scheme corresponds to functionals of the form

$$\mathsf{F}_\tau(n, u, x) = \varphi(x) + \frac{1}{2\tau}|x - u|^2.$$

In fact, $u \in \mathsf{MM}(\mathsf{F}, \mathbb{R}^d)$ if and only if $u \in \mathrm{Lip}(\mathbb{R}^d)$ solves $(\mathfrak{G})$.

De Giorgi also already noticed that, weakening the conditions on the potential, we can only say that any $u \in \mathsf{MM}(\mathsf{F}, \mathbb{R}^d)$ is a Lipschitz function satisfying the Cauchy problem $(\mathfrak{G})$, but the opposite implication does not hold. More notably, he also remarked that "*many interesting and difficult problems can be formulated if* $\boldsymbol{X}$ *is a space of functions and F is an integral functional, giving problems of* gradient flow *type.*" [De 93]

Example. Consider the Sobolev space[§] $\mathbf{X} = \mathbf{W}^{1,2}(\mathbb{R}^d)$ and the Moreau envelope [‡]

$$\mathsf{F}_\tau(n, \varphi, g) = \mathsf{E}(\varphi) + \frac{1}{\tau} \int_{\mathbb{R}^d} |\varphi - g|^2$$

[§]Refer to the brief Appendix B.

[‡]The fact that it attains a strict minimum follows directly from convex duality as discussed in Appendix A.

of the Dirichlet energy

$$\mathsf{E}(\varphi) = \int_{\mathbb{R}} |\nabla\varphi(u)|^2 \,\mathrm{d}u,$$

then $u \in \mathsf{MM}(\mathsf{F}, \mathbf{X})$ if and only if it solves the heat equation $\dot{u}_t = \Delta\varphi(u_t)$.

1.3 Convergence in the subgradient

We shall now show existence and uniqueness in $\mathsf{MM}(\Xi, \mathbb{R}^d)$, that is, of the time-incremental minimization (MM), under suitable conditions, mainly adapting results from [Mie23] in our simplified setting.

Definition 1.4. Based on the discrete solution $(u_n)_n$, we are now able to define the continuous *piecewise affine interpolant*

$$\hat{u}_\tau((n+\theta-1)\tau) \equiv (1-\theta)u_{n-1} + \theta u_n, \quad \theta \in [0,1]$$

with piecewise constant derivative

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{u}_\tau(t) = \frac{1}{\tau}(u_n - u_{n-1}), \quad t \in ((n-1)\tau, n\tau)$$

and the left-continuous *piecewise constant interpolants*

$$\overline{u}_\tau(t) \equiv u_n, \quad \overline{\xi}_\tau(t) \equiv \xi_n \quad \text{for} \quad t \in ((n-1)\tau, n\tau].$$

We will now prove convergence to a solution for non-necessarily smooth potentials, in fact, we will only assume λ-convexity and, for a trajectory to be a solution, we just require its derivative to be *a gradient*[§] of the potential, namely $-\dot{u}_t \in \partial\varphi(u_t)$.

Theorem 1.5. Suppose $\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}_\infty$ is a proper, lower semicontinuous, λ-convex functional with compact sublevels $S_r = \{u \in \mathbb{R}^d : \varphi(u) \leqslant r\}$, then, for all $u_0 \in \mathrm{dom}(\varphi)$, the solutions $\hat{u}_\tau : [0,T] \longrightarrow \mathbb{R}^d$, obtained from the incremental minimizing scheme, converge to the unique solution $u \in \mathbf{W}^{1,2}([0,T];\, \mathbb{R}^d)$ of ($\mathfrak{G}$), i.e.

$$\forall t \in [0,T]: \quad \hat{u}_\tau(t) \longrightarrow u(t) \quad \text{in } \mathbb{R}^d.$$

[§] See App. A.3 for the definition of subdifferential.

PROOF. Exploiting the compactness of the sublevels, $\inf_{u\in\mathbb{R}^d}\varphi(u)$ is attained by Weierstraß extreme value theorem, thus the potential is bounded from below: along with lower semicontinuity, this implies that its minimizers exist and similarly u_n as minimizer of Ξ exists, i.e.

$$\frac{1}{2\tau}|u_n-u_{n-1}|^2+\varphi(u_n)=\Xi(u_n)\leqslant\Xi(u_{n-1})=\varphi(u_{n-1})$$

hence $\varphi_{\min}\leqslant\varphi(u_n)\leqslant\varphi(u_0)$ for $n\in\{0,\dots,N\}$ and

$$\int_0^t\left|\frac{\mathrm{d}}{\mathrm{d}t}\hat{u}_\tau(t)\right|^2\,\mathrm{d}t=\sum_{n=1}^{N}\tau\left|\frac{1}{\tau}(u_n-u_{n-1})\right|^2\leqslant 2(\varphi(u_0)-\varphi_{\min})$$

by adding up the incremental estimate for $n=1,\dots,N$. By Bolzano-Weierstraß, because $\hat{u}_\tau$ is bounded in $\mathbf{W}^{1,2}$, we can extract a subsequence such that [‡]

$$\hat{u}_\tau\longrightarrow u\text{ and }\frac{\mathrm{d}}{\mathrm{d}t}\hat{u}_\tau\longrightarrow\dot{u}\text{ in }\mathbf{L}^2([0,T];\,\mathbb{R}^d)$$

for a limit $u\in\mathbf{W}^{1,2}([0,T];\,\mathbb{R}^d)$. Since for all $t\in[0,T]$ and $\tau=T/N$, the values $\hat{u}_\tau(t)$ lie in the compact sublevel $S_{\varphi(u_0)}$, and since each interpolant is equicontinuous:

$$|\hat{u}_\tau(t)-\hat{u}_\tau(s)|\leqslant\sqrt{|t-s|}\left\|\frac{\mathrm{d}}{\mathrm{d}t}\hat{u}_\tau\right\|_{\mathbf{L}^2}\leqslant\sqrt{2|t-s|}(\varphi(u_0)-\varphi_{\min})^{1/2},$$

by Arzelà-Ascoli, up to extraction of a further subsequence, we get the uniform convergence:

$$\hat{u}_\tau\longrightarrow u\text{ in }\mathbf{C}^0([0,T];\,\mathbb{R}^d)\quad\text{ and }\quad\overline{u}_\tau\longrightarrow u\text{ in }\mathbf{L}^\infty([0,T];\,\mathbb{R}^d).$$

Moreover, because $\hat{u}_\tau(n\tau)=\overline{u}_\tau(n\tau)$ and $\overline{u}_\tau$ is piecewise constant, we get the second convergence as $\tau\downarrow 0$:

$$\|\hat{u}_\tau-\overline{u}_\tau\|_{\mathbf{L}^\infty}\leqslant\sqrt{2\tau(\varphi(u_0)-\varphi_{\min})}.$$

We now show that the limit we found satisfies ($\mathfrak{G}$), letting

$$\xi_\tau=-\frac{\mathrm{d}}{\mathrm{d}t}\hat{u}_\tau\in\mathbf{L}^2([0,T];\,\mathbb{R}^d)\quad\text{such that}\quad\xi_\tau\longrightarrow\xi\equiv-\dot{u}\text{ in }\mathbf{L}^2([0,T];\,\mathbb{R}^d),$$

[‡] Refer to App. B for weak convergence in Hilbert spaces.

together with $\overline{u}_\tau \longrightarrow u$ in $\mathbf{L}^2([0,T];\,\mathbb{R}^d)$ and $\xi_\tau \in \partial\varphi(\overline{u}_\tau(t))$ for all $t \in [0,T]$, by (strong-weak) energy closedness of the subdifferential[§], also the limit $\xi \in \partial\varphi(u(t))$ i.e.

$$-\dot{u}_t \in \partial\varphi(u_t).$$

We now need only to show uniqueness and full convergence. Since the auxiliary function

$$\tilde{\varphi}(u) \equiv \varphi(u) - \frac{\lambda}{2}|u|^2 \quad \text{is convex with} \quad \partial\tilde{\varphi}(u) = -\lambda u + \partial\varphi(u)$$

then, for arbitrary $u, \tilde{u} \in \mathrm{dom}(\partial\varphi)$ and $\xi_j \in \partial\varphi(u_j)$, setting $\eta_j \equiv \xi_j - \lambda u_j \in \partial\tilde{\varphi}(u_j)$,

$$\langle \tilde{\xi} - \xi, \tilde{u} - u\rangle = \langle \tilde{\eta} - \eta, \tilde{u} - u\rangle + \lambda\langle \tilde{u} - u, \tilde{u} - u\rangle \geqslant \lambda|\tilde{u} - u|^2$$

by monotonicity of subdifferentials of $\tilde{\varphi}$. Assuming $u, \tilde{u} \in \mathbf{W}^{1,2}([0,T];\,\mathbb{R}^d)$, the latter inequality yields the Grönwall-like estimate:

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|\tilde{u}(t) - u(t)| = \left\langle \frac{\mathrm{d}}{\mathrm{d}t}\tilde{u}(t) - \frac{\mathrm{d}}{\mathrm{d}t}u(t), \tilde{u}(t) - u(t) \right\rangle \leqslant -\lambda|\tilde{u}(t) - u(t)|^2$$

from which we obtain the desired Lipschitz continuity:

$$|\tilde{u}(t) - u(t)| \leqslant e^{-\lambda(t-s)}|\tilde{u}(s) - u(s)| \quad \text{for } 0 \leqslant s < t \tag{LC}$$

and hence uniqueness of solutions if $\tilde{u}(0) = u(0) = u_0$. ∎

Corollary 1.6 (λ-contractivity). Any two solutions satisfy (LC) and, in particular, for every $u_0 \in \mathrm{dom}(\varphi)$, there is at most one (EVI$_\lambda$) solution satisfying the initial condition $\lim_{t\downarrow 0} u_t = u_0$.

However, this scheme fails to respect the "energy-dissipating" properties, already discussed for λ-convex functions, at the discrete level. [JST19]

In particular, from above, we easily get that the "energy drop" must be bounded from below:

$$\varphi(v) - \varphi(u_t) \geqslant \frac{1}{2\tau}|u_t - v|^2$$

for all $v \in \mathbb{R}^d$ thus, at each time step of the discretization:

$$\varphi(u_{n-1}) - \varphi(u_n) \geqslant \frac{1}{2\tau}|u_n - u_{n-1}|^2$$

and it does not correspond to the "dissipation" $|(u_n - u_{n-1})|^2$ of the n^{th} step.

[§] See Def. 2.2 in App. B.

2. Riemannian gradient flows

Before introducing the new discretization scheme, however, we first need to address a fundamental problem in the very definition of ($\mathfrak{G}$) we provided, that is, it is not coordinate-independent[‡]. For the sake of clarity, let us first discuss an example.

Example (Damped particle in a radial potential). Suppose that we have the Newton-like potential

$$V(r) = -\frac{k}{r}, \quad k > 0, \quad r = |u| \quad \text{such that} \quad \nabla V = \frac{ku}{r^3}$$

and ($\mathfrak{G}$) becomes

$$\dot{u}_t = -\gamma^{-1}\frac{k}{r^3}u_t.$$

Consider a spherical coordinate change

$$u = (x, y, z) = (r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta) \longmapsto \tilde{u} = (r, \theta, \phi) \quad \text{in} \quad \mathbb{R}^3$$

then the structure of the gradient flow equation is only preserved if we account for the Jacobian $\Gamma(r, \theta, \phi)$ of the transformation, namely:

$$\dot{\tilde{u}}_t = -\gamma^{-1}(\Gamma_{r,\theta,\phi})^{-1}\nabla\tilde{V}_{r,\theta,\phi} = -\gamma^{-1}\begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta \end{pmatrix}^{-1}\nabla\tilde{V}_{r,\theta,\phi}.$$

We thus need a more sophisticated definition of gradient, already encoding the geometry of the coordinate-space we are working in and, no matter the frame considered, ending up with the same trajectories.

The key geometrical ideas to achieve this were already developed by the the theory of canonical transformations in the well-established framework of Hamiltonian mechanics[§].

2.1 Flows on manifolds

The natural setting where to study (gradient) flows is geometric [Lee12, Ch. 9] in nature: in ($\mathfrak{G}$), we are just imposing the velocity of a curve equal to a vector field, specifically the gradient $\nabla\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}^d$. Consider differentiable manifolds $\mathcal{M}$ and $\mathcal{N}$, where, differentiating

[‡] Cfr. with the very first observations of [Mie23].

[§] For a comprehensive and detailed survey, see [MR13] and our discussion in Ch. II.

a smooth function $F : \mathcal{M} \longrightarrow N$, induces the linear map

$$\mathrm{D}F_p : \mathrm{T}_p\mathcal{M} \longrightarrow \mathrm{T}_{F(p)}\mathcal{N}$$

that for any tangent vector $v \in \mathrm{T}_p\mathcal{M}$ and smooth curve $\gamma : I = (-\epsilon, \epsilon) \subset \mathbb{R} \longrightarrow \mathcal{M}$, $\epsilon > 0$, on the manifold such that $\gamma(0) = m$ and $\dot{\gamma}(0) = v$, i.e. whose initial velocity is the tangent vector v, takes the form

$$\mathrm{D}F_p(v) \equiv \frac{\mathrm{d}}{\mathrm{d}t} F(\gamma(t))\bigg|_{t=0}$$

and it does not depend on the choice of γ. This definition is natural if we acknowledge that, among the many ways of constructing the *tangent space* $\mathrm{T}_p\mathcal{M}$ *to a manifold* $\mathcal{M}$ *at a point* p, the most intuitive is that of taking the collection of all possible initial speeds of curves starting at p and traveling on the manifold. This provides our manifold with a vector space, of *tangent vectors*, at each point. This definition of differential is, in particular, valid for real-valued maps $f : \mathcal{M} \longrightarrow \mathbb{R}$ that are differentiable, which we will denote as $f \in \mathscr{F}(\mathcal{M})$: it's enough to consider $\mathbb{R}$ as a manifold and canonically identify $\mathrm{T}_{f(p)}\mathbb{R} \equiv \mathbb{R}$. In this case, then, $\mathrm{D}f_p \in \mathrm{T}_p^*\mathcal{M}$. Consider the *tangent bundle*

$$\mathrm{T}\mathcal{M} \equiv \bigsqcup_{p\in\mathcal{M}} \mathrm{T}_p\mathcal{M} = \{(p, v) : p \in \mathcal{M}, v \in \mathrm{T}_p\mathcal{M}\}$$

i.e., the collection of all tangent vectors at every point on the manifold, and *sections* of it, $X : \mathcal{M} \longrightarrow \mathrm{T}\mathcal{M}$, associating to each point: $p \longmapsto X_p \in \mathrm{T}_p\mathcal{M}$ a tangent vector. This is a *vector field*, and we denote by $\mathfrak{X}(\mathcal{M})$ the collection of them. We say that γ is an *integral curve* of the vector field $X \in \mathfrak{X}(\mathcal{M})$ if:

$$\dot{\gamma}(t) = X_{\gamma(t)} \text{ for all } t \in I$$

that is, at each point p, its velocity is exactly X_p. Now, by the existence theorem for ODEs on the coordinate representation of the smooth vector field X, for each point $p \in \mathcal{M}$, there exists an integral curve γ_p starting at p, i.e. $\gamma_p(0) = p$. For each $t \in \mathbb{R}$, this gives a endomorphism $\gamma_t : p \in \mathcal{M} \longmapsto \gamma_t(p) \equiv \gamma_p(t) \in \mathcal{M}$ "sliding" the manifold along integral curves. Since, by the chain rule, $\gamma_p(t + s)$ is an integral curve at $q \equiv \gamma_p(s)$, assuming also uniqueness of integral curves, $\gamma_q(t) = \gamma_p(t + s)$ and thus $\gamma_t \circ \gamma_s(p) = \gamma_{t+s}(p)$, meaning that we can define a *global flow on* $\mathcal{M}$ as the continuous map $\gamma : \mathbb{R} \times \mathcal{M} \longrightarrow \mathcal{M}$ such that $\gamma(t, \gamma(s, p)) = \gamma(t + s, p)$ and

$\gamma(0,p) = p$. The collection of homeomorphisms $\gamma_t \equiv \gamma(t,p)$ satisfies the *semigroup property*

$$\gamma_t \circ \gamma_s = \gamma_{t+s} \quad \gamma_0 = \mathrm{Id}_{\mathcal{M}}$$

as the image of the curve $\gamma_p(t) \equiv \gamma(t,p)$ is the orbit of p under the group action γ of the additive group $\mathbb{R}$ on $\mathcal{M}$.

2.2 Onsager dynamics

Now, as promised, we describe a coordinate-independent formulation of a vector field representing the gradient of our potential.

Definition 2.1. A *Riemannian manifold* is a pair $(\mathcal{M}, \Gamma)$ where $\mathcal{M}$ is a differentiable manifold, the *Riemannian structure* $\Gamma_u : \mathrm{T}_u\mathcal{M} \longrightarrow \mathrm{T}_u^*\mathcal{M}$ is symmetric and positive, and the *Riemannian metric* g, defined as follows:

$$g(v,w)_u = \Gamma_u(v)[w],$$

is a symmetric 2-tensor [‡].

The motivation for introducing such structures[§] is that we can define the *Onsager operator* [‡] as

$$\mathrm{K}_p \equiv \Gamma_p^{-1} : \mathrm{T}_p^*\mathcal{M} \longrightarrow \mathrm{T}_p\mathcal{M}$$

and the gradient of a potential $\varphi \in \mathscr{F}(\mathcal{M})$, if $\mathcal{M}$ is a Riemannian manifold, as:

$$\nabla\varphi(p) \equiv \mathrm{K}_p \mathrm{D}\varphi_p \tag{$\sharp_g$}$$

so that ($\mathfrak{G}$) can be expressed in:

- *rate form*:

$$\dot{u}_t = -\mathrm{K}_{u_t}\mathrm{D}\varphi_{u_t} \in \mathrm{T}_{u_t}\mathcal{M};$$

- *force-balance form*:

$$0 = \Gamma_{u_t}\dot{u}_t + \mathrm{D}\varphi_{u_t} \in \mathrm{T}_{u_t}^*\mathcal{M}$$

  where we can think of $\xi \equiv \Gamma_u\dot{u}$ as a "*viscous force induced by the rate and of the differential as a potential restoring force.*" [Mie23]

[‡] Refer to Appendix C for the differential geometric tools not discussed in Sect. 2.1 of Chp. I.
[§] Also see the discussion on musical isomorphisms in App. C.
[‡] The reference is to the discussion in [OM53, p. 1507].

Hence, a *dissipative system* is a triple $(\mathcal{M}, \Gamma, \varphi)$, where we can represent the Riemannian structure Γ by a matrix, as it is a linear map between vector spaces: for instance, in $\mathbb{R}^d$, we just have the *flat metric* $\Gamma = \mathrm{Id}_d$.

Example. For the damped particle in a radial potential, $\Gamma_{x,y,z} = \mathrm{Id}_d$ was just the flat metric, while

$$\tilde{\Gamma}_{r,\theta,\phi} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{pmatrix} \quad \text{but} \quad \dot{u}_t = \nabla_g V_{x,y,z} \Longleftrightarrow \dot{\tilde{u}}_t = \nabla_{\tilde{g}} \tilde{V}_{r,\theta,\phi}.$$

We consider (structurally) the same equation, but the Onsager operator K becomes $\tilde{\mathrm{K}}$ as we go from the flat Euclidean space:

$$g_{\mathbb{R}^3} = \sum_{i=1}^{3} \mathrm{d}x_i \otimes \mathrm{d}x_i$$

to the (spherical) warped product space:

$$\tilde{g} = g_{\mathbb{R}} + r^2 g_{\mathbb{S}^2}$$

with flat base manifold and the unit sphere as fiber.

This construction is easily extendable to a convex setting.

2.3 Riemannian subgradients

With a Riemannian structure, we can naturally measure length of curves and angles between them, as we are providing each tangent space with an inner product g_u, even though we call it a "metric", and hence a norm $|v|_g \equiv \sqrt{g(v,v)_p}$ to measure the length of any tangent vector $v \in \mathrm{T}_p\mathcal{M}$. In fact, for any two points $x_0, x_1 \in \mathcal{M}$, calling Υ the collection of smooth curves $\gamma : t \in [0,1] \longmapsto \gamma_t \in \mathcal{M}$ connecting them, $\gamma_0 = x_0, \gamma_1 = x_1$, the distance

$$d_g(x_0, x_1) \equiv \inf_{\gamma \in \Upsilon} \int_0^1 \sqrt{|\dot{\gamma}_t|_g}\, \mathrm{d}t$$

makes the manifold a metric space, supposing it is connected.

As we need a notion of convexity on a Riemannian manifold, let's call a *geodesic* a curve γ such that

$$d_g(\gamma_s, \gamma_t) \leqslant |t-s| d_g(\gamma_0, \gamma_1) \quad \text{for all} \quad 0 \leqslant s \leqslant t \leqslant 1$$

and define a set $U \subseteq \mathcal{M}$ to be *geodesically convex* if every couple of points $x_0, x_1 \in U$ can be connected by a geodesic $\{\gamma_t\}_t \subset U$.

Definition 2.2 ([DS10]). The function $\varphi : \mathcal{M} \longrightarrow \mathbb{R}$ is λ-*convex along a curve* γ if $\forall s \in [0,1]$:

$$\varphi(\gamma_s) \leqslant (1-s)\varphi(\gamma_0) + s\varphi(\gamma_1) - \frac{\lambda}{2}s(1-s)d_g^2(\gamma_0,\gamma_1)$$

and it is *geodesically* λ-*convex* if every couple of points x_0, x_1 can be connected by a geodesic along which φ is λ-convex.

Again, we can characterize solutions of the gradient flow equation on $\mathcal{M}$ by the curves solving

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}d_g^2(u_t,v) + \frac{\lambda}{2}d_g^2(u_t,v) \leqslant \varphi(v) - \varphi(u_t) \tag{EVI$_\lambda$}$$

for every other tangent vector v.
Now, we take the flow on $\mathcal{M}$ generated by these curves as definition of *gradient flow on a Riemannian manifold*, and justify the restriction to (geodesically) λ-convex potentials by the following.

Proposition 2.3 ([DS10]). If there exists a gradient flow for $\varphi : \mathcal{M} \longrightarrow \mathbb{R}$, then φ is (geodesically) λ-convex.

Proof. If there exists a gradient flow u_t for $\varphi : \mathcal{M} \longrightarrow \mathbb{R}$, any curve $\gamma_s^t \equiv u_t(\gamma_s)$ will solve (EVI$_\lambda$) for each point of the geodesic γ_s connecting x_0, x_1, that is, the differential inequality (DEVI$_\lambda$) but now for the function $\Phi(t) \equiv d_g^2(\gamma_s^t, \gamma_s)/2$, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t) + \lambda\Phi(t) \leqslant \varphi(\gamma_s) - \varphi(\gamma_s^t)$$

and integrating for some $t > 0$:

$$\int_0^t e^{\lambda r}\frac{\mathrm{d}}{\mathrm{d}r}\Phi(r)\,\mathrm{d}r + \lambda\int_0^t e^{\lambda r}\Phi(r)\,\mathrm{d}r \leqslant \int_0^t e^{\lambda r}(\varphi(\gamma_s) - \varphi(\gamma_s^t))\,\mathrm{d}r,$$
$$\int_0^t \frac{\mathrm{d}}{\mathrm{d}r}(e^{\lambda r}\Phi(r))\,\mathrm{d}r = e^{\lambda t}\Phi(t) - \Phi(0) \leqslant (\varphi(\gamma_s) - \varphi(\gamma_s^t))\frac{e^{\lambda t}-1}{\lambda}.$$

Let $E_\lambda(t) \equiv (e^{\lambda t}-1)/\lambda$ so that for every $s \in [0,1]$:

$$\frac{1}{2}e^{\lambda t}d_g^2(\gamma_s^t,\gamma_0) - \frac{1}{2}d_g^2(\gamma_s,\gamma_0) \leqslant E_\lambda(t)(\varphi(\gamma_0) - \varphi(\gamma_s)), \tag{1}$$

$$\frac{1}{2}e^{\lambda t}d_g^2(\gamma_s^t,\gamma_1)-\frac{1}{2}d_g^2(\gamma_s,\gamma_1)\leqslant E_\lambda(t)(\varphi(\gamma_1)-\varphi(\gamma_s)). \tag{2}$$

Now, because γ is a geodesic: $(1-s)d_g^2(\gamma_s,\gamma_0)+sd_g^2(\gamma_s,\gamma_1)=s(1-s)d_g^2(\gamma_0,\gamma_1)$, multiplying (1) by $\tilde{s}\equiv(1-s)$ and (2) by s and then adding, we get

$$\frac{1}{2}e^{\lambda t}(\tilde{s}d_g^2(\gamma_s^t,\gamma_0)+sd_g^2(\gamma_s^t,\gamma_1))-s\tilde{s}d_g^2(\gamma_0,\gamma_1)\leqslant 2E_\lambda(t)(\tilde{s}\varphi(\gamma_0)+s\varphi(\gamma_1)-\varphi(\gamma_s))$$

where

$$(1-s)d_g^2(\gamma_s^t,\gamma_0)+sd_g^2(\gamma_s^t,\gamma_1)\geqslant s(1-s)(d_g(\gamma_s^t,\gamma_0)+d_g(\gamma_s^t,\gamma_1))^2\geqslant s(1-s)d_g^2(\gamma_0,\gamma_1)$$

by the triangle inequality, hence

$$\frac{\lambda}{2}s(1-s)d_g^2(\gamma_0,\gamma_1)\leqslant(1-s)\varphi(\gamma_0)+s\varphi(\gamma_1)-\varphi(\gamma_s)$$

which is exactly (geodesic) λ-convexity. ∎

Finally, we can take subgradients and the flow they generate.

Definition 2.4. Let $(\mathcal{M},g)$ be a Riemannian manifold, and $f:\mathcal{M}\to\mathbb{R}$ be geodesically λ-convex. A tangent vector $v\in \mathrm{T}_p\mathcal{M}$ at a point $p\in\mathcal{M}$ is called a *subgradient* of f at p if, for every geodesic γ on $\mathcal{M}$ with $\gamma(0)=p$:

$$(f\circ\gamma)(t)\geqslant f(x)+tg(v,\dot{\gamma}(0))_p$$

for any $t\geqslant 0$, and the *subdifferential* $\partial f_p\subset \mathrm{T}_p\mathcal{M}$ is the set of all subgradients of f at p.

Thus, it makes sense to write

$$\Gamma_{u_t}\dot{u}_t\in\partial\varphi_{u_t}$$

and the product $\mathrm{K}_{u_t}\partial\varphi_{u_t}$ in the rate form can be intended pointwisely.

3. Dissipation potentials

We need to develop an "energy-dissipating" scheme, as variants of the Euler scheme respecting the inequalities we studied, even for, non-necessarily smooth, λ-convex functions and consider

the flow generated by the subgradients, for which ($\mathfrak{G}$) becomes:

$$\dot{u}_t \in -\partial\varphi(u_t), \quad \lim_{t\downarrow 0} u_t = u_0 \in \mathbb{R}$$

and the previous results hold because the chain rule holds for the subdifferential (even if we don't assume smoothness of the potential, an "upper chain rule" holds for *smooth perturbations of a convex function* [AGS08, Cor. 1.4.5 and Rem. 1.4.6] and, in particular, λ-convex functions belong to this class) of $\varphi = \varphi_c + \varphi_d$ where φ_c is convex and lower semicontinuous, and φ_d is differentiable, for which: $\partial\varphi = \partial\varphi_c + \nabla\varphi_d$.
Indeed, we would like to consider time-incremental minimization schemes in the form

$$u_n = \operatorname*{arg\,min}_{u\in\mathbb{R}^d} \left(\tau\Psi\left(\frac{1}{\tau}(u - u_{n-1})\right) + \varphi(u) \right)$$

for a convex function $\Psi : \mathbb{R}^d \longrightarrow \mathbb{R}$, because the Euler-Lagrange equation[§] reads as the following doubly nonlinear evolution equation:

$$0 \in \partial\Psi\left(\frac{1}{\tau}(u - u_{n-1})\right) + \partial\varphi(u)$$

which will be interpreted as the backward-Euler, fully implicit, discretization of the evolutionary inclusion $0 \in \partial\Psi(\dot{u}_t) + \partial\varphi(u_t)$, which is equivalent to the *double subdifferential inclusion*:

$$\exists \xi_n \quad \text{such that} \quad \xi_n \in \partial\varphi(u_n), \quad -\xi_n \in \partial\Psi\left(\frac{1}{\tau}(u_n - u_{n-1})\right) \tag{DSI}$$

Definition 3.1 ([Mie23]). A function $\Psi : \mathbb{R}^d \longrightarrow \mathbb{R}_\infty$ is a *Rayleigh dissipation potential* if it is lower semicontinuous, convex and satisfies $\Psi(u_t) \geqslant \Psi(0) = 0$.

Because the Legendre transform Ψ^* of a dissipation potential is automatically convex and lower semicontinuous, we shall call it *dual dissipation potential*, and they together satisfy the *Fenchel optimality condition*[‡]

$$\Psi(v) + \Psi^*(\xi) \geqslant \langle \xi, v \rangle \tag{FOC}$$

[§] See Appendix D for a geometric-mechanical derivation.
[‡] See App. A to refresh the elegant convex-analytic results used throughout.

for all $\xi, v \in \mathbb{R}$, with equality if and only if $\xi \in \partial\Psi(v)$. Applying (FOC) to (DSI), we get:

$$\Psi\left(\frac{1}{\tau}(u_n - u_{n-1})\right) + \Psi^*\left(-\xi_n\right) = -\left\langle \xi_n, \frac{1}{\tau}(u_n - u_{n-1})\right\rangle$$

and it is now that we invoke the crucial λ-convexity assumption for the potential:

$$\tau\left(\Psi\left(\frac{1}{\tau}(u_n - u_{n-1})\right) + \Psi^*\left(-\xi_n\right)\right) \leqslant \varphi(u_{n-1}) - \varphi(u_n) - \frac{\lambda}{2}|u_n - u_{n-1}|^2.$$

Now, we can rewrite the incremental Euler-Lagrange equation as an approximate equation as:

$$\overline{\xi}_\tau \in \partial\varphi(\overline{u}_\tau(t)), \quad 0 \in \partial\Psi\left(\frac{\mathrm{d}}{\mathrm{d}t}\hat{u}_\tau(t)\right) + \overline{\xi}_\tau(t)$$

and the inequality it generates, summed over n, as:

$$\varphi(\overline{u}_\tau(t)) + \int_0^t \left(\Psi\left(\frac{\mathrm{d}}{\mathrm{d}t}\hat{u}_\tau\right) + \Psi^*(-\overline{\xi}_\tau)\right) \mathrm{d}r \leqslant \varphi(u_0) - \frac{\tau\lambda}{2}\int_0^t \left|\frac{\mathrm{d}}{\mathrm{d}t}\hat{u}_\tau(r)\right|^2 \mathrm{d}r.$$

3.1 Energy-Dissipation principle

On a manifold, we define dissipation potentials as maps $\Psi : \mathrm{T}\mathcal{M} \longrightarrow \mathbb{R}$ such that each local $\Psi_u : \mathrm{T}_u\mathcal{M} \longrightarrow \mathbb{R}$ is itself a dissipation potential with $\partial\Psi_u(v) \subset \mathrm{T}_u^*\mathcal{M}$, and its dual $\Psi^* : \mathrm{T}^*\mathcal{M} \longrightarrow \mathbb{R}$ on the cotangent bundle is obtained at fixed $u \in \mathcal{M}$ by $\Psi_u^*(\xi)$ so that $\partial\Psi_u^*(\xi) \subset \mathrm{T}_u\mathcal{M}$. For instance,

$$\Psi_u(v) = \frac{1}{2}g(v,v)_u = \frac{1}{2}\Gamma_u(v)[v] \quad \Longleftrightarrow \quad \Psi_u^*(\xi) = \frac{1}{2}\xi_u[\mathrm{K}_u(\xi_u)] \tag{DP}$$

allow us to write both geometric forms of the gradient flow equation for smooth perturbations of convex potentials:

$$0 \in \partial\Psi_u(\dot{u}) + \mathrm{D}\varphi_u \subset \mathrm{T}_u^*\mathcal{M} \quad \Longleftrightarrow \quad \dot{u}_t \in \partial\Psi_u^*(-\mathrm{D}\varphi_u) \subset \mathrm{T}_u\mathcal{M}$$

and yield the optimality condition

$$\Psi_u(\dot{u}) + \Psi_u^*(-\xi) \geqslant -\xi[\dot{u}] \quad \text{with equality if and only if} \quad \xi \in \partial\Psi_u(\dot{u}) \tag{FOC}$$

providing the analogous of the energy identity:

$$\Psi_{u_t}(\dot{u}_t) + \Psi_{u_t}^*(-\mathrm{D}\varphi_{u_t}) \overset{\text{(FOC)}}{=} -\mathrm{D}\varphi_{u_t}[\dot{u}_t] = -\frac{\mathrm{d}}{\mathrm{d}t}\varphi(u_t)$$

which can be integrated to obtain the *energy-dissipation balance*

$$\varphi(u_t) + \int_0^t (\Psi_{u_r}(\dot{u}_r) + \Psi^*_{u_r}(-\mathrm{D}\varphi_{u_r}))\,\mathrm{d}r = \varphi(u_0)$$

for all $t > 0$. It simply states that the potential at a later time t, added to the dissipated energy in the time interval, give exactly the energy at the earlier time.

Definition 3.2. We say that a geodesic u_t is a *curve of maximal slope* if it satisfies:

$$\varphi(u(t_2)) + \int_{t_1}^{t_2} (\Psi_{u_r}(\dot{u}_r) + \Psi^*_{u_r}(-\mathrm{D}\varphi_{u_r}))\,\mathrm{d}r = \varphi(u(t_1)) \tag{EDB}$$

for all $t_1, t_2 \in [0, T]$ with $t_1 < t_2$.

This already provides us the time-incremental minimization scheme for step $\tau = T/N > 0$ and given initial value $u_0^\tau = u_0 \in \mathcal{M}$:

$$u_n^\tau = \operatorname*{arg\,min}_{u \in \mathcal{M}} \left(\tau \Psi_{u_{n-1}} \left(\frac{1}{\tau}(u - u_{n-1}) \right) + \varphi(u) \right),$$

that is, we are looking for solutions to the Euler-Lagrange equation:

$$0 \in \partial \Psi_{u_{n-1}} \left(\frac{1}{\tau}(u - u_{n-1}) \right) + \partial \varphi_{u_n}.$$

This, defining the right-continuous piecewise constant interpolant $\underline{u}_\tau$ for $t \in [n\tau, (n-1)\tau]$, similarly to and considered together with, the interpolants in Def. 1.4, we can rewrite as:

$$\overline{\xi}_\tau \in \partial \varphi_{\overline{u}_\tau(t)}, \quad 0 \in \partial \Psi_{\underline{u}_\tau(t)} \left(\frac{\mathrm{d}}{\mathrm{d}t} \hat{u}_\tau(t) \right) + \overline{\xi}_\tau \quad \text{for} \quad t \in [0, T]$$

and sending $\tau \downarrow 0$ in the following version of the 'Euler-Lagrange inequality':

$$\varphi(\overline{u}_\tau(t)) + \int_0^t \left(\Psi_{\underline{u}_\tau(t)} \left(\frac{\mathrm{d}}{\mathrm{d}t} \hat{u}_\tau \right) + \Psi^*_{\underline{u}_\tau(t)}(-\overline{\xi}_\tau) \right) \mathrm{d}r \leqslant \varphi(u_0) - \frac{\tau\lambda}{2} \int_0^t \left| \frac{\mathrm{d}}{\mathrm{d}t} \hat{u}_\tau(r) \right|_g^2 \mathrm{d}r$$

we get the *energy-dissipation inequality*

$$\varphi(u_T) + \int_0^T (\Psi_u(\dot{u}) + \Psi^*_u(-\mathrm{D}\varphi_u))\,\mathrm{d}t \leqslant \varphi(u_0). \tag{EDI}$$

LEMMA 3.3. A proper, lower semicontinuous, λ-convex functional $\varphi : \mathcal{M} \longrightarrow \mathbb{R}_\infty$ satisfies the *chain rule inequality*

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi(u_t) \leqslant \xi_t[\dot{u}_t] \tag{CRI}$$

if $\xi \in \mathbf{L}^1([0,T];\ \mathrm{T}^*\mathcal{M})$ and $\xi_t \in \partial\varphi_{u_t}$.

ENERGY-DISSIPATION PRINCIPLE. Each pair (u, ξ) satisfies (EDI) if and only if it satisfies the gradient-flow equation

$$0 \in \xi_t + \partial\Psi_{u_t}(u_t) \quad \text{and} \quad \xi_t \in \partial\varphi(u_t) \quad \text{for} \quad t \in [0,T] \tag{$\mathfrak{G}$}$$

and (EDB) for $0 \leqslant s < t \leqslant T$, namely, u is a curve of maximal slope.

PROOF. The backward direction is trivial. For the necessary condition we consider the "energy gap"

$$\Delta \equiv \varphi(u_0) - \varphi(u_T) - \int_0^T (\Psi_u(\dot{u}) + \Psi_u^*(-\mathrm{D}\varphi_u))\,\mathrm{d}t \overset{\text{(EDI)}}{\geqslant} 0$$

and combine the upper chain rule:

$$\varphi(u_T) - \varphi(u_0) \overset{\text{(CRI)}}{\leqslant} \int_0^T \xi_t[\dot{u}_t]\,\mathrm{d}t$$

with (EDI), to obtain:

$$\int_0^T (\Psi_u(\dot{u}) + \Psi_u^*(-\mathrm{D}\varphi_u) + \xi_t[\dot{u}_t])\,\mathrm{d}t = -\Delta \leqslant 0.$$

The integrand is nonnegative, hence $\Delta = 0$ and (EDI) is indeed equivalent to (EDB) on $[0,T]$ and the integrand vanishes, that is (FOC), integrated on $[s,t]$ yields the desired balance, or, equivalently, ($\mathfrak{G}$). ∎

Thus, we only need to prove existence of curves of maximal slope, and to do it we will use the Minimizing Movements scheme.

COROLLARY 3.4. If $\varphi : \mathcal{M} \longrightarrow \mathbb{R}_\infty$ is geodesically λ-convex and has compact sublevels, for all $u_0 \in \mathcal{M}$, there exists a curve of maximal slope $u(t)$ satisfying $u(0) = u_0$, ($\mathfrak{G}$) and (EDB) for $0 \leqslant s < t \leqslant T$, namely, $u(t)$ is a curve of maximal slope.

SKETCH OF THE PROOF. The argument is a replica of the proof given for Thm. 1.5, except that we show existence of minimizers for the functional

$$\Xi(u) \equiv \tau \Psi_{u_{n-1}}\left(\frac{1}{\tau}(u-u_{n-1})\right) + \varphi(u)$$

namely u_n such that $\forall w \in \mathcal{M}$:

$$\tau \Psi_{u_{n-1}}\left(\frac{1}{\tau}(u_n-u_{n-1})\right) + \varphi(u_n) \leqslant \tau \Psi_{u_{n-1}}\left(\frac{1}{\tau}(w-u_{n-1})\right) + \varphi(w)$$

and by the sum rule for subdifferentials:

$$0 \in \partial \Xi(u) = \partial \Psi_{u_{n-1}}\left(\frac{1}{\tau}(u_n-u_{n-1})\right) + \partial \varphi(u)$$

or, equivalently,

$$-\xi_n \in \partial \Psi_{u_{n-1}}\left(\frac{1}{\tau}(u_n-u_{n-1})\right) \quad \text{and} \quad \xi_n \in \partial \varphi(u_n), \quad n=1,\dots,N.$$

To get equicontinuity, the key observation is that we can always bound from below any dissipation potentials, because they are convex, by an increasing convex superlinear function $\psi : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ such that, for all $(u,v,\xi) \in S_r \times \mathcal{M} \times \mathrm{T}^*\mathcal{M}$: $\Psi_u(v) \geqslant \psi(|v|_g)$ and $\Psi_u(\xi) \geqslant \psi(|\xi|_g)$. In fact,

$$\tau \psi\left(\frac{1}{\tau} d_g(u_n-u_{n-1})\right) = \tau \Psi_{u_{n-1}}\left(\frac{1}{\tau}(u_n-u_{n-1})\right) \leqslant \varphi(u_n) - \varphi(u_{n-1})$$

i.e., summing over n:

$$\int_0^T \psi\left(\left|\frac{\mathrm{d}}{\mathrm{d}t}\hat{u}_\tau\right|_g\right) \mathrm{d}t = \sum_{n=1}^N \tau \psi\left(\frac{1}{\tau} d_g(u_n-u_{n-1})\right) \leqslant \varphi_0 - \varphi_{\min} < \infty$$

and by Arzelà-Ascoli, up to the extraction of a subsequence:

$$\hat{u}_\tau \longrightarrow u, \quad \overline{u}_\tau \longrightarrow u, \quad \underline{u}_\tau \longrightarrow u \quad \text{in} \quad \mathbf{C}^0([0,T];\, \mathcal{M}).$$

We derive (EDI) by passing to the limit $\tau \downarrow 0$ in the discrete approximate energy-dissipation inequality:

$$\varphi(\hat{u}_\tau(T)) + \int_0^T \left(\Psi_{\underline{u}_\tau}\left(\frac{\mathrm{d}}{\mathrm{d}t}\hat{u}_\tau\right) + \Psi_{\underline{u}_\tau}\left(-\overline{\xi}_\tau\right)\right) \mathrm{d}t \leqslant \varphi(u_0) - \frac{\lambda}{2}\tau \int_0^T \left|\frac{\mathrm{d}}{\mathrm{d}t}\hat{u}_\tau\right|_g^2 \mathrm{d}t$$

for $\overline{\xi}_\tau \in \partial\varphi(\overline{u}_\tau)$, to obtain (EDI), as $\overline{\xi}_\tau \longrightarrow \xi$ in $\mathbf{L}^1([0,T];\mathrm{T}^*\mathcal{M})$. Exploiting the upper chain rule and applying the Energy-Dissipation principle, the constructed pair (u,ξ) satisfies both ($\mathfrak{G}$) and (EDB). ∎

Notice that uniqueness is yielded by the λ-contractivity of solutions to (EVI$_\lambda$), which provides the very definition of our gradient flow.

Remark. The functional exploited in the sketch of the proof of Cor. 3.4 corresponds to the functional generating the Minimizing Movement scheme with the additional introduction of a convex function ψ, with *superlinear growth at infinity*, modeling dissipation, i.e.

$$\mathsf{F}_\tau(n,u,x) = \tau\psi\left(\frac{1}{\tau}\mathsf{d}(u,x)\right) + \varphi(x),$$

now considering the case of a general metric space $(\mathbf{X},\mathsf{d})$. This induces a sort of *action functional* $\mathsf{a}_\tau : \mathbf{X}\times\mathbf{X} \longrightarrow \mathbb{R}$, "*measuring the cost for moving between the two point in the amount of time* $\tau > 0$" [RS23], which would need not to depend on the metric in general:

$$\mathsf{F}_\tau(n,u,x) = \mathsf{a}_\tau(u,x) + \varphi(x).$$

This structure still "*preserves the natural splitting between the driving energy functional and a metric-like dissipation functional*". Restricting to actions that are symmetric and vanish only on the diagonal of $\mathbf{X}\times\mathbf{X}$, we get a full operational generalization of a metric space, imposing as third axiom the *concatenation inequality*:

$$\mathsf{a}_{\tau_1+\tau_2}(u_1,u_3) \leqslant \mathsf{a}_{\tau_1}(u_1,u_2) + \mathsf{a}_{\tau_2}(u_2,u_3) \quad \text{for all } u_1,u_2,u_3 \in \mathbf{X}, \tau_1,\tau_2 > 0,$$

to be seen as a dynamic version of the triangle inequality, so that, we recover the metric setting only when a is τ-invariant, in which case the action cost itself is the metric. Conversely, whenever we have an underlying distance, we recover the expression $\mathsf{a} = \tau\psi(\mathsf{d}/\tau)$. In the gradient flow case, we just have $\psi(r) = r^2/2$. In this context, action-minimizing curves play the role of geodesics.

CHAPTER II

Hamiltonian systems

As already said, for Newton's second law, a particle of mass $m \equiv 1$ in a potential $V(q)$ moves along a curve $q(t)$ in $\mathbb{R}^3$ such that $m\ddot{q} = -\nabla V(q)$. Introducing the momentum $p = \dot{q}$ and the *Hamiltonian*

$$H(q,p) = \frac{1}{2}m|p|^2 + V(q),$$

it is equivalent to *Hamilton's equations of motion*:

$$\begin{cases} \dot{q} = \dfrac{\partial H}{\partial p} \\ \dot{p} = -\dfrac{\partial H}{\partial q} \end{cases} \quad \text{i.e. letting } \xi = (p,q), \quad \dot{\xi} = J\nabla H(\xi) \equiv X_H(\xi), \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (\mathscr{H})$$

that is, $\dot{\xi}$ is an integral curve of the vector field $X_H \in \mathfrak{X}(\mathbb{R}^6)$.
Define the skew-symmetric bilinear form $\omega(v_1, v_2) = \langle v_1, Jv_2 \rangle$ on $\mathbb{R}^3 \times \mathbb{R}^3$, called *symplectic form*, so that for any $v \in \mathbb{R}^3$:

$$\omega(X_H(\xi), v) = \langle \nabla H(\xi), v \rangle.$$

Since we are interested in what happens to the equations of motion when we perform an arbitrary coordinate change $\eta \equiv f(\xi)$, let's consider any $f : \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \mathbb{R}^3$, and let the matrix A represent the differential of f, namely $A_{ij} = \partial\eta_i/\partial\xi_j$, and see how the equations of motion change:

$$\dot{\eta} = A\dot{\xi} = AJ\nabla_\xi H(\xi) = AJA^\mathsf{T}\nabla_\eta H(\xi(\eta))$$

meaning that, they are preserved under a coordinate transformation if and only if $AJA^\mathsf{T} = J$. We will call such coordinate changes under which the equations for η have Hamiltonian $\tilde{H}(\eta) = H(\xi(\eta))$, *canonical transformations*. [AM08, Preview]

Example (Rayleigh-Lorentz pendulum). For a small-angle pendulum, taking q to be the total horizontal displacement of the mass m from equilibrium and p its conjugate momentum:

$$H = \frac{1}{2m}p^2 + \frac{mg}{2\ell}q^2$$

and thus it behaves as an oscillator of amplitude A. At the turning points: $p = 0$ and thus $H = mgA^2/2\ell \equiv E$ and the locus of allowed (q, p) values of an orbit must lie on an ellipse

$$1 = \frac{p^2}{gm^2A^2/\ell} + \frac{q^2}{A^2} \equiv \frac{p^2}{b^2} + \frac{q^2}{a^2}$$

where a, b are respectively its semi-major and semi-minor axes. The area enclosed by this ellipse

$$\mathcal{A} = \pi ab = \pi A\left(Am\sqrt{\frac{g}{\ell}}\right) \equiv \pi A^2 m\omega$$

corresponds to the *Action*

$$J = \oint_{\text{orbit}} p\,\mathrm{d}q$$

because

$$J = \frac{2\pi E}{\omega} = 2\pi\frac{mgA^2}{2\ell}\sqrt{\frac{\ell}{g}} = \pi A^2 m\omega = \mathcal{A}.$$

In classical mechanics, the classical action J is an adiabatic invariant, meaning that it stays constant under "slow" transformations, such as the slow variation in the length of the rope of a pendulum ℓ, hence the area doesn't change as k varies.

We will see that, in a more general setting, transformations preserving volumes in the (p, q)-space are exactly canonical transformations: in this specific example, under a slow (adiabatic) change of ℓ we can view the evolution of the system in an extended space (including the length as a slowly-varying parameter), so that the net effect on the (q, p)-subspace is a sequence of infinitesimal canonical transformations, each of which preserves volume, that is, in two degrees of freedom, area.

1. Phase space

It is worth noticing that the right way of 'doing' classical mechanics is to treat both position and momentum equally as coordinates (of purely geometric nature) in some 'common room',

picturesquely called the *phase space*. This is the Hamiltonian formalism [‡] and, to make it more rigorous, we can identify each position-momentum tuple (q, p) with its complexification $q+\mathrm{i}p$. Using the notation $z^j, j = 1, \ldots, 2n$ to collectively denote the coordinates $q^i, p_i, i = 1, \ldots, n$, Hamilton's equations read:

$$\dot{z} = -2\mathrm{i}\frac{\partial H}{\partial \overline{z}}$$

where we mean to use the differential operator $\partial_{\overline{z}} \equiv (\partial_q + \mathrm{i}\partial_p)/2$. Denoting by $\mathbf{Z}$ the vector space of the (q, p)'s, by $\mathbf{Z}^*$ its dual and by $R : \mathbf{Z}^* \longrightarrow \mathbf{Z}$ the linear map represented by the *symplectic matrix*

$$J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix},$$

an Hamiltonian vector field is one of the form

$$X_H(z) = R \circ \mathrm{D}H(z) \in \mathfrak{X}(\mathbf{Z}),$$

meant as composition of operators i.e., using Einstein's summation convention:

$$X^i = R^{ij}\frac{\partial H}{\partial z^j},$$

resembling the very first definition given in ($\mathscr{H}$). We now build the *symplectic form* Ω as the bilinear form associated to $R^{-1} : \mathbf{Z} \longrightarrow \mathbf{Z}^*$, i.e.

$$\Omega(\alpha, \beta) \equiv \langle R^{-1}(\alpha), \beta\rangle = \langle \alpha, J\beta\rangle.$$

For a bilinear form $B : V \times W \longrightarrow \mathbb{R}$, represented by the matrix B_{ij}, we define the linear map $B^\flat : V \longrightarrow W^*$ as $B^\flat(v)(w) \equiv B(v, w)$ represented by $(B^\flat)_{ji} = B_{ij}$ [MR13, sect. 2.2], so that the definition of the symplectic form is equivalent to: $\Omega^\flat = R^{-1}$.
Denoting by $\Omega^\sharp \equiv (\Omega^\flat)^{-1} = R$, ($\mathscr{H}$) becomes $\dot{z} = X_H(z) = \Omega^\sharp \mathrm{D}H(z)$ and multiplying on both sides by $\Omega^\flat$ we again end up with:

$$\Omega(X_H(z), v) = \langle \nabla H(z), v\rangle.$$

[‡] As 'opposed' to the Lagrangian formalism, the bridge between the two being the Legendre transform (App. D).

Let us now consider a transformation of coordinates $(\tilde{q}, \tilde{p}) = \phi(q, p)$ given by $\phi : \mathbf{Z} \longrightarrow \mathbf{Z}$ i.e. $\tilde{z} = \phi(z)$ in the phase space, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{z}(t) = \frac{\mathrm{d}\phi^i}{\mathrm{d}z^j}\dot{z}^j \equiv A^i_j \dot{z}^j = A^i_j B^{jk}\frac{\partial H}{\partial z^k} = A^i_j B^{jk} A^\ell_k \frac{\partial H}{\partial \tilde{z}^\ell}$$

is Hamiltonian if and only if $A^i_j B^{jk} A^\ell_k = B^{i\ell}$, that being, $A \circ \Omega^\sharp \circ A^\mathsf{T} = \Omega^\sharp$. These specific linear transformations that preserve a skew-symmetric bilinear Ω, form a group of matrices, which was first given by Hermann Weyl the name of "symplectic".

Definition 1.1 (Symplectic group). $\mathrm{Sp}(n) = \{A \in \mathbf{M}_{2n}(\mathbb{R}) : A\Omega A^\mathsf{T} = \Omega\}$.

1.1 Hamiltonian flows

Definition 1.2. We say that a vector field $X_H \in \mathfrak{X}(\mathbf{Z})$ is a *Hamiltonian vector field* if

$$\Omega^\flat(X_H(z)) = \nabla H(z)$$

for all $z \in \mathbf{Z}$ and for some scalar *Hamiltonian*, or *energy*, *function* $H : \mathbf{Z} \longrightarrow \mathbb{R}$.

Now, $(\mathscr{H})$ becomes:

$$\dot{u}_t = X_H(u_t) = J\nabla H(u_t) \qquad (\mathscr{H})$$

in analogy with our notation for Gradient flows. We even say that the collection of maps $H_t : \mathbf{Z} \longrightarrow \mathbf{Z}$, $t \in [0, +\infty)$, such that for any $u_0 \in \mathbb{R}^d$, $H_0(u_0) \equiv u_0$ and the curve $u_t \equiv H_t(u_0)$ is the only $\mathbf{C}^1$ solution of $(\mathscr{H})$, is a *Hamiltonian*, or *phase*, flow of H.
Henri Poincaré discovered that any phase flow satisfies

$$\Omega(\mathrm{D}H_t\, v, \mathrm{D}H_t\, w) = \Omega(v, w) \quad \text{for all vectors } v, w \in \mathbf{Z}$$

namely, that the total areas spanned by the projections $v_t \equiv \mathrm{D}H_t\, v$ generated by the flow, is constant, i.e. it preserves the symplectic structure. [DR05]
Here, however, differently from the gradient flow case, the Hamiltonian does not decrease along its trajectories, like a potential φ would, but instead we have a rather important principle.

Conservation of energy. If u_t is an integral curve of X_H, meaning that it solves $(\mathscr{H})$, then $H(u_t)$ is constant along trajectories.

PROOF. Again, the chain rule is all we need:

$$\frac{\mathrm{d}}{\mathrm{d}t}H(u_t) = \langle \nabla H(u_t), \dot{u}_t \rangle \overset{\text{def}\Omega}{=} \Omega(X_H(u_t), \dot{u}_t) \overset{(\mathscr{H})}{=} \Omega(X_H(u_t), X_H(u_t)) = 0$$

because Ω is skew-symmetric. ∎

Thus, $H \circ H_t = H$, meaning that trajectories are confined to the level sets of the Hamiltonian.

REMARK. An arbitrary two-dimensional system of ODEs defined by $f, g \in \mathbf{C}^1$ is a Hamiltonian flow if and only if

$$\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} = 0$$

and, if the Hamiltonian has an isolated critical point at (x_*, y_*):

- if it is a local minimum/maximum of H, then (x_*, y_*) is a (nonlinear) center for the system;
- if it is a saddle point of H, then (x_*, y_*) is a saddle node for the system.

This result is helpful when proving that a system has a (nonlinear) center by building a Hamiltonian function for it. In fact, one cannot conclude anything from a (linear) center, as such a prediction would not be hyperbolic and the Hartman–Grobman theorem does not apply.

DEFINITION 1.3. Given any pair of functions $f, g \in \mathscr{F}(\mathbf{Z})$, their *Poisson bracket*[§]

$$\{f, g\} = \Omega(\nabla f, \nabla g) = \sum_{i=1}^{n} \frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i}$$

and the *Liouville operator*, or *Liouvillian*

$$\mathscr{L}_H f \equiv \{f, H\}.$$

We can thus rewrite $(\mathscr{H})$ as

$$\dot{u}_t = \mathscr{L}_H u_t$$

in terms of the Liouvillian, with formal solution

$$u(t) = e^{t\mathscr{L}_H} u(0).$$

[§]Despite the name, these were first 'discovered' by Lagrange, along with the invariance of the symplectic form, when he was already 73 years old. In fact, according to [MR13, Sect. 8.2], "*it is quite probable that Lagrange generously gave some of these bracket ideas to Poisson at this time. In any case, it is clear that Lagrange had a surprisingly large part of the symplectic picture of classical mechanics.*"

Example ([DR05]). For the 'classical' Hamiltonian

$$H(q,p) = V(q) + T(p), \quad \mathscr{L}_H = T'(p)\frac{\partial}{\partial q} - V'(q)\frac{\partial}{\partial p}$$

so $\mathscr{L}_H u_t = J\nabla H_{u_t}$. In the case of the classical harmonic oscillator $H(q,p) = (q^2+p^2)/2$, we get the linear system

$$\dot{u}_t = \mathscr{L}_H u_t = (p,q)J = Ju_t$$

with phase flow solution

$$u(t) = e^{tJ} = U(t), \quad \text{where } \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}.$$

If we try to approximate the Hamiltonian system above via first-order methods, we stumble into the same problem we met with gradient flow system: the explicit Euler method

$$\begin{pmatrix} q_{n+1} \\ p_{n+1} \end{pmatrix} = \tilde{U} \begin{pmatrix} q_n \\ p_n \end{pmatrix}, \quad \tilde{U} = \begin{pmatrix} 1 & \tau \\ -\tau & 1 \end{pmatrix}$$

does not conserve the total energy as the numerical solution produced by this method progressively diverges from the true solution, the oscillation amplitude grows larger, and both the system's energy and the corresponding phase space area increase consistently by a factor $H_{n+1}/H_n = 1+\tau^2$. Moreover, this is not a so-called *symplectic method*, those for which the symplectic invariance, $\tilde{U}J\tilde{U}^{\mathsf{T}} = J$, holds at the discrete level.
We can "*symplectify*" Euler, blending explicit and implicit methods:

$$q_{n+1} = q_n + \tau\partial_p H(q_n, p_{n+1}), \quad p_{n+1} = p_n - \tau\partial_q H(q_n, p_{n+1}).$$

For the harmonic oscillator

$$q_{n+1} = q_n + p_{n+1}\tau, \quad p_{n+1} = p_n - q_n\tau$$

i.e.

$$q_{n+1} = q_n(1-\tau^2) + p_n\tau, \quad \tilde{U} = \begin{pmatrix} 1-\tau^2 & \tau \\ -\tau & 1 \end{pmatrix} \in \mathrm{Sp}(2).$$

For separable, high-school physics, Hamiltonians,

$$\mathscr{L}_H u_t = \mathscr{L}_T u_t + \mathscr{L}_V u_t, \quad \mathscr{L}_T = \begin{pmatrix} T'(p) \\ 0 \end{pmatrix}, \quad \mathscr{L}_V = \begin{pmatrix} 0 \\ -V'(q) \end{pmatrix}$$

and we first evolve the state with respect to the potential energy and then kinetically, resembling symplectic Euler:

$$u_{n+1} = e^{\tau \mathscr{L}_T} e^{\tau \mathscr{L}_V} u_n.$$

Integrations in the positions and momenta are not in phase and, as a consequence, energy is not conserved yet. However, symplectic Euler behaves more nicely, as there is energy production only along trajectories that diverge from the equilibrium position: the potential energy peaks before the kinetic energy is minimized, as functions of q.

To reach the perfection of energy conservation, we can find a perturbed Hamiltonian function

$$H_\epsilon(p,q) = \sum_{k=0}^{\infty} \tau^k H_k$$

such that

$$e^{\tau \mathscr{L}_{H_\epsilon}} = e^{\tau \mathscr{L}_T} e^{\tau \mathscr{L}_V}$$

and compute each term H_k in the expansion using the Baker-Campbell-Hausdorff formula

$$\begin{aligned} \tau \mathscr{L}_H = \tau(\mathscr{L}_T + \mathscr{L}_V) + \frac{1}{2}\tau^2[\mathscr{L}_T + \mathscr{L}_V] \\ + \frac{1}{12}\tau^3 \left([\mathscr{L}_T, [\mathscr{L}_T, \mathscr{L}_V]] + [\mathscr{L}_V, [\mathscr{L}_V, \mathscr{L}_T]]\right) + \dots \end{aligned}$$

where the *commutator*

$$[\mathscr{L}_T, \mathscr{L}_V](f) \equiv \mathscr{L}_T(\mathscr{L}_V f) - \mathscr{L}_V(\mathscr{L}_T f) = \mathscr{L}_{\{T,V\}} f$$

so that

$$\begin{aligned} H_0 &= T + V, \quad H_1 = \frac{1}{2}\{T,V\} = \frac{\partial T}{\partial q}\frac{\partial V}{\partial p} - \frac{\partial T}{\partial p}\frac{\partial V}{\partial q}, \\ H_2 &= \frac{1}{12}\left(\{T,\{T,V\}\} - \{V,\{T,V\}\}\right), \quad \dots \end{aligned}$$

or, for the harmonic oscillator

$$H_0 = \frac{1}{2}(q^2 + p^2), \quad H_1 = -\frac{pq}{2}, \quad \dots$$

so, approximating just up to second order, choosing $\tau = 2$,

$$H_\epsilon(p, q) \approx \frac{1}{2}(p - q)^2$$

and updating $(q_{n+1}, p_{n+1}) = e^{2\mathscr{L}_T} e^{2\mathscr{L}_V}(q_n, p_n)$, we get the full energetic stability:

$$H_\epsilon(q_{n+1}, p_{n+1}) = H_\epsilon(q_n, p_n).$$

2. Symplectic geometry

Once we have solved the problem of making the flow equations independent of the choice of coordinates, by encoding the geometry of such choice in the very definition of the vector field giving our ODE, namely in the symplectic matrix J, we need just to make a small step to further generalize this concept to differentiable manifolds, endowed with an appropriate symplectic structure. [MR13, Sect. 5.1]

Definition 2.1. A *symplectic manifold* is a pair $(\mathcal{M}, \Omega)$ where $\mathcal{M}$ is a differentiable manifold and Ω is a closed [‡] two-form that is *non-degenerate*, i.e.

$$\Omega(z, \tilde{z}) = 0 \ \forall \tilde{z} \in \mathcal{M} \implies z = 0, \quad \text{or,} \quad \Omega^\flat \text{ is injective.}$$

The two-form assigns to any point $p \in \mathcal{M}$ a skew-symmetric bilinear form on the tangent space, $\Omega_p : \mathrm{T}_p\mathcal{M} \times \mathrm{T}_p\mathcal{M} \longrightarrow \mathbb{R}$, and because it is non-degenerate, it defines an isomorphism $\Omega_p^\flat : \mathrm{T}_p\mathcal{M} \longrightarrow \mathrm{T}_p^*\mathcal{M}$, between coordinates and momenta.

Notice that a vector space compatible with an anti-symmetric non-degenerate form must necessarily be even-dimensional and, because the dimension of the manifold is the same as all of its tangent spaces, every symplectic manifold must be even-dimensional (physically, this ensures that each position q^j has its conjugate momentum p_j) and we can always find local canonical coordinates for which $\Omega = \mathrm{d}q^i \wedge \mathrm{d}p_i$.

[‡] We will see later that the closedness of Ω is necessary to ensure that the induced Poisson bracket satisfies the Jacobi identity and the Hamiltonian flows consist of canonical transformations.

Moreover, any $2n$-symplectic manifold is oriented by the so-called *Liouville form*[§]

$$\Lambda = \frac{(-1)^{n(n-1)/2}}{n!} \underbrace{\Omega \wedge \cdots \wedge \Omega}_{n \text{ copies}}$$

that in canonical coordinates: $\Lambda = \mathrm{d}q^1 \wedge \cdots \wedge \mathrm{d}q^n \wedge \mathrm{d}p_1 \wedge \cdots \wedge \mathrm{d}p_n$, which can be thought of as a volume element. Thus, any symplectic manifold is orientable and has a well-defined notion of volume, as one can integrate with respect to the Liouville form and use it as a volume measure.

Definition 2.2. Let $(\mathcal{M}, \Omega)$ and $(\mathcal{N}, \tilde{\Omega})$ be two symplectic manifolds, then, the smooth map $\phi : \mathcal{M} \longrightarrow \mathcal{N}$ is a *symplectomorphism*, or *canonical transformation*, if

$$\phi^*\Omega = \tilde{\Omega} \quad \text{i.e.} \quad \Omega_p(v, w) = \tilde{\Omega}_{\phi(p)}(\mathrm{D}\phi_p(v), \mathrm{D}\phi_p(w)).$$

It follows that canonical transformations preserve orientation and because, by the naturality of the exterior product, $\phi^*(\omega \wedge \eta) = \phi^*\omega \wedge \phi^*\eta$, they also preserve volume: $\phi^*\Lambda = \Lambda$, thanks to the expression of Λ in local coordinates.

Remark. The corresponding Riemannian notion is that of an *isometry* $(\mathcal{M}, g) \xrightarrow{\phi} (\mathcal{N}, \tilde{g})$, that is, a transformation inducing the pullback metric $\phi^* g = \tilde{g}$ on the destination manifold.

2.1 Poisson structures

In introducing symplectic/Poisson structures, we are mainly interested in canonical transformations, "*which leave the respective structures invariant*" and are "*the key to the implementation of the structure of reversible dynamics in* time-discretized dynamical systems, *clearly important for computer simulations: every discrete time step must be described by a canonical transformation*", as remarked by Öttinger. [Ött05, App. B.2]

Definition 2.3. A *Poisson manifold* is a differentiable manifold $\mathcal{P}$ endowed with a bilinear, antisymmetric bracket operation $\{\cdot, \cdot\} : \mathscr{F}(\mathcal{P}) \times \mathscr{F}(\mathcal{P}) \longrightarrow \mathscr{F}(\mathcal{P})$ satisfying the:

- *Jacobi identity*: $\{A, \{B, C\}\} + \{C, \{A, B\}\} + \{B, \{C, A\}\} = 0$;
- *Leibniz rule*: $\{A \circ B, C\} = A \circ \{B, C\} + \{A, C\} \circ B$.

[§]Differential forms, their pullbacks and its interaction with exterior products of them, and the orientation they induce are discussed in App. C.

The binary operation basically mimics the behavior, with products, of a derivation, thanks to the Leibniz rule, which must be intended to hold both with respect to the first and second variable. We can represent the Poisson bracket as follows:

$$\{F, G\}(p) \equiv \mathrm{D}F(p) \circ \mathrm{L}(p)\mathrm{D}G(p) \tag{PB}$$

where $\mathrm{L}(p) : \mathrm{T}^*_p\mathcal{P} \longrightarrow \mathrm{T}_p\mathcal{P}$ is the skew-symmetric linear *Poisson operator* [Ött05, App. B.2]. It is the most convenient tool for the introduction of dynamical systems because it directly maps differentials (of Hamiltonians) into vector fields. The Jacobi identity can be re-expressed in coordinates:

$$L^{i\ell}\frac{\partial L^{jk}}{\partial \xi^{(\ell)}} + L^{j\ell}\frac{\partial L^{ki}}{\partial \xi^{\ell}} + L^{k\ell}\frac{\partial L^{ij}}{\partial \xi^{(\ell)}} = 0$$

for $L^{\alpha\beta} = \mathrm{D}\xi^\alpha \circ L(\mathrm{D}\xi^\beta)$.

Remark. Defining for $F, G \in \mathscr{F}(\mathcal{P})$ the Poisson bracket:

$$\{F, G\}(z) \equiv \Omega(X_F(z), X_G(z))$$

corresponding to the Poisson operator $L^{\alpha\beta} \equiv (\Omega^{-1})^{\beta\alpha}$, every symplectic manifold is evidently (and trivially) a Poisson manifold. For this bracket, the Jacobi identity follows from the integrability condition: $\mathrm{d}\Omega = 0$, and that's why we need to impose the symplectic form to be closed. In canonical coordinates:

$$\{F, G\} = \frac{\partial F}{\partial \xi^j}\frac{\partial G}{\partial \xi^*_j} - \frac{\partial F}{\partial \xi^*_j}\frac{\partial G}{\partial \xi^j}$$

which resembles the usual form of the Poisson bracket from classical mechanics:

$$\{A, B\} \equiv \frac{\partial A}{\partial q^j}\frac{\partial B}{\partial p_j} - \frac{\partial A}{\partial p_j}\frac{\partial B}{\partial q^j}.$$

On a symplectic manifold, ($\mathscr{H}$) reads:

$$\dot{u}_t = \mathrm{L}(u_t)\mathrm{D}H(u_t) \in \mathrm{T}_{u_t}\mathcal{M} \quad \Longleftrightarrow \quad \Omega(u_t)\dot{u}_t = \mathrm{D}H(u_t) \in \mathrm{T}^*_{u_t}\mathcal{M}. \tag{$\mathscr{H}$}$$

Hence, an *Hamiltonian system* is a triple $(\mathcal{M}, \Omega, H)$, where the Poisson structure, employed to characterize reversible dynamics, serves as an abstract extension of classical mechanics. In this framework, the energy H emerges as the fundamental generator of time evolution for

reversible systems. The defining properties of anti-symmetry and the Jacobi identity (which can be thought of as the criterion for the possibility of introducing canonical coordinates) encapsulate the core principles of the classical Poisson bracket, ensuring a formulation that inherently preserves reversibility. Furthermore, Poisson structures encompass a broader class of mathematical frameworks than symplectic structures, as they permit degeneracies that are explicitly ruled out in the definition of symplectic forms. This intrinsic degeneracy plays a fundamental role in incorporating entropy, thereby enabling the formulation of irreversible contributions within dynamical systems.

On a Poisson manifold, there exists a unique vector field $X_H \in \mathfrak{X}(\mathcal{P})$: $X_H(F) = \{F, H\}$, because for any Hamiltonian $H \in \mathscr{F}(\mathcal{P})$, the map $F \longmapsto \{F, H\}$ satisfies the Leibniz rule and thus uniquely determines the *Hamiltonian vector field* X_H.

LEMMA 2.4. If u_t is a flow on $\mathcal{P}$ and $H \in \mathscr{F}(\mathcal{P})$, then for every $F \in \mathscr{F}(U)$, with U open in $\mathcal{P}$:

$$\frac{\mathrm{d}}{\mathrm{d}t}(F \circ u_t) = \{F \circ u_t, H\}$$

if and only if u_t is the flow of X_H.

PROOF. Let $z \in \mathcal{P}$, then

$$\frac{\mathrm{d}}{\mathrm{d}t}F(u_t(z)) = \mathrm{D}F_{u_t(z)}\left(\frac{\mathrm{d}}{\mathrm{d}t}u_t(z)\right)$$

and

$$\{F, H\}(u_t(z)) = \mathrm{D}F_{u_t(z)}(X_H(u_t(z)))$$

so, to get equality of the two expressions, we need:

$$\mathrm{D}F_{u_t(z)}\left(\frac{\mathrm{d}}{\mathrm{d}t}u_t(z)\right) - \mathrm{D}F_{u_t(z)}(X_H(u_t(z))) = \mathrm{D}F_{u_t(z)}\left(\frac{\mathrm{d}}{\mathrm{d}t}u_t(z) - X_H(u_t(z))\right) = 0$$

that being, we need the tangent vector

$$v \equiv \frac{\mathrm{d}}{\mathrm{d}t}u_t(z) - X_H(u_t(z)) \in \mathrm{T}_z\mathcal{P}$$

to vanish when coupled with any differential and thus, by the Hahn-Banach theorem, $v = 0$ or,

$$\frac{\mathrm{d}}{\mathrm{d}t}u_t(z) = X_H(u_t(z))$$

i.e. u_t is the integral curve of X_H with initial condition z, or, u_t is the flow of X_H.
Conversely, if u_t is the flow of X_H, then $X_H(u_t) = \mathrm{D}u_t(X_H)$ and by the chain rule:

$$\frac{\mathrm{d}}{\mathrm{d}t}F(u_t) = \mathrm{D}F_{u_t}(X_H(u_t)) = \mathrm{D}F_{u_t}\mathrm{D}u_t(X_H) = \mathrm{D}(F \circ u_t)(X_H) = \{F \circ u_t, H\}$$

and moreover $H \circ u_t = H$. ∎

We can rewrite ($\mathscr{H}$) as:

$$\dot{u}_t = \{u_t, H\} \tag{$\mathscr{H}$}$$

and the Hamiltonian being constant along the integral curves of X_H as the following statement of the conservation of energy. [MR13, Sect. 10.2]

Proposition 2.5. F is constant along the integral curves of X_H if and only if $\{F, H\} = 0$.

We shall call *Casimir functions* $C \in \mathscr{F}(\mathcal{P})$ those that are constant along the flow of all Hamiltonian vector fields:

$$\{C, F\} = 0 \text{ for all } F \in \mathscr{F}(\mathcal{P})$$

and generate the trivial dynamics: $X_C = 0$.

Example. If $\mathcal{P}$ is symplectic, because $X_C = 0 \implies \mathrm{d}C = 0$, Casimir functions are locally constant (on the connected components of $\mathcal{P}$).

Now, $\phi : \mathcal{P} \longrightarrow \tilde{\mathcal{P}}$ is a canonical transformation if and only if $\{F, G\} \circ \phi = \{F \circ \phi, G \circ \phi\}$.

CHAPTER III

GENERIC

1. Reversible and irreversible thermodynamics

In thermodynamics, the variation in the internal energy of a system obeys [Ött05]:

$$\begin{aligned} \mathrm{D}\mathcal{U} &= \underbrace{\mathrm{D}W}_{\text{doing work on the system}} + \underbrace{\mathrm{D}Q}_{\text{transferring heat to the system}} \\ &= \underbrace{-p\mathrm{D}V}_{\text{changing volume under a certain pressure}} + \underbrace{T}_{\text{temperature,}}\ \underbrace{\mathrm{D}S}_{\text{entropy}}. \end{aligned}$$

This is Gibbs' fundamental equation of thermodynamics, which splits contribution in a reversible (the former) and an irreversible (the latter) process. By means of a Legendre transform [‡], we get the *Helmholtz free energy*

$$F = \mathcal{U}[T] = \mathcal{U} - TS$$

and the *Gibbs free energy*

$$G = \mathcal{U}[T, P] = \mathcal{U} - TS + PV$$

that serve as generalized potentials for measuring "how far" the system is from equilibrium in terms of the "slopes", that are the conjugate variables. In fact,

$$\mathrm{D}F = -S\mathrm{D}T - P\mathrm{D}V, \quad \mathrm{D}G = -S\mathrm{D}T + V\mathrm{D}P$$

and a set of equations of state is associated with each of the potentials:

$$S(T,V) = \left(\frac{\partial F}{\partial T}\right)_V, \quad P(T,V) = \left(\frac{\partial F}{\partial V}\right)_T$$

[‡] See Appendix A.

and since

$$\frac{\partial^2 F}{\partial T \partial V} = \frac{\partial^2 F}{\partial V \partial T}$$

one can obtain the so-called *Maxwell's relations*:

$$\left(\frac{\partial P}{\partial T}\right)_V = \left(\frac{\partial S}{\partial V}\right)_T .$$

"*These interrelations are heavily used in thermodynamic calculations. Maxwell, whose turn of mind was geometrical, obtained them, and similar relations, from the properties of a parallelogram; he concluded*", by his own words in *The Theory of Heat*, "*that such a relation* «is a merely geometrical truth, and does not depend upon thermodynamic principles», *whereas he offered the formulation in the language of differential calculus only as a footnote.*" [Ött05, A.5]

The fundamental assumption here is the *Maximum entropy postulate*, that there exists a function, called the Entropy, of the extensive parameters, defined for all equilibrium states, and having the following property: the values assumed by the extensive parameters in the absence of a constraint are those that maximize the entropy over the manifold of constrained equilibrium states.

A consequence of this postulate is that the hypersurface $S = S(U, X_1, \ldots, X_n)$ in the thermodynamic configuration space should have the property that it lies everywhere below its tangent planes. This implies conditions for the derivatives of the fundamental relation. For example, in the case $S = S(U, V, N)$, we have

$$\left(\frac{\partial^2 S}{\partial U^2}\right)_{V,N} \leqslant 0, \left(\frac{\partial^2 S}{\partial V^2}\right)_{U,N} \leqslant 0 \quad \text{and} \quad \frac{\partial^2 S}{\partial U^2}\frac{\partial^2 S}{\partial V^2} - \left(\frac{\partial^2 S}{\partial U \partial V}\right)^2 \geqslant 0.$$

The positiveness of physical quantities like specific heats, compressibilities, and expansion coefficients is a direct consequence of these thermodynamic stability requirements.

2. Formalism

We want to develop a common framework for a time-evolution equation encoding both a reversible and irreversible component:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \dot{x}_{\text{rev.}} + \dot{x}_{\text{irr.}}$$

in analogy with Gibbs' fundamental equation.
While Poisson structures provide a mathematical formalism to study reversible processes as Hamiltonian systems, also gradient flows naturally arise in the study of irreversible thermodynamical processes.

Example (Ginzburg-Landau equation). Identifying the constant parameter T with the equilibrium temperature of a final time independent state defined by a vanishing gradient of the Helmholtz free energy $F(x,T)$, a system tries to minimize its free energy, or to maximize its entropy, by following the gradients of these generators. However, according to Ginzburg and Landau, the system does not develop directly in the direction of the gradients, but frictional effects are rather taken into account:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{1}{T} M \frac{\partial F(x,T)}{\partial x} = M \frac{\partial S}{\partial x}$$

through the matrix M, here playing the role of the Onsager operator.

The first proposal for a general evolution equation for beyond-equilibrium systems [GÖ97], known as GENERIC (general equation for the non-equilibrium reversible-irreversible coupling) is:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mathrm{L}(x) \frac{\partial E(x)}{\partial x} + \mathrm{M}(x) \frac{\partial S(x)}{\partial x}$$

where x completely describes the non-equilibrium system, "*the functionals $E(x)$ and $S(x)$ are the total energy and entropy expressed in terms of the state variables x, and $L(x)$ and $M(x)$ are the Poisson and friction matrices representing geometric structures and dissipative material properties in terms of linear operators*", in the words of Öttinger. [Ött05]
This is however rather rough. But we have developed all the mathematical tools needed to study both reversible and irreversible dynamics, respectively through the introduction of the Poisson and Onsager operators, and just need to (literally) "sum up" ($\mathscr{H}$) together with ($\mathfrak{G}$). What we obtain is a *GENERIC system*, which is just a dissipative-Hamiltonian system, i.e. a quintuple $\{\mathcal{M}, \Gamma, \varphi, \Omega, H\}$, referring to the notions, introduced in Chapters I and II, of Riemannian and Poisson structures and the dynamics they induce.[§]

[§] Also see Appendix C.3 for a fully differential geometric formulation of the GENERIC evolution, in terms of bundle isomorphisms induced by the metric and symplectic forms.

Definition 2.1. A *GENERIC flow* is a flow u_t with time-evolution:

$$\dot{u}_t \in \mathrm{L}(u_t)\mathrm{D}H_{u_t} - \mathrm{K}(u_t)\partial\varphi(u_t) \tag{GENERIC}$$

for some Hamiltonian $H \in \mathscr{F}(\mathcal{M})$, called *energy*, and a (geodesically) λ-convex potential $\varphi : \mathcal{M} \longrightarrow \mathbb{R}$, called *entropy*, satisfying the *non-degeneracy*, or *non-interaction*, *conditions*

$$\Omega(u_t)\partial\varphi(u_t) = \Gamma(u_t)\mathrm{D}H(u_t) = 0 \in \mathrm{T}_{u_t}\mathcal{M}. \tag{$\star$}$$

The definition (PB) implements in ($\mathscr{H}$) the idea that the reversible contribution to time evolution should be well structured, namely, in the form of a Hamiltonian dynamical system in the geometrical setting of Poisson manifolds. Analogously, we can define the Lie "*dissipative*" *bracket*

$$[F, G](u) \equiv \mathrm{D}F(u) \circ \mathrm{K}(u)\mathrm{D}G(u), \quad F, G \in \mathscr{F}(\mathcal{M})$$

with which the operator version of GENERIC can be rewritten in the equivalent bracket form [‡]

$$\dot{F} = \{F, H\} + [F, \varphi].$$

This, together with ($\star$), yields a rather strong formulation of energy conservation:

$$\frac{\mathrm{d}}{\mathrm{d}t}H = \{H, H\} + [H, \varphi] = 0$$

and of the very hallmark of reversibility, that is, entropy production:

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi = [\varphi, \varphi] \geqslant 0$$

by non-negativity of the (dissipative) Lie bracket. In fact, the non-degeneracy requirements are equivalent to the entropy being always a Casimir function:

$$\{\varphi, H\} = 0, \quad \text{or, in coordinates,} \quad \forall j : \frac{\partial\varphi}{\partial q^i}L^{ij} = 0$$

and the Hamiltonian being automatically conserved:

$$\dot{H} = [H, H] = 0$$

[‡] In this formulation of the GENERIC time-evolution of some arbitrary smooth observable, we consider 'plus' a gradient flow to be coherent with the physical interpretation of an ever-ascending gradient for the entropy.

and respectively correspond to the First and Second Principle of Thermodynamics. They express the reversible nature of the L-contribution to the dynamics (the functional form of the entropy is such that it cannot be affected by the operator generating the reversible dynamics) and the conservation of the total energy of an isolated system by the irreversible contribution to the dynamics: all dissipated mechanical energy is converted into internal energy.
Thanks to the geometric nature given to the formalism, it is invariant by coordinate changes, namely, under isometric and canonical transformations.

Remark. If $\varphi \in \mathscr{F}(\mathcal{M})$, then $\partial\varphi = \{\mathrm{D}\varphi\}$ and (GENERIC) becomes

$$\dot{u}_t = \mathrm{L}_{u_t}\mathrm{D}H_{u_t} - \mathrm{K}_{u_t}\mathrm{D}\varphi_{u_t}.$$

Suppose $\mathcal{M} = \mathbb{R}^2$, then the standard Euclidean metric and Poisson structure are:

$$\mathrm{L} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad \mathrm{K} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

i.e. any GENERIC system takes the form

$$\dot{x} = -\frac{\partial\varphi}{\partial x} + \frac{\partial H}{\partial y},$$
$$\dot{y} = -\frac{\partial\varphi}{\partial y} - \frac{\partial H}{\partial x}.$$

We can generalize the formalism in Euclidean non-even dimension by just letting L be any skew-symmetric *Poisson matrix* and K some symmetric *friction matrix*.

3. Discrete gradients

Supposing we had not developed any integrator theory for gradient flows and Hamiltonian systems, there are two immediate abstract time-stepping algorithms[§] that preserve both the dissipative and conservative structure of GENERIC, without resorting to minimizing movements or symplectic integrators.

[§] First proposed by [Rom09].

The first idea is to couple both structures in a unique *monolithic integrator*

$$\frac{u_{n+1}-u_n}{\tau} = \mathrm{L}_{u_\tau}\mathrm{D}H_{u_\tau} - \mathrm{K}_{u_\tau}\mathrm{D}\varphi_{u_\tau}$$

where $u_\tau = (u_{n+1}-u_n)/2$. Namely, we want our discrete gradients to be at least second-order approximations to their continuous counterparts at the midpoint $\tau/2$. Hence, we impose the non-interaction condition

$$\mathrm{L}_{u_\tau}\mathrm{D}\varphi_{u_\tau} = \mathrm{K}_{u_\tau}\mathrm{D}H_{u_\tau} = 0$$

and directions

$$\mathrm{D}H_{u_\tau}(u_{n+1}-u_n) = H_{u_{n+1}} - H_{u_n}, \quad \mathrm{D}\varphi_{u_\tau}(u_{n+1}-u_n) = \varphi_{u_{n+1}} - \varphi_{u_n}. \tag{dir.}$$

This may seem rather arbitrary, but, considering the solely-conservative system

$$\frac{u_{n+1}-u_n}{\tau} = X_H(u_\tau)$$

which is nothing more than the discretization of ($\mathscr{H}$), then it was noticed [Gon96] that, assuming (dir.),

$$H_{u_{n+1}} - H_{u_n} = \mathrm{D}H_{u_\tau}(u_{n+1}-u_n) \overset{(\mathscr{H})}{=} \tau\mathrm{D}H_{u_\tau}(u_{n+1}-u_n)X_H(u_\tau)$$

and it's enough to take an orthogonal Hamiltonian vector field for H to be the searched integral. Robustly enough, energy is conserved:

$$\begin{aligned} H_{u_{n+1}} - H_{u_n} &= \mathrm{D}H_{u_\tau}(u_{n+1}-u_n) \\ &= \tau\mathrm{D}H_{u_\tau}(\mathrm{L}_{u_\tau}\mathrm{D}H_{u_\tau} - \mathrm{K}_{u_\tau}\mathrm{D}\varphi_{u_\tau}) = 0 \end{aligned}$$

and entropy is never-decaying:

$$\begin{aligned} \varphi_{u_{n+1}} - \varphi_{u_n} &= \mathrm{D}\varphi_{u_\tau}(u_{n+1}-u_n) \\ &= \tau\mathrm{D}\varphi_{u_\tau}(\mathrm{L}_{u_\tau}\mathrm{D}H_{u_\tau} - \mathrm{K}_{u_\tau}\mathrm{D}\varphi_{u_\tau}) \geqslant 0. \end{aligned}$$

Alternatively, we can separately discretize the energetic gradient vector field generated by the reversible Poissonian dynamics and entropy gradient flow generated by the irreversible

Onsager dynamics, by means of a *staggered integrator*

$$\frac{\tilde{u}_{n+1}-u_n}{\tau}=\mathrm{L}_{\tilde{u}_\tau}\mathrm{D}H_{\tilde{u}_\tau},\quad \frac{u_{n+1}-\tilde{u}_{n+1}}{\tau}=\mathrm{K}_{v_n}\mathrm{D}\varphi_{v_n}$$

where $\tilde{u}_\tau=(\tilde{u}_{n+1}-u_n)/2$, $v_n=(u_n-\tilde{u}_n)/2$ and $\tilde{u}_n$ represents the state of the system after it evolves reversibly, and both energy and entropy are conserved. The irreversible iterative updates, at each step, conserve the energy and dissipate the negative entropy, resulting in a global robustness.

3.1 Minimizing Movements approach

We consider again the dissipation potentials (DP), so that $\partial\Psi_u(\partial\varphi_u)=\mathrm{K}_u\partial\varphi_u$ and we can rewrite (GENERIC) as

$$\dot{u}_t\in\mathrm{L}_{u_t}\mathrm{D}H_{u_t}-\partial\Psi_{u_t}(\partial\varphi_{u_t}) \tag{GENERIC}$$

and thus, as the inequality [JST19]:

$$-\varphi(u_t)+\int_0^t\Psi^*_{u_t}(\dot{u}_t-\mathrm{L}_{u_t}\mathrm{D}H_{u_t})\,\mathrm{d}r+\int_0^t\Psi_{u_t}(-\partial\varphi_{u_t})\,\mathrm{d}r+\varphi(u_0)\leqslant 0 \tag{GENERIC$_\geqslant$}$$

because

$$\begin{aligned}\text{(GENERIC)}&\iff\dot{u}_t-\mathrm{L}_{u_t}\mathrm{D}H_{u_t}\in\partial\Psi_{u_t}(-\partial\varphi_{u_t})\\&\overset{\text{(FOC)}}{\iff}\Psi_{u_t}(\dot{u}_t-\mathrm{L}_{u_t}\mathrm{D}H_{u_t})+\Psi^*_{u_t}(-\partial\varphi_{u_t})-(\dot{u}_t-\mathrm{L}_{u_t}\mathrm{D}H_{u_t})[-\partial\varphi_{u_t}]\leqslant 0\\&\overset{(\star)}{\iff}\frac{\mathrm{d}}{\mathrm{d}t}\varphi\circ u+\Psi_{u_t}(\dot{u}_t-\mathrm{L}_{u_t}\mathrm{D}H_{u_t})+\Psi^*_{u_t}(-\partial\varphi_{u_t})\leqslant 0\overset{\text{(FOC)}}{\iff}\text{(GENERIC}_\geqslant\text{)}.\end{aligned}$$

In the spirit of De Giorgi's Energy-Dissipation Principle, (GENERIC) can equivalently be formulated, in terms of the scalar equation [JST22]:

$$\varphi(u_t)+\int_0^t\Psi_{u_r}(\dot{u}_r-\mathrm{L}_{u_r}\mathrm{D}H_{u_r})\,\mathrm{d}r+\int_0^t\Psi^*_{u_r}(-\partial\varphi_{u_r})\,\mathrm{d}r=\varphi(u_0) \tag{EDP}$$

for all $t\in[0,T]$, because

$$\begin{aligned}\text{(GENERIC)}&\overset{\text{(FOC)}}{\iff}\Psi_{u_t}(\dot{u}_t-\mathrm{L}_{u_t}\mathrm{D}H_{u_t})+\Psi^*_{u_t}(-\partial\varphi_{u_t})-(\dot{u}_t-\mathrm{L}_{u_t}\mathrm{D}H_{u_t})[-\partial\varphi_{u_t}]=0\\&\overset{(\star)}{\iff}\frac{\mathrm{d}}{\mathrm{d}t}\varphi\circ u_t+\Psi_{u_t}(\dot{u}_t-\mathrm{L}_{u_t}\mathrm{D}H_{u_t})+\Psi^*_{u_t}(-\partial\varphi_{u_t})=0\iff\text{(EDP)}.\end{aligned}$$

In analogy with the gradient flow case, we will tackle a discrete version of ($\star$), extending De Giorgi's Minimizing Movements scheme: given a time partition $0 = t_0 < t_1 < \cdots < t_N = T$ with diameter τ, consider the time-discrete trajectory $\{u_n\}_{n=0}^N \in \mathcal{M}^{N+1}$ from the initial state u_0, and subsequently perform the time-incremental minimization of the functional

$$\Xi_{\tau,u_{n-1}}(u) \equiv \varphi(u) + \tau\Psi^*_{u^\theta}\left(\frac{u-u_{n-1}}{\tau} - \mathrm{L}_u \mathrm{D}H_{u^\theta}\right) + \tau\Psi_{u^\theta}(-\partial\varphi_u) - \varphi(u_{n-1})$$

where $u^\theta \equiv \theta u + (1-\theta)u_{n-1}$ for some fixed $\theta \in [0,1]$.
For instance, trajectories solving the Euler scheme:

$$\frac{u_n - u_{n-1}}{\tau} = \mathrm{L}_{u_n}\mathrm{D}H_{u_n^\theta} - \mathrm{K}_{u_n^\theta}\partial\varphi_{u_n}$$

enable us to solve the inequality

$$\Xi_{\tau,u_{n-1}}(u_n) \leqslant \tau\Psi^*_{u_n^\theta}\left(\frac{u_n-u_{n-1}}{\tau} - \mathrm{L}_{u_n}\mathrm{D}H_{u_n^\theta}\right) + \tau\Psi_{u_n^\theta}(-\partial\varphi_{u_n}) - \partial\varphi_{u_n}[u_n - u_{n-1}] \overset{\text{(FOC)}}{=} 0$$

and this *a priori* bound will compose the assumption behind the convergence of discrete solution to the variational problem.

Theorem 3.1 ([JST22]). Under the above assumptions and the usual convention for the variational interpolants, if

$$\lim_{\tau\downarrow 0}\sum_{n=1}^N \Xi_{\tau,u_{n-1}}(u_n) = 0,$$

each $\{\hat{u}_n\}_n \in \mathbf{W}^{1,2}([0,T];\ \mathcal{M})$, each conservative interpolant $\{\mathrm{L}_{\overline{u}_n}\mathrm{D}H_{\overline{u}_n^\theta}\}_n \in \mathbf{L}^2([0,T];\ \mathcal{M})$ and each dissipative slope $\{\partial\varphi_{\overline{u}_n}\}_n \in \mathbf{L}^2([0,T];\mathrm{T}^*\mathcal{M})$ is bounded, then, up to extracting a subsequence, there is the (weak) convergence: $\hat{u}_n \longrightarrow u \in \mathbf{W}^{1,2}([0,T];\ \mathcal{M})$.

Sketch of the proof. Our discrete sequences of solutions converge (weakly, upon extracting subsequences) and fulfill

$$\sum_{n=1}^N \Xi_{\tau,u_{n-1}}(u_n) \geqslant \varphi(\overline{u}(t)) + \int_0^t \Psi^*_{\overline{u}^\theta}\left(\frac{\mathrm{d}}{\mathrm{d}t}\hat{u} - \mathrm{L}_{\hat{u}}\mathrm{D}H_{\overline{u}^\theta}\right)\mathrm{d}r + \int_0^t \Psi_{\overline{u}^\theta}(-\partial\varphi_{\overline{u}})\,\mathrm{d}r - \varphi(u(0))$$

for all $t \in [0,T]$. By lower semicontinuity of the sum of the entropy-production potential and its dual, namely, if $y_n \xrightarrow{\mathcal{M}} y, \eta_n \xrightarrow{\mathcal{M}} \eta$ and $\xi_n \xrightarrow{\mathrm{T}^*\mathcal{M}} \xi$, then

$$\liminf_{n\longrightarrow+\infty}(\Psi^*_{y_n}(\eta_n) + \Psi_{y_n}(\xi_n)) \geqslant \Psi^*_y(\eta) + \Psi_y(\xi)$$

since refining the partition ($\tau \downarrow 0$) is equivalent to adding more and more interpolants (i.e. $n \uparrow +\infty$), our inequality converges to (GENERIC$_{\geqslant}$). ∎

3.2 GENERIC Formalism Informed Neural Networks

Given a dynamical system

$$\dot{u}_t = F(u_t), \quad u_t \in \Omega \subseteq \mathbb{R}^d, \quad t \in (0, T],$$

we want to approximate F through sampled data, by employing a Neural Network $F^{\text{NN}}(u, \vartheta)$, parametrized by ϑ, such that the trajectories

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{u}_t = F^{\text{NN}}(u, \vartheta)$$

resemble as close as possible the observed trajectory data. Let M be the number of trajectories and T the number of times steps

$$t_1, \dots, t_j, \dots, t_T$$

then the μ^{th} trajectory is

$$\left\{ t_j, u_{t_j}\left(u_0^{(\mu)}\right) \right\}_{\mu=0}^{T}, \quad \text{starting from } u_0^{(\mu)}$$

and we want to minimize

$$\mathcal{L}(\vartheta) = \frac{1}{M}\sum_{\mu=1}^{M}\frac{1}{T}\sum_{j=1}^{T}\left| \tilde{u}_{t_j}^{\vartheta}\left(u_0^{(\mu)}\right) - u_{t_j}\left(u_0^{(\mu)}\right)\right|^2$$

where

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{u}_{t_j}^{\vartheta}\left(u_0^{(\mu)}\right) = F^{\text{NN}}(z, \vartheta), \quad \text{starting at } u_{t_j}\left(u_0^{(\mu)}\right).$$

Since we know physics, we can either modify the loss with a penalization function for not following the principles, or we can encode it in the very structure of the Neural Network [‡].

If $f(x, \vartheta) : \mathbb{R}^d \longrightarrow \mathbb{R}$ is a L-layer NN, then we can define the gradient of $f \circ g$ with respect to x as

$$\nabla f(g(x), \vartheta) = \mathrm{D}g(x)\left[W^{(L)}\prod_{j=1}^{L-1}\Sigma_j W^{(j)}\right] = \mathrm{D}g(x)[\nabla f \circ g(x)].$$

[‡] We follow the derivations of [ZSE22].

Our goal is to design Neural Networks that satisfy symmetry and degeneracy of GENERIC, yet are sufficiently expressive to learn the underlying dynamics from data. The GENERIC formalism comprises two orthogonal modules, each of which contains two components:

$$\mathrm{D}H \perp \mathrm{K}, \quad \partial\varphi \ni \xi \perp \mathrm{L}.$$

Example. If L is known, we ought to find some proper NN such that $\mathrm{L}\partial\varphi^{\mathrm{NN}} = 0$. If φ was given, we want to find skew-symmetric matrix L^{NN} such that $\mathrm{L}\partial\varphi^{\mathrm{NN}} = 0$. In both cases, there are infinitely many choices.

Suppose the generators L and K are known, and, for any matrix-valued map

$$u \in \Omega \longmapsto \mathsf{A}_u \in \mathbf{M}_d(\mathbb{R}), \quad n = \mathrm{rank}(\mathsf{A})$$

with $\{\mathsf{q}^1, \dots, \mathsf{q}^n\}$ an orthonormal basis of $\ker \mathsf{A}_u$, define the *invariant kernel*

$$\widetilde{\ker}\mathsf{A}_u = \mathrm{span}\{\mathsf{q}^1, \dots, \mathsf{q}^{\tilde{n}}\}, \quad \tilde{n} \leqslant n$$

as the longest subspace of $\ker \mathsf{A}_u$ such that $\mathrm{Im}(\mathrm{D}\mathsf{q}_u) \subset \ker \mathsf{A}_u$. We will throughout assume there exist $\mathsf{F}_u \in \mathscr{F}^{n-\tilde{n}}(\Omega)$ such that

$$\mathrm{span}\nabla\mathsf{F}_u \oplus \widetilde{\ker}\mathsf{A}_u = \ker \mathsf{A}_u.$$

Moreover, letting $\mathsf{Q}_u = (\mathsf{q}_u^1, \dots, \mathsf{q}_u^{\tilde{n}})$, define

$$\mathsf{P}_u \equiv (\mathsf{Q}_u u, \mathsf{F}_u) \in \mathbb{R}^n$$

namely, the first $\tilde{n}$ components of P_u are the orthogonal projection coefficients of u onto $\widetilde{\ker}\mathsf{A}_u$ and the remaining components are the F_u^j's.
Given G, representing either the Hamiltonian energy or the entropy potential, we can reinterpret the degeneracy conditions ($\star$) as:

$$\nabla \mathrm{G} \in \ker \mathsf{A}_u$$

with A being either the Onsager K or the Poisson operator L. In fact, the above assumptions provide the key information on how to control $\nabla \mathrm{G}^{\mathrm{NN}}$ inside $\ker \mathsf{A}$ as a function of u.

Let

$$\mathsf{G}^{\text{NN}}(u;\vartheta) \equiv f(\mathsf{P}_u,\vartheta)$$

so that any given G^{NN} satisfies $\mathsf{A}_u\nabla\mathsf{G}_u = 0$, $\forall u \in \Omega$.
Let $\mathfrak{F}_{\mathsf{A}}$ be the collection of $\mathbf{C}^1$ functions $\mathsf{G} : \Omega \longrightarrow \mathbb{R}$ such that

$$\nabla\mathsf{G}_u = \mathrm{D}\mathsf{P}_u[\mathsf{c}_u \circ \mathsf{P}_u]$$

for some continuous function $\mathsf{c}_u : \mathbb{R}^n \longrightarrow \mathbb{R}^n$.

Lemma 3.2. Suppose A is constant, then

$$\mathfrak{F}_{\mathsf{A}} = \{f \in \mathbf{C}^1(\Omega) : \nabla f \in \ker \mathsf{A}\}.$$

Proof. Since A is constant, so is Q and $n = \tilde{n}$, $\mathsf{P}_u = \mathsf{Q}_u$ and $\mathrm{D}\mathsf{P}_u = \mathsf{Q}$. Let $\tilde{\mathsf{G}}(\xi) \equiv \mathsf{G}(\mathsf{Q}_\xi)$, then

$$\nabla\mathsf{G}_u = \mathsf{Q}\nabla\tilde{\mathsf{G}}|_{\xi=\mathsf{Q}_u} = \mathsf{Q}\nabla\tilde{\mathsf{G}} \circ \mathsf{P}_u$$

and we conclude by letting $\mathsf{c}_\xi = \nabla\tilde{\mathsf{G}}(\xi)$. ∎

Theorem 3.3. For all $\mathsf{G}_u \in \mathfrak{F}_{\mathsf{A}}$ and $\epsilon > 0$, there exists G_u^{NN} such that

$$\sup_{u\in\Omega} |\nabla\mathsf{G}_u - \nabla\mathsf{G}_u^{\text{NN}}| < \epsilon$$

and $\mathrm{D}\mathsf{G}_u^{\text{NN}} \in \ker \mathsf{A}$ for all $u \in \Omega$.

Proof. By the Universal Approximation Theorem[§], there exists a neural network $f_u(\vartheta)$ such that

$$\sup_{u\in\Omega} |\nabla f_u \circ \mathsf{P}_u - \mathsf{c}_u \circ \mathsf{P}_u| < \epsilon$$

because $\mathrm{D}\mathsf{P}_u$ is full rank in Ω. ∎

For GENERIC: $H \in \mathfrak{F}_{\mathsf{K}}$ and $\varphi \in \mathfrak{F}_{\mathsf{L}}$. If H and φ are unknown, the goal is to approximate L and K: just set $\mathsf{G} \equiv \mathsf{G}^{\text{NN}}$. If instead the entire quadruple $(\mathsf{L},\mathsf{K},H,\varphi)$ is unknown, we do not need to control the gradient of G^{NN} to remain inside $\ker \mathsf{A}_u$ anymore, and we can rather design a matrix-valued neural network A^{NN} to satisfy $\nabla\mathsf{G}^{\text{NN}} \in \ker \mathsf{A}_u^{\text{NN}}$.

[§] See App. E for a proof and the implications of the theorem on the architecture of deep neural networks.

Let $S \in \mathbf{M}_d(\mathbb{R})$ be any skew-symmetric matrix, that is, $\langle v, Sv\rangle = 0$ for all $v \in \mathbb{R}^d$, then $\langle S\nabla g, \nabla g\rangle = 0$ for any differentiable function $g : \mathbb{R}^d \longrightarrow \mathbb{R}$. Consider $S^1, \dots, S^k, \dots$ skew-symmetric such matrices and define the matrix Q_g as the 2-tensor

$$(Q_g)_{ij} = S^{ijk}\nabla g_k$$

so that $Q_g \nabla g \equiv 0$. Simply let

$$A^{NN} = Q_{G^{NN}}^{T} B^{NN} Q_{G^{NN}}$$

where B^{NN} is skew-symmetric if $A = L$ and is a symmetric positive semi-definite matrix if $A = K$. With this model, motivated by the spectral decomposition of either skew-symmetric or symmetric positive semi-definite matrix, degeneracy conditions ($\star$) are respected and, owing to the universal approximation theorem, the proposed neural network is expressive enough for approximating the underlying target function [‡].

In this case, the trainable parameters of the Neural Networks H^{NN}, φ^{NN}, B_L^{NN}, B_K^{NN} are the skew-symmetric matrices used in the covariant definition of $Q_{H^{NN}}$ and $Q_{\varphi^{NN}}$.

[‡]The full proof, which we do not provide for the sake of brevity, can be found in [ZSE22].
The bulk of the idea is to conclude, similarly to our previous argument, the existence of a skew-symmetric matrix-valued neural network distant enough, in the Frobenius norm, from the spectral decomposition of $Q_{G^{NN}}$, so that the approximation is universal.

CHAPTER IV

Examples

We shall now illustrate some examples of GENERIC systems.

1. Thermoelastic double pendulum

Let's start from the simple thermodynamical model system of two point masses m_1, m_2 connected by thermoelastic springs of entropy S_1, S_2 and tied to a fixed point. Let $q^i, p_i \in \mathbb{R}^2$, $i = 1, 2$, denote positions and momenta of the particles, as we are assuming the motion to take place in the plane, and $\lambda_1 = |p_1|, \lambda_2 = |\lambda_2|$ the lengths of the springs; then, choosing

$$u = (q^1, q^2, p_1, p_2, S_1, S_2) \in \mathbb{R}^{10}$$

as statevector, we obtain the the potential $\varphi = S_1 + S_1$ and the Hamiltonian

$$H = \sum_{i=1}^{2} K_i + E_i, \qquad \text{where the kinetic energies } K_i = \frac{1}{2}|p_i|^2,$$

$$\text{and the internal energies } E_i = \frac{1}{2}(\log \lambda_i)^2 + \log \lambda_i + e^{S_i - \log \lambda_i} - 1$$

with absolute temperatures

$$\vartheta_i = \frac{\partial E_i}{\partial S_i}$$

such that the time evolution governing the system is the GENERIC flow equation

$$\begin{pmatrix} \dot{q}^1 \\ \dot{q}^2 \\ \dot{p}_1 \\ \dot{p}_2 \\ \dot{S}_1 \\ \dot{S}_2 \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ -\partial_{q^1}(E_1 + E_2) \\ -\partial_{q^2}(E_1 + E_2) \\ \vartheta_2/\vartheta_1 - 1 \\ \vartheta_1/\vartheta_2 - 1 \end{pmatrix} = \mathrm{L}\nabla H - \mathrm{K}\nabla\varphi$$

thus

$$\mathrm{L} = \begin{pmatrix} 0_4 & S \\ -S^{\mathsf{T}} & 0_6 \end{pmatrix} \in \mathbf{M}_{10}(\mathbb{R}), \quad S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

and

$$\mathrm{K} = \begin{pmatrix} 0_8 & 0 \\ 0 & -T \end{pmatrix} \in \mathbf{M}_{10}(\mathbb{R}), \quad T = \begin{pmatrix} \vartheta_2/\vartheta_1 & -1 \\ -1 & \vartheta_1/\vartheta_2 \end{pmatrix}.$$

1.1 Monolithic symmetry

This may appear very rough, but a, energy-preserving and monotonic entropy, monolithic integrator for this system is implemented in [Rom09], exploiting the symmetries of the problem. If the n-dimensional state space has symmetries, that rigorously means that the orbits of the symmetry group have dimension $s \neq n$, and thus there are $n - s$ "*degrees of freedom*", or invariants of the symmetry, which we will denote by $\pi = (\pi_1, \pi_2, \ldots, \pi_{n-s})$, meaning that each $\pi_i(u)$ is invariant under the action of the symmetry group. Intuitively, it "collapses" each symmetry orbit (which is an s-dimensional manifold itself) to a single point in its image, reducing to only the essential degrees of freedom. Then, considering the *reduced energy* and *entropy*

$$\tilde{H}_u = H_u \circ \pi_u^{-1}, \quad \tilde{\varphi}_u = \varphi_u \circ \pi_u^{-1}$$

we can directly define discrete gradients as follows:

$$\mathrm{D}H_{u_\tau} = \nabla\pi(u_\tau)\mathrm{D}\tilde{E}_{\pi_\tau}, \quad \mathrm{D}\varphi_{u_\tau} = \nabla\pi(u_\tau)\mathrm{D}\tilde{\varphi}_{\pi_\tau}.$$

In particular, these additionally conserve angular momentum, suitably defined as the momentum map [‡] associated to the Lie group whose orbits fix the Hamiltonian when acting on the state space.

For the double pendulum, it's enough to approximate the absolute temperatures as $\vartheta_i \approx \mathrm{D}_{\pi_i} E_i$ at each step.

[‡] See [MR13, Ch. 11].

2. The minimizing-movements scheme for the damped harmonic oscillator

Consider a one-dimensional harmonic oscillator in the presence of friction. We let $q \in \mathbb{R}$ be the position of the particle and $m > 0$ its mass, $\nu > 0$ the viscosity of a homogeneous medium causing its friction, $\kappa > 0$ the Hookean spring constant, $\vartheta > 0$ the absolute temperature of the oscillator and λ a thermal-exchange coefficient. Its mechanical motion with thermal effects is thus governed by

$$m\ddot{q} + \nu\dot{q} + \kappa q + \lambda\vartheta = 0$$

and its thermal energetic evolution by

$$c\dot{\vartheta} = \nu\dot{q}^2 + \lambda\dot{q}\vartheta.$$

Introducing the canonical momentum $p \in \mathbb{R}$ we get the three-dimensional system

$$\begin{aligned}
\dot{q} &= \frac{p}{m} \\
\dot{p} &= -\frac{\nu p}{m} - \kappa q - \lambda\vartheta \\
\dot{\vartheta} &= \frac{\nu p^2}{m^2 c} + \frac{\lambda p \vartheta}{mc}
\end{aligned}$$

which can be formulated as a (Euclidean) GENERIC flow, with $u \equiv (q, p, \vartheta)$ and state space $\mathcal{M} \equiv \mathbb{R}^2 \times (0, +\infty)$, of the Helmholtz free energy function

$$\Phi(q, p, \vartheta) = \frac{\kappa q^2}{2} + \lambda q\vartheta - c\vartheta\log\vartheta$$

that is, with entropy $\varphi = -\partial_\vartheta\Phi$ and energy

$$H = \frac{1}{2m}p^2 + \Phi + \vartheta\varphi$$

and Poisson and friction matrices

$$\mathrm{L} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & -\lambda\vartheta/c \\ 0 & \lambda\vartheta/c & 0 \end{pmatrix}, \quad \mathrm{K} = \nu\vartheta \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -p/(mc) \\ 0 & -p/(mc) & p^2/(mc)^2 \end{pmatrix}.$$

For this system, Öttinger proposed in [Ött18] an integrator inspired by Moser's symplectic Euler, with fully perturbed Hamiltonian and friction matrix. However, in [JST22] the construction of a dissipation potential and the implementation of the minimizing movements scheme can be found.

3. GENERIC formalism under fluctuations

GENERIC is equally powerful in the mathematical description of the time-evolution of random systems. We just let $\{u_t\}_{t\in\mathbf{T}}$ be a sequence of random variables $u_t \in \mathbf{X}$, for continuous space $\mathbf{X}$ and time $\mathbf{T}$. We want to study the infinitesimal evolution of the random quantity

$$\Delta_h \equiv (u_h - u_0 | u_0 = x), \qquad \text{w.l.o.g. } x = 0.$$

The simplest assumptions would be that, as $h \downarrow 0$, our process has constant zero mean: $\mathbb{E}[\Delta_h/h] \downarrow 0$, i.e. that it as no "drift", has a fixed variance: $\mathbb{E}[\Delta_h^2/h] = \sigma^2$, i.e. that it possesses some constant volatility, and that it makes no jumps[§]: $\mathbb{P}(|\Delta_h| > \epsilon) \downarrow 0$ for all $\epsilon > 0$. We immediately understand that trajectories $t \longmapsto u_t$ must be a.s. everywhere continuous, that is, they are continuous with prob. 1 for all times, and that the stationary independent increments are normally distributed:

$$u_t - u_s \sim N(0, (t-s)\sigma^2), \quad 0 < s < t$$

meaning that it is a Markov process: intuitively, the conditional probability of future states only depends on the current spatial configuration and not on the entire history of events. Wiener proved that such a process exists, and corresponds to the long-before studied Brownian motion, governing the dynamics of diffusing particles inside viscous media. We will denote by $(B_t)_{t\in[0,+\infty)}$ the Brownian motion with $\sigma = 1$, so that we can impose the formal (stochastic) differential equation

$$\mathrm{d}u_t = \mu(u,t)\,\mathrm{d}t + \Sigma(u,t)\,\mathrm{d}B_t, \tag{SDE}$$

in which case we say that $(u_t)_{t\in\mathbf{T}}$ is a *diffusion process* [Pav14, Sect. 2.5], for drift $\mu(u,t)$ and diffusion $\Sigma(u,t)$ coefficients, provided that the process has continuous trajectories. The

[§]After all, by introducing stochasticity we would like to extend our model to almost all physically-relevant phenomena and "*Natura non facit saltus*"!

conditional density $\varrho(u,s)$, acting on test functions $f : \mathbf{X} \longrightarrow \mathbb{R}$ as follows:

$$\varrho(u_s)[f] = \mathbb{E}[f(u_t)|u_s = u]$$

evolves according to the boundary value problem

$$\frac{\partial \varrho}{\partial s} = \mathscr{L}\varrho, \quad \varrho(u,0) = f(u) \tag{KBE}$$

known as the Kolmogorov (backward) equation, driven by the so-called *generator*

$$\mathscr{L} = \langle \mu(u,s), \nabla \rangle + \frac{1}{2}\Sigma(u,s)\nabla^2$$

for $s \in [0,t]$.

Example. In the standard Brownian motion, $\mu \equiv 0$ and $\Sigma = \mathrm{Id}_{\mathbf{X}}$, hence the cross terms of the Hessian vanish, leaving the diagonal terms summing up to the Laplacian, namely, (KBE) is the heat equation $\dot{\varrho} = \frac{1}{2}\Delta\varrho$, zero-drift equation with constant diffusion.

3.1 Fokker–Planck in the phase space

Consider now the *transition probability density* $\rho(u,t)$, which satisfies the Fokker–Planck, or Kolmogorov (forward), equation

$$\frac{\partial \rho}{\partial t} = \mathscr{L}^\dagger \rho = \mathrm{div}\left(-\mu_u \rho + \frac{1}{2}\mathrm{div}(\Sigma_u \rho)\right).$$

Remark. It may come as a surprise, but our choice to use the same notation $\mathscr{L}$ for the generator of diffusions and the Liouvillian (Ch. II, 1.1) of classical mechanical systems was not fortuitous! Suppose introducing stochasticity in a Hamiltonian system and consider the *phase space distribution* $\rho(u = (\dot{p},\dot{q}),t)$, then Liouville's Theorem asserts that the distribution function is constant along any trajectory in phase space, namely, that density is conserved in the phase space:

$$\begin{aligned}\frac{\partial \rho}{\partial t} &= -\mathrm{div}(\rho u)\\ &= -\rho\partial_i u^i - u^i\partial_i\rho\\ &= \{H,\rho\} = \mathscr{L}_H\rho\end{aligned}$$

where $\partial_i u^i$ vanishes by assumption. If we let $\Sigma \equiv 0$ in (SDE), we obtain a classical ODE with no diffusion, and, similarly, we recover Liouville's equation from Fokker-Planck: the former totally described the (reversible) dynamics and, as soon as a heat kernel driven by a diffusive term is added, irreversibility manifests and we get a full dissipative-conservative coupling.

The previous remark justifies the Liouville equation corresponding to GENERIC [GÖ97]:

$$\frac{\partial \rho(u,t)}{\partial t} = -\mathrm{div}\left[\rho(u,t)\left(\mathrm{L}\nabla H - \mathrm{K}\nabla\varphi\right)\right],$$

responsible for the overall dynamics, while we confine the diffusive term to the Onsager dynamics: we want to introduce a noise term and thus consider how perturbed our system's dissipation stands from the base (Shannon-Boltzmann-Gibbs) entropy $h(u) = -k_B u \log u$, i.e. the Fokker-Planck equation

$$\frac{\partial \rho(u,t)}{\partial t} = -\mathrm{div}\left[\rho(u,t)\left(\mathrm{L}\nabla H - \mathrm{K}\nabla\varphi\right)\right] + k_B \mathrm{div}\left[\mathrm{K}\,\mathrm{div}\rho\right]$$

governs the coupled time-evolution of the distribution function and is the particular realization of GENERIC with entropy

$$S = \int_{\mathbf{X}} u(\varphi(u) - k_B \log \rho(u)). \tag{Ent}$$

This is equivalent to the (SDE) with

$$\Sigma\Sigma^{\mathsf{T}} \equiv 2k_B \mathrm{K}^{\ddagger}, \quad \mu \equiv \mathrm{L}\nabla H - \mathrm{K}\nabla\varphi + k_B\ \mathrm{div}\ \mathrm{K}$$

so that, together with the usual interaction condition ($\star$), we recover (GENERIC) in the deterministic limit $k_B \downarrow 0$. GFINNs, for such a system [ZSE22], reduce to the drift and diffusion components

$$\begin{aligned}
\mu^{\mathrm{NN}}_{u_t} &\equiv \mathrm{L}^{\mathrm{NN}}_{u_t}\mathrm{D}H^{\mathrm{NN}}_{u_t} - \mathrm{K}^{\mathrm{NN}}_{u_t}\mathrm{D}\varphi^{\mathrm{NN}}_{u_t} + k_B\ \mathrm{div}\ \mathrm{K}^{\mathrm{NN}}_{u_t}, \\
\Sigma^{\mathrm{NN}}_{u_t} &\equiv \sqrt{2k_B}\left(\sqrt{B^{\mathrm{NN}}_{\mathrm{K}}}H^{\mathrm{NN}}\right)^{\mathsf{T}}.
\end{aligned}$$

[‡] This is an instance of the Fluctuation-Dissipation Theorem, if we once more interpret K as a "friction operator".

3.2 Langevin dynamics

Consider again Newton's equation, now with an additional linear dissipation term, driven by some friction coefficient γ as before, and a stochastic forcing driven by the Brownian noise $\xi(t) = \sigma\dot{B}_t \equiv \sqrt{2\gamma\beta^{-1}}\dot{B}_t$:

$$\ddot{q}_t = -\nabla\varphi(q_t) - \gamma\dot{q}_t + \xi(t)$$

this is the Langevin's equation[§] governing the diffusion of a particle with a thermal reservoir at temperature $\beta = 1/k_B T$. Dissipation and noise are related through the *Fluctuation–Dissipation Theorem*:

$$\mathbb{E}[\xi(t)\xi(s)] = \gamma\beta^{-1}\delta(t-s)$$

and, introducing the momentum $p_t = \dot{q}_t$, the evolution decouples as:

$$\mathrm{d}q_t = p_t\mathrm{d}t, \quad \mathrm{d}p_t = -\nabla\varphi(q_t)\,\mathrm{d}t - \gamma p_t\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$$

that is, a process with the generator

$$\mathscr{L} = \mathscr{L}_H + \gamma\left(-p\nabla_p + \beta^{-1}\Delta_p\right)$$

which is itself decoupled in a first-order differential operator $\mathscr{L}_H$ i.e. the Liouvillian corresponding to the Hamiltonian

$$H(p,q) = \frac{1}{2}|p|^2 + \varphi(q),$$

plus a second-order diffusive kernel.

Assuming a smooth confining potential $\varphi \in \mathscr{F}(\mathbb{R}^{2d})$, this generator induces an (ergodic) Markov process with (unique) stationary distribution given by the Boltzmann:

$$\rho_B(p,q) = \frac{1}{Z}e^{-\beta H(p,q)}, \quad Z = \int_{\mathbb{R}^{2d}} e^{-\beta H(p,q)}\,\mathrm{d}p\,\mathrm{d}q.$$

Again, we can consider the overdamped limit:

$$\mathrm{d}u_t = -\frac{1}{\gamma}\nabla\varphi(u_t)\,\mathrm{d}t + \frac{\sigma}{\gamma}\,\mathrm{d}B_t,$$

[§]Langevin's original proposal (*C. R. Acad. Sci. Paris.*, 1908) for an equation governing the diffusion of a particle under Brownian motion: $\dot{v} = -\gamma v + \eta(t)$ was inspired by Einstein's studies of the phenomenon and had an evident 'viscosity flavor', with the correlation requirement: $\langle\eta_i(t)\eta_j(s)\rangle = \sigma^2\delta(t-s)$.
For our treatment, see [Pav14, Sect. 6.1].

recovering exactly (GF) as $k_B \downarrow 0$. This is particularly interesting since imposing detailed balance, i.e. stationary in (KBE): $\mathscr{L}\rho = 0$, we get that any stationary distribution ρ must obey $2\mu = \nabla \log \rho$, with unitary volatility, so the overdamped Langevin dynamics with potential $\varphi(u) = -\frac{1}{2}\log \rho(u)$ induces a π-invariant diffusion process[‡], and we can think of it as a 'gradient-flow plus noise' which defines a canonical stochastic process used to draw samples from ρ. The former equation can be formulated as a GENERIC system under fluctuations, with entropy $\varphi = S$ and Hamiltonian

$$H = \frac{1}{2}|p|^2 + S$$

for state variables $(q, p, S) \in \mathbb{R}^3$, and generators

$$\mathrm{L} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathrm{K} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/2 & -p/2 \\ 0 & -p/2 & p^2/2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0 \\ 1 \\ -p \end{pmatrix}.$$

A GFINN-simulation of its diffusing trajectories can be found in [ZSE22].

[‡]In fact, the stationary distribution for the inertial regime is the Boltzmann, which maximizes entropy. As γ grows, our system undergoes diffusion: the gap in the expression (Ent) loosens up and the invariant distribution diverges from the Boltzmann.

APPENDIX A

Convex analysis

Given a function $f : \mathrm{dom}(f) \longrightarrow \mathbb{R}$, we want to find a (possibly constrained, i.e. $\mathrm{dom}(f) \subsetneq \mathbb{R}^d$) minimum point. If the function is smooth, we can simply start from any point and look at the negative gradient flow: once discretized, this is the "gradient descent" algorithm, which typically leads us to a local minimum point. Convexity [‡] upgrades local properties to global ones: in a convex setting, gradient descent always converges to a global minimum point, with some theoretical guarantees on the speed of convergence. Moreover, the usual optimality conditions which are necessarily satisfied by a (local or global) minimum point turn out to be also sufficient (also in the constrained case) for convex functions.

1. Convex functions

Definition 1.1. Given a nonempty convex set $\mathrm{dom}(f) \subseteq \mathbb{R}^d$, we say that a function $f : \mathrm{dom}(f) \longrightarrow \mathbb{R}$ is convex if

$$f((1-\lambda)p + \lambda q) = \lambda f(p) + (1-\lambda) f(q)$$

for all $p, q \in \mathrm{dom}(f)$ and $\lambda \in [0, 1]$.

It follows easily that $f \in \mathbf{C}^2(\mathrm{dom}(f))$ is convex if and only if $\nabla f : \mathrm{dom}(f) \longrightarrow \mathbb{R}^d$ is monotone, in the sense that

$$\langle \nabla f(a) - \nabla f(b), a - b \rangle \geqslant 0 \quad \text{for all} \quad a, b \in \mathrm{dom}(f)$$

and the Hessian $\nabla^2 f$ is positive semidefinite at all points. Also notice that every convex function is locally Lipschitz (and, in particular, continuous) on the interior of its domain.

[‡] Here, the reference must be [HL04].

Another equivalent formulation of convexity is that the *epigraph*

$$\operatorname{epi}(f) \equiv \{(x,t) \in \mathbb{R}^d \times \mathbb{R} : x \in \operatorname{dom}(f),\ t \geqslant f(x)\}$$

is a convex set in $\mathbb{R}^d \times \mathbb{R}$. If instead the epigraph of our function is closed, we say that the function is *lower semicontinuous*. Hence, if f is both convex and lower semicontinuous, namely its epigraph is closed and convex, it turns out that it can be written as the supremum of a family of affine functions. This is a necessary and sufficient condition, and counterexamples can be found for both only-convex and only-semicontinuous functions. Assume now that we have an affine function $\ell : \mathbb{R}^n \to \mathbb{R}$, of the form $\ell(x) = \langle \xi, x \rangle + a$ for some $\xi \in \mathbb{R}^n$ and $a \in \mathbb{R}$, with $\ell \leqslant f$ at every point.

2. Legendre transform

We can then "optimize" the value of a: namely, we can replace it with the highest value such that the inequality $\langle \xi, x \rangle + a \leqslant f(x)$ still holds for all $x \in \mathbb{R}^n$. The latter is the same as

$$a \leqslant f(x) - \langle \xi, x \rangle \quad \text{for all } x \in \mathbb{R}^n$$

the highest value of a such that this holds is precisely the infimum of $f(x) - \langle \xi, x \rangle$. In other words, letting

$$a_\xi \equiv \inf_{x \in \mathbb{R}^n} \big[f(x) - \langle \xi, x \rangle \big],$$

we have that f is convex and lower semicontinuous if and only if

$$f(x) = \sup_\xi \{ \langle \xi, x \rangle + a_\xi \} \quad \text{for all} \quad x \in \mathbb{R}^d.$$

Note that we have $a_\xi = +\infty$, for some $\xi \in \mathbb{R}^d$, if and only if $f(x) = +\infty$ for all $\xi \in \mathbb{R}^d$, i.e., $f \equiv +\infty$. So, for the sake of regularity, we say that $f : \operatorname{dom}(f) \longrightarrow \mathbb{R}_\infty \equiv \mathbb{R} \cup \{+\infty\}$ is *proper* if $f \not\equiv +\infty$.

Definition 2.1. The *Fenchel–Legendre–Moreau transform* f^*, called the *convex conjugate*, of f is given by the transformation

$$f[\xi] = f^*(\xi) \equiv \sup_\xi \{ \langle \xi, x \rangle - f(x) \}.$$

The formula

$$f(x) = \sup_{\xi \in \mathbb{R}^n} \{\langle \xi, x \rangle - f^*(\xi)\}$$

is in a sense analogous to reconstructing a function from its Fourier transform: in that context, a function is written as a "linear combination of pure waves" $\exp(2\pi i \langle \xi, x \rangle)$ (indexed by the frequency ξ) which are viewed as the fundamental building blocks; the Fourier transform $\hat{f}(\xi)$ gives the (infinitesimal) coefficient to be put in front of $\exp(2\pi i \langle \xi, x \rangle)$. In this context, the building blocks are affine functions of slope ξ, the convex conjugate $f^*(\xi)$ gives (negative) the intercept, and the way we combine them is by taking a supremum, rather than a linear combination.

3. Subgradients

Looking at the *Fenchel inequality*

$$f^*(\xi) \geqslant \langle \xi, x \rangle - f(x)$$

it follows very naturally to consider the set of all slopes such that the previous inequality holds true, as this would allow for a (lower) extension of the concept of derivative.

Definition 3.1. Given $x_0 \in \mathrm{dom}(f)$, the *subdifferential* of f at x_0 is the set

$$\partial f(x_0) = \{\xi \in \mathbb{R}^d : f(x) \geqslant f(x_0) - \langle \xi, x - x_0 \rangle \quad \text{for all } x \in \mathrm{dom}(f)\}.$$

From the definition of subdifferential, we have $0 \in \partial f(x_0)$ if and only if x_0 is a global minimum point, hence, for a convex function, local minimum points are automatically global minimum points.

Fenchel optimality condition. If $f : \mathrm{dom}(f) \longrightarrow \mathbb{R}$ is proper, lower semi-continuous and convex, then,

$$\langle \xi, x \rangle = f(x) + f^*(\xi) \quad \iff \quad \xi \in \partial f(x)$$

for all $x, y \in \mathrm{dom}(f)$.

PROOF. Since $\langle \xi, x\rangle \leqslant f(x) + f^*(\xi)$,

$$\begin{aligned}\langle \xi, z\rangle = f(x) + f^*(\xi) &\iff \langle \xi, x\rangle \geqslant f(x) + f^*(\xi)\\ &\overset{(2.1)}{\iff} \langle \xi, x\rangle \geqslant f(x) + \langle \xi, x_0\rangle - f(x_0) \quad \forall x_0 \in \mathrm{dom}(f)\\ &\iff f(x_0) \geqslant f(x) + \langle \xi, x_0 - x\rangle \quad \forall x_0 \in \mathrm{dom}(f)\\ &\overset{(3.1)}{\iff} \xi \in \partial f(x)\end{aligned}$$

for all $x, \xi \in \mathrm{dom}(f)$. ∎

The subdifferential is moreover always convex and closed (possibly empty), it is, additionally, compact if taken at a point in the interior of the domain. In such situation, it can thus be a singleton, but that would correspond to the smooth case:

$$\partial f(x_0) = \{\nabla f(x_0)\}.$$

As a further consequence of compactness, we have the following convergence result:

$$x_n \longrightarrow x \in \mathrm{dom}(f) \quad \text{and} \quad \xi_n \longrightarrow \xi \quad \text{with each} \quad \xi_n \in \partial f(x_n), \quad \text{then} \quad \xi \in \partial f(x). \quad (\rightarrow_\partial)$$

EXAMPLE. Consider the convex function $\mu(x) \equiv |x|$, defined on $\mathbb{R}^d$, which is differentiable on $\mathbb{R}^d \backslash \{0\}$, so that

$$\partial\mu(x) = \left\{ \frac{x}{|x|} \right\} \quad \text{for all} \quad x \neq 0$$

and, by $(\rightarrow_\partial)$, all unit vectors belong to $\partial\mu(0)$. Moreover,

$$\xi \in \partial\mu(0) \quad \Longrightarrow \quad \forall x \in \mathbb{R}^d : |x| \geqslant \langle \xi, x\rangle \quad \text{hence} \quad |\xi| \geqslant |\xi|^2 \quad \text{i.e.} \quad |\xi| \leqslant 1$$

so $\partial\mu(0) = \mathbb{D}$.

APPENDIX B

Sobolev spaces

When we take subgradient flows, we are considering the subdifferential of a functional, because the actual input of the potential are the trajectories u_t. So we would like to have a functional analytic framework where all the precise and delicate apparatus of convex analysis can be conveniently extended.

1. Weak derivatives

Let $u \in \mathbf{C}^1(\Omega), \eta \in \mathbf{C}^\infty(\Omega)$ with compact support in the region $\Omega \subseteq \mathbb{R}^d$, then integration by parts yields

$$\int_\Omega u \frac{\partial \eta}{\partial x_i}\, \mathrm{d}x = -\int_\Omega \frac{\partial u}{\partial x_i}\, \eta\, \mathrm{d}x, \quad i = 1, \ldots, d$$

and there are no boundary terms because η is compactly supported and vanishes near $\partial\Omega$. This motivates our definition for a generalized derivative.

Definition 1.1. Let $u, \nu \in \mathbf{L}^p(\Omega)$, then ν is the *weak derivative* of u if

$$\int_\Omega u \frac{\partial \eta}{\partial x_i}\, \mathrm{d}x = -\int_\Omega \nu_i\, \eta\, \mathrm{d}x, \quad i = 1, \ldots, d$$

for all compactly-supported $\eta \in \mathbf{C}^\infty(\Omega)$.

Example. Integrating by parts on $\mathbb{R}^d \backslash \mathrm{B}_\epsilon(0)$ and taking the limit $\epsilon \downarrow 0$ to make the boundary contributions vanish, the modulus $\mu(x) \equiv |x|$ has $\nu(x) = \mathrm{sign}(x)$ as weak derivative

The Sobolev space $\mathbf{W}^{1,p}(\Omega)$ consists of all summable functions $u \in \mathbf{L}^p(\Omega)$ that are square-integrable along with their (first) weak derivatives. The definition of the most general Sobolev space $\mathbf{W}^{k,p}(\Omega)$ allows for the existence of higher (up to the k^{th}) order weak derivatives, but this is beyond our scope. We are instead interested in square-integrable weak derivatives of smooth curves $u_t : [0, T] \longrightarrow \mathbf{X}$, and Def. 1.1 can be easily (pointwise) generalized to vector valued function if $\mathbf{X} = \mathbb{R}^d$, in which case we work with $\mathbf{W}^{1,2}([0, T];\ \mathbb{R}^d)$, or, if $\mathbf{X} = \mathcal{M}$ is

a differentiable manifold by considering $\mathbf{W}^{1,2}([0,T];\ \mathcal{M})$ as the collection of curves whose coordinate representations belong to $\mathbf{W}^{1,2}([0,T];\ \mathbb{R}^m)$.

2. Weak convergence

$\mathbf{W}^{1,2}(\Omega)$ is a Hilbert space when equipped with the inner product

$$\langle u, \tilde{u} \rangle_{1,2} = \int_\Omega u\, \tilde{u}\ + \sum_{i=1}^{d} \int_\Omega \nu_i\, \tilde{\nu}_i$$

justified by the norm

$$\|u\|_{1,2} = \left(\int_\Omega |u|^2\ + \sum_{i=1}^{d} \int_\Omega |\nu_i|^2 \right)^{½}$$

it induces, which is well-defined because everything is square-integrable. For curves u_t,

$$\|u\|_{1,2} = \left(\int_0^T g(\nu_t, \nu_t)_{u_t}\, \mathrm{d}t \right)^{½}$$

where g is the Riemannian metric. Actually, contrarily to the definition we gave, geodesics are defined as the curves that minimize the Energy functional $E(u) = ½\|u\|_{1,2}$. The Hilbertian characterization is very powerful, because infinitely-many dimensions allow sequences of functions to have limit points even when norm convergence fails. Indeed, in a general Hilbert space $\mathbf{H}$, we say that $(x_n)_n \subset \mathbf{H}$ *converges weakly* to $x \in \mathbf{H}$ if

$$\forall y \in \mathbf{H}: \quad \lim_{n \uparrow \infty} \langle x_n, y \rangle = \langle x, y \rangle$$

Definition 2.1. A sequence u_n in $\mathbf{W}^{1,2}$ is said to *weakly converge*:

$$u_n \longrightarrow u \quad \text{in} \quad \mathbf{W}^{1,2},$$

if $u_n \longrightarrow_{\mathbf{L}^2} u$ and $\nu_n \longrightarrow_{\mathbf{L}^2} \nu$, or, equivalently, if

$$\langle u_n, \eta \rangle_{1,2} \longrightarrow \langle u, \eta \rangle_{1,2}$$

for any compactly-supported η.

A simple example in $\mathbf{W}^{1,2}(\mathbb{R})$ is

$$u_n(x) = \frac{\sin(nx)}{n},$$

in fact, $u_n \longrightarrow 0$ since

$$\int_{\mathbb{R}} u_n \eta \, dx = -\frac{1}{n^2} \int_{\mathbb{R}} \cos(nx) \dot{\eta} \, dx \longrightarrow 0,$$

but it does not converge strongly because $\|\dot{u}_n\|_{\mathbf{L}^2} = 1/\sqrt{2}$.
Now we are ready to weaken ($\rightarrow_\partial$), once again, following [Mie23].

Definition 2.2. A functional $f : \mathrm{dom}(f) \longrightarrow \mathbb{R}$ has a (*strong-weak*) *energy closed subdifferential* if

$$x_n \longrightarrow x \quad \text{and} \quad \xi_n \longrightarrow \xi \quad \text{with each} \quad \xi_n \in \partial f(x_n), \quad \text{then} \quad \xi \in \partial f(x).$$

Again confirming how powerful it is, the class of λ-convex functions always admits energy closed subdifferentials.

Lemma 2.3. If $f : \mathrm{dom}(f) \longrightarrow \mathbb{R}_\infty$ is λ-convex, then for all $u \in \mathrm{dom}(f)$:

$$\xi \in \partial f(u) \quad \Longleftrightarrow \quad \forall w \in \mathrm{dom}(f) : \; f(w) \geqslant f(u) + \langle \xi, w - u \rangle + \frac{\lambda}{2} |w - u|^2.$$

Proposition 2.4. If $\varphi : \mathrm{dom}(\varphi) \longrightarrow \mathbb{R}_\infty$ is proper, lower semicontinuous, and λ-convex, then its subdifferential is (strong-weak) energy closed.

Proof. If $x_n \longrightarrow x$ and $\xi_n \longrightarrow \xi$ with each $\xi_n \in \partial f(x_n)$, by the Lemma we have, for all $n \in \mathbb{N}$ and all $w \in \mathrm{dom}(\varphi)$, the estimate

$$\varphi(w) \geqslant \varphi(u_n) + \langle \xi_n, w - u_n \rangle + \frac{\lambda}{2} |w - u_n|^2$$

where we can pass to the limit $n \uparrow +\infty$, using lower semicontinuity, the weak-strong continuity of the inner product and the strong continuity of the norm we find the desired characterization (Lemma 2.3) of $\xi \in \partial \varphi(u)$. ∎

APPENDIX C

Differential forms

Finally, let us introduce the basic differential geometric tools needed to do calculus in the general setting of manifolds, with the standard notation of [Lee12]. We will take as a motivation our early remark about the curl-free characterization of gradient flows, with the aim of obtaining concepts like being irrotational or conservative, without making reference to any additional (Riemannian or Poissonian) structure whatsoever.

Recall the concepts introduced in Ch. I, 2.1, and consider the dual space $\mathrm{T}_p^*\mathcal{M}$ of the tangent space, the *cotangent space* of *covectors* $\omega = \omega_i \mathrm{d}x^i$ (using Einstein's summation convention), where $\{\mathrm{d}x^i\}_{i=1}^m$ is used to denote the dual basis of the tangent space.

1. Pullback

The differential, of a smooth map $F : \mathcal{M} \longrightarrow \mathcal{N}$ between smooth manifolds, yields a dual linear map

$$\mathrm{D}F_p^* : \mathrm{T}_{F(p)}^*\mathcal{N} \longrightarrow \mathrm{T}_p^*\mathcal{M}$$

called *pullback* and, for each tangent vector $v \in \mathrm{T}_p^*\mathcal{M}$ and covector $\omega \in \mathrm{T}_{F(p)}^*\mathcal{N}$, defined by:

$$\mathrm{D}F_p^*(\omega)[v] \equiv \omega(\mathrm{D}F_p(v)).$$

Similarly to vector field, we can define covector fields ω on the *cotangent bundle* $\mathrm{T}^*\mathcal{M}$, and their pullback by a smooth function F:

$$F_p^*\omega = \mathrm{D}F_p^*(\omega_{F(p)}).$$

To generalize all of this concepts to higher dimensions, we call a *k-tensor* the real-valued multilinear function

$$\alpha_p : \underbrace{\mathrm{T}_p\mathcal{M} \times \ldots \times T_p\mathcal{M}}_{k \text{ copies}} \longrightarrow \mathbb{R}$$

so that, for instance, if $k = 2$, α is just a bilinear form such as a dot product. We denote by $\mathfrak{T}^k(\mathrm{T}^*_p\mathcal{M})$ the set of k-tensors, as it can be thought as the tensor product $\bigotimes_k \mathrm{T}^*_p\mathcal{M}$ of k copies of the cotangent space. If $F : \mathcal{M} \longrightarrow \mathcal{N}$ is smooth and $\alpha \in \mathfrak{T}^k(\mathrm{T}^*_{F(p)}\mathcal{N})$, its *pullback* is the tensor $F^*_p(\alpha) \in \mathfrak{T}^k(\mathrm{T}^*_p\mathcal{M})$ given by:

$$F^*_p(\alpha)[v_1, \ldots, v_k] = \alpha(\mathrm{D}F_p(v_1), \ldots, \mathrm{D}F_p(v_k)) \tag{i}$$

for any $v_1, \ldots, v_k \in \mathrm{T}_p\mathcal{M}$.

2. Tensors

In analogy with vector fields, we can define *tensor fields* as sections of the bundle

$$\mathfrak{T}^k(\mathrm{T}^*\mathcal{M}) = \bigsqcup_{p\in\mathcal{M}} \mathfrak{T}^k(\mathrm{T}^*_p\mathcal{M}).$$

Alternating, i.e. vanishing for linearly dependent arguments, covariant k-tensor fields are called (*differential*) k*-forms*. While symmetric 2-tensors provide a manifold with a Riemannian metric that allows to measure distances, angles, calculate norms, and define concepts like curvature, differential forms can be thought of as assigning to k-dimensional submanifolds. For instance, the easiest example of multilinear alternating map is the determinant on $\mathbb{R}^d$, that exactly measures the n-dimensional change in volume of a given linear function. Also, an alternating 1-tensor field, is a covector field on the manifold, which assigns a number to each tangent vector: it assigns a length to each one-dimensional subspace of each tangent space (how "far" along some direction a vector goes, rather than the vector's full magnitude, which is a Riemannian "business"). To mimic, and generalize, the behavior of the determinant, let us introduce the operation that *alternates* any k-tensor whatsoever:

$$\mathrm{Alt}(\alpha) = \frac{1}{k!} \sum_{\sigma\in\mathrm{S}_k} \mathrm{sgn}(\sigma)\, \alpha \circ \sigma$$

where S_k is the symmetric groups of permutations on k elements. In $\mathbb{R}^3$, the magnitude of the cross product $\vec{v}_1 \times \vec{v}_2$ of any two vectors $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^3$ is the (bidimensional) area of the parallelogram formed by the vectors, because indeed: $\det(\vec{v}_1, \vec{v}_2, \vec{w}) = \langle \vec{v}_1 \times \vec{v}_2, \vec{w} \rangle$. This concept generalizes, beyond dimension three, to the *exterior product* of a k-form ω with an

ℓ-form η:

$$\omega \wedge \eta = \mathrm{Alt}(\omega \otimes \beta)$$

which is a $k+\ell$-form, where the *tensor product* is the multilinear function

$$\omega_1 \otimes \cdots \otimes \omega_k(v_1, \ldots, v_k) \equiv \omega_1(v_1) \ldots \omega_k(v_k)$$

i.e.,

$$(\omega \wedge \eta)(v_1, \ldots, v_{k+\ell}) = \frac{1}{k!\ell!} \sum_{\sigma \in S_{k+\ell}} \mathrm{sgn}(\sigma)\, \omega(v_{\sigma(1)}, \ldots, v_{\sigma(k)})\, \eta(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}). \qquad \text{(ii)}$$

For example, if ω and η are both 1-forms, then $(\omega \wedge \eta)(v_1, v_2) = \omega(v_1)\eta(v_2) - \omega(v_2)\eta(v_1)$.

Naturality of the exterior product. For any smooth function $F : \mathcal{M} \longrightarrow \mathcal{N}$,

$$F^*(\omega \wedge \eta) = (F^*\omega) \wedge (F^*\eta) \qquad (\mathrm{nat}_\wedge)$$

i.e., the pullback of a smooth function distributes over the wedge product.

Proof. At any point $p \in \mathcal{M}$,

$$\begin{aligned}
F_p^*(\omega \wedge \eta)[v_n]_{n=1}^{k+\ell} &\overset{\text{(i)}}{=} (\omega \wedge \eta)_{F(p)}\big(\mathrm{D}F_p(v_1), \ldots, \mathrm{D}F_p(v_{k+\ell})\big) = (\omega \wedge \eta)_{F(p)}\big(\mathrm{D}F_p(v_n)\big)_{n=1}^{k+\ell} \\
&\overset{\text{(ii)}}{=} \frac{1}{k!\,\ell!} \sum_{\sigma \in S_{k+\ell}} \mathrm{sgn}(\sigma)\, \omega_{F(p)}\big(\mathrm{D}F_p(v_{\sigma(n)})\big)_{n=1}^{k}\, \eta_{F(p)}\big(\mathrm{D}F_p(v_{\sigma(n)})\big)_{n=k+1}^{k+\ell} \\
&\overset{\text{(i)}}{=} \frac{1}{k!\,\ell!} \sum_{\sigma \in S_{k+\ell}} \mathrm{sgn}(\sigma)\, F_p^*(\omega)[v_{\sigma(1)}, \ldots, v_{\sigma(k)}]\, F_p^*(\eta)[v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}] \\
&\overset{\text{(ii)}}{=} \big(F^*\omega \wedge F^*\eta\big)_p(v_1, \ldots, v_{k+\ell})
\end{aligned}$$

for any collection of tangent vectors $v_1, \ldots, v_{k+\ell} \in \mathrm{T}_p\mathcal{M}$. ∎

3. Exterior differentiation

A smooth covector field ω is said to be *conservative* if there exists a *potential* $f \in \mathscr{F}(\mathcal{M})$ s.t. $\omega = \mathrm{D}f$, that is, in coordinates:

$$\omega_i = \frac{\partial f}{\partial x^i} \quad \text{hence} \quad \frac{\partial \omega_j}{\partial x^i} = \frac{\partial \omega_i}{\partial x^j},$$

we will say that ω is *closed* if it satisfies this last condition and define the *exterior derivative* to be the 2-form

$$\mathrm{d}\omega = \sum_{i<j} \left(\frac{\partial \omega_j}{\partial x^i} - \frac{\partial \omega_i}{\partial x^j} \right) \mathrm{d}x^i \wedge \cdots \wedge \mathrm{d}x^i$$

that vanishes exactly when ω is closed. Similarly, $\mathrm{d}\omega$ will be a $(k+1)$-form if ω is a k-form, and it is closed if $\mathrm{d}\omega = 0$. Notice that

$$\mathrm{d}^2 = \mathrm{d} \circ \mathrm{d} \equiv 0$$

and thus any exact form is closed. The converse implication is only true locally, or under sufficient conditions first established by Poincaré.

Poincaré lemma. A closed differential k-form ω, on a contractible region Ω of $\mathcal{M}$, is exact.

The intuition for a space being contractible is that it can be "deformed in a continuous way" to a single point and thus its k^{th} *de Rham cohomology group*

$$H^k(\Omega) = \frac{\ker \mathrm{d}}{\operatorname{im} \mathrm{d}}$$

is trivial, and this is actually the result Poincaré proved using his celebrated lemma and why we still refer to it as such. For 1-forms, contractibility is indeed equivalent to the differential form being supported on a simply connected domain, as we need only to look to $H^1(\Omega)$ and check that it is trivial: that is, its group rank, which we call the *first Betti number*, is zero. In higher dimensions, i.e. for a k-form with $k > 1$, this is true over regions with "no k-dimensional holes".

However, here the catch is that when a differential form is exact, then its integral is independent of the path chosen: say $\omega = \mathrm{d}\eta$, then

$$\int_\Omega \omega \overset{\text{Stokes}}{=} \int_{\partial\Omega} \mathrm{d}\omega = \int_{\partial\Omega} \mathrm{d}^2\eta \equiv 0,$$

and we have thus, finally, provided a full justification for our very early remark on the characterization of gradient flows as closed forms. For deeper implications in topology, see [Ful97], and in geometry, notice how the definition of exterior derivative of a differential form is a measure of how much suitable generalizations of (GFC) "fail".

Removing topological assumptions, it is very easy to construct, for instance, (scalar) curl-free vector fields that serve as a counterexample for punctured, not simply connected, domains, such as the *vortex*

$$V(x,y) = \frac{1}{2\pi(x^2+y^2)}(-y,x), \quad (x,y) \in \mathbb{R}^2_* \equiv \mathbb{R}^2 \backslash \{0\},$$

which is not conservative because

$$0 \neq \oint_{\partial \mathbb{D}} V = \int_0^{2\pi} \frac{\sin^2 t + \cos^2 t}{2\pi} \, \mathrm{d}t = 1$$

and this is the most basic rotational non-conservative component on the punctured plane $\mathbb{R}^2_*$, because any other irrotational vector field F can be Helmholtz-decomposed as follows:

$$F = aV + \nabla\varphi, \quad a \in \mathbb{R}, \quad \varphi \in \mathbf{C}^2(\mathbb{R}^2_*),$$

namely, curl-free vector fields on the punctured plane are gradients up to a one-degree-of-freedom error term. If we "add more holes", say the first Betti number of Ω is n, then any curl-free vector field F has the decomposition

$$F = \sum_{k=1}^{n} a_k V_i + \nabla\varphi, \quad \text{for each hole } (x^*,y^*), V_i(x,y) \equiv \frac{((x-x^*),(y-y^*))}{2\pi((x-x^*)^2+(y-y^*)^2)}.$$

3.1 GENERIC *musical* formalism

Given any map $f \in \mathscr{F}(\mathcal{M})$, then $\mathrm{d}f \in \mathrm{T}^*\mathcal{M}$ is a one-form and, supposing our manifold is endowed with some non-degenerate bilinear form $B : \mathrm{T}\mathcal{M} \times \mathrm{T}\mathcal{M} \longrightarrow \mathbb{R}$ on the tangent bundle, we can consider the so-called *musical isomorpshism* $\sharp : \mathrm{T}^*\mathcal{M} \longrightarrow \mathrm{T}\mathcal{M}$ of bundles sending any covector ω to the unique vector $\omega_B^\sharp$ such that $B(\omega_B^\sharp, X) = \omega(X)$ for any tangent $X \in \mathrm{T}\mathcal{M}$. In the respective cases in which $B = g$ is a Riemannian metric, then $(\mathrm{d}f)_g^\sharp$ corresponds precisely to ($\sharp_g$), and when $B = \Omega$ is the symplectic form, then $(\mathrm{d}f)_\Omega^\sharp$ is precisely equivalent to the "symplectic gradient" $X_f = \{\cdot, f\}$ as in ($\mathscr{H}$). Thus we can, rather obscurely, rewrite (GENERIC) as:

$$\frac{\mathrm{d}}{\mathrm{d}t}u_t = (\mathrm{d}H)_\Omega^\sharp u_t - (\mathrm{d}\varphi)_g^\sharp u_t$$

and ($\star$): $0 = (\mathrm{d}H)_g^\sharp = (\mathrm{d}\varphi)_\Omega^\sharp \in \mathrm{T}_{u_t}\mathcal{M}$.

4. Orientation

If the manifold is d-dimensional, any non-vanishing d-form determines a unique orientation as follows: to each $\mathrm{T}_p\mathcal{M}$ the set of ordered bases $E_1, \ldots, E_n$ such that $\omega(E_1, \ldots, E_n) > 0$ gives the pointwise orientation and, because in coordinates $\omega = f\,\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$ for a non-vanishing $f \in \mathscr{F}(\mathcal{M})$, every pointwise orientation can be chosen to be positive, hence it is continuous. If $F : \mathcal{M} \longrightarrow \mathcal{N}$ is a local diffeomorphism and $\mathcal{N}$ is oriented by the form ω, then $F^*\omega$ is an orientation form for $\mathcal{N}$.

APPENDIX D

Lagrangian and Hamiltonian formalisms of mechanics

Since we are assuming the Principle of Least Action together with Netwon's equations from the very start of our treatment, it would be nice for the Hamiltonian formalism of mechanics, depicted in Chapter II, to be coherent with such axioms. One of the advantages of working with some *configuration manifold*, in mechanics coordinatized by $(q^1, \dots, q^n)$, is that we get a convenient geometric description of the phase space in the cotangent bundle $\mathrm{T}^*\mathcal{M}$ with induced local coordinates $(q^1, \dots, q^n, p_1, \dots, p_n)$. In fact, $(\mathrm{d}q^1, \dots, \mathrm{d}q^n)$ is a basis of each $\mathrm{T}^*_q\mathcal{M}$, and thus, writing each covector $\omega \in \mathrm{T}^*\mathcal{M}$ as the linear combination $\omega = p_i \mathrm{d}q^i$ [‡], the (momentum) phase space has the "intrinsic" symplectic form

$$\hat{\Omega} = \mathrm{d}q^i \wedge \mathrm{d}p_i,$$

which we shall call *Poincaré form* on the cotangent bundle. [MR13, Ch. 6]

1. Euler-Lagrange equations

The Lagrangian formulation of classical mechanics is instead naturally rooted in the (velocity) phase space $\mathrm{T}\mathcal{M} \ni (q^1, \dots, q^n, \dot{q}^1, \dots, \dot{q}^n)$ of a configuration space, often pointwise endowed with a Riemannian metric $g_{ij}(q)$ expressing some quadratic kinetic energy on each $\mathrm{T}_q\mathcal{M}$. The bulk of Lagrange's approach is to introduce the momentum as follows:

$$p = \frac{\partial L}{\partial \dot{q}}$$

for some *Lagrangian* $\mathcal{L} \in \mathscr{F}(\mathrm{T}\mathcal{M})$ given by:

$$\mathcal{L}(q, \dot{q}) = \frac{1}{2} m |\dot{q}|^2 - V(q) = \frac{1}{2} g_{ij}(q) \dot{q}^i \dot{q}^j - V(q).$$

[‡] The fact that coordinates live on the underlying manifold and the momenta live up in the cotangent bundle is why these are conventionally and conveniently (for Einstein's notation) written with upper and lower indexes.

Hence, we ought to minimize the action

$$J(q) = \int_\gamma p \, \mathrm{d}q$$

among paths γ on $\mathcal{M}$ with fixed endpoints. Suppose it attains a local minimum at q, then $J(q) \leqslant J(q + \delta q)$ for any *variation* δq and thus imposing stationarity:

$$\begin{aligned} 0 = \delta J &= \delta \int_\gamma \mathcal{L}(q^i, \dot{q}^i) \\ \text{(chain rule)} \ &= \int_\gamma \left(\frac{\partial \mathcal{L}}{\partial q^i} \delta q^i + \frac{\partial \mathcal{L}}{\partial \dot{q}^i} \frac{\mathrm{d}}{\mathrm{d}t} \delta q^i \right) \mathrm{d}t \\ \text{(bd. cond.)} \ &= \int_\gamma \left(\frac{\partial \mathcal{L}}{\partial q^i} - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}^i} \right) \right) \delta q^i \end{aligned}$$

i.e. we get the so-called *Euler-Lagrange equations*:

$$\frac{\partial \mathcal{L}}{\partial q^i} - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}^i} \right) = 0.$$

These are the equivalent of (GFC) in Netwonian mechanics and ($\mathscr{H}$) in Hamiltonian mechanics. Under conditions described in Appendix A, these are solved by trajectories $q(t)$ minimizing the given action functional and they generate some dissipative "Lagrangian flow" involving the Riemannian metric: because each momentum

$$p_i = \frac{\partial}{\partial \dot{q}^i} \left(\frac{1}{2} g_{jk}(q) \dot{q}^j \dot{q}^k \right) = \frac{1}{2} \left(g_{jk}(q) \delta_i^j \dot{q}^k + g_{jk} \dot{q}^j \delta_i^k \right) = g_{ij}(q) \dot{q}^j$$

i.e., in terms of musical isomorphism, $\dot{q} = p_g^\sharp$ and the Euler–Lagrange equations become:

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}^i} \right) - \frac{\partial \mathcal{L}}{\partial q^i} \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \left(g_{ij}(q) \dot{q}^j \right) - \frac{1}{2} \frac{\partial g_{jk}}{\partial q^i} \dot{q}^j \dot{q}^k + \frac{\partial V}{\partial q^i} \\ -\frac{\partial V}{\partial q^i} &= g_{ij} \ddot{q}^j + \left(\frac{\partial g_{ij}}{\partial q^k} - \frac{1}{2} \frac{\partial g_{jk}}{\partial q^i} \right) \dot{q}^j \dot{q}^k \end{aligned}$$

which, using the *Christoffel symbols*

$$\Gamma_{ij}^k = \frac{1}{2} g^{k\ell} \left(\frac{\partial g_{\ell i}}{\partial q^j} + \frac{\partial g_{\ell j}}{\partial q^i} - \frac{\partial g_{ij}}{\partial q^\ell} \right)$$

of the metric g, can be rearranged to give:

$$\ddot{q}^i + \Gamma^i_{jk}(q)\dot{q}^j\dot{q}^k = -g^{ij}\frac{\partial V}{\partial q^j}$$

where g^{ij} is to be intended as the inverse of the metric or the $\sharp$ isomorphism: $g^{ij}g_{ij} \equiv \delta^j_i$. What we get at the end is the so-called *geodesic flow* equation

$$\ddot{q}^i + \Gamma^i_{jk}\dot{q}^j\dot{q}^k = 0$$

which admits as solutions exactly the geodesics of the manifold $\mathcal{M}$. When $V \neq 0$ the system just feels a force $-\mathrm{d}V^\sharp_g$ and we have thus shown that, assuming the Least Action Principle, (free) motion is restricted along geodesics!

This full describes how a particle moves "straight" (inertially) in curved space, following the gradient flow of the potential. It is striking that, solely by choosing as ambient space the tangent bundle, the Lagrangian formulation of mechanics we get fully describes the dissipative dynamics of GENERIC, while the other half od the dynamics, the Hamiltonian formulation's conservative one, is relegated to the cotangent bundle. Additionally, convex analysis provides us with the optimal tool to pass from one to the other anytime we want.

2. Fiber derivatives

Suppose we start with some Lagrangian $\mathcal{L} \in \mathscr{F}(\mathrm{T}\mathcal{M})$, then we can move to the cotangent bundle by means of the map $\mathcal{FL} : \mathrm{T}\mathcal{M} \longrightarrow \mathrm{T}^*\mathcal{M}$, given by:

$$\mathcal{FL}(v)[w] = \frac{\mathrm{d}}{\mathrm{d}s}\mathcal{L}(v + sw)\bigg|_{s=0}, \quad v, w \in \mathrm{T}_q\mathcal{M}$$

representing the directional derivative of the Lagrangian along the fiber $\mathrm{T}_q\mathcal{M}$, which gets mapped to its dual fiber $\mathrm{T}_q\mathcal{M}$. In coordinates,

$$\mathcal{FL}(q^i, \dot{q}^i) = \left(q^i, \frac{\partial \mathcal{L}}{\partial \dot{q}^i}\right).$$

Notice that now is rather natural to define the action as the map $J \in \mathscr{F}(\mathrm{T}\mathcal{M})$ given by:

$$J(v) = \mathcal{FL}(v)[v], \quad v \in \mathrm{T}_q\mathcal{M}$$

so that the energy functional associated to our Lagrangian

$$E(v) = J(v) - \mathcal{L}(v) = \mathcal{FL}(v)[v] - \mathcal{L}(v)$$

is its Legendre transform!
In coordinates,

$$J(q^i, \dot{q}^i) = \dot{q}^i \frac{\partial \mathcal{L}}{\partial \dot{q}^i} = p_i \dot{q}^i, \quad E(q^i, \dot{q}^i) = p_i \dot{q}^i - \mathcal{L}(q^i, \dot{q}^i).$$

Now, similarly to the natural symplectic form with which we can endow the cotangent bundle, which is the arena where Hamiltonian dyanmics rules, there is also a straightforward way to "lift" the Riemannian metric on the configuration $\mathcal{M}$ to its velocity phase space $\mathrm{T}\mathcal{M}$, by means of the so-called *Sasaki metric* [Sas58], which in local flat coordinates is given by:

$$\hat{g} \equiv g_{ij}(q) \mathrm{d}q^i \otimes \mathrm{d}q^i + g_{ij}(q) \nabla_i \dot{q}^i \otimes \nabla_j \dot{q}^i,$$

for the *covariant derivative*

$$\nabla_k \dot{q}^k \equiv \mathrm{d}\dot{q}^k + \Gamma^k_{ij} \dot{q}^i \, \mathrm{d}q^j,$$

measuring how the velocity changes covariantly while moving along the base manifold, so that, intuitively, the Sasaki metric separates horizontal and vertical component inside the tangent bundle. This allows us to define some *Lagrangian vector field* as the actual gradient flow

$$\mathrm{X}_E \equiv (\mathrm{d}E)^\sharp_{\hat{g}} \in \mathfrak{X}(\mathrm{T}\mathcal{M}).$$

Let us now instead consider the so-called *Lagrange two-form* [AM08, Def. 3.5.5]

$$\Omega_{\mathcal{L}} \equiv (\mathcal{FL})^* \, \hat{\Omega}$$

associated to the Lagrangian $\mathcal{L} \in \mathscr{F}(\mathrm{T}\mathcal{M})$ as the form induced by the pullback of the canonical symplectic form on $\mathrm{T}^*\mathcal{M}$ via its fiber derivative $\mathcal{FL} : \mathrm{T}\mathcal{M} \longrightarrow \mathrm{T}^*\mathcal{M}$. In coordinates:

$$\Omega_{\mathcal{L}} = (\mathcal{FL})^* \, \hat{\Omega} = (\mathcal{FL})^* \, (\mathrm{d}q^i \wedge \mathrm{d}p_i) \overset{(\mathrm{nat}_\wedge)}{=} (\mathcal{FL})^* \, \mathrm{d}q^i \wedge (\mathcal{FL})^* \, \mathrm{d}p_i = \mathrm{d}q^i \wedge \mathrm{d}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}^i} \right).$$

Lemma 2.1. The Lagrange two-form $\Omega_{\mathcal{L}}$ is closed.

Proof. $\mathrm{d}\Omega_{\mathcal{L}} \equiv \mathrm{d} \, (\mathcal{FL})^* \, \hat{\Omega} \overset{(\mathrm{nat}_\wedge)}{=} (\mathcal{FL})^* \, \mathrm{d}\hat{\Omega} = 0$, because $\hat{\Omega}$ is symplectic. ∎

By the inverse function theorem, assuming the Lagrangian $\mathcal{L}$ is enough regular so that $\mathcal{FL}$ is a local diffeomorphism, then the Lagrange two-form is a symplectic form on $\mathrm{T}\mathcal{M}$. With a global assumption, we can build the desired bridge between the Lagrangian and Hamiltonian formalisms.

THEOREM 2.2. If $\mathcal{L}$ is a Lagrangian such that its fiber derivative $\mathcal{FL}$ is a (global) diffeomorphism then $\mathcal{FL} : (\mathrm{T}\mathcal{M}, \Omega_{\mathcal{L}}) \longrightarrow (\mathrm{T}^*\mathcal{M}, \hat{\Omega})$ is a symplectomorphism.

Consider now the inverse $\mathcal{F}^\sharp\mathcal{L} : \mathrm{T}^*\mathcal{M} \longrightarrow \mathrm{T}\mathcal{M}$ of the Legendre transform $\mathcal{FL}$, given by:

$$\mathcal{F}^\sharp\mathcal{L}(q,p) \equiv (q, p_g^\sharp)$$

and define the Hamiltonian as the pullback of the energy via this inverse Legendre transform:

$$H(q,p) = (\mathcal{F}^\sharp\mathcal{L})^* E(q,\dot{q})$$

meaning that we take as Hamiltonian the unique map $H \in \mathscr{F}(\mathrm{T}^*\mathcal{M})$ such that the following diagram commutes:

$$\begin{array}{ccc} \mathrm{T}\mathcal{M} & \xrightarrow{\mathcal{FL}} & \mathrm{T}^*\mathcal{M} \\ & \searrow^{E} & \downarrow{\scriptstyle H} \\ & & \mathbb{R} \end{array}$$

so H is the pushforward of the Lagrangian energy through the symplectomorphism $\mathcal{FL}$:

$$H : (q,p) \in \mathrm{T}^*\mathcal{M} \overset{\mathcal{F}^\sharp\mathcal{L}}{\longmapsto} (q,\dot{q}) = (q, p_g^\sharp) \in \mathrm{T}\mathcal{M} \overset{E}{\longmapsto} E(q,\dot{q}) \in \mathbb{R}.$$

Let us now restrict to the purely kinetic case:

$$\mathcal{L}(q,\dot{q}) = \frac{1}{2} g_q(\dot{q},\dot{q})$$

so that we can reinterpret the Lagrangian vector field

$$X_E = (\mathrm{d}E)^\sharp_{\hat{g}} \equiv (\mathrm{d}E)^\sharp_{\Omega_{\mathcal{L}}}$$

as a symplectic vector field[§].

[§]This does not hold in general, and corresponds to the assumption that the Sasaki metric $\hat{g}$ and the Lagrange symplectic form $\Omega_{\mathcal{L}}$ endow $\mathrm{T}\mathcal{M}$ with a Kähler structure.

Corollary 2.3. The integral curves of $X_E = (\mathrm{d}E)^\sharp_{\Omega_\mathcal{L}} \in \mathfrak{X}(\mathrm{T}\mathcal{M})$ are mapped by $\mathcal{FL}$ onto the integral curves of $X_H = (\mathrm{d}H)^\sharp_{\hat{\Omega}} \in \mathfrak{X}(\mathrm{T}\mathcal{M})$.

Proof. Because by (nat$_\wedge$) and the definition of H via the Legendre transform

$$\mathrm{d}E = (\mathcal{FL})^*\mathrm{d}H \tag{E*H}$$

for any $v \in \mathrm{T}\mathcal{M}$ and $v^* = \mathrm{D}\mathcal{FL}(v)[w]$ with $w \in \mathrm{T}_v\mathrm{T}\mathcal{M}$ we have that

$$\begin{aligned}\hat{\Omega}(\mathrm{D}\mathcal{FL}(X_E(v)), v^*) &= \Omega_\mathcal{L}(X_E(v), w) = \mathrm{d}E(v)[w] \\ &= \mathrm{d}(\mathcal{FL})^*(v)[w] \overset{(\mathrm{E^*H})}{=} \mathrm{d}H(\mathcal{FL}(v))[v^*] \\ &= \hat{\Omega}(X_H(\mathcal{FL}(v)), v^*).\end{aligned}$$

This proves that the following diagram commutes:

$$\begin{array}{ccc} \mathrm{T}\mathcal{M} & \xrightarrow{\mathcal{FL}} & \mathrm{T}^*\mathcal{M} \\ {\scriptstyle X_E}\downarrow & & \downarrow{\scriptstyle X_H} \\ \mathrm{TT}\mathcal{M} & \xrightarrow[\mathrm{D}\mathcal{FL}]{} & \mathrm{TT}^*\mathcal{M} \end{array}$$

i.e.

$$\mathrm{D}\mathcal{FL}(X_E(v)) = X_H(\mathcal{FL}(v)) \quad \forall v \in \mathrm{T}\mathcal{M}$$

which is the desired claim, as it implies that $\mathrm{D}\mathcal{FL}(X_E) = X_H$. ∎

Hence, the dynamics on $\mathrm{T}\mathcal{M}$ and $\mathrm{T}^*\mathcal{M}$ describe the same physical system, the Legendre transform acts as a coordinate change between the tangent and cotangent bundles and, because $\mathcal{FL} \in \mathrm{Sp}(2n)$, it maps solutions of the Euler–Lagrange equations [‡] to solutions of Hamilton's equations.

3. The Lagrange–d'Alembert Principle

Now, suppose having a mechanical system with a Lagrangian $\mathcal{L}$ and an exterior force one-form η acting on it. The dissipation induced by the external environment via η is exactly the type of behavior we would like to model using GENERIC. Before the introduction of Öttinger's

[‡] Or, equivalently, in the kinetic energy case, solutions to the Sasaki-gradient and geodesic flow equations.

framework, the most general proposal of dynamics for such a system was given by vector fields of the following form:

$$X = (\mathrm{d}E - \eta)^\sharp_{\Omega_\mathcal{L}}$$

which has a striking similarity to our musical reformulation of GENERIC. [MR13, Sect. 7.8] The limitation of this symplectic formulation is that it lacks a Riemannian metric g or, rather, that the choice of g cannot be made arbitrary: as before, through a Legendre transform, the only gradient flow we can model is the one associated to geodesic motion. This rough principle's power in modelling dissipative Lagrangian systems can be strengthened by considering associated Rayleigh dissipation potentials and it serves as a formal justifications for introducing the latters in the context of gradient flows. The crucial observation is that the Lagrange–d'Alembert Principle is equivalent to the Euler-Lagrange equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}}\right) - \frac{\partial \mathcal{L}}{\partial q} = -\eta(q, \dot{q})$$

and, in particular, we say that η is a *dissipative form* if:

$$\mathrm{d}E(\eta^\sharp) \leqslant 0$$

in which case it induces the force field $F : \mathrm{T}\mathcal{M} \longrightarrow \mathrm{T}^*\mathcal{M}$ given by:

$$F(v)[\eta] = -\Omega_\mathcal{L}(\eta^\sharp(v), V_v), \quad v \equiv \frac{\mathrm{d}q}{\mathrm{d}t}$$

and where $V_v \in \mathrm{T}_v\mathrm{T}\mathcal{M}$ such that $\mathrm{D}V_v = \eta$.

Example ([MR13]). Consider a Riemannian manifold $\mathcal{M}$ and a potential $\varphi : \mathcal{M} \longrightarrow \mathbb{R}$ such that $\eta = -\mathrm{d}\varphi$ is dissipative, then the Lagrange-d'Alembert principle for the kinetic energy K with external forces given by the one-form η produces the same dynamics as the standard kinetic minus potential Lagrangian.

In this context, a dissipation potential is a function $R : \mathrm{T}\mathcal{M} \longrightarrow \mathbb{R}$ such that the force field F is the fiber derivative of $-R$. As a consequence, the Euler–Lagrange equations with forcing become

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}^i}\right) - \frac{\partial \mathcal{L}}{\partial q^i} = -\frac{\partial R}{\partial \dot{q}^i}.$$

We get a dissipative system exactly when $\nabla^2 R(q, v)[v] > 0$.

Remark ([MR13]). If R is linear in the fiber variable, the Rayleigh dissipation potential takes on the classical form $\mathcal{R}_q(v)[v]$, where $\mathcal{R}_q : \mathrm{T}\mathcal{M} \longrightarrow \mathrm{T}^*\mathcal{M}$ is a bundle map over the identity that defines a symmetric positive definite form on each fiber of $\mathrm{T}\mathcal{M}$.

Luckily enough, we find that under the flow of the Euler–Lagrange equations with forcing of Rayleigh dissipation type, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}E(q,v) = F(v)[v] = -\mathcal{F}R(q,v)[v] < 0$$

that is, energy dissipation. Namely, we get a satisfactory model of irreversible dynamics within a symplectic framework.

APPENDIX E

Neural Networks

Here, we would like to briefly explain why it is a good idea to approximate dynamical systems using neural networks, following [Ber+21].

1. Universal approximators

Neural networks are conceived to solve classification tasks. Consider the following model for an artificial neuron: some input vector x is weighted by the conductivity-*weights* vector w, this voltage $\langle w, x\rangle$ of the soma is compared to some threshold *bias* b. If it exceeds the threshold, the neuron is active and it fires:

$$y = \sigma(\langle w, x\rangle + b)$$

for some *activation function* $\sigma : \mathbb{R} \longrightarrow \mathbb{R}$, and this is the so-called *perceptron* model. In the original proposal by McCulloch and Pitts, σ is chosen to be the Heaviside step function, so that a binary output is obtained; lately, smooth versions (of very dense around 1 and -1 functions) were preferred: most famously, the sigmoid

$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$

Through training, the parameters w, b can be fine-tuned so that the perceptron can shatter any linear classification task, namely, if the convex hulls of datapoints corresponding to different labels are separable by means of some hyperplane. If this is not the case, we could in principle implement a layered architecture, in which the input is sent to an array of different perceptrons, whose output is then collected and sent to an output perceptron.

Definition 1.1. Let $K \subset \mathbb{R}^d$ be compact, then a continuous function $\sigma : \mathbb{R} \longrightarrow \mathbb{R}$ is said to be *discriminatory* if the only signed measure μ on K such that

$$\int_K \sigma(\langle w, x\rangle + b)\, \mathrm{d}\mu(x) = 0 \quad \text{for all} \quad w \in \mathbb{R}^d, b \in \mathbb{R}$$

is the zero measure $\mu \equiv 0$.

Example. Any sigmoidal function ϱ, i.e. one that, for blowing-up weights:

$$\varrho(\langle w, x\rangle + b) \longrightarrow (\chi_{(0,+\infty)} + \varrho(b)\chi_{\{0\}})(x),$$

is discriminatory. In fact, by superposition, in the limit:

$$\int_K \chi_{[a,b]}(\langle w, x\rangle + b)\, \mathrm{d}\mu(x) = 0$$

and since we can uniformly approximate the exponential function: $x \longmapsto e^{-2\pi i x}$ by linear combinations of indicator functions $\chi_{[a,b]}$, we obtain:

$$\int_K e^{-2\pi i(\langle w,x\rangle + b)}\, \mathrm{d}\mu(x) = 0$$

namely, the Fourier transform of μ vanishes and hence $\mu \equiv 0$.

Precisely, we speak of a 1-hidden-layer neural network to mean a map $f : K \longrightarrow \mathbb{R}$ of the form

$$f(x) = \sum_{j=1}^{h} c_j \sigma(\langle w^{(j)}, x\rangle + b^{(j)})$$

for a collection of vectors $w^{(j)} \in \mathbb{R}^d$, biases $b^{(j)}$, $j = 1, \ldots, h$, and coefficients $c \in \mathbb{R}^h$. Luckily, the solution of a layered architecture is mathematically sound, thanks to the following result.

Universal Approximation Theorem. For any continuous function $g : K \longrightarrow \mathbb{R}$, discriminatory $\sigma : \mathbb{R} \longrightarrow \mathbb{R}$ and any $\epsilon > 0$, there exists a 1-hidden-layer network $f : K \longrightarrow \mathbb{R}$, with activation functions σ, such that

$$\sup_{x \in K} |f(x) - g(x)| < \epsilon. \tag{UAT}$$

Proof. Technically, we would like to prove that the set

$$\mathcal{F} := \left\{ f(x) = \sum_{j=1}^{h} c_j \, \sigma(w^{(j)} \cdot x + b^{(j)}) \;\middle|\; h \in \mathbb{N},\ c_j \in \mathbb{R},\ w^{(j)} \in \mathbb{R}^d,\ b^{(j)} \in \mathbb{R} \right\}$$

is dense in $\mathcal{C}(K)$, the vector space of continuous real-valued functions on K. Suppose, by contradiction, that $\mathcal{F}$ is not dense in $\mathcal{C}(K)$. Then, by the Hahn–Banach theorem, there exists a non-zero bounded linear functional $\Psi \in \mathcal{C}^*(K)$ such that

$$\Psi(f) = 0 \quad \text{for all } f \in \mathcal{F}.$$

By the Riesz Representation Theorem, there exists a non-zero finite signed Borel measure μ on K such that

$$\Psi(f) = \int_K f(x) \, \mathrm{d}\mu(x).$$

In particular, for any $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$, the function $x \mapsto \sigma(\langle w, x \rangle + b)$ belongs to $\mathcal{F}$, so we must have

$$\int_K \sigma(\langle w, x \rangle + b) \, \mathrm{d}\mu(x) = 0 \quad \text{for all } w \in \mathbb{R}^d,\ b \in \mathbb{R}.$$

However, since σ is discriminatory, the only measure μ that satisfies this condition for all w, b is $\mu \equiv 0$, contradicting our earlier assumption that μ is non-zero.
Therefore, $\mathcal{F}$ must be dense in $\mathcal{C}(K)$. ∎

The theorem gives assurance there is a way to approximate any function but it does not tell us how: the training of the architecture's weights and biases will give us that information.

2. Multi-Layer Perceptrons

There is, however, a canard regarding implementation: the theorem does not bound the size of this hidden layer of perceptrons and that could, in principle, grow impractically large. This last problem is easily solved by considering deeper multi-layered architectures, namely, the so-called Multi-Layer Perceptrons.
Formally, a L-layers neural network is given by

$$u^{(\ell)}(x) = W^{(\ell)} \sigma(u^{(\ell-1)}(x)) + b^{(\ell)}, \quad 1 \leqslant \ell \leqslant L + 1$$

where each $W^{(\ell)}$ is a *weight* matrix and each $b^{(\ell)} \in \mathbb{R}^{n_\ell}$ some bias vector, while σ is the (nonlinear) *activation function*. Now, we can give a name to our parametrization: namely, it is the collection $\vartheta = \left\{(W^{(\ell)}, b^{(\ell)})\right\}_\ell$ of weights and biases and, in a similarly compact way, given any differentiable function $g : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, we can store the slopes of its nonlinearities in a diagonal matrix-valued function Σ_j defined by

$$(\Sigma_j)_{ii} = \dot{\sigma}(u_i^{(j)}(g(x))), \quad i \leqslant j \leqslant L, \quad 1 \leqslant i \leqslant n_j$$

where d has to match the dimensionality of the input.

Bibliography

[AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Birkhäuser Verlag, 2008.

[AM08] R. Abraham and J. E. Marsden. *Foundations of mechanics*. AMS, 2008.

[Ber+21] Berner et al. *The modern mathematics of deep learning*. 2021. arXiv: 2105.04026.

[BV04] Boyd and Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[De 93] Ennio De Giorgi. "New problems on minimizing movements". In: *Ennio de Giorgi: selected papers* (1993), pp. 699–713.

[DR05] Denis Donnelly and Edwin Rogers. "Symplectic integrators: An introduction". In: *American Journal of Physics* 73.10 (2005), pp. 938–945.

[DS10] Sara Daneri and Giuseppe Savaré. *Lecture notes on gradient flows and optimal transport*. 2010. arXiv: 1009.3737.

[Ful97] William Fulton. *Algebraic topology: a first course*. Springer, 1997.

[GÖ97] M. Grmela and H. C. Öttinger. "Dynamics and thermodynamics of complex fluids. I. Development of a general formalism". In: *Phys. Review E* 56.6 (1997).

[Gon96] Oscar Gonzalez. "Time integration and discrete Hamiltonian systems". In: *Journal of Nonlinear Science* 6 (1996), pp. 449–467.

[HL04] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of convex analysis*. Springer, 2004.

[JST19] Ansgar Jüngel, Ulisse Stefanelli, and Lara Trussardi. "Two structure-preserving time discretizations for gradient flows". In: *Appl. Math. & Optim.* 80 (2019), pp. 733–764.

[JST22] Ansgar Jüngel, Ulisse Stefanelli, and Lara Trussardi. "A minimizing-movements approach to GENERIC systems". In: *Mathematics in Engineering* 4.1 (2022), pp. 1–18.

[Lee12] John M. Lee. *Introduction to Smooth Manifolds*. Springer, 2012.

[Mie23] Alexander Mielke. *An introduction to the analysis of gradients systems*. 2023. arXiv: 2306.05026.

[MR13] Jerrold E Marsden and Tudor S Ratiu. *Introduction to mechanics and symmetry: a basic exposition of classical mechanical systems*. Springer-Verlag, 2013.

[OM53] Lars Onsager and Stefan Machlup. "Fluctuations and irreversible processes". In: *Physical Review* 91.6 (1953), pp. 1505–1512.

[Ött05] Hans Christian Öttinger. *Beyond Equilibrium Thermodynamics*. Wiley, 2005.

[Ött18] Hans Christian Öttinger. "GENERIC integrators: structure preserving time integration for thermodynamic systems". In: *Journal of Non-Equilibrium Thermodynamics* 43.2 (2018), pp. 89–100.

[Pav14] Grigorios A. Pavliotis. *Stochastic processes and applications*. Springer, 2014.

[Rom09] Ignacio Romero. "Thermodynamically consistent time-stepping algorithms for nonlinear thermomechanical systems". In: *International journal for numerical methods in engineering* 79.6 (2009), pp. 706–732.

[RS23] Riccarda Rossi and Giuseppe Savaré. "Abstract action spaces and their topological and dynamic properties". In: *Discrete and Continuous Dynamical Systems - S* (2023).

[Sas58] Shigeo Sasaki. "On the differential geometry of tangent bundles of Riemannian manifolds". In: *Tohoku Mathematical Journal, Second Series* 10.3 (1958), pp. 338–354.

[ZSE22] Zhen Zhang, Yeonjong Shin, and George Em Karniadakis. "GFINNs: GENERIC formalism informed neural networks for deterministic and stochastic dynamical systems". In: *Philosophical Transactions of the Royal Society A* 380 (2022).