Information geometry of diffeomorphism groups

Boris Khesin and Gerard Misiołek

Abstract The study of diffeomorphism groups and their applications to problems in analysis and geometry has a long history. Geometric hydrodynamics pioneered by V. Arnold in the 1960s considers an ideal fluid flow as a geodesic motion on the infinite-dimensional group of volume-preserving diffeomorphisms of a flow domain equipped with the energy metric. Similar considerations on the space of densities led to a geometric description of the optimal mass transport with the Kantorovich-Wasserstein metric. In the same vein information geometry associated with the Fisher-Rao metric and Hellinger distance can be viewed as an analogue of optimal transportation equipped with a higher-order Sobolev metric. In the present chapter we describe various metrics on diffeomorphism groups, introduce appropriate topology, smooth structures, and describe the dynamics on such infinite-dimensional manifolds. One of the goals in this chapter is to explain how, alongside with topological hydrodynamics, symplectic dynamics and mass transport problems, information geometry with its sophisticated toolbox and techniques has become yet another area for potential applications of geometric analysis on diffeomorphism groups.

Boris Khesin
Department of Mathematics, University of Toronto, Toronto, ON M5S 2E4, Canada, e-mail: khesin@math.toronto.edu

Gerard Misiołek
Department of Mathematics, University of Notre Dame, Notre Dame, IN 46556, USA e-mail: gmisiole@nd.edu

Contents

1 Introduction

One of the main ingredients of the geometric approach to hydrodynamics pioneered by V. Arnold [5] was the consideration of the group of volume-preserving diffeomorphisms of a flow domain as a natural infinite-dimensional Riemannian manifold and a configuration space for ideal fluids. The geodesic flow on that group related to the L^2 metric induced by the fluid kinetic energy describes solutions of the Euler equation of an incompressible inviscid fluid. The general framework of Arnold turned out to include a variety of other nonlinear partial differential equations of mathematical physics, now often referred to as the *Euler-Arnold* equations. The L^2 metric on the diffeomorphism group also happened to be related to another flourishing mathematical domain, namely, the problem of optimal mass transport.

Subsequently, it turned out that a similar geometric approach to diffeomorphism groups, albeit related to Sobolev H^1-type metrics, sheds new light on many notions in geometric statistics and information theory. In our survey we focus on these metrics on diffeomorphism groups and their quotient spaces viewed as spaces of densities.

We start with a very detailed background introducing appropriate tame Fréchet topology and smooth structures on infinite-dimensional manifolds, for the future use on diffeomorphism groups and density spaces. This allows us to put on a firm ground all the relevant differential-geometric and dynamical considerations. We continue by presenting the Euler-Arnold equations for the geodesic flow in one-sided invariant metrics on Lie groups, both in finite and infinite dimensions. We also show how the L^2 metric on the group of diffeomorphisms naturally descends to the Wasserstein metric on the space of densities and whose geodesics provide the way of optimal mass transfer.

On the other hand, geometric statistics turns out to be closely related to degenerate right-invariant $\dot{H}^1$ Riemannian metrics on the full diffeomorphism group. In a sense, in the framework of diffeomorphism groups, information geometry associated with the Fisher-Rao metric and its spherical Hellinger distance can be viewed as an $\dot{H}^1$-analogue of standard optimal transport associated with the metric on the density space induced by the (non-invariant) L^2-metric on the group of all diffeomorphisms and whose Riemannian distance is the celebrated Wasserstein distance, see [30].

We describe this geometry in detail and discuss properties of solutions of the associated geodesic equations. It turns out that the corresponding geometry on the space of densities is *spherical* for *any* compact manifold M, while the corresponding Euler-Arnold equation is a natural generalization of the completely integrable one-dimensional Hunter-Saxton equation. Lastly, we present geometric constructions of the so-called α-connections introduced in geometric statistics by Chentsov [15] and Amari [1] as well as their generalizations to diffeomorphism groups of higher-dimensional manifolds, following [36].

The approach developed in this chapter places information geometry squarely within the general differential-geometric framework of diffeomorphism groups envisioned at various times by Cartan, Kolmogorov, Kantorovich and Arnold which includes hydrodynamics, symplectic geometry and optimal transport. It provides a foundation for a natural infinite dimensional generalization of this fast growing field and may hopefully lead to new insights and further developments. For these reasons and for the benefit of those readers approaching these subjects for the first time we included a more extensive background on nonlinear functional analysis.

2 Infinite-dimensional manifolds

It has long been recognized that many of the function spaces that arise in analysis and geometry possess a natural structure of infinite-dimensional differentiable manifolds. This includes various groups of diffeomorphisms of compact manifolds, which are of interest in this chapter. As mathematical objects these spaces are both very interesting and very complicated, and any researcher planning to take full advantage of their properties as manifolds and groups faces a problem at the outset of choosing a suitable topology.

For our purposes a convenient and natural functional-analytic framework of tame Fréchet spaces, as introduced by Sergeraert and further developed by Hamilton, provides the most convenient setting. An excellent expository article [27] can be consulted for details regarding the constructions needed in the sequel.

In this section we recall the basic notions of differential calculus in Fréchet spaces and then introduce the group of diffeomorphisms of a compact Riemannian manifold, its subgroup of diffeomorphisms preserving the volume form, and the quotient space of smooth probability densities, as tame Fréchet manifolds.

2.1 Differential calculus in Fréchet spaces

We begin with a brief review of the fundamentals of the calculus in Fréchet spaces. Most of the basic definitions and properties of Fréchet spaces can be found in the monographs of Dunford and Schwartz [19] and Rudin [48].

2.1.1 Fréchet spaces

Definition 1 A *Fréchet space* $\mathfrak{X}$ is a complete Hausdorff topological vector space whose topology is defined by a countable collection of seminorms $\|\cdot\|_k$ where $k = 0, 1, 2....$ A sequence u_n converges to u in $\mathfrak{X}$ if and only if for all k one has $\|u_n - u\|_k \to 0$ as $n \to \infty$.

The seminorms are separating in the sense that to each $u \neq 0$ there corresponds at least one k for which $\|u\|_k \neq 0$. Furthermore, the topology on $\mathfrak{X}$ is locally convex and metrizable — it has a countable local base (at the origin 0) consisting of convex sets and there is a compatible translation-invariant distance function obtained directly from the seminorms by setting

$$d(u,v) := \sum_{k=0}^{\infty} 2^{-k} \frac{\|u-v\|_k}{1+\|u-v\|_k} \qquad \text{for any } u, v \in \mathfrak{X}.$$

Bounded subsets of $\mathfrak{X}$ are precisely those which are bounded with respect to all the seminorms defining the topology. Thus, continuous linear transformations between Fréchet spaces can be characterized as mapping bounded sets to bounded sets.

Example 1 A canonical example of a Fréchet space is the space $C^\infty(M)$ of smooth functions on a compact Riemannian manifold M with a family of seminorms given by the uniform C^k norms.

Example 2 More generally, let E be a vector bundle over M equipped with a metric and a compatible connection ∇. If u is a smooth section of E then $\nabla u \in C^\infty(T^*M \otimes E)$ and, using the induced connection on the tensor product $T^*M \otimes E$, we also have $\nabla^2 u \in C^\infty(T^*M \otimes T^*M \otimes E)$. Continuing this process we obtain a countable collection of the uniform C^k norms

$$\|u\|_{C^k} = \sum_{j=0}^{k} \sup_{x\in M} |\nabla^j u(x)|, \qquad k = 0, 1, 2 \ldots, \tag{1}$$

which turn the space $C^\infty(M, E)$ of smooth sections of E into a Fréchet space. Other norms, which are often used in this setting, are the Sobolev H^k norms

$$\|u\|_{H^k}^2 = \sum_{j=0}^{k} \int_M |\nabla^j u|^2 d\mu, \qquad k = 0, 1, 2 \ldots, \tag{2}$$

where μ is the Riemannian volume form on M, and the Hölder $C^{k,\alpha}$ norms

$$\|u\|_{C^{k,\alpha}} = \|u\|_{C^k} + \sup_{x\in M, 0<r<i_M} r^{-\alpha} \omega_r(x, \nabla^k u), \qquad k = 0, 1, 2 \ldots \tag{3}$$

with $0 < \alpha < 1$ and the modulus of smoothness given by

$$\omega_r(x, u) = \sup_{y_1 \neq y_2 \in B_r(x)} \left| \nabla^k u(y_1) - \Pi_{y_1}^{y_2} \nabla^k u(y_2) \right|.$$

Here $B_r(x)$ is a geodesic ball at x of radius less than the injectivity radius $i_M > 0$ of M and $\Pi_{y_1}^{y_2}$ is the parallel translation operator in the tensor bundle $T^*M^{\otimes k} \otimes E$ along the unique minimal geodesic between y_2 and y_1.

Example 3 Another example of a Fréchet space which is of particular importance in Fourier analysis is the Schwartz space $\mathcal{S}(\mathbb{R}^n)$. Its elements are the rapidly decreasing smooth functions on $\mathbb{R}^n$ and its topology is defined this time by seminorms

$$\|u\|_k = \sup_{x\in\mathbb{R}^n, |\alpha|\le k} |x^\alpha D^\alpha u(x)| \qquad \text{for } k = 0, 1, 2 \ldots,$$

that are not norms, where $\alpha = (\alpha_1, \ldots, \alpha_n)$ is an n-tuple of nonnegative integers of length $|\alpha| = \sum_{j=1}^{n} \alpha_j$, $x^\alpha = x^{\alpha_1} \cdots x^{\alpha_n}$ and $D^\alpha = i^{-|\alpha|}\partial^\alpha/\partial x^\alpha$.

Many of the usual operations on normed spaces can be applied to construct further examples of Fréchet spaces. A closed linear subspace of a Fréchet space, the direct sum of Fréchet spaces and the quotient of a Fréchet space by a closed subspace are all Fréchet spaces. Fundamental results of abstract functional analysis such as the Hahn-Banach theorem, the closed graph theorem, and the open mapping theorem continue to hold in the Fréchet setting with proofs requiring only minor adjustments as compared with the standard Banach case. Thus, for example, an immediate consequence of the open mapping theorem is that any continuous linear bijection between Fréchet spaces has a continuous inverse, i.e., it is a topological isomorphism (homeomorphism) of Fréchet spaces. For proofs of such facts see e.g., [19], [31] or [48].

An important exception arises when constructing the dual of a Fréchet space $\mathfrak{X}$, that is, the space $\mathfrak{X}^*$ of continuous linear functionals on $\mathfrak{X}$. Although the assumption of local convexity ensures, via the Hahn-Banach theorem, a good supply of such functionals, this is insufficient to guarantee that $\mathfrak{X}^*$ is a Fréchet space. In general, the dual space of a locally convex space does not carry any distinguished topology and it will not be Fréchet unless $\mathfrak{X}$ itself is normable, cf. [31]. For example, the dual of $C^\infty(M)$ is the space $\mathcal{D}'(M)$ of distributions on a compact manifold M. The same problem arises for more general continuous linear transformations between Fréchet spaces and hence some care must be taken when working with notions involving families of such maps. In particular, this entails the use of a notion of differentiability based on the Gateaux (directional) derivative.

2.1.2 The Gateaux derivative and its properties

Definition 2 Let $\mathfrak{X}$ and $\mathfrak{Y}$ be Fréchet spaces. A continuous map $f : \mathfrak{X} \supset \mathscr{U} \to \mathfrak{Y}$ is of class C^1 (continuously differentiable) on an open subset $\mathscr{U} \subset \mathfrak{X}$ if the limit

$$df(u)h = f'(u)h := \lim_{t\to 0} \frac{f(u+th) - f(u)}{t} \tag{4}$$

exists for all $u \in \mathscr{U}$ and $h \in \mathfrak{X}$ and if the map $df : \mathscr{U} \times \mathfrak{X} \to \mathfrak{Y}$ is continuous as a function of both variables. Partial derivatives of functions depending on two or more variables are defined in the usual manner.

It should be noted that if $\mathfrak{X}$ and $\mathfrak{Y}$ are Banach spaces then this notion of differentiability is weaker than the standard one based on the Fréchet derivative which requires that the map $u \to df(u)$ from $\mathscr{U}$ to the space of bounded linear maps $L(\mathfrak{X}, \mathfrak{Y})$ be continuous in the operator norm topology. In fact, the two derivatives do not coincide even in finite dimensions.

Example 4 The real-valued function

$$f(x,y) = \begin{cases} 0 & \text{if} \quad (x,y) = (0,0) \\ x + y + \dfrac{x^3 y}{x^4 + y^2} & \text{if} \quad (x,y) \neq (0,0) \end{cases}$$

is Gateaux but not Fréchet differentiable at the origin in $\mathbb{R}^2$.

The second derivative $d^2 f$ is defined as the derivative of the first order derivative, and we say that f is of class C^2 if $d^2 f$ exists and is continuous jointly as a function on the product space. Now C^r functions are defined by induction and we say that f is smooth of class C^∞ if it is C^r for all r. A function of two (or more) variables that is jointly continuous, C^∞ smooth with respect to one of the variables and linear with respect to the other is easily seen to be jointly smooth in both variables.

Notwithstanding differences between the two notions,[1] the Gateaux derivative of any continuously differentiable function enjoys most of the expected properties and obeys the usual rules of calculus.

Proposition 1 *Let $f : \mathfrak{X} \supset \mathscr{U} \to \mathfrak{Y}$ be a function of class C^r between Fréchet spaces and let $\mathscr{U} \subset \mathfrak{X}$ be an open set.*

1. *(Linearity) For any $u \in \mathscr{U}$ and any $h, k \in \mathfrak{X}$ and for any scalar c we have*

$$df(u)(ch + k) = c\,df(u)h + df(u)k.$$

2. *(Fundamental theorem of calculus) For any $h \in \mathscr{X}$ we have*

$$f(u + th) - f(u) = \int_0^1 df(u + th)h\,dt$$

provided that $\mathscr{U}$ is convex so that the entire segment $u + th$ $(0 \le t \le 1)$ lies in $\mathscr{U}$; in particular, f is locally constant if and only if $df = 0$.

3. *For any $u \in \mathscr{U}$ the map $(h_1, \dots, h_r) \to (d^r f)(u)(h_1, \dots, h_r)$ is symmetric and r-linear.*
4. *(Chain rule) If g is another function of class C^r then so is the composition $g \circ f$, and we have*

$$d(g \circ f)(u) = dg(f(u)) \cdot df(u),$$

as well as analogous formulas for the iterated derivatives.

5. *(Taylor's formula) For any $h \in \mathfrak{X}$ we have*

$$f(u + h) = f(u) + df(u)h + \cdots + \frac{1}{(r-1)!} \int_0^1 (1-t)^{r-1} (d^r f)(u + th)(h, \dots, h)\,dt$$

provided that $r \ge 2$ and the segment $u + th$ lies in $\mathscr{U}$.

Proof The proofs of all these facts are carried out in a routine manner with the help of the Hahn-Banach theorem by reducing to the real-valued case and applying the classical finite-dimensional calculus, see e.g., [27]. □

[1] It is not hard to show that any function of class C^2 in the above sense is continuously differentiable in the sense of the standard Fréchet differentiability.

2.1.3 Manifolds modelled on Fréchet spaces

We now have enough tools of the Fréchet calculus to generalize a number of standard constructions of differential topology such as manifolds, vector bundles, principal bundles etc. to the Fréchet setting.

Definition 3 A *Fréchet manifold* $\mathfrak{M}$ modelled on a Fréchet space $\mathfrak{X}$ is a Hausdorff topological space equipped with a maximal atlas of (pairwise compatible) coordinate charts $\{\mathscr{U}_\alpha, \varphi_\alpha\}$ where $\{\mathscr{U}_\alpha\}$ form an open cover of $\mathfrak{M}$ and $\varphi_\alpha : \mathscr{U}_\alpha \to \mathscr{V}_\alpha$ are homeomorphisms onto open subsets $\mathscr{V}_\alpha$ of $\mathfrak{X}$ such that for all α, β the coordinate transition maps $\varphi_\alpha \circ \varphi_\beta^{-1}$ defined on $\varphi_\beta(\mathscr{U}_\alpha \cap \mathscr{U}_\beta)$ are of class C^∞.

Remark 1 $\mathfrak{M}$ is locally compact if and only if the model space $\mathfrak{X}$ is finite-dimensional. In this case $\mathfrak{M}$ is a smooth manifold in the usual sense.

Many standard constructions can be now carried over from finite dimensions to the Fréchet setting without much difficulty.

(i). A subset $\mathfrak{S} \subset \mathfrak{M}$ is a *Fréchet submanifold* of $\mathfrak{M}$ if $\mathfrak{X} = \mathfrak{X}_1 \times \mathfrak{X}_2$ is a product of Fréchet spaces and if around each point in $\mathfrak{S}$ there is a coordinate chart $(\mathscr{U}, \varphi)$ with $\varphi : \mathscr{U} \subset \mathfrak{M} \to \mathfrak{X}$ such that $\varphi(\mathscr{U} \cap \mathfrak{S}) = \varphi(\mathscr{U}) \cap (\mathfrak{X}_1 \times \{0\})$. An atlas for $\mathfrak{S}$ is now obtained from that of $\mathfrak{M}$ by restriction, i.e., $\{(\mathscr{U}_\alpha \cap \mathfrak{S}, \varphi_\alpha|_{\mathscr{U}_\alpha \cap \mathfrak{S}})\}$.

(ii). A continuous map $f : \mathfrak{M} \to \mathfrak{N}$ between two Fréchet manifolds modelled on $\mathfrak{X}$ and $\mathfrak{Y}$ is of class C^r (resp. C^∞) if for every $p \in \mathfrak{M}$ there exist charts $(\mathscr{U}, \varphi)$ at $p \in \mathscr{U}$ and $(\mathscr{V}, \psi)$ at $f(p) \in \mathfrak{N}$ such that the map $\psi \circ f \circ \varphi^{-1}$ from the open set $\varphi(f^{-1}(\mathscr{V}) \cap \mathscr{U})$ into $\mathfrak{Y}$ is of class C^r (resp. C^∞). If the map f is bijective and if both f and f^{-1} are of class C^r (resp. C^∞) then f is a C^r (resp. C^∞) *diffeomorphism*. Furthermore, any such map induces a well-defined tangent map $df : T\mathfrak{M} \to T\mathfrak{N}$ which for each $p \in \mathfrak{M}$ carries the fibre $T_p\mathfrak{M}$ linearly to $T_{f(p)}\mathfrak{N}$. If, in addition, the differential $df(p)$ is surjective then f is a *submersion*.

(iii). In particular, let $t \to c(t)$ be a curve through a point p in a Fréchet manifold $\mathfrak{M}$, that is, a differentiable map $c : \mathbb{R} \supset I \to \mathfrak{M}$ from an open interval I containing zero with $c(0) = p$. Two curves c_1 and c_2 are tangent at p if $c_1(0) = p = c_2(0)$ and $(\varphi \circ c_1)'(0) = (\varphi \circ c_2)'(0)$ in some chart $(\mathscr{U}, \varphi)$ (hence, by the chain rule, in every chart) around p. An equivalence class of such curves defines a *tangent vector* to $\mathfrak{M}$ at p. As in finite dimensions, this establishes a bijection between the model space $\mathfrak{X}$ and the set $T_p\mathfrak{M}$ of all such equivalence classes (the tangent space at p) by means of which the latter acquires the structure of an isomorphic Fréchet space. The disjoint union $T\mathfrak{M} = \bigcup_{p \in \mathfrak{M}} T_p\mathfrak{M}$ with the natural smooth projection map $\pi : T\mathfrak{M} \to \mathfrak{M}$ given by $\pi(v) = p$ if $v \in T_p\mathfrak{M}$ becomes another Fréchet manifold, the *tangent bundle* of $\mathfrak{M}$ modelled on the product space $\mathfrak{X} \times \mathfrak{X}$.

(iv). More generally, one defines *vector bundles* and *fibre bundles* over $\mathfrak{M}$ in the usual way as another Fréchet manifold $\mathfrak{F}$ together with a smooth projection map $\pi : \mathfrak{F} \to \mathfrak{M}$ whose derivative is surjective (i.e., a submersion). Each point p of the base manifold lies in some coordinate chart $(\mathscr{U}, \varphi)$ on $\mathfrak{M}$ with $\pi^{-1}(\mathscr{U}) \simeq \varphi(\mathscr{U}) \times \mathfrak{Y} \subset \mathfrak{X} \times \mathfrak{Y}$ and the fibre $\pi^{-1}(p)$ has the structure of a linear Fréchet space in the former and a Fréchet manifold in the latter case. A (cross-) section of either bundle is a smooth map $s : \mathfrak{M} \to \mathfrak{F}$ satisfying $\pi \circ s = \mathrm{id}_{\mathfrak{M}}$. When $\mathfrak{F}$ is the tangent bundle $T\mathfrak{M}$ then its sections are just smooth vector fields on $\mathfrak{M}$.

2.1.4 Geometric tools: connections, curvature, geodesics and metrics

A *connection* ∇ on a Fréchet vector bundle $\mathfrak{F}$ with (standard) fibre $\pi^{-1}(p) \simeq \mathfrak{Y}$ over a manifold $\mathfrak{M}$ modelled on a Fréchet space $\mathfrak{X}$ is a smooth map that assigns to each point of $\mathfrak{F}$ a subspace of the tangent space which is complementary to the null space of $d\pi$ at that point. This amounts to assigning to each coordinate chart of the bundle $\mathfrak{F}$ a family of bilinear maps, called the *Christoffel symbol* (or *connection coefficient*) map

$$\Gamma : (\mathscr{U} \subset \mathfrak{M}) \times \mathfrak{Y} \times \mathfrak{X} \to \mathfrak{Y} \qquad p, w, v \mapsto \Gamma_p(w, v) \tag{5}$$

which is jointly continuous as a function on the product of Fréchet spaces, smooth in p and linear in v and w (hence, jointly smooth in all three variables). The *curvature* of a connection on a Fréchet vector bundle $\mathfrak{F}$ is the trilinear map $\mathcal{R} : T\mathfrak{M} \times T\mathfrak{M} \times \mathfrak{F} \to \mathfrak{F}$ whose local representation in a coordinate chart is

$$\mathcal{R}_p(u, v)w = d\Gamma(p)(w, u, v) - d\Gamma(p)(w, v, u) + \Gamma_p(\Gamma_p(w, v), u) - \Gamma_p(\Gamma_p(w, u), v)$$

where $p \in \mathscr{U} \subset \mathfrak{M}$, $u, v \in T_x\mathfrak{M}$ and $w \in \mathfrak{F}_p$. The curvature $\mathcal{R}$ is independent of the choice of a chart.

A connection on $\mathfrak{M}$ is by definition a connection on the tangent bundle $T\mathfrak{M}$ which gives rise to the notion of differentiation of vector fields on $\mathfrak{M}$. Namely, the *covariant derivative* of a vector field W in the direction of V is the vector field $\nabla_V W$ whose expression in a coordinate chart $\mathscr{U} \subset \mathfrak{M}$ is

$$\nabla_V W(p) = dW(p) \cdot V(p) + \Gamma_p(V(p), W(p)) \qquad p \in \mathscr{U} \subset \mathfrak{M}. \tag{6}$$

Furthermore, a connection ∇ is said to be *symmetric* if its Christoffel symbols Γ_p are symmetric with respect to its two entries at any point p.

If $\gamma(t)$ is a smooth curve in $\mathfrak{M}$ and if V is a vector field along[2] γ then a connection ∇ induces covariant differentiation along γ by the formula $\frac{D}{dt}V = \nabla_{\gamma'}V$. As in classical (finite dimensional) geometry, a curve $\gamma(t)$ in $\mathfrak{M}$ is a *geodesic* of ∇ if $\frac{D}{dt}\gamma' = 0$ (i.e., if the acceleration along γ is zero) which in a local chart takes the form of a second order differential equation

$$\gamma'' = \Gamma_\gamma(\gamma', \gamma'). \tag{7}$$

Remark 2 It should be pointed out that, in contrast to the case when $\mathfrak{M}$ is a Banach manifold, prescribing an initial position $\gamma(0) = p_0$ and velocity $\gamma'(0) = u_0$ does not imply that the corresponding Cauchy problem for the geodesic equations (7) admits a local (in time) unique solution. It is not hard to construct examples displaying either non-existence or non-uniqueness of solutions.

[2] That is, the map $\mathbb{R} \ni t \mapsto V(t) \in T_{\sigma(t)}\mathfrak{M}$ depends differentiably on t.

In order to proceed it is also important to allow for generalizations of the classical Riemannian geometric notions.

Definition 4 A *weak-Riemannian* (or *pre-Riemannian*) metric on a Fréchet manifold $\mathfrak{M}$ modelled on $\mathfrak{X}$ is a smooth assignment to each point in $\mathfrak{M}$ of a positive definite bilinear form g_p on the tangent space at p. In coordinate charts this yields a jointly continuous function $(\mathscr{U} \subset \mathfrak{M}) \times \mathfrak{X} \times \mathfrak{X} \to \mathbb{R}$ on the product:

$$p, v, w \mapsto g_p(v, w) = \langle v, w \rangle_p$$

which is smooth in p and linear in v and w.

The prefix *weak-* (or *pre-*) is meant to indicate that each tangent space $T_p\mathfrak{M}$ is an inner product (pre-Hilbert) space with the topology induced by $g_p = \langle \cdot, \cdot \rangle_p$ which is weaker than the Fréchet topology induced from the model space $\mathfrak{X}$. Likewise, the Riemannian distance function of a pre-Riemannian metric on $\mathfrak{M}$ defined in the usual manner as the infimum of lengths

$$\int_a^b \langle \eta'(t), \eta'(t) \rangle^{1/2} dt$$

of all piecewise smooth curves joining two points $p = \eta(a)$ and $q = \eta(b)$, induces a topology on $\mathfrak{M}$ that is also strictly weaker than its original Fréchet topology.

Example 5 For any integer $r \geq 0$ the Sobolev H^r inner product

$$\langle u, v \rangle_{L^2} = \sum_{j=0}^{r} \int_M \langle \nabla^j u(x), \nabla^j v(x) \rangle \, d\mu \tag{8}$$

on the space $C^\infty(M, TM)$ of smooth vector fields on a compact Riemannian manifold M (with volume form μ) is a weak-Riemannian metric.

Let $\mathfrak{M}$ be a Fréchet manifold equipped with a weak-Riemannian metric $g = \langle \cdot, \cdot \rangle$. As in the finite-dimensional case, we say that an affine connection ∇ is *Levi-Civita* if it is symmetric and satisfies

$$X\langle V, W \rangle = \langle \nabla_X V, W \rangle + \langle V, \nabla_X W \rangle \tag{9}$$

for any vector fields X, V and W. Unlike in the finite-dimensional case, the existence of a Levi-Civita connection on $\mathfrak{M}$ (compatible in the above sense with a weak-Riemannian metric) is not guaranteed.[3] However, if such a connection can be defined then it is necessarily unique.

In their geometric treatments of statistics Amari [3] and Chentsov [15] found it useful to generalize the metric property (9) and work systematically with a more adequate concept of (squared) distance from one point to another. These notions can be defined also in our present setting.

[3] Essentially, this is because not all continuous linear functionals on a pre-Hilbert space can be represented by the inner product.

More precisely, let $\mathfrak{M}$ be a Fréchet manifold with a pre-Riemannian metric g as above. Two connections ∇ and ∇^* on $\mathfrak{M}$ are said to be *dual* (or *conjugate*) with respect to g if

$$X\langle V, W\rangle = \langle \nabla_X V, W\rangle + \langle V, \nabla^*_X W\rangle \tag{10}$$

for any vector fields X, V and W on $\mathfrak{M}$. It is clear that in this case $(\nabla + \nabla^*)/2$ is a connection satisfying (9). The triple (g, ∇, ∇^*) is said to define a *dualistic structure* on $\mathfrak{M}$.

Similarly, we say that a smooth function $D : \mathfrak{M} \times \mathfrak{M} \to \mathbb{R}$ is a *contrast function* (also called a *divergence*) if

(i) for any $p, q \in \mathfrak{M}$ we have $D(p,q) \geq 0$ with equality only if $p = q$
(ii) the matrix of second derivatives $-(\partial^2 D(p,q)/\partial p\, \partial q)|_{p=q}$ is strictly positive definite at every $p \in \mathfrak{M}$.

Condition (ii) de facto defines a Riemannian metric on $\mathfrak{M}$ together with a covariant derivative whose Christoffel symbols can be obtained from the matrix of its third order derivatives. In the finite dimensional case these formulas take the form

$$g_{ij}(p) = -\frac{\partial^2 D(p,q)}{\partial p_i \partial q_j}\Big|_{p=q}, \quad \Gamma_{ij,k}(p) = -\frac{\partial^3 D(p,q)}{\partial p_i \partial p_j \partial q_k}\Big|_{p=q} \quad 1 \leq i, j, k \leq n.$$

A symmetric connection ∇ is said to be *flat* if its curvature tensor vanishes. A Fréchet manifold $\mathfrak{M}$ equipped with a dualistic structure is called *dually flat* if both ∇ and ∇^* are flat.

2.2 The tame category of Sergeraert and Hamilton

One important item conspicuously missing from the list in Sect. 2.1.2 and 2.1.3 is the inverse function theorem — perhaps the most fundamental result of differential calculus. It shows that the study of many nonlinear problems in analysis can be effectively accomplished by linearization. It is also a useful tool in geometric analysis when it comes to constructing nontrivial examples of manifolds. Such a tool will be needed to endow the group of volume preserving diffeomorphisms with the structure of a Fréchet manifold.

2.2.1 Tame Fréchet spaces

There is a good reason for this omission. While it is well known that this theorem holds in the category of Banach spaces (see e.g., [18] or [34]) a straightforward generalization fails spectacularly for Fréchet spaces with any reasonable notion of differentiability. For a simple example consider the map $f \mapsto e^f$ from the space $C(\mathbb{R})$ of continuous real-valued functions on the line into itself with the topology of uniform convergence on compact sets. Clearly, its differential at 0 is the identity and the function is injective. However, it is not locally invertible because any neighbourhood of 1 contains functions which can assume negative values. A more elaborate example of this phenomenon of greater geometric interest involves the (Lie group) exponential mapping of the group of diffeomorphisms, see Example 20 below. Other interesting counterexamples can be found in [37] or [27].

A satisfactory replacement for that is the Nash-Moser inverse function theorem whose formulation requires introducing some extra structure. First, note that, without loss of generality, the seminorms in any Fréchet space $\mathfrak{X}$ can be assumed to be graded by strength

$$\|u\|_k \leq \|u\|_{k+1} \qquad \text{for } k = 0, 1, 2 \dots \text{ and } u \in \mathfrak{X}.$$

This can be achieved simply by adding to each seminorm all the seminorms of lower index. There may be many different collections of gradings defining the same topology on $\mathfrak{X}$. Once a specific choice of a grading is made, the space $\mathfrak{X}$ is referred to as a graded Fréchet space.

Example 6 Any Banach space is a graded Fréchet space.

Example 7 The sequential space

$$\Sigma\mathfrak{B} = \Big\{ \{x_n\}_{n=1,2\dots} \subset \mathfrak{B} : \|\{x_n\}\|_k = \sup_n e^{nk} \|x_n\| < \infty \text{ for all } k \Big\}$$

consisting of exponentially decreasing sequences in a fixed Banach space $\mathfrak{B}$ with norm $\|\cdot\|$ is a graded Fréchet space with seminorms $\|\cdot\|_k$.

Example 8 The Fréchet space $C^\infty(M, E)$ of smooth sections of a vector bundle E over a compact manifold M is graded by either the uniform C^k norms (1) or the Sobolev H^s norms (2).

Definition 5 Let $\mathfrak{X}$ and $\mathfrak{Y}$ be graded Fréchet spaces and let $\mathscr{U} \subset \mathfrak{X}$ be an open set. A continuous map $f : \mathfrak{X} \supset \mathscr{U} \to \mathfrak{Y}$ is said to be a *tame map* if there are integers r (the degree) and b (the base) such that

$$\|f(u)\|_k \leq C(1 + \|u\|_{k+r}) \qquad \text{for all } u \in \mathscr{U} \text{ and } k \geq b, \tag{11}$$

where $C > 0$ depends on k. (Note that the numbers r and b may be different for different open sets $\mathscr{U}$.) The map is said to be a *smooth* tame map if f is of class C^∞ and all of its Gateaux derivatives are tame.

Remark 3 In the special case when f is a linear map $L : \mathfrak{X} \to \mathfrak{Y}$ we have

$$\|Lu\|_k \leq C\|u\|_{k+r} \qquad \text{for any } k \geq b$$

by applying (11) to $\varepsilon u/\|u\|_b$ with sufficiently small $\varepsilon > 0$ and any $u \neq 0$ (increasing the base b if necessary to ensure that $\|u\|_b \neq 0$) and using linearity of L.

It is not hard to see that any tame linear map is continuous in the Fréchet topology. Furthermore, a linear isomorphism of Fréchet spaces which is tame and has a tame inverse establishes an equivalence of gradings. Consequently, in order to show that a map f is tame it is enough to establish the estimates in (11) for any pair among the equivalent gradings on $\mathfrak{X}$ and $\mathfrak{Y}$.

Example 9 The gradings of the space of smooth sections $C^\infty(M, E)$ of a bundle E given by the uniform C^k norms, the Sobolev H^s norms or the Hölder $C^{k,\alpha}$ norms are equivalent.[4]

Example 10 A routine verification of the definitions shows that products and compositions of tame Fréchet maps are tame and, furthermore, any (linear or nonlinear) partial differential operator of order r between smooth sections of vector bundles over a compact manifold M is a tame map of degree r (cf. also Section 3 below).

Definition 6 A graded Fréchet space $\mathfrak{X}$ is *tame* if there is a Banach space $\mathfrak{B}$ such that the identity operator on $\mathfrak{X}$ factors through $\Sigma\mathfrak{B}$, that is

$$\begin{tikzcd}
\mathfrak{X} \arrow[rr, "\mathrm{Id}"] \arrow[dr, "L"'] & & \mathfrak{X} \\
& \Sigma\mathfrak{B} \arrow[ur, "M"'] &
\end{tikzcd}$$

for some tame linear maps $L : \mathfrak{X} \to \Sigma\mathfrak{B}$ and $M : \Sigma\mathfrak{B} \to \mathfrak{X}$. A *tame Fréchet manifold* is a Fréchet manifold $\mathfrak{M}$ modelled on a tame Fréchet space equipped with an atlas whose coordinate transition functions are smooth tame maps.

Example 11 Any submanifold of a tame Fréchet manifold is tame.

Example 12 Let M be a compact manifold. The Fréchet space of smooth sections of any fibre (or vector) bundle over M is a tame Fréchet manifold.

Remark 4 Although Def. 6 looks somewhat technical, the property it captures can be described perhaps more intuitively as the space $\mathfrak{X}$ admitting a family of "smoothing operators" $S_\theta : \mathfrak{X} \to \mathfrak{X}$ with $\theta \geq 0$, that is, linear maps satisfying certain estimates that single out a preferred grading among those defining the same topology on $\mathfrak{X}$.[5] They can be explicitly constructed on the sequential spaces $\mathfrak{X} = \Sigma\mathfrak{B}$ and shown to satisfy for any $m \leq n$ the estimates

$$\|S_\theta u\|_n \leq Ce^{(n-m)}\theta\|u\|_m \quad \text{and} \quad \|(Id - S_\theta)u\|_m \leq Ce^{-(n-m)\theta}\|u\|_n\,, \tag{12}$$

[4] These gradings are easily shown to be equivalent with the help of the Sobolev lemma.

[5] Such operators (defined by convolutions with smooth functions and Fourier truncation methods) were used in various contexts, e.g. by Nash [43], Moser [41] and Hormander [28] in the work on isometric embeddings of Riemannian manifolds and in designing iteration schemes for solving nonlinear partial differential equations.

where $C > 0$ is a constant depending only on m and n. In turn, these estimates yield useful interpolation inequalities[6]

$$\|u\|_m^{n-l} \leq C\|u\|_n^{m-l}\|u\|_l^{n-m} \tag{13}$$

for any $l \leq m \leq n$ with $C > 0$ depending on l, m and n. Such estimates are then used to implement a rapidly converging iteration scheme (a modified Newton algorithm) needed in the proof.

In the tame Fréchet setting we can state now the following version of the inverse function theorem.

Proposition 2 (The Nash-Moser-Hamilton theorem) *Let $\mathfrak{X}$ and $\mathfrak{Y}$ be tame Fréchet spaces and let $\mathscr{U} \subset \mathfrak{X}$ be an open set. Let $f : \mathfrak{X} \supset \mathscr{U} \to \mathfrak{Y}$ be a smooth tame map. Suppose that there is an open subset $\mathscr{V} \subset \mathscr{U}$ such that $df(u) : \mathfrak{X} \to \mathfrak{Y}$ is a linear isomorphism for all $u \in \mathscr{V}$ whose inverse $(df)^{-1} : \mathscr{V} \times \mathfrak{Y} \to \mathfrak{X}$ is a smooth tame map. Then f is locally invertible on $\mathscr{V}$ and the inverse is also a smooth tame map.*

Proof For a detailed proof we refer to [27]. Shorter expositions with a somewhat different emphasis can be found in [37] and [32]. □

Remark 5 The inverse function theorem may be used to solve differential equations and in particular to find integral curves of (possibly time-dependent) vector fields on Fréchet manifolds. Outside of Banach spaces the situation becomes much less clear as the main tool for such purposes, namely, Banach's contraction mapping principle, is no longer available and examples showing that solutions may neither exist nor be unique are not difficult to construct. However, in the tame Fréchet setting the presence of the smoothing operators S_θ, such as those in (12), makes it possible to first mollify the differential equation and then produce a solution by passing to the limit with $\theta \to \infty$.

2.2.2 Manifolds of maps

The next example of a Fréchet manifold is of considerable geometric interest. Let M be a compact manifold (without boundary) and let N be a finite dimensional manifold. Without loss of generality we may assume N to be Riemannian.

Proposition 3 *The space $\mathfrak{M}(M, N)$ of all smooth maps of M into N is an infinite-dimensional tame Fréchet manifold.*

Proof There are different ways of exhibiting a differentiable manifold structure of this space. We sketch a proof based on an idea developed by Eells [21]. It consists of four steps.

---

[6] Such interpolation inequalities are well-known for $C^\infty(M)$ equipped with either the Sobolev H^k or the Hölder $C^{k,\alpha}$ norms in (2) or (1).

Step 1: The first objective is to find a suitable candidate for the model space. As indicated already, this space should be isomorphic to the tangent space $T_f\mathfrak{M}(M,N)$ at each point f in $\mathfrak{M}(M,N)$ and it can be therefore identified with the Fréchet space $C^\infty\big(f^{-1}(TN)\big)$ of smooth sections of the pull-back of the bundle TN to M by the map f and topologized by the uniform norms (cf. Example 2).

Step 2: Next, in order to define local charts at $f \in \mathfrak{M}(M,N)$ we pick a Riemannian metric on N (any Riemannian metric will do) and recall that for any $p \in N$ the associated Riemannian exponential map $\exp_p : T_pN \to N$ is a local diffeomorphism near zero in T_pN onto a neighbourhood of p. Since f is continuous, $f(M)$ is a compact subset of N and thus the injectivity radius of the target manifold has a global lower bound $\varepsilon > 0$ on $f(M)$. Setting

$$\mathscr{U}_f(\varepsilon) = \Big\{M \ni x \to \exp_{f(x)}(w(x)) : w \in C^\infty\big(f^{-1}(TN)\big),\ \|w\|_k < \varepsilon_f\Big\}$$

and

$$w \to \varphi_f(w) = \exp_f \circ w \tag{14}$$

gives now a coordinate chart $(\mathscr{U}_f, \varphi_f)$ at f.

Step 3: It remains to verify that given any points f and g in $\mathfrak{M}(M,N)$ the coordinate transition maps

$$\varphi_g \circ \varphi_f^{-1} : \varphi_f(\mathscr{U}_f \cap \mathscr{U}_g) \to \varphi_g(\mathscr{U}_f \cap \mathscr{U}_g)$$

are of class C^∞, which follows essentially from the chain rule etc. — after localizing the transition maps to functions defined on open sets in Fréchet spaces of smooth sections of appropriate vector bundles. This shows that $\mathfrak{M}(M,N)$ is a smooth Fréchet manifold.

Step 4: The fact that it is tame is an immediate consequence of the fact that $\mathfrak{M}(M,N)$ can be viewed as the space of sections of the fibre bundle $F = M \times N$ over M (see Example 12). □

The following special cases of this example introduce objects that will play an important role in what follows.

Example 13 The set $\mathfrak{D}(M,N)$ of all smooth diffeomorphisms between two compact manifolds M and N or, more generally, the set $\mathfrak{E}(M,N)$ of all smooth embeddings of a compact manifold M into a manifold N, is a tame Fréchet manifold being an open subset of $\mathfrak{M}(M,N)$.

Example 14 (The space of probability densities on $\mathbb{T}$) If M is the unit circle $\mathbb{T}$ then diffeomorphisms of M it can be viewed as 2π-pseudo-periodic functions, $\phi(x+2\pi) = \phi(x) + 2\pi$ such that $\phi'(x) > 0$ for all $x \in \mathbb{R}$. It follows that any ϕ' is 2π-periodic and satisfies the integral (fixed volume) constraint

$$\int_0^{2\pi} \phi'(x)\, dx = \phi(2\pi) - \phi(0) = 2\pi$$

so that the set of normalized derivatives $\phi'/2\pi$ — which we shall denote by $\mathfrak{Dens}(\mathbb{T})$ — can be viewed as the Fréchet space of smooth (probability) densities on $\mathbb{T}$. Since ϕ' determines ϕ uniquely up to a constant in $2\pi\mathbb{Z}$ we find that the set of smooth orientation-preserving diffeomorphisms of $\mathbb{T}$, denoted by $\mathfrak{D}(\mathbb{T})$, is diffeomorphic (as a Fréchet manifold) to the product $\mathbb{T}\times\mathfrak{Dens}(\mathbb{T})$. Thus, the space of smooth probability densities on $\mathbb{T}$ is a quotient space of $\mathfrak{D}(\mathbb{T})$. Moreover, it is clearly contractible as a convex open subset of a closed affine subspace (of codimension 1) of the Fréchet space of 2π-periodic functions.

Remark 6 (Banach completions of manifolds of maps) The construction in the proof of Prop. 3 is quite general and can be readily adapted to other function spaces, including Banach spaces of functions with finite smoothness conditions such as $C^k(M,N)$ with the uniform norm (1) or the space $H^k(M,N)$ of maps of Sobolev class with the norm (2). In these cases the corresponding manifold of maps admits the structure of a smooth Banach manifold. However, to carry this out one requires that (i) the topology of the modelling space be stronger than the uniform topology[7] and (ii) the functions are "well-behaved" under compositions and inverses. The latter, in particular, leads to serious analytical obstacles concerning derivative loss when performing various operations involving compositions of functions of finite smoothness, since e.g. any iteration scheme (such as Newton's algorithm, Picard's method of successive approximations etc.) would quickly degenerate in complete loss of differentiability. See also Appendix 1 below.

3 Diffeomorphism groups and their quotients

Groups of diffeomorphisms of compact manifolds arise naturally as symmetry groups of various geometric structures (such as volume forms or symplectic forms) carried by the underlying manifold or as configuration spaces of dynamical systems characterized by infinitely many degrees of freedom. As already mentioned, we will view them as infinite-dimensional Fréchet manifolds (in the sense of the previous section). However, in the vast literature on the subject other function space topologies are also used and in many cases provide a more suitable setting depending on the analytical or geometric tasks at hand.[8]

3.1 Fréchet Lie groups

Definition 7 A *tame Fréchet Lie group* is a tame Fréchet manifold $\mathfrak{G}$ whose group operations of multiplication $g,h \mapsto g\cdot h$ and inversion $g \mapsto g^{-1}$ are C^∞ smooth tame

[7] This means that in the two cases above we require $k \geq 0$ or $k > \dim M/2$, respectively.

[8] For example, when studying nonlinear PDE where Hilbert space techniques may provide more precise tools to derive the necessary *a priori* estimates.

maps of $\mathfrak{G}\times\mathfrak{G}$ and $\mathfrak{G}$ into $\mathfrak{G}$, respectively. The *Lie algebra* $\mathfrak{g}$ of $\mathfrak{G}$ is the tangent space $T_e\mathfrak{G}$ at the identity element e.

Despite the complications resulting from infinite dimensions Fréchet Lie groups can be very effectively studied using standard Lie-theoretic tools. The Lie algebra $\mathfrak{g}$ is naturally isomorphic (as a vector space) to the space of left- (resp. right-) invariant vector fields on $\mathfrak{G}$, since any element v of the algebra generates a unique vector field V on the group by the formula $V(g) = dL_g(e)v$ where $L_g(h) = g{\cdot}h$ (resp. $dR_g(e)v$, where $R_g(h) = h{\cdot}g$) is the left- (resp. right-) translation.

The *group adjoint* action $\mathrm{Ad} : \mathfrak{G}\times\mathfrak{g} \to \mathfrak{g}$ of $\mathfrak{G}$ on its Lie algebra is defined in the standard way as the derivative at $h = e$ of the smooth map of $\mathfrak{G}$ to itself given by inner automorphisms $h \mapsto g{\cdot}h{\cdot}g^{-1}$ for any fixed group element g. Namely, we have

$$v \mapsto \mathrm{Ad}_g v = d(L_g \circ R_{g^{-1}})(e)v \qquad \text{for any} \quad g \in \mathfrak{G}.$$

Since $\mathrm{Ad}_g v$ is smooth in g and linear in v, we can define similarly the *algebra adjoint* action $\mathrm{ad} : \mathfrak{g}\times\mathfrak{g} \to \mathfrak{g}$ as the derivative of the group adjoint at $g = e$

$$v \mapsto \mathrm{ad}_u v = \frac{d}{dt}\Big|_{t=0}\mathrm{Ad}_{g(t)}v \qquad \text{for any} \quad v \in \mathfrak{g}\,,$$

where $g(t)$ is a smooth curve in $\mathfrak{G}$ with $g(0) = e$ and $\dot{g}(0) = u$. As in finite dimensions it induces a commutation operation $\mathrm{ad}_v w = [V, W]$ on the Lie algebra, which coincides with the Lie bracket of left-invariant vector fields on the group $\mathfrak{G}$ generated by $v, w \in \mathfrak{g}$. (It also coincides with the negative of the Lie bracket of right-invariant vector fields.)

Example 15 The Lie algebra of the general linear group $\mathfrak{G} = GL(n,\mathbb{R})$ of invertible $n{\times}n$ matrices is the space $\mathfrak{g} = \mathfrak{gl}(n,\mathbb{R})$ of all square $n{\times}n$ matrices. The corresponding adjoint actions are given by $\mathrm{Ad}_g B = gBg^{-1}$ and $\mathrm{ad}_A B = AB - BA$.

Let $\mathfrak{H}$ be a subgroup of a Fréchet Lie group $\mathfrak{G}$. Then $\mathfrak{H}$ acts on $\mathfrak{G}$ by left multiplications

$$\mathfrak{H}\times\mathfrak{G} \to \mathfrak{G} \qquad h, g \mapsto h{\cdot}g.$$

The orbits in $\mathfrak{G}$ under this action are the *right cosets*[9] of $\mathfrak{H}$ in $\mathfrak{G}$, that is

$$\mathfrak{H}{\cdot}g = \{h{\cdot}g : h \in \mathfrak{H}\} \qquad \text{for} \quad g \in \mathfrak{G}$$

and the quotient space consisting of all such cosets is denoted by $\mathfrak{H}\backslash\mathfrak{G}$. Similarly, using right multiplications one defines the quotient space of *left cosets* $\mathfrak{G}/\mathfrak{H}$ whose elements are $g{\cdot}\mathfrak{H}$.

[9] Some authors refer to such orbits as left cosets, which may cause some confusion.

3.2 Riemannian structures and Euler-Arnold equations on Lie groups and quotient spaces

In this chapter we are primarily interested in those pre-Riemannian metrics on Fréchet Lie groups $\mathfrak{G}$ and their quotient spaces that are *invariant*, i.e., for which $\mathfrak{G}$ acts by isometries. Metrics invariant under actions given by right- (resp. left-) multiplications are called *right-* (resp. *left-*) *invariant*; those invariant under both right and left actions are called *bi-invariant*.

Thus, given an inner product $\langle\cdot,\cdot\rangle_e$ on the Lie algebra $\mathfrak{g}$, a right-invariant metric on the group $\mathfrak{G}$ can be defined simply by setting

$$\langle V, W\rangle_g = \langle dR_{g^{-1}}V, dR_{g^{-1}}W\rangle_e \tag{15}$$

for any vectors $V, W \in T_g\mathfrak{G}$ and any $g \in \mathfrak{G}$.

3.2.1 The Euler-Arnold equations

In 1960's Arnold [5] proposed a differential-geometric framework to study the Euler equations of ideal hydrodynamics. It is based on the observation that motions of an ideal (that is, incompressible and non-viscous) fluid in a bounded domain M trace out curves in the group $\mathfrak{D}_\mu(M)$ of volume-preserving diffeomorphisms of M which correspond to geodesics of the right-invariant pre-Riemannian metric defined by the kinetic energy of the fluid. This approach is very general and applies to numerous partial differential equations of interest in mathematical physics and geometry. Such equations arise within this framework through a general reduction procedure which starts with a given geodesic system on the group to produce a dynamical system on the tangent space at the identity. In this section we describe Arnold's framework for general Lie groups and homogeneous spaces.

Let $\mathfrak{G}$ be a (finite or infinite dimensional Banach or Fréchet) Lie group which carries a right-invariant pre-Riemannian metric $g = \langle\cdot,\cdot\rangle$ induced by an inner product on its Lie algebra $T_e\mathfrak{G}$ as in (15). The *Euler-Arnold equation* on the Lie algebra associated with the geodesic flow of g has the form

$$u_t = -\mathrm{ad}_u^* u \tag{16}$$

where $u(t)$ is a curve in $T_e\mathfrak{G}$ and the bilinear operator ad^* on the right-hand side is an operator on $T_e\mathfrak{G}$ defined by

$$\langle \mathrm{ad}_v^* u, w\rangle_e = \langle u, \mathrm{ad}_v u\rangle_e \qquad \text{for any} \quad u, v, w \in \mathfrak{g} \tag{17}$$

called the *coadjoint operator*.

When equation (16) is augmented by an initial condition

$$u(0) = u_0 \tag{18}$$

then solutions of the resulting Cauchy problem (16)-(18) describe evolution in the Lie algebra of the dynamical system $t \mapsto u(t) = \dot{g}(t)\cdot g^{-1}(t)$ obtained by right-translating the velocity field of the corresponding geodesic $g(t)$ in the group $\mathfrak{G}$ starting at the identity e in the direction $\dot{g}(0) = u_0$.

Observe that, conversely, if in turn $u(t)$ is known then the geodesic can be obtained by solving the Cauchy problem for the *flow equation*, namely

$$\frac{dg(t)}{dt} = dR_{g(t)}(e)u(t)\,, \qquad g(0) = e.$$

Remark 7 The Cauchy problem (16)-(18) can be rewritten in the form

$$\frac{d}{dt}\big(\mathrm{Ad}^*_{g(t)}u\big) = 0\,, \qquad u(0) = u_0\,,$$

where

$$\langle \mathrm{Ad}^*_g u, v\rangle_e = \langle u, \mathrm{Ad}_g v\rangle_e \qquad \text{for any } \; v \in T_e\mathfrak{G} \; \text{ and } \; g \in \mathfrak{G}\,,$$

which immediately yields a *conservation law*

$$\mathrm{Ad}^*_{g(t)}u(t) = u_0.$$

This last equation expresses the fact that solutions $u(t)$ of the Euler-Arnold equation are confined to one and the same orbit during the evolution.

Example 16 (The Rigid Body) In the important special case when $\mathfrak{G} = SO(3)$ this procedure yields the classical Euler equations describing rotations of a rigid body in the internal coordinates of the body. In vector notation they have the form

$$\frac{d}{dt}P = P \times \Omega$$

where P is the vector of angular momentum and Ω is the vector of angular velocity - the two are related by the so-called inertia operator of the system.

Example 17 (The Euler equations of ideal hydrodynamics) Another special case involves the group of volume-preserving diffeomorphisms $\mathfrak{G} = \mathfrak{D}_\mu(M)$ of a compact Riemannian manifold M — see Sect. 3.3 below. This group can be equipped with a right-invariant metric which is essentially the fluid's kinetic energy and which at the identity diffeomorphism is given by the L^2 inner product of vector fields on M. In this case the Euler-Arnold equation (16) becomes the Euler equations of ideal hydrodynamics

$$u_t + \nabla_u u = -\nabla p\,, \qquad \operatorname{div} u = 0\,,$$

where u is the vector field on M representing the velocity field and p is the function on M representing the pressure in the fluid, see [5].

Example 18 (Integrable systems and circle diffeomorphisms) If the group of circle diffeomorphisms $\mathfrak{G} = \mathfrak{D}(\mathbb{T})$ is equipped with the right-invariant metric generated by the L^2 inner product

$$\langle u, v\rangle_{L^2} = \int_{\mathbb{T}} uv\, dx \qquad \text{for any} \quad u, v \in T_e\mathfrak{D}(\mathbb{T}),$$

then the Euler-Arnold equation is the (scaled) inviscid Burgers equation

$$u_t + 3uu_x = 0.$$

If the metric is generated by the Sobolev H^1 inner product

$$\langle u, v\rangle_{H^1} = \int_{\mathbb{T}} (uv + u_x v_x)\, dx$$

then the Euler-Arnold equation yields the Camassa-Holm equation

$$u_t - u_{txx} + 3uu_x - 2u_x u_{xx} - uu_{xxx} = 0.$$

Both of these equations are well-known examples of infinite-dimensional completely integrable systems in that they are bi-hamiltonian, possess infinitely many conserved integrals etc., for more details see [29]; see also Sect. 4.5.3.

Remark 8 Many other conservative dynamical systems in mathematical physics also describe geodesic flows on appropriate Lie groups. In Figure 1 we list several examples of such systems to demonstrate the range of applications of this approach. The choice of a group $\mathfrak{G}$ (column 1) and an energy metric $\langle\cdot,\cdot\rangle$ (column 2) defines the corresponding Euler equations (column 3). (Note that the L^2 and H^s in the column 2 refer to various Sobolev inner products on vector fields in the corresponding Lie algebra.) This list is by no means complete, and we refer to [6] for more details.

Group $\mathfrak{G}$	Metric $\langle\cdot,\cdot\rangle$	Equation
$SO(3)$	$\langle\omega, A\omega\rangle$	Euler top
$E(3) = SO(3) \ltimes \mathbb{R}^3$	quadratic forms	Kirchhoff equation for a body in a fluid
$SO(n)$	Manakov's metrics	n-dimensional top
$\mathfrak{D}(\mathbb{T})$	L^2	Hopf (or, inviscid Burgers) equation
$\mathfrak{D}(\mathbb{T})$	$\dot{H}^{1/2}$	Constantin-Lax-Majda-type equation
Virasoro	L^2	KdV equation
Virasoro	H^1	Camassa–Holm equation
Virasoro	$\dot{H}^1$	Hunter–Saxton (or Dym) equation
$\mathfrak{D}_\mu(M)$	L^2	Euler ideal fluid
$\mathfrak{D}_\mu(M)$	H^1	averaged Euler flow
$\mathfrak{D}_\omega(M)$	L^2	symplectic fluid
$\mathfrak{D}(M)$	L^2	EPDiff equation
$\mathfrak{D}_\mu(M) \ltimes T_e\mathfrak{D}_\mu(M)$	$L^2 \oplus L^2$	Magnetohydrodynamics
$C^\infty(S^1, SO(3))$	H^{-1}	Heisenberg magnetic chain

Fig. 1 Euler–Arnold equations related to various Lie groups and metrics.

Remark 9 (Euler-Arnold equation on quotient spaces) More generally, let $\mathfrak{G}$ be a Fréchet Lie group equipped with a right-invariant metric as above and let $\mathfrak{H}$ be a closed subgroup. The metric on $\mathfrak{G}$ descends to an invariant (under the right action of $\mathfrak{G}$) metric on the quotient $\mathfrak{H}\backslash\mathfrak{G}$ if and only if its projection onto the orthogonal complement $T_e^{\perp}\mathfrak{H} \subset T_e\mathfrak{G}$ is bi-invariant with respect to the action of $\mathfrak{H}$. In particular, if the metric on $\mathfrak{G}$ is degenerate along the subgroup $\mathfrak{H}$ then this condition reduces to the metric bi-invariance with respect to the $\mathfrak{H}$ action, see e.g., [29] and the section below. In this case the corresponding Euler-Arnold equation is defined as before as long as the metric on the quotient $\mathfrak{H}\backslash\mathfrak{G}$ is nondegenerate.

3.2.2 Quotient spaces and Riemannian submersions

Let $\mathfrak{F}$ be a smooth fibre bundle over a Fréchet manifold $\mathfrak{M}$ with projection π and fibres which are modelled on a Fréchet space $\mathfrak{Y}$. Assume that both $\mathfrak{F}$ and $\mathfrak{M}$ carry (possibly weak) Riemannian metrics and let $\mathfrak{Y}^{\perp}$ denote the orthogonal complement. The projection $\pi : \mathfrak{F} \to \mathfrak{M}$ defines an (infinite-dimensional) *Riemannian submersion* if $d\pi|_{\mathfrak{Y}^{\perp}}$ is an isometry at each point of $\mathfrak{F}$.

In particular, if $\mathfrak{H}$ is a closed subgroup of a Fréchet Lie group $\mathfrak{G}$ equipped with a right-invariant (possibly weak) Riemannian metric then the following general result characterizes those metrics that descend to the base manifold of right cosets.

Proposition 4 *A right-invariant metric $\langle\cdot,\cdot\rangle$ on $\mathfrak{G}$ descends to a right-invariant metric on the quotient space $\mathfrak{H}\backslash\mathfrak{G}$ if and only if the inner product $\langle\cdot,\cdot\rangle_e$ restricted to the orthogonal complement $T_e^{\perp}\mathfrak{H}$ is bi-invariant with respect to the action of $\mathfrak{H}$, that is*

$$\langle u, \mathrm{ad}_w v\rangle_e + \langle \mathrm{ad}_w u, v\rangle_e = 0 \tag{19}$$

for any $u, v \in T_e^{\perp}\mathfrak{H}$ and any $w \in T_e\mathfrak{H}$.

Proof The proof repeats, with obvious modifications, the arguments in the finite-dimensional case, see [30]. See also [14]. □

Example 19 (Circle diffeomorphisms and the periodic Hunter-Saxton equation) Let $\mathfrak{G}$ be the group of circle diffeomorphisms $\mathfrak{D}(\mathbb{T})$ from Example 14 and let $\mathfrak{H} \simeq \mathbb{T}$ be the subgroup of rotations. Equip $\mathfrak{G}$ with a right-invariant metric which at the identity is given by the (homogeneous) Sobolev H^1 inner product

$$\langle u, v\rangle_{\dot{H}^1} = \int_{\mathbb{T}} u_x v_x \, dx \qquad u, v \in T_e\mathfrak{D}(\mathbb{T}) \tag{20}$$

The tangent space to the quotient space $\mathfrak{Dens}(\mathbb{T})$ at the identity coset $[e]$ can be identified with the space of smooth periodic mean-zero functions and the Euler-Arnold equation for the geodesic flow of the metric (20) (right-translated to the tangent space at the identity) is the completely integrable Hunter-Saxton equation

$$u_{txx} + 2u_x u_{xx} + u u_{xxx} = 0\,, \tag{21}$$

see [29]. We shall considerably generalize this example later on.

3.3 Diffeomorphism groups as Fréchet manifolds and Lie groups

Our principal examples are groups of diffeomorphisms and their quotient spaces. In the sequel we shall always assume that all diffeomorphisms are orientation preserving. Let $\mathfrak{G} = \mathfrak{D}(M)$ denote the set of all smooth diffeomorphisms of a compact n-dimensional manifold M. As an open subset of the tame Fréchet manifold $\mathfrak{M}(M, M)$ it is itself a smooth tame Fréchet manifold. Its model space (identified with the tangent space $\mathfrak{g} = T_e\mathfrak{D}(M)$ at the identity diffeomorphism e) is just the space $C^\infty(M,TM)$ of all smooth vector fields on M.

The group operations on $\mathfrak{D}(M)$ are given by compositions $\eta, \xi \mapsto \mathrm{c}(\eta, \xi) = \eta \circ \xi$ and inversions $\eta \mapsto \mathrm{i}(\eta) = \eta^{-1}$ of diffeomorphisms. The group adjoint action is given by the change of coordinates map $Ad_\varphi v = \varphi_* v \circ \varphi^{-1}$ and the algebra adjoint is the standard Lie derivative $ad_v w = \mathcal{L}_v w = [v, w]$ of vector fields on M.

Proposition 5 *Let M be a compact manifold. The set $\mathfrak{D}(M)$ of all diffeomorphisms of M is a smooth tame Fréchet Lie group.*

Proof We already noted that $\mathfrak{D}(M)$ is a tame Fréchet manifold. To show that the composition map $\mathrm{c}(\eta, \xi)$ is smooth we need to compute its derivatives on the product manifold $\mathfrak{D}(M) \times \mathfrak{D}(M)$. If $t \to \eta(t)$ is a smooth curve through η with the tangent vector $\partial_t \eta(0) = V \in T_\eta \mathfrak{D}$ and $s \to \xi(s)$ is a curve through ξ with $\partial_s \xi(0) = W \in T_\xi \mathfrak{D}$, then applying the rules of calculus from Sect. 2 we find that the derivative at $t = s = 0$ is the sum of its two partial derivatives

$$d\mathrm{c}(\eta, \xi)(V, W) = V \circ \xi + D\eta \circ \xi \cdot W\,,$$

which shows that $\mathrm{c}(\eta, \xi)$ is differentiable. Existence of the higher order differentials follows similarly.

To show that $\mathrm{c}(\eta, \xi)$ is tame it suffices to work in local charts on $\mathfrak{D}(M)$ and establish tame estimates (11) for the local representatives of the diffeomorphisms satisfying $\|\eta\|_1, \|\xi\|_1 \le 1$ where $\|\cdot\|_k$ is any one of the equivalent gradings of the model space $C^\infty(M,TM)$ given by the Hölder $C^{k,\alpha}$, the Sobolev H^k, or the uniform C^k norms. For example, we have

$$\|\eta \circ \xi\|_0 = \sup_x |\eta \circ \xi(x)| \le \|\eta\|_{C^1} = \|\eta\|_1 \le 1.$$

Furthermore, for any $k \ge 1$ successive applications of the chain rule together with the interpolation inequalities (13) for the uniform norms yield

$$\|\eta\circ\xi\|_k = \sum_{|\alpha|\le k}\|D^\alpha(\eta\circ\xi)\|_0 \le \sum_{l=0}^{k}\sum_{j=0}^{l}\sum_{i_1+\cdots+i_j=l} C_{l,i_1,\dots,i_l}\|\eta\|_{C^l}\|\xi\|_{C^{i_1}}\cdots\|\xi\|_{C^{i_l}}$$
$$\le C_k\sum_{l=0}^{k}\big(\|\eta\|_{C^l}\|\xi\|_{C^1}+\|\eta\|_{C^1}\|\xi\|_{C^l}\big)\|\xi\|_{C^1}^{l-1}$$
$$\le C_k\big(1+\|\eta\|_{C^k}+\|\xi\|_{C^k}\big).$$

Thus, the composition map satisfies a tame estimate of degree 0 and base 1. Since all of its Gateaux derivatives are linear combinations of compositions and products of derivatives of η and ξ, it follows that $\mathrm{c}(\eta,\xi)$ is a smooth tame map.

Turning to the inversion map we first note that differentiating in t the identity $\eta^{-1}(t)\circ\eta(t)=e$ and evaluating at $t=0$ gives

$$d\mathrm{i}(\eta)V = -D\eta^{-1}\cdot V\circ\eta^{-1} = -(D\eta)^{-1}\circ\eta^{-1}\cdot V\circ\eta^{-1}.$$

This shows that $\mathrm{i}(\eta)$ is differentiable and, in fact, smooth as a map from $\mathfrak{D}(M)$ to itself, since all the higher differentials are computed analogously. Showing that $\mathrm{i}(\eta)$ is a tame map involves once again local charts and successive applications of the chain rule and interpolation estimates in a manner similar to that for the composition map. For further details we refer to [27]. □

The group of diffeomorphisms $\mathfrak{D}(M)$ has several important subgroups. Of particular importance is the stabilizer subgroup of the Riemannian volume form $\mu\in\Omega^n M$, namely, the group of volume-preserving diffeomorphisms

$$\mathfrak{D}_\mu(M) = \big\{\eta\in\mathfrak{D}(M) : \eta^*\mu=\mu\big\}$$

where $\eta^*\mu = \mathrm{Jac}_\mu\eta\,\mu$ and the Jacobian is computed with respect to the Riemannian reference volume μ. The tangent space at the identity map e consists of divergence-free vector fields on M, that is

$$T_e\mathfrak{D}_\mu(M) = \big\{u\in T_e\mathfrak{D}(M) : \mathrm{div}_\mu u = 0\big\},$$

which is evident by differentiating the identity $\eta_t^*\mu=\mu$ at $t=0$ to get

$$0 = \frac{d}{dt}\Big|_{t=0}\eta_t^*\mu = \eta_t^*\Big(\mathcal{L}_{\frac{d\eta_t}{dt}\circ\eta_t^{-1}}\mu\Big) = \mathrm{div}_\mu u,$$

where $t\mapsto\eta_t$ is a curve of volume-preserving diffeomorphisms issuing from e in the direction u.

Proposition 6 *The group $\mathfrak{D}_\mu(M)$ of smooth volume-preserving diffeomorphisms of a compact Riemannian manifold M is a closed tame Fréchet Lie subgroup of $\mathfrak{D}(M)$.*

Proof Since the defining pullback condition is nonlocal, constructing local charts for $\mathfrak{D}_\mu(M)$ is more complicated than for $\mathfrak{D}(M)$. It will be sufficient to describe such a chart near the identity. Let φ_e be the diffeomorphism from a neighbourhood $\mathcal{U}$ of the zero section in $C^\infty(M,TM)$ to a neighbourhood of e in $\mathfrak{D}(M)$ defined in (14). Consider the pullback map

$$\phi_\mu : u \mapsto \varphi_e(u)^*\mu/\mu \qquad \text{for} \quad u \in \mathcal{U} \subset C^\infty(TM).$$

Writing it out in local charts and using the estimates on compositions and products shows that ϕ_μ is a tame nonlinear PDO of order 1 in u that maps smoothly to $C^\infty(M)$ and whose differential at zero is

$$w \mapsto d\phi_\mu(0)(w) = \mathcal{L}_w\mu/\mu = \operatorname{div}_\mu w\,, \tag{22}$$

where div_μ is the divergence operator of the Riemannian mertic on M. Using the Helmholtz-Weyl-Hodge decomposition of vector fields into L^2-orthogonal components $C^\infty(TM) = \operatorname{div}_\mu^{-1}(0) \oplus \nabla C^\infty(M)$ (see e.g., [33] or [40]) and the natural identification[10] of $\nabla C^\infty(M)$ with the closed subspace $C_0^\infty(M)$ of mean-zero functions on M we now define a map

$$\Phi_\mu : u \mapsto \big(v, \phi_\mu(u) - 1\big) \qquad \text{for} \quad u \in \mathcal{U} \subset C^\infty(TM)$$

on the product of Fréchet spaces $\operatorname{div}_\mu^{-1}(0) \oplus C_0^\infty(M)$. The task of constructing a local chart at e for $\mathfrak{D}_\mu(M)$ reduces now to showing that Φ_μ is invertible near the zero section. Its derivative at $u = v + \nabla f$ is readily computed to be

$$d\Phi_\mu(u)(w,g) = \big(w, d\phi_\mu(v + \nabla f)(w + \nabla g)\big) \qquad w \in \operatorname{div}_\mu^{-1}(0),\ g \in C_0^\infty(M).$$

In particular, if $v = 0$ and $f = 0$ then $d\Phi_\mu(0)(g,w) = (w, \operatorname{div}_\mu(w + \nabla g)) = (w, \Delta g)$ so that its second component is an elliptic operator, which remains elliptic under suitably small perturbations of v and f. It follows that the derivative of Φ_μ is a linear isomorphism with tame inverse at all points near $u = 0$ and, consequently, it is locally invertible by the Nash-Moser-Hamilton theorem (Prop. 2). In order to obtain a chart near $e \in \mathfrak{D}_\mu(M)$ it suffices now to compose Φ_μ^{-1} with the local chart φ_e for $\mathfrak{D}(M)$ and set $\varphi_e \circ \Phi_\mu^{-1}\big|_{\mathcal{V}\cap(0\oplus\operatorname{div}^{-1}(0))}$ where $\mathcal{V}$ is some neighbourhood of 0 on which Φ_μ is a diffeomorphism.

Finally, observe that $\mathfrak{D}_\mu(M)$ is a subgroup whose group operations enjoy the same regularity properties as those of the ambient group $\mathfrak{D}(M)$. This shows that $\mathfrak{D}_\mu(M)$ is a tame Fréchet Lie group. □

Remark 10 A subtle but important point to keep in mind when constructing smooth tame families of elliptic inverses is that to make it work one needs to make a judicious choice of an appropriate grading (e.g., using the $C^{k,\alpha}$ Hölder or the H^s Sobolev norms) for the Fréchet spaces involved in the argument. The reason for this is that elliptic estimates do not hold in certain functional settings, such as e.g., the uniform C^k spaces.

---

[10] For example, using a special case of de Rham's theorem which states that a closed form is exact if all of its periods vanish, see e.g., [55].

Remark 11 As remarked in 2.2.2, slightly modifying the procedure described there it is possible to equip $\mathfrak{D}(M)$ with the structure of a Banach (or Hilbert) manifold, e.g., by enlarging it to Sobolev H^s class diffeomorphisms $\mathfrak{D}^s(M)$ with $s > n/2 + 1$ or Hölder $C^{1,\alpha}$ diffeomorphisms $\mathfrak{D}^{1,\alpha}(M)$ with $0 < \alpha < 1$. Both options are very convenient from the point of view of analysis[11] - this is, once again, in contrast to the C^k case where this procedure seems to fail, as it depends crucially on the Hodge decomposition theorem.

Remark 12 The theory of (infinite-dimensional) Banach Lie groups is well developed and goes back to Birkhoff [12], see also [13]. However, repeating the construction in the proof of Prop. 5 immediately runs into serious obstacles, since compositions with diffeomorphisms on the left (i.e., left translation maps in the group), as well as the Lie bracket of vector fields on M (i.e., the commutator in the "Lie algebra"), lose derivatives with respect to both topologies. Consequently, neither $\mathfrak{D}^s(M)$ nor $\mathfrak{D}^{1,\alpha}(M)$ is a Banach Lie group in Birkhoff's sense. Of course, this problem disappears if the modelling space is a Fréchet space of smooth functions.

Example 20 Consider the group $\mathfrak{D}(\mathbb{T})$ of smooth diffeomorphisms of the unit circle. Any vector field u on the circle, viewed as an element of the Lie algebra $T_e\mathfrak{D}(\mathbb{T})$, gives rise to a one-parameter subgroup $\phi_t(x)$ obtained as a solution of the corresponding flow equation $d\phi_t/dt = u \circ \phi_t$ subject to the initial condition $\phi_0(x) = x$. The Lie group exponential map at the identity element e in $\mathfrak{D}(\mathbb{T})$ is now defined in complete analogy to the finitie-dimensional case as a "time-one map" by the formula

$$\exp_e : T_e\mathfrak{D}(\mathbb{T}) \to \mathfrak{D}(\mathbb{T}), \qquad \exp_e u = \phi_1.$$

It is not hard to see that this map is continuous in the Fréchet (or any sufficiently strong Banach) topology on $\mathfrak{D}(\mathbb{T})$, but it is not of class C^1. In fact, observe that even though its derivative at $t = 0$ is $d\exp_e(0) = \mathrm{id}$, the map fails to be invertible in any neighbourhood of the identity e.[12] This "bad" behaviour of the (Lie group) exponential map on $\mathfrak{D}(\mathbb{T})$ stands in sharp contrast to the case of the classical, as well as Banach, Lie groups such as loop groups. We refer to [44], [27] or [26] for further details.

3.4 The quotient space of probability densities

A fixed volume form on a compact n-manifold M can be used to define a natural fibration of $\mathfrak{D}(M)$ over the Fréchet manifold of smooth positive normalized measures on M. To describe the latter, recall that any volume form ν defines a Borel measure, which in any coordinate chart $\big(U, x = (x_1, \ldots, x_n)\big)$ on M has the form $d\nu = \rho_U(x)dx$

[11] See Appendix 1.

[12] Note that in Prop. 2 invertibility of the derivative is assumed to hold in a small neigbourhood rather than at a single point.

where $\rho_U(x)$ is a positive C^∞ function and $dx = dx_1 \dots dx_n$ is the Lebesgue measure on $\mathbb{R}^n$. Its expressions in different charts are related by Jacobian functions

$$\rho_U(x) = |\det(\partial\eta^i/\partial x^j)|\rho_V \circ \eta(x), \qquad \eta \in \mathfrak{D}(U, V),$$

which, in the geometric language, represent the cocycle of transition functions of the corresponding line bundle of volume forms over M. Cross-sections of this vector bundle are called *densities* and can be identified with smooth measures on M. If the manifold carries a Riemannian metric g then there is a canonical choice of a positive density on M, which in a local chart is given by $\sqrt{\det g}(x)$. In this case the corresponding volume form μ can be used as a fixed reference measure on M.

We set

$$\mathfrak{Dens}(M) = \left\{ \nu \in \Omega^n M : \nu > 0, \int_M d\nu = 1 \right\}, \tag{23}$$

where $\Omega^n M$ denotes the tame Fréchet space of smooth n-forms on M (i.e., the cross-sections of the one-dimensional vector bundle $\Lambda^n T^* M$). This set is clearly open and connected. In fact, if ν and λ are in $\mathfrak{Dens}(M)$ then so is their convex combination $t\nu + (1-t)\lambda$ where $0 \le t \le 1$. Of course, we can also think of $\mathfrak{Dens}(M)$ as an open convex subset of positive density functions $\rho > 0$ in the Fréchet manifold of C^∞ functions with average value 1 on M, so that we can write $\nu = \rho\mu$.

The tangent space at any $\nu \in \mathfrak{Dens}(M)$ is

$$T_\nu \mathfrak{Dens}(M) = \left\{ \beta \in \Omega^n M : \int_M d\beta = 0 \right\}, \tag{24}$$

which follows at once by differentiating the condition in (23).

Moser [42] proved that for any two volume forms ν and λ on a compact manifold M with $\int_M d\nu = \int_M d\lambda$ there is a smooth diffeomorphism ξ such that $\xi^*\nu = \lambda$. He did this by constructing a map χ_μ from the space of densities $\mathfrak{Dens}(M)$ to the diffeomorphism group $\mathfrak{D}(M)$ which is a smooth inverse of the pullback map $\xi \to \xi^*\mu$. Subsequently, Ebin and Marsden [20] observed that $\mathfrak{D}(M)$ is diffeomorphic[13] to the product $\mathfrak{D}_\mu(M) \times \mathfrak{Dens}(M)$, as verified directly upon setting $(\eta, \nu) \mapsto \xi = \eta \circ \chi_\mu(\nu)$ with inverse $\xi \mapsto \left(\xi \circ \chi_\mu(\xi^*\mu)^{-1}, \xi^*\mu\right)$. Reformulating these somewhat gives

Proposition 7 *The group of diffeomorphisms $\mathfrak{D}(M)$ is a smooth tame Fréchet principal bundle over $\mathfrak{Dens}(M)$ with fibre $\mathfrak{D}_\mu(M)$ and projection given by the pullback map $\pi(\xi) = \xi^*\mu$. In particular, there is a short exact sequence of tame Fréchet spaces*

$$0 \to T_e\mathfrak{D}_\mu(M) \to T_e\mathfrak{D}(M) \to T_\mu\mathfrak{Dens}(M) \to 0$$

which splits in a canonical way.

Proof This is essentially an immediate consequence of the bundle structure theorem since Moser's map χ_e defines a (global) cross-section of $\mathfrak{D}_\mu(M)$ in $\mathfrak{D}(M)$, cf. [52]. For details see Hamilton's paper [27]. □

[13] In particular, $\mathfrak{D}_\mu(M)$ is a deformation retract of $\mathfrak{D}(M)$, since $\mathfrak{Dens}(M)$ is convex.

Observe that the action of the diffeomorphism group on $\mathfrak{Dens}(M)$ is transitive (this is Moser's result above) and the fibres of the bundle $\mathfrak{D}(M) \to \mathfrak{Dens}(M)$ are precisely the *right cosets* of the subgroup $\mathfrak{D}_\mu(M)$ in $\mathfrak{D}(M)$, that is

$$\mathfrak{Dens}(M) \simeq \mathfrak{D}_\mu(M)\backslash\mathfrak{D}(M) = \left\{[\xi] = \mathfrak{D}_\mu \circ \xi : \xi \in \mathfrak{D}(M)\right\},$$

see Figure 2. This fact will play an important role below.

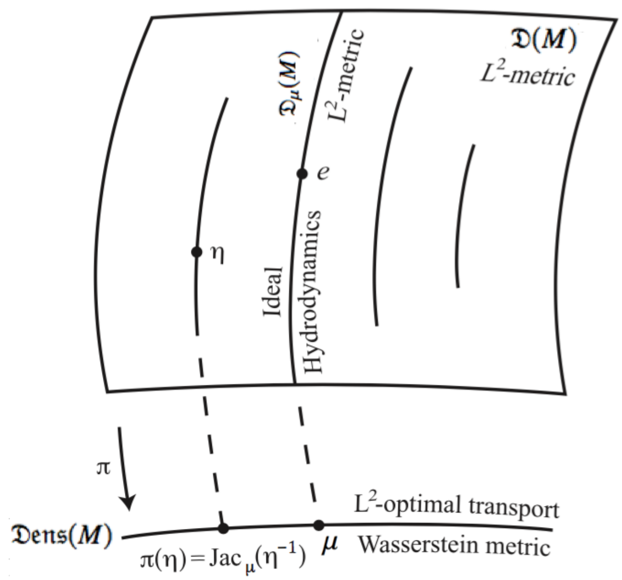

Fig. 2 The fibration of $\mathfrak{D}(M)$ with fiber $\mathfrak{D}_\mu(M)$ determined by the reference density μ together with the L^2-metric.

Remark 13 (Right cosets versus left cosets) In a similar manner one can consider the quotient space of densities as a space of *left cosets* and identify

$$\mathfrak{Dens}(M) \simeq \mathfrak{D}(M)/\mathfrak{D}_\mu(M) = \left\{[\zeta] = \zeta \circ \mathfrak{D}_\mu(M) : \zeta \in \mathfrak{D}(M)\right\}.$$

It is clear that in general a right cosets need not be a left coset. Nonetheless many constructions and arguments concerning the former apply, mutatis mutandis, to the latter as well.

It is interesting to note that identifying $\mathfrak{Dens}(M)$ with *left* cosets provides the setting for the Wasserstein (or Kantorovich-Rubinstein) geometry of Optimal Transport. See Appendix 2 below.

4 The Fisher-Rao metric in infinite dimensions

As configuration spaces of various physical systems diffeomorphism groups often come equipped with natural (pre-) Riemannian structures. Historically, of most interest were those structures related to the L^2 inner product corresponding to the kinetic energy of the system - the prime example being hydrodynamics of ideal fluids. However, in recent years there appeared many new examples from analysis, to geometry, to mathematical physics where the appropriate Riemannian structures came from higher order Sobolev inner products, see Example 5.

In this section we shall show that the Fisher-Rao (information) metric, which plays a fundamental role in geometric statistics, is closely related to an H^1-type Sobolev inner product of the space of diffeomorphisms. The approach developed below places information geometry squarely within the general differential-geometric framework for diffeomorphism groups envisioned by Cartan, Kolmogorov and Arnold.

4.1 A right-invariant homogeneous H^1 Sobolev metric on $\mathfrak{D}(M)$

We shall continue to assume that M is an n-dimensional compact Riemannian manifold without boundary whose volume form is normalized so that $\mu(M) = 1$. The triple $(M, \mathcal{B}, \mu)$ where $\mathcal{B}$ is the σ-algebra of Borel sets in M will be the fixed background sample space and the Fréchet manifold of right cosets $\mathfrak{Dens}(M) = \mathfrak{D}_\mu(M)\backslash\mathfrak{D}(M)$ will serve as an *infinite-dimensional statistical model* whose points are (smooth) probability measures[14] on M that are absolutely continuous with respect to μ. We shall equip the principal bundle $\mathfrak{D}(M)$ over $\mathfrak{Dens}(M)$ with the structure of a Riemannian submersion. To that end, observe that the condition stated in Prop. 4 is precisely what one needs.

Consider the homogeneous Sobolev H^1 inner product on the Fréchet Lie algebra of divergence free vector fields on M. Define the corresponding (degenerate) right-invariant H^1 metric on the total space $\mathfrak{D}(M)$ using (15), namely

$$g_{\dot{H}^1}(V, W) = \langle V, W \rangle_{\dot{H}^1} = \frac{1}{4} \int_M \operatorname{div} v \cdot \operatorname{div} w \, d\mu, \qquad V, W \in T_\eta \mathfrak{D}(M), \qquad (25)$$

such that $V = v \circ \eta$ and $W = w \circ \eta$ where $v, w \in T_e\mathfrak{D}(M)$ and $\eta \in \mathfrak{D}(M)$.

Clearly, the metric (25) generalizes the one-dimensional case (20). Moreover, it satisfies the condition (19) and therefore descends to a (non-degenerate) metric on $\mathfrak{Dens}(M)$. The attendant geometry is particularly remarkable. As we will see below, the space of densities $\mathfrak{Dens}(M)$ viewed as the space of right cosets equipped with this metric is isometric to a subset of the unit sphere in the Hilbert space and its Riemannian distance coincides with the spherical Hellinger distance. Furthermore,

[14] Alternatively, they are (smooth) density functions given by the corresponding Radon-Nikodym derivatives.

(25) is related to the so-called Bhattacharyya coefficient (affinity) in probability and statistics.

4.2 The square root map

We begin by constructing a map between the quotient space of right cosets $\mathfrak{Dens}(M)$ and a subset of the unit sphere in the Lebesgue space of square integrable functions on M with induced metric

$$S^{\infty}_{L^2} = \Big\{u \in L^2(M, d\mu) : \int_M |u|^2 d\mu = 1\Big\}.$$

As before we let $\mathrm{Jac}_\mu \eta$ denote the Jacobian of η with respect to μ.

Theorem 1 (Isometry theorem) *The map*

$$\Psi : \mathfrak{D}(M) \to L^2(M, d\mu) \qquad \xi \mapsto \Psi(\xi) = \sqrt{\mathrm{Jac}_\mu \xi}$$

defines an isometry from the space of densities $\mathfrak{Dens}(M) = \mathfrak{D}_\mu(M)\backslash\mathfrak{D}(M)$ with the $\dot{H}^1$ metric (25) *to a subset of the unit sphere $S^{\infty}_{L^2} \subset C^\infty(M) \cap L^2(M, d\mu)$ with the standard L^2 metric.*

Proof First, note that the Jacobian of any orientation preserving smooth diffeomorphism is a (strictly) positive smooth function. Since we are viewing $\mathfrak{Dens}(M)$ as the space of right cosets with the projection π given by the pullback map (cf. Prop. 7), given any ξ in $\mathfrak{D}(M)$ and η in $\mathfrak{D}_\mu(M)$ we have

$$(\eta \circ \xi)^* \mu = \pi(\eta \circ \xi) = \pi(\xi) = \xi^* \mu,$$

which implies that $\mathrm{Jac}_\mu(\eta \circ \xi) = \mathrm{Jac}_\mu \xi$. Furthermore, using the change of variable formula we find

$$\int_M \Psi^2(\xi)\, d\mu = \int_M \mathrm{Jac}_\mu \xi \, d\mu = \mu(M) = 1.$$

From the above it follows that Ψ descends to a well-defined map from the quotient space $\mathfrak{Dens}(M)$ into the unit sphere in $L^2(M, d\mu)$.

Next, if $\mathrm{Jac}_\mu \xi = \mathrm{Jac}_\mu \zeta$ for some ξ and ζ in $\mathfrak{D}(M)$ then we have $(\xi \circ \zeta^{-1})^* \mu = \mu$ and, consequently, we find that Ψ is injective.

Finally, differentiating the identity $\mathrm{Jac}_\mu \xi \; \mu = \xi^* \mu$ with respect to ξ in the direction $V \in T_\xi \mathfrak{D}(M)$ we obtain

$$d_\xi \mathrm{Jac}_\mu(V) = \mathrm{div}_\mu(V \circ \xi^{-1}) \circ \xi \; \mathrm{Jac}_\mu \xi.$$

Recall that tangent vectors V, W at any $\xi \in \mathfrak{D}(M)$ can be always represented in the form $V = v \circ \xi$ and $W = w \circ \xi$ where $v, w \in T_e \mathfrak{D}(M)$. Thus, applying Fréchet calculus as before, changing variables and using (25), we compute

$$
\begin{aligned}
\langle d_\xi(\Psi\circ\pi)(V), d_\xi(\Psi\circ)(W)\rangle_{L^2} &= \frac{1}{4}\int_M (\mathrm{div}_\mu v\circ\xi)\cdot(\mathrm{div}_\mu w\circ\xi)\,\mathrm{Jac}_\mu\xi\, d\mu \\
&= \frac{1}{4}\int_M \mathrm{div}_\mu v\cdot \mathrm{div}_\mu w\, d\mu \\
&= \langle v, w\rangle_{\dot{H}^1}
\end{aligned}
$$

This shows that Ψ is an isometry. □

Remark 14 If $s > n/2+1$ then the map Ψ defines a diffeomorphism from the Sobolev completion $\mathfrak{D}^s_\mu(M)\backslash\mathfrak{D}^s(M)$ onto a subset of $S^\infty_{L^2}\cap H^{s-1}(M)$ of positive functions on M and the above proof extends with minor modifications. The fact that any positive function in $S^\infty_{L^2}\cap H^{s-1}(M)$ lies in the image of Ψ follows directly from Moser's lemma, whose generalization to the setting of Sobolev H^s spaces can be found in [20]; see also Appendix 1.

4.3 The infinite-dimensional Fisher-Rao metric on $\mathfrak{Dens}(M)$

The appearance of diffeomorphism groups in our formulation of the infinite-dimensional generalization of information geometry should not be entirely surprising. In some sense it could be discerned from various results in the finite-dimensional setting such as invariance properties of the Fisher information with respect to sufficient statistics[15] — in particular, with respect to invertible mappings of the sample space M.

Attempts to find conceptually natural and useful approaches to mathematical statistics go back to the pioneering work of Fisher, Rao and Kolmogorov. Their ideas were subsequently further developed by Chentsov, Morozova, Efron, Amari, Barndorff-Nielsen, Lauritzen, Nagaoka among others; see e.g., [15], [16], [23], [1], [3], [4]. Along the way various infinite-dimensional generalizations were also considered by David [17], Friedrich [24], Pistone and Sempi [46], Gibilisco and Pistone [25], Ay, Jost, Le and Schwachhofer [7]. For an informative historical discussion, as well as for an excellent introduction to the field and its wide range of applications, we refer to the recent book of Amari [2], as well as a survey paper of Morozova and Chentsov [16].

In the classical approach one considers finite dimensional families of probability density functions on M whose elements are (smoothly) parametrized by open subsets Σ of the euclidean space

$$
\mathcal{S} = \left\{\rho = \rho_{s_1,\dots,s_k} \in \mathfrak{Dens}(M) : (s_1,\dots,s_k)\in\Sigma\subset\mathbb{R}^k\right\}.
$$

Points of M are viewed as random samples from some (typically unknown) distribution $\rho_{s_1,\dots,s_k}$ where $s_1,\dots,s_k$ are certain statistical parameters. When equipped with the structure of a smooth manifold the set $\mathcal{S}$ is referred to as a *k-dimensional*

[15] This suggests a possibility of further generalizations, which we will not pursue here.

statistical model with $s_1, \ldots, s_k$ playing the role of local coordinates. Rao [47] introduced further structure by defining at each point a $k \times k$ symmetric matrix

$$\{g_F\}_{ij} = \int_M \frac{\partial \log \rho}{\partial s_i} \frac{\partial \log \rho}{\partial s_j} \rho \, d\mu \qquad 1 \le i, j \le k, \tag{26}$$

which in the positive definite case defines a Riemannian metric g_F on $\mathcal{S}$ called the *Fisher-Rao (information) metric*. The significance of this metric for mathematical statistics was immediately recognized by mathematicians, see e.g., [15].

In our setting we shall regard a statistical model $\mathcal{S}$ as a k-dimensional Riemannian submanifold of the Fréchet manifold of smooth probability densities $\mathfrak{Dens}(M)$ on the underlying compact n-dimensional sample space manifold M. The next result shows that the metric g_F defined via (26) is in fact the same as the metric induced on $\mathcal{S}$ by the homogeneous Sobolev metric defined in the previous section on the full diffeomorphism group $\mathfrak{D}(M)$.

Theorem 2 *The right-invariant Sobolev $\dot{H}^1$ metric* (25) *on the Fréchet Lie group $\mathfrak{D}(M)$ of diffeomorphisms of a compact Riemannian manifold M descends to a (weak) Riemannian metric g_F^∞ on the quotient space of right cosets $\mathfrak{Dens}(M)$. Furthermore, it coincides (up to a constant multiple) with the standard Fisher-Rao metric g_F on any statistical submanifold $\mathcal{S}$.*

We shall refer to g_F^∞ above as the *infinite-dimensional Fisher-Rao metric*.

Proof First, we check that the homogeneous Sobolev $\dot{H}^1$ metric satisfies the required descent condition of Prop. 4. That is, we need to verify the formula (19) for $\mathfrak{G} = \mathfrak{D}(M)$ and $\mathfrak{H} = \mathfrak{D}_\mu(M)$ where $\mathrm{ad}_w v = [v, w]$ is the Lie bracket of vector fields on M. Given any vector fields u, v, w with $\mathrm{div}_\mu w = 0$, we compute

$$\begin{aligned}
\langle \mathrm{ad}_w v, u \rangle_{\dot{H}^1} + \langle v, \mathrm{ad}_w u \rangle_{\dot{H}^1} &= -\frac{1}{4} \int_M \big(\mathrm{div}_\mu [w, v] \, \mathrm{div}_\mu u + \mathrm{div}_\mu [w, u] \, \mathrm{div}_\mu v \big) d\mu \\
&= -\frac{1}{4} \int_M \Big\{ \big(\langle w, \nabla \mathrm{div}_\mu v \rangle - \langle v, \nabla \mathrm{div}_\mu w \rangle \big) \, \mathrm{div}_\mu u \\
&\qquad + \big(\langle w, \nabla \mathrm{div}_\mu u \rangle - \langle u, \nabla \mathrm{div}_\mu w \rangle \big) \, \mathrm{div}_\mu v \Big\} d\mu \\
&= \frac{1}{4} \int_M \mathrm{div}_\mu w \cdot \mathrm{div}_\mu v \cdot \mathrm{div}_\mu u \; d\mu = 0.
\end{aligned}$$

This shows that the homogeneous Sobolev (degenerate) metric (25) on $\mathfrak{D}(M)$ descends to a non-degenerate metric on the quotient $\mathfrak{D}_\mu(M) \backslash \mathfrak{D}(M)$.

For the second statement it will be convenient to carry out the calculations directly in the group $\mathfrak{D}(M)$. Given any vectors v, w in $T_e \mathfrak{D}(M)$ consider a two-parameter family of diffeomorphisms $s_1, s_2 \mapsto \xi_{s_1, s_2}$ in $\mathfrak{D}(M)$ such that $\xi(0,0) = e$ with $\partial \xi / \partial s_1 (0,0) = v$ and $\partial \xi / \partial s_2 (0,0) = w$. Their right-translations $v \circ \xi_{s_1, s_2}$ and $w \circ \xi_{s_1, s_2}$ are the corresponding variation vector fields along the surface defined by the family.

If ρ is the Jacobian of ξ_{s_1, s_2} computed with respect to the reference volume μ then (26) assumes the form

$$\{g_F\}(v,w) = \int_M \frac{\partial}{\partial s_1}\big(\log \mathrm{Jac}_\mu \xi_{s_1,s_2}\big)\frac{\partial}{\partial s_2}\big(\log \mathrm{Jac}_\mu \xi_{s_1,s_2}\big)\,\mathrm{Jac}_\mu \xi_{s_1,s_2}\, d\mu.$$

Since

$$\frac{\partial}{\partial s_1}\big(\mathrm{Jac}_\mu \xi_{s_1,s_2}\big) = \mathrm{div}_\mu v \circ \xi_{s_1,s_2} \cdot \mathrm{Jac}_\mu \xi_{s_1,s_2}$$

and similarly for the other partial derivative, using these formulas and changing variables in the integral we now obtain

$$\begin{aligned}\{g_F\}(v,w) &= \int_M \frac{\frac{\partial}{\partial s_1}\mathrm{Jac}_\mu \xi_{s_1,s_2}\frac{\partial}{\partial s_2}\mathrm{Jac}_\mu \xi_{s_1,s_2}}{\mathrm{Jac}_\mu \xi_{s_1,s_2}}\Big|_{s_1=s_2=0} d\mu\\ &= \int_M (\mathrm{div}_\mu v \circ \xi)\cdot(\mathrm{div}_\mu \circ \xi)\,\mathrm{Jac}_\mu \xi\, d\mu\\ &= \int_M \mathrm{div}_\mu v \cdot \mathrm{div}_\mu w\, d\mu\\ &= 4\langle v,w\rangle_{\dot{H}^1} = g_{\dot{H}^1}(v,w),\end{aligned}$$

which proves the theorem. □

As an immediate consequence we have

Corollary 1 *The space of densities* $\mathfrak{Dens}(M) = \mathfrak{D}_\mu(M)\backslash\mathfrak{D}(M)$ *equipped with the infinite-dimensional Fisher-Rao metric* g_F^∞ *has positive constant curvature* $1/\mu(M)$.

Proof Since Ψ is an isometry by Theorem 1 the corollary follows directly from the previous theorem and the fact that the sectional curvature of a sphere in a Hilbert space of radius $r > 0$ is precisely $1/r^2$. □

It is perhaps worth noting that if the volume of M grows to infinity then the above corollary implies that the space of densities $\mathfrak{Dens}(M)$ becomes "flatter" in the Fisher-Rao metric g_F^∞.

There is an analogue for g_F^∞ of the well-known Chentsov uniqueness theorem for the Fisher-Rao metric g_F^∞, according to which the former is essentially unique among those metrics on $\mathfrak{D}(M)$ that descend to the base $\mathfrak{Dens}(M)$ of right cosets.

Theorem 3 *Let* M *be a compact Riemannian manifold of dimension* $n \geq 2$ *without boundary. Any (weak) Riemannian right-invariant metric on* $\mathfrak{D}(M)$ *which descends to the quotient space of right cosets* $\mathfrak{Dens}(M) = \mathfrak{D}_\mu(M)\backslash\mathfrak{D}(M)$ *is a multiple of* g_F^∞.

Proof As in the finite-dimensional case this is essentially a consequence of invariance properties under the action of diffeomorphisms. A concise proof can be found in [9]. □

4.4 The metric space structure of the space of densities

Consider two smooth measures λ and ν on M that are absolutely continuous with respect to the reference measure μ and have the same total volume $\mu(M)$. Let $d\lambda/d\mu$ and $d\nu/d\mu$ be the respective Radon-Nikodym derivatives.

Theorem 4 *The Riemannian distance between λ and ν induced by the infinite-dimensional Fisher-Rao metric g_F^∞ on $\mathfrak{Dens}(M)$ is*

$$\mathrm{dist}_{g_F^\infty}(\lambda,\nu) = \sqrt{\mu(M)}\arccos\left(\frac{1}{\mu(M)}\int_M \sqrt{\frac{d\lambda}{d\mu}\frac{d\nu}{d\mu}}\,d\mu\right). \tag{27}$$

Equivalently, this distance can be computed as the Riemannian distance of (25) between two diffeomorphisms ξ and ζ in $\mathfrak{D}(M)$ which map the volume form μ to λ and ν, respectively, from the formula

$$\mathrm{dist}_{\dot{\mathcal{H}}^1}(\xi,\zeta) = \sqrt{\mu(M)}\arccos\left(\frac{1}{\mu(M)}\int_M \sqrt{\mathrm{Jac}_\mu\xi\cdot\mathrm{Jac}_\mu\zeta}\,d\mu\right).$$

Proof Let $f^2 = d\lambda/d\mu$ and $g^2 = d\nu/d\mu$. If $\lambda = \xi^*\mu$ and $\nu = \zeta^*\mu$ then using the isometry given by the square root map Ψ of Theorem 1 it suffices to compute the distance between the functions $\Psi(\xi) = f$ and $\Psi(\zeta) = g$ considered as points on the sphere $S^\infty_{L^2}(r)$ of radius $r = \sqrt{\mu(M)}$ with the metric induced from $L^2(M, d\mu)$. However, since geodesics of this metric are precisely the great circles, it follows that the length of the corresponding arc joining f and g is

$$r\arccos\left(r^{-2}\int_M fg\,d\mu\right),$$

which is formula (27). □

Recall that the diameter of a Riemannian manifold is defined as the supremum of the Riemannian distances between its points. Thus, in particular

$$\mathrm{diam}_{g_F^\infty}\big(\mathfrak{Dens}(M)\big) = \sup\Big\{\mathrm{dist}_{g_F^\infty}(\lambda,\nu) : \lambda,\nu\in\mathfrak{Dens}(M)\Big\}.$$

Corollary 2 *The diameter of $\mathfrak{Dens}(M)$ with the metric g_F^∞ equals $\pi\sqrt{\mu(M)}/2$, i.e., it is a quarter of the circumference of the sphere in $L^2(M, d\mu)$ of radius $\sqrt{\mu(M)}$.*

Proof The upper bound follows easily from the formula (27) since the argument of the arccos function is always between 0 and 1. To see that it can be arbitrarily close to 0 it suffices to choose the functions f and g as in the proof of Theorem 4 with supports in disjoint subsets. □

The Riemannian distance of the metric g_F^∞ on $\mathfrak{Dens}(M)$ is closely related to the *Hellinger distance*. Recall that the latter is defined to be

$$\mathrm{dist}_H^2(\lambda, \nu) = \int_M \left(\sqrt{d\lambda/d\mu} - \sqrt{d\nu/d\mu}\right)^2 d\mu$$

for any probability measures λ and ν on M which are absolutely continuous with respect to μ. It is readily checked that if λ and ν coincide then $\mathrm{dist}_H(\lambda, \nu) = 0$ and if they are mutually singular then $\mathrm{dist}_H(\lambda, \nu) = \sqrt{2}$. Of course, when comparing it with (27) one needs to normalize all the measures involved.

Remark 15 The Hellinger distance is also related to the so-called *Bhattacharyya affinity BC* by the formula

$$\mathrm{dist}_H^2(\lambda, \nu) = 2\big(1 - BC(\lambda, \nu)\big)$$

see e.g., [15] for more information.

The following two corollaries can be readily verified using the isometry property of Theorem 1.

Corollary 3 *The Hellinger distance between two normalized densities $d\lambda = f^2 d\mu$ and $d\nu = g^2 d\mu$ is equal to the distance in the Hilbert space $L^2(M, d\mu)$ between the functions f and g considered as points on the unit sphere $S^\infty_{L^2}$.*

Corollary 4 *The Bhattacharyya coefficient $BC(\lambda, \nu)$ of two normalized densities $d\lambda = f^2 d\mu$ and $d\nu = g^2 d\mu$ is equal to the inner product of the corresponding (positive) functions f and g in $L^2(M, d\mu)$, that is $BC(\lambda, \nu) = \int_M fg\, d\mu$.*

Let $0 < \alpha < \pi/2$ denote the angle between two vectors f and g of the unit sphere in $L^2(M, d\mu)$. In this case we have

$$\mathrm{dist}_H(\lambda, \nu) = 2 \sin \alpha/2 \qquad \text{and} \qquad BC(\lambda, \nu) = \cos \alpha$$

while

$$\mathrm{dist}_{g_F^\infty}(\lambda, \nu) = \arccos BC(\lambda, \nu).$$

We can therefore refer to the Riemannian distance of the infinite-dimensional Fisher-Rao metric on $\mathfrak{Dens}(M)$ as the *spherical Hellinger distance*; see Figure 3.

4.5 Geodesic equations as the Euler-Arnold equations and complete integrability

For a deeper insight into the infinite-dimensional analogue g_F^∞ of the Fisher-Rao metric on $\mathfrak{Dens}(M)$ we can turn to the study of its geodesics. Since the metric is invariant the associated geodesic equation can be derived via a reduction procedure as an Euler-Arnold equation on the quotient space $\mathfrak{D}_\mu(M)\backslash\mathfrak{D}(M)$. In fact, it will be convenient to work with the right-invariant Sobolev $\dot{H}^1$ metric "upstairs" on the total space $\mathfrak{D}(M)$ of all diffeomorphisms.

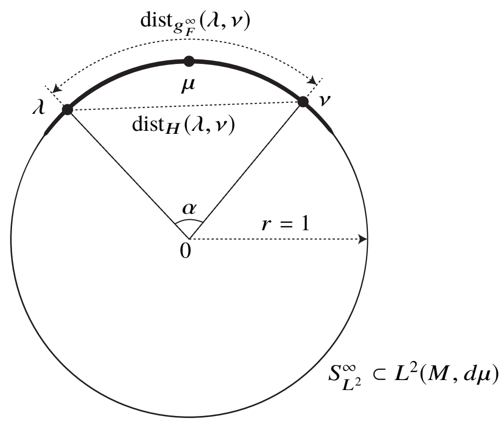

Fig. 3 The Hellinger distance $\mathrm{dist}_H(\lambda, \nu)$ and the spherical Hellinger distance $\mathrm{dist}_{g_F^\infty}(\lambda, \nu)$ between two densities $d\lambda = f^2 d\mu$ and $d\nu = g^2 d\mu$ in $S^\infty_{L^2}$. The thick arc represents the image of $\mathfrak{D}(M)$ under the square root map Ψ.

Theorem 5 *The Euler-Arnold equation of the homogeneous metric* (25) *has the form*

$$\nabla \mathrm{div}_\mu u_t + \mathrm{div}_\mu u \cdot \nabla \mathrm{div}_\mu u + \nabla \langle u, \nabla \mathrm{div}_\mu u \rangle = 0 \tag{28}$$

or, equivalently,

$$h_t + \langle u, \nabla h \rangle + \frac{1}{2} h^2 = -\frac{1}{2\mu(M)} \int_M h^2 d\mu \tag{29}$$

where $h = \mathrm{div}_\mu u$.

Proof Using the general Euler-Arnold equation and Remark 9 of Sec. 3.2.1 we only need to compute the coadjoint operator with respect (25). On the one hand, from (25) for any u, v and w in $T_e \mathfrak{D}(M)$ we have

$$\langle \mathrm{ad}^*_v u, w \rangle_{\dot{H}^1} = -\frac{1}{4} \int_M \langle \nabla \mathrm{div}_\mu \mathrm{ad}^*_v u, w \rangle d\mu.$$

On the other hand, using (17) we compute

$$
\begin{aligned}
\langle \mathrm{ad}_v^* u, w\rangle_{\dot H^1} &= \langle u, \mathrm{ad}_v w\rangle_{\dot H^1} = \frac{1}{4}\int_M \mathrm{div}_\mu u \cdot \mathrm{div}_\mu [v,w]\, d\mu \\
&= \frac{1}{4}\int_M \big(\mathrm{div}_\mu u \cdot \langle v, \nabla \mathrm{div}_\mu w\rangle - \mathrm{div}_\mu u \cdot \langle \nabla \mathrm{div}_\mu v, w\rangle\big) d\mu \\
&= -\frac{1}{4}\int_M \mathrm{div}_\mu\big(\mathrm{div}_\mu u \cdot v\big)\cdot \mathrm{div}_\mu w\, d\mu - \frac{1}{4}\int_M \langle \mathrm{div}_\mu u \cdot \nabla \mathrm{div}_\mu v, w\rangle d\mu \\
&= \frac{1}{4}\int_M \big\langle \nabla \mathrm{div}_\mu(\mathrm{div}_\mu u\cdot v) - \mathrm{div}_\mu u \cdot \nabla \mathrm{div}_\mu v, w\big\rangle d\mu.
\end{aligned}
$$

Since w is an arbitrary vector field on M, comparing the two integral expressions above we obtain

$$
\begin{aligned}
\nabla \mathrm{div}_\mu \mathrm{ad}_v^* u &= -\nabla \mathrm{div}_\mu(\mathrm{div}_\mu u \cdot v) + \mathrm{div}_\mu u \cdot \nabla \mathrm{div}_\mu v \\
&= -\nabla \langle \nabla \mathrm{div}_\mu u, v\rangle - \nabla(\mathrm{div}_\mu u \cdot \mathrm{div}_\mu v) + \mathrm{div}_\mu u \cdot \nabla \mathrm{div}_\mu v.
\end{aligned}
$$

Substituting into (16) yields the desired Euler-Arnold equation (28). □

Observe that in the one-dimensional case when $M = \mathbb{T}$ differentiating equation (29) with respect to x gives the Hunter-Saxton equation (21) of Example 19.

4.5.1 The Cauchy problem: explicit solutions

The question of wellposedness of the Cauchy problem for a nonlinear evolution equation subject to an appropriate initial data involves constructing a unique solution which belongs to a given function space, satisfies both the equation and the initial condition, and depends at least continuously on the data. In the case of the general Euler-Arnold equation (16) this question can be studied either by working directly with the *partial differential equation* or indirectly by reformulating it in terms of the associated geodesic flow in the group (or the homogeneous space). One advantage of the latter approach is that in a suitable Banach space setting (such as Sobolev H^s with $s > n/2 + 1$ or Hölder $C^{1,\alpha}$ with $0 < \alpha < 1$) the geodesics can be often constructed using Banach-Picard iterations as solutions of an *ordinary differential equation*. We point out however that the two formulations of the Cauchy problem (one in the Lie algebra and the other in the group) are in general not equivalent since in the latter case the data-to-solution map is typically smooth, while in the former it is at best continuous in any reasonable Banach space topology.

As it turns out, in our case we can solve the Euler-Arnold equations of Thm. 5 by deriving explicit formulas for the corresponding solutions.

Theorem 6 *Let $h = h(t,x)$ be the solution of* (29) *with the initial condition*

$$h(0,x) = \mathrm{div}_\mu u_0(x). \tag{30}$$

Let $t \mapsto \xi(t)$ be the flow of the corresponding velocity field $u = u(t,x)$, that is

$$\frac{d}{dt}\xi(t,x) = u(t,\xi(t,x)), \qquad \xi(0,x) = x.$$

Then, we have

$$h(t,\xi(t,x)) = 2\kappa \tan\Big(\arctan\frac{\mathrm{div}_\mu u_0(x)}{2\kappa} - \kappa t\Big), \tag{31}$$

$$\textit{where} \quad \kappa^2 = \frac{1}{4\mu(M)}\int_M (\mathrm{div}_\mu u_0)^2 d\mu. \tag{32}$$

Furthermore, the Jacobian of the flow is

$$\mathrm{Jac}_\mu(\xi(t,x)) = \Big(\cos\kappa t + \frac{\mathrm{div}_\mu u_0(x)}{2\kappa}\sin\kappa t\Big)^2. \tag{33}$$

Proof If $\phi(t,x)$ is a smooth real-valued function then the chain rule gives

$$\frac{d}{dt}\phi(t,\xi(t,x)) = \frac{\partial\phi}{\partial t}(t,\xi(t,x)) + \big\langle u(t,\xi(t,x)), \nabla\phi(t,\xi(t,x))\big\rangle.$$

From this formula and from (29) we obtain an equation for $\phi = h\circ\xi$, namely

$$\frac{d\phi}{dt} + \frac{1}{2}\phi^2 = -C(t) \tag{34}$$

where $C(t) = (2\mu(M))^{-1}\int_M h^2 d\mu$. Observe that $C(t)$ is in fact independent of the time variable t since

$$\begin{aligned}\mu(M)\frac{dC}{dt}(t) &= \int_M hh_t\,d\mu = \int_M \mathrm{div}_\mu u\cdot\mathrm{div}_\mu u_t\,d\mu \\ &= -\int_M \langle u\nabla\mathrm{div}_\mu u\rangle\cdot\mathrm{div}_\mu u\,d\mu - \frac{1}{2}\int_M (\mathrm{div}_\mu u)^3 d\mu = 0,\end{aligned}$$

where the last step follows at once by integrating by parts.

Set $C = 2\kappa^2$. Then, for any fixed $x\in M$ the solution of the ODE in (34) has the form

$$\phi(t) = 2\kappa\tan\left(\arctan\left(f(0)/2\kappa\right) - \kappa t\right),$$

which is precisely (31).

Finally, to find the formula for the Jacobian we first compute the time derivative of $\mathrm{Jac}_\mu\xi\;\mu$ to get

$$\frac{d}{dt}(\mathrm{Jac}_\mu\xi\;\mu) = \frac{d}{dt}\xi^*\mu = \xi^*(\mathcal{L}_u\mu) = (\phi\circ\xi)\,\mathrm{Jac}_\mu\xi\;\mu,$$

which gives a differential equation, whose solution can be now easily verified to be (33) by making use of the formulas (31) . □

Remark 16 The formula (33) for the Jacobian of the flow $\xi(t)$ can be viewed in light of the correspondence between the geodesics of g_F^∞ in $\mathfrak{Dens}(M)$ and those on the round sphere in a Hilbert space established in Thm. 1. Indeed, the map

$$t \mapsto \sqrt{\mathrm{Jac}_\mu(\xi(t,x))} = \cos \kappa t + \frac{\mathrm{div}_\mu u_0(x)}{2\kappa} \sin \kappa t$$

describes the great circle on the unit sphere $S^\infty_{L^2} \subset L^2(M, d\mu)$.

4.5.2 The Cauchy problem: breakdown of solutions

Using the explicit formulas of Thm. 6 it possible to make conclusions regarding long time behaviour of solutions. For example, it turns out that all smooth (classical) solutions of the Euler-Arnold equation (29) must break down in finite time.

Theorem 7 *The lifespan of any (smooth) solution of the Cauchy problem* (29)-(30) *constructed in Thm. 6 is*

$$0 < T_{\max} = \frac{\pi}{2\kappa} + \frac{1}{\kappa} \arctan\left(\frac{1}{2\kappa} \inf_{x\in M} \mathrm{div}_\mu u_0(x)\right). \tag{35}$$

Furthermore, $\|u(t)\|_{C^1} \nearrow +\infty$ *as* $t \to T_{\max}$.

Proof The theorem follows at once from formula (31) and $\mathrm{div}_\mu u = h$. Alternatively, observe that formula (33) implies that the flow of $u(t,x)$ ceases to be a diffeomorphism at the critical time $t = T_{\max}$. □

This result can be also interpreted geometrically as saying that the corresponding geodesics of the Fisher-Rao metric g^∞_F leave the set of positive densities $\mathfrak{Dens}(M)$ and can be no longer lifted to a smooth curve of diffeomorphisms in $\mathfrak{D}(M)$.

4.5.3 Complete integrability

The Euler-Arnold systems are a special class of hamiltonian systems. In the finite dimensional case, given a smooth $2n$-dimensional manifold M equipped with a nondegenerate closed 2-form $\omega \in \Omega^2(M)$ a vector field v on M is called *hamiltonian* if $\iota_v\omega$ is exact, $\iota_v\omega = dH$ for a certain function H, and such a function H is called a *hamiltonian function* of v. Then a *hamiltonian system* is a triple (M, ω, H).

In this setting the Fréchet space $C^\infty(M)$ of smooth functions on M acquires a structure of an associative commutative algebra with a Poisson bracket given by $\{f, g\} = \omega(v_f, v_g)$ where $\iota_{v_f}\omega = df$ and $\iota_{v_g}\omega = dg$. The equations of motion of a hamiltonian system assume a simple and elegant form

$$\frac{df}{dt} = \{f, H\} \tag{36}$$

for any function $f \in C^\infty(M)$.

Any smooth function f such that $\{f, H\} = 0$ is called a *constant of motion* (or a *first integral*) of the hamiltonian system since (as an easy consequence of (36)) it necessarily assumes a constant value along any orbit of v_H. A (finite-dimensional) hamiltonian system is said to be *(completely) integrable* if it possesses n (almost everywhere) functionally independent first integrals $f_1 = H, f_2, \ldots, f_n$ which are pairwise in involution, i.e., $\{f_i, f_j\} = 0$ for any i and j.

In infinite dimensions the situation is more subtle. Possessing infinitely many first integrals may be insufficient to determine the motion of the system and the relations between various competing definitions of integrability are not yet fully understood. Nevertheless, there are infinite-dimensional systems for which most of these notions agree. Perhaps, the most famous example is the Korteweg-de Vries equation of mathematical shallow water theory. Other examples include the one-dimensional Hunter-Saxton equation, the two-dimensional Kadomtsev-Petviashvili equation and the Camassa-Holm type equations.

Suppose now that $\mathfrak{G}$ is a Lie group (finite or infinite dimensional) with Lie algebra $\mathfrak{g}$. The (smooth part of its) dual $\mathfrak{g}^*$ carries a natural bracket (the Kirillov-Kostant bracket) defined by

$$\{f, g\}_{KK}(m) = \big([df(m), dg(m)], m\big)\,,$$

where $f, g \in C^\infty(\mathfrak{g}^*)$ and the differentials $df(m)$ and $dg(m)$ at $m \in \mathfrak{G}^*$ are identified with the corresponding elements of the Lie algebra via the natural pairing of $\mathfrak{g}$ and $\mathfrak{g}^*$ given by $(\cdot, \cdot)$. In this setting the hamiltonian equation corresponding to the function H and the Kirillov-Kostant structure on $\mathfrak{g}^*$ takes the form

$$\frac{dm}{dt} = -\mathrm{ad}^*_{dH(m)} m.$$

If the hamiltonian system is driven by a kinetic energy E (a quadratic function of its argument) then choosing $H = E$ yields the Euler-Arnold equation (16) of Sect. 3.2.1 above.

It turns out that the equations induced by the $g_{\dot{H}^1}$ metric in Thm. 5 also belong to this class.

Theorem 8 *The Euler-Arnold equation of the Fisher-Rao metric g_F^∞ on the space of densities $\mathfrak{Dens}(M)$ is an infinite-dimensional completely integrable dynamical system.*

Proof This is essentially a consequence of the fact that (29) describes the geodesic flow on the sphere $S^\infty_{L^2}(r) \subset L^2(M, d\mu)$ of radius r and it therefore admits infinitely many first integrals in direct analogy with the finite dimensional case $S^{n-1}(r) \subset \mathbb{R}^n$. We refer to [30] for further details. $\square$

Remark 17 It is natural to expect that in any dimension n the Euler-Arnold equation (29) is integrable in that it admits a bi-hamiltonian structure. In the one-dimensional case of the Hunter-Saxton equation (21) this fact is well known, see e.g., [29].

5 Amari-Chentsov connections and their geodesics in the space of densities

One of the questions posed in the monograph [4] asked for an infinite-dimensional theory of Amari-Chentsov connections. So far several proposals have been developed in various levels of generality. In this section we present an approach that fits well within the framework of diffeomorphism groups and their quotients.

Let M be a compact Riemannian manifold without boundary. On the product $\mathfrak{D}(M) \times \mathfrak{D}(M)$ consider the following family of real-valued functions

$$D^{(\alpha)}(\xi,\eta) = \frac{1}{1-\alpha^2}\left(1 - \int_M (\mathrm{Jac}_\mu \xi)^{\frac{1-\alpha}{2}} (\mathrm{Jac}_\mu \eta)^{\frac{1+\alpha}{2}}\, d\mu\right), \tag{37}$$

$$D^{(-1)}(\xi,\eta) = D^{(1)}(\eta,\xi) = \frac{1}{4}\int_M \big(\log \mathrm{Jac}_\mu \xi - \log \mathrm{Jac}_\mu \eta\big) \mathrm{Jac}_\mu \xi \, d\mu\,, \tag{38}$$

where $-1 < \alpha < 1$. These functions are clearly well defined on $\mathfrak{D}_\mu(M)\backslash\mathfrak{D}(M)$ and satisfy $D^{(\alpha)}(\xi,\eta) \geq 0$ with equality if and only if ξ and η project onto the same density on M. They can be naturally viewed as diffeomorphism group analogues of the contrast functions (*α-divergences*) considered by Amari and Chentsov in the classical setting of finite-dimensional statistical models.

Although for the sake of clarity we will focus on the one-dimensional case, all the constructions can be readily generalized to diffeomorphism groups of higher-dimensional manifolds.

If M is the unit circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ then $\mathfrak{D}_\mu(\mathbb{T})$ is simply the set of rigid rotations $\mathrm{Rot}(\mathbb{T}) \simeq \mathbb{T}$. In this case it will be convenient to identify the quotient space of densities $\mathfrak{Dens}(\mathbb{T}) = \mathrm{Rot}(\mathbb{T})\backslash\mathfrak{D}(\mathbb{T})$ with the subgroup of all those circle diffeomorphisms which fix a prescribed point, e.g. $\mathfrak{Dens}(\mathbb{T}) \simeq \big\{\xi \in \mathfrak{D}(\mathbb{T}) : \xi(0) = 0\big\}$. Its tangent space at the identity map can then be identified with the space of smooth periodic functions that vanish at $x = 0$. Furthermore, for any such function $u(x)$ the inverse operator of $A = -\partial_x^2$ can be written explicitly in the form

$$A^{-1}u(x) = -\int_0^x \int_0^y u(z)\, dz dy + x \int_0^1 \int_0^y u(z)\, dz dy. \tag{39}$$

We are now in a position to prove the following result.

Theorem 9 (Reduced α-geodesic equations)

1. *Each contrast function $D^{(\alpha)}$ induces on $\mathfrak{Dens}(\mathbb{T})$ the homogeneous $\dot{H}^1$ Sobolev metric and an affine connection $\nabla^{(\alpha)}$ whose Christoffel symbols are given by*

$$\Gamma^{(\alpha)}_\xi(W,V) = -\frac{1+\alpha}{2}\Big\{A^{-1}\partial_x\Big((V\circ\xi^{-1})_x (W\circ\xi^{-1})_x\Big)\Big\}\circ\xi \tag{40}$$

   where $-1 \leq \alpha \leq 1$.
2. *For any α the connections $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ are dual with respect to the right invariant Sobolev $\dot{H}^1$ metric given at the identity by*

$$\langle u,v\rangle_{\dot H^1}=\frac{1}{4}\int_{\mathbb{T}}u_xv_x\,dx.$$

$\nabla^{(0)}$ *is the corresponding self-dual Levi-Civita connection.*

3. *The geodesic equations of* $\nabla^{(\alpha)}$ *on* $\mathfrak{Dens}(\mathbb{T})$ *correspond to the generalized Proudman-Johnson equations*

$$\partial_t\partial_x^2u+(2-\alpha)\partial_xu\partial_{xx}u+u\partial_x^3u=0. \tag{41}$$

In particular, the case $\alpha=0$ *yields the completely integrable Hunter-Saxton equation*

$$u_{txx}+2u_xu_{xx}+uu_{xxx}=0\,,$$

while the case $\alpha=-1$ *yields the completely integrable* μ*-Burgers equation*

$$u_{txx}+3u_xu_{xx}+uu_{xxx}=0.$$

The equation corresponding to $\alpha=1$ is also integrable in that its solutions can be written down explicitly, see Thm. 10 below.

Proof As in finite dimensions the functions $D^{(\alpha)}$ induce metrics and connections, see Sect. 2.1.4. Assume first that $\alpha\neq\pm1$. Given any vectors V,W tangent at $\xi\in\mathfrak{D}(\mathbb{T})$ let $\xi_{s,t}$ be a two-parameter family of diffeomorphisms in $\mathfrak{D}(\mathbb{T})$ such that $\xi|_{s=t=0}=\xi$ with $\frac{\partial}{\partial s}\xi|_{s=t=0}=V$ and $\frac{\partial}{\partial t}\xi|_{s=t=0}=W$. Then from (37) we have

$$\begin{aligned}\langle V,W\rangle_\alpha&=-\frac{\partial}{\partial s}\Big|_{s=0}\frac{\partial}{\partial t}\Big|_{t=0}D^{(\alpha)}(\xi_{s,0},\xi_{0,t})\\&=\frac{1}{1-\alpha^2}\frac{\partial}{\partial s}\Big|_{s=0}\frac{\partial}{\partial t}\Big|_{t=0}\int_{\mathbb{T}}(\partial_x\xi_{s,0})^{\frac{1-\alpha}{2}}(\partial_x\xi_{0,t})^{\frac{1+\alpha}{2}}\,dx\\&=\frac{1}{4}\int_{\mathbb{T}}\partial_xV\partial_xW\,(\partial_x\xi)^{-\frac{1+\alpha}{2}}(\partial_x\xi)^{\frac{-1+\alpha}{2}}\,dx\\&=\frac{1}{4}\int_{\mathbb{T}}\frac{V_xW_x}{\xi_x}\,dx=\langle V,W\rangle_{\dot H^1}.\end{aligned}\tag{42}$$

Suppose that W is a vector field on $\mathfrak{D}(\mathbb{T})$ defined in some neighbourhood of ξ. Let $\xi_{s,t,r}$ be a three-parameter family of diffeomorphisms such that $\xi|_{s=t=r=0}=\xi$ with $\frac{\partial}{\partial s}\xi|_{s=t=r=0}=V$, $\frac{\partial}{\partial r}\xi|_{s=t=r=0}=Z$ and $\frac{\partial}{\partial t}\xi_s|_{t=r=0}=W_{\xi_{s,0,0}}$ for sufficiently small s. Now, using (37) and (42) we compute

$$
\begin{aligned}
\langle\nabla_V^{(\alpha)}W,Z\rangle_\alpha &= \frac{1}{4}\int_{\mathbb{T}}\frac{(\nabla_V^{(\alpha)}W)_x Z_x}{\xi_x}\,dx\\
&= -\frac{\partial}{\partial s}\Big|_{s=0}\frac{\partial}{\partial t}\Big|_{t=0}\frac{\partial}{\partial r}\Big|_{r=0}D^{(\alpha)}(\xi_{s,t,0},\xi_{0,0,r})\\
&= \frac{1}{1-\alpha^2}\frac{\partial}{\partial s}\Big|_{s=0}\frac{\partial}{\partial t}\Big|_{t=0}\frac{\partial}{\partial r}\Big|_{r=0}\int_{\mathbb{T}}(\partial_x\xi_{s,t,0})^{\frac{1-\alpha}{2}}(\partial_x\xi_{0,0,r})^{\frac{1+\alpha}{2}}\,dx\\
&= \frac{1}{4}\int_{\mathbb{T}}\left\{\Big((dW\cdot V)\circ\xi^{-1}\Big)_x-\frac{1+\alpha}{2}\big(V\circ\xi^{-1}\big)_x\big(W\circ\xi^{-1}\big)_x\right\}\big(Z\circ\xi^{-1}\big)_x dx\,.
\end{aligned}
$$

Now integrating by parts and using the fact that Z is arbitrary, we obtain

$$\big(\nabla_V^{(\alpha)}W\big)_\xi = \big(dW\cdot V\big)(\xi)-\Gamma_\xi^\alpha(W,V)$$

where the Christoffel map is given by the formula (40).

The computations in the remaining two cases are analogous and for $\alpha=-1$ and $\alpha=1$ yield

$$\Gamma_\xi^{(-1)}(W,V)=0\,, \tag{43}$$

$$\Gamma_\xi^{(1)}(W,V)=-A^{-1}\partial_x\Big((V\circ\xi^{-1})_x(W\circ\xi^{-1})_x\Big)\circ\xi\,, \tag{44}$$

which establishes the first part of the theorem.

To establish the second part we need to verify that for any vector fields V,W and Z on $\mathrm{Rot}(\mathbb{T})\backslash\mathfrak{D}(\mathbb{T})$ we have

$$V\cdot\langle W,Z\rangle_{\dot H^1}=\langle\nabla_V^{(\alpha)}W,Z\rangle_{\dot H^1}+\langle W,\nabla_V^{(\alpha)}Z\rangle_{\dot H^1}. \tag{45}$$

This is done by a direct calculation as above. Alternatively, it can be deduced from general properties of contrast functions of the type (37) and (38) as discussed e.g. in Chap. 3 of [4]. The fact that $\nabla^{(0)}$ is a Levi-Civita connection of the $g_{\dot H^1}$ metric follows at once from (45).

The equation for geodesics of $\nabla^{(\alpha)}$ on $\mathfrak{Dens}(\mathbb{T})$ has the form

$$\frac{d^2\gamma}{dt^2}=\Gamma_\gamma^{(\alpha)}\Big(\frac{d\gamma}{dt},\frac{d\gamma}{dt}\Big). \tag{46}$$

Let $d\gamma/dt=u\circ\gamma$ where u is a time-dependent vector field on $\mathbb{T}$ (i.e., a periodic function vanishing at $x=0$). Differentiating in the time variable and substituting into (46) we obtain the corresponding nonlinear PDE

$$u_t+uu_x=-\frac{1+\alpha}{2}A^{-1}\partial_x(u_x^2)\,,$$

which we can rewrite as

$$-u_{txx}-3u_xu_{xx}-uu_{xxx}=-(1+\alpha)u_xu_{xx}\,,$$

which is precisely (41). □

Remark 18 The Hunter-Saxton equation (21) (cf. Thm. 9 part 3.) can be alternatively derived by observing that it is the Euler-Arnold equation of $\nabla^{(0)}$ on tangent space to $\mathfrak{Dens}(\mathbb{T})$ at the identity map and as such it is obtained from the geodesic equation of the right-invariant metric $g_{\dot{H}^1}$ by a standard procedure, see [29].

Remark 19 (α-curvature) Using the Christoffel symbols in (40) it is possible to calculate the curvature of the α-connections. It turns out to be proportional to the curvature of the $\dot{H}^1$ metric, i.e.,

$$\mathcal{R}^{(\alpha)}(V,W)Z = (1-\alpha^2)\Big(\langle V\cdot\langle W,Z\rangle_{\dot{H}^1} + W\cdot\langle V,Z\rangle_{\dot{H}^1}\Big) \tag{47}$$

for any vector fields V, W and Z on $\mathfrak{Dens}(\mathbb{T})$. This formula can be computed as in finite dimensions, see [16] where a different choice of parameters is made.

As already mentioned, it turns out that the geodesic equation corresponding to $\alpha = 1$ can be integrated as well. This is done indirectly by constructing affine coordinates for $\nabla^{(1)}$. Observe that from (47) we already know that the connections $\nabla^{(-1)}$ and $\nabla^{(1)}$ are flat. In the former case this is also evident from (43).

Theorem 10 *The geodesic equations of $\nabla^{(1)}$ corresponding to the Euler-Arnold equation*

$$u_{txx} + u_x u_{xx} + u u_{xxx} = 0 \tag{48}$$

is integrable with solutions given explicitly by

$$u = \frac{d\xi}{dt}\circ\xi^{-1} \qquad \textit{where} \quad \xi(t,x) = \frac{\int_0^x e^{a(y)t+b(y)}dy}{\int_0^1 e^{a(x)t+b(x)}dx} \tag{49}$$

and a and b are smooth mean-zero functions on $\mathbb{T}$.

Proof We will construct a chart on $\mathfrak{Dens}(\mathbb{T}) = \mathrm{Rot}(\mathbb{T})\backslash\mathfrak{D}(\mathbb{T})$ in which the Christoffel symbols of $\nabla^{(1)}$ vanish. Consider the map

$$\xi \mapsto \varphi(\xi) = \log\xi_x - \int_{\mathbb{T}}\log\xi_x\,dx \tag{50}$$

from the quotient space to the space of smooth periodic mean-zero functions. To determine how the Christoffel symbols transform under the change of variables $\xi \mapsto \tilde{\xi} = \varphi(\xi)$ we first compute

$$d_\xi\varphi(W) = \frac{W_x}{\xi_x} - \int_{\mathbb{T}}\frac{W_x}{\xi_x}dx$$

and

$$d_\xi^2\varphi(W,V) = -\frac{V_xW_x}{\xi_x^2} + \int_{\mathbb{T}}\frac{V_xW_x}{\xi_x^2}dx$$

for $V, W \in T_\xi(\mathfrak{Dens}(\mathbb{T})$. Using (39) and (44) with extra work we now find that

$$\widetilde{\Gamma}^{(1)}_{\varphi(\xi)}\big(d_\xi\varphi(W), d_\xi\varphi(V)\big) = d^2_\xi\varphi(W,V) + d_\xi\varphi\big(\Gamma^{(1)}_\xi(W,V)\big) = 0$$

where $v = V \circ \xi^{-1}$ and $w = W \circ \xi^{-1}$.

We can now construct explicit solutions of (48) as follows. Since $\widetilde{\Gamma}^{(1)} = 0$ all geodesics of $\nabla^{(1)}$ in the affine coordinates are straight lines

$$t \mapsto \tilde{\xi}(t,x) = a(x)t + b(x) \qquad \text{for} \quad x \in \mathbb{T}$$

where a and b are smooth periodic functions of mean zero. To find a general solution $u(t,x)$ it now suffices to invert the map φ in (50) to obtain the flow $t \mapsto \xi(t) = \varphi^{-1}\tilde{\xi}(t)$ and then right-translate the velocity vector of the curve $\xi(t)$ to the tangent space at the identity in $\mathfrak{Dens}(\mathbb{T})$. This yields the explicit formulas in (49). □

The proof of Thm. 10 shows that the equation (48) is integrable. In fact, the explicit change of coordinates linearizes the flow in the same spirit as the formalism of the inverse scattering transform. For further details we refer to the paper [36].

Appendix 1: Banach completions of manifolds of maps

Even though for our purposes it was convenient to work with C^∞ maps, most of the constructions presented in this chapter could be (and, in the general literature on the subject, typically are) carried out in the framework of Banach spaces, such as Sobolev spaces. We describe this setup briefly and refer the reader to e.g., [20] or [44] for further details.

Sobolev spaces of arbitrary order

As before let M be a closed Riemannian manifold and let E be a hermitian vector bundle with fibre over M and connection ∇. To define Sobolev sections of E of arbitrary order $s \in \mathbb{R}$ it will be convenient to use the Fourier transform. By compactness we pick a trivializing cover by charts $\varphi_i : U_i \subset M \to \{x \in \mathbb{R}^n : |x| \le 1\}$ where $i = 1, \dots, N$ so that $E|_{U_i} \simeq U_i \times \mathbb{C}^m$ with a smooth extension to a neighbourhood of each U_i. We can further arrange things so that $M = \bigcup_{i=1}^N B_i(1/\sqrt{2})$ where $B_i(1/\sqrt{2}) = \big\{x \in U_i : |\varphi(x)| < 1/\sqrt{2}\big\}$ and setting $\psi_i = \varphi_i/(1-|\varphi_i|^2)^{1/2}$ produce a coordinate cover with the property that $U_i \simeq \mathbb{R}^n$ and the restriction to U_i of any smooth section of E can be viewed as a function $u : \mathbb{R}^n \to \mathbb{C}^m$ that is bounded together with its weighted derivatives $x \mapsto |x^\alpha D^\alpha u(x)|$ for any multi-index α. Let $\{\varrho_i\}_{i=1,\dots,N}$ be a smooth partition of unity subordinate to the U_i's. Then a section of E can be written as $u = \sum_i \varrho_i u$ whose terms are smooth functions of compact support in the unit ball in $\mathbb{R}^n$. This effectively reduces the study of sections of E to that of $\mathbb{C}^m$ valued Schwartz class functions (cf. Ex. 3 of Sec. 2.1) and makes it possible to utilize the *Fourier transform* on $\mathbb{R}^n$

$$\hat{u}(\xi) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} u(x)\, dx$$

and its inversion on $\mathcal{S}(\mathbb{R}^n)$. For any $s \in \mathbb{R}$ we let

$$\|u\|^2_{H^s} = \int_{\mathbb{R}^n} \left(1 + |\xi|^2\right)^s |\hat{u}(\xi)|^2 d\xi \tag{51}$$

and define the *Sobolev space* of H^s sections of E to be the completion of $\mathcal{S}(\mathbb{R}^n)$ in the norm (51). Clearly, this construction generalizes the spaces of Sobolev sections introduced in Example 2 for any integer $s = k$ with norms given by (2).

The importance of Sobolev spaces lies in the fact that they account for the "size" of functions in terms of both the "mass" and "height" of their derivatives in a way that relates the two. This is borne out by the following result.

Theorem 11 (Sobolev Lemma) *If $s > n/2 + k$ then any $u \in H^s(E)$ can be modified (on a set of μ-measure zero) to a function of class C^k. Furthermore, we have*

$$\|u\|_{C^k} \leq C_{n,s} \|u\|_{H^s}$$

for some constant $C_{n,s}$ depending only on s and n.

Proof Observe that the function $\xi \mapsto (1 + |\xi|^2)^{-s}$ is integrable on $\mathbb{R}^n$ whenever $s > n/2$. For any $u \in \mathcal{S}(\mathbb{R}^n)$ using the Fourier inversion formula and applying the Cauchy-Schwartz inequality we find

$$|u(x)|^2 \leq (2\pi)^{-n} \int_{\mathbb{R}^n} \left(1 + |\xi|^2\right)^{-s} d\xi \int_{\mathbb{R}^n} \left(1 + |\xi|^2\right)^s |\hat{u}(\xi)|^2 d\xi \leq C_{n,s} \|u\|^2_{H^s}$$

for any $x \in \mathbb{R}^n$ which settles the case $k = 0$. Repeating this argument for each derivative $D^\alpha u$ with $|\alpha| < s - n/2$ yields the result. $\square$

Sobolev manifolds of maps

The set $H^s(M, M)$ is now defined as consisting of maps f of M into itself such that for any $x \in M$ there are local charts (U, φ) at x and (V, ψ) at $f(x)$ for which the composition $\psi \circ f \circ \varphi^{-1}$ belongs to $H^s(\phi(U), \mathbb{R}^n)$. If $s > n/2$ then using the Sobolev lemma one shows that this definition is independent of the choice of charts on M. The tangent space at $f \in H^s(M, M)$ is the set of all H^s cross-sections of the pull-back bundle $T_f H^s(M, M) = H^s(f^{-1}TM)$ and is used as the model space.

A differentiable atlas for $H^s(M, M)$ can be constructed using the Riemannian exponential map on M. For example, to find a chart at the identity $f = e$ we may consider the map $\mathrm{Exp} : TM \to M \times M$ given by

$$v \mapsto \mathrm{Exp}(v) = \left(\pi(v), \exp_{\pi(v)} v_{\pi(v)}\right)$$

where $\pi : TM \to M$ is the tangent bundle projection. Exp is clearly a diffeomorphism from a neighbourhood U of the zero section onto a neighbourhood of the diagonal in $M \times M$. Using this map one defines a bijection from the set

$$\mathcal{U}_e = \{v \in H^s(TM) : \ v(M) \subset U\}$$

onto a neighbourhood of the identity map in $H^s(M, M)$ by

$$\Psi : T_e H^s(M, M) \supset \mathcal{U}_e \to H^s(M, M) \qquad v \mapsto \Psi(v) = \mathrm{Exp} \circ v.$$

One checks now that the pair $(\mathcal{U}_e, \Psi)$ defines a chart in $H^s(M, M)$ around e. Compactness, properties of Riemannian exponential maps and standard facts about compositions of Sobolev functions and diffeomorphisms ensure that the charts are well-defined and independent of the metric on M and their transition functions are smooth on the overlaps.

For any $s > n/2 + 1$ the group of H^s diffeomorphisms of M is now defined as

$$\mathfrak{D}^s(M) = \mathfrak{D}^1(M) \cap H^s(M, M),$$

where $\mathfrak{D}^1(M)$ is the set of C^1 diffeomorphisms of M. Since $\mathfrak{D}^1(M)$ forms an open set in $C^1(M, M)$, it follows that $\mathfrak{D}^s(M)$ is also open as a subset of the Hilbert manifold $H^s(M, M)$ (Sobolev lemma) and hence a smooth manifold. Furthermore, it is a topological group under composition of diffeomorphisms. In fact, right multiplications $R_\eta(\xi) = \xi \circ \eta$ are smooth in the H^s topology, but left multiplications $L_\eta(\xi) = \eta \circ \xi$ and inversions $\mathrm{i}(\xi) = \xi^{-1}$ are continuous, but not Lipschitz continuous.

Example 21 Let $f(x) = x + ah(x)$ and $g(x) = x + bh(x)$, where $-1 \le x \le 1$, $0 < b = 2a \ll 1$ and $h(x)$ is smooth function of compact support in $(-1, 1)$ such that $h(0) = 0$ and $\|h\|_{H^2} \le 1$. Clearly, both f and g are increasing functions.

Furthermore, on the one hand we have

$$\|f - g\|_{H^2} = \|ah\|_{H^2} \le a.$$

On the other hand

$$\begin{aligned} \|f^{-1} - g^{-1}\|_{H^2} &\ge \left|(f^{-1})'(0) - (g^{-1})'(0)\right| \\ &= \left|\frac{1}{1 + ah'(0)} - \frac{1}{1 + bh'(0)}\right| = a^{-1}\frac{|1 - a|}{|1 - 2a|}. \end{aligned}$$

Letting now $a \searrow 0$ we find that the functions converge in H^2, while the norm of their inverses blows up.

The subgroup of volume-preserving H^s diffeomorphisms

$$\mathfrak{D}^s_\mu(M) = \left\{\eta \in \mathfrak{D}(M) : \eta^*\mu = \mu\right\}$$

is a closed C^∞ submanifold of $\mathfrak{D}^s(M)$. This follows directly from the implicit function theorem for Banach manifolds and the Hodge decomposition.

Appendix 2: Optimal Transport and Wasserstein distance

Consider the space of densities $\mathfrak{Dens}(M)$ as the space $\mathfrak{D}(M)/\mathfrak{D}_\mu(M)$ of left cosets described in Rem. 13 above. The projection of $\mathfrak{D}(M)$ onto this quotient given by the pushforward map $\Pi(\xi) = (\xi^{-1})^*\mu = \mathrm{Jac}_\mu \xi^{-1}\mu$ defines (as in the case of right cosets) a smooth principal bundle with fibre $\mathfrak{D}_\mu(M)$.

The group $\mathfrak{D}(M)$ carries a natural L^2 metric

$$\langle u \circ \xi, v \circ \xi \rangle_{L^2} = \int_M \langle u, v \rangle \, \mathrm{Jac}_\mu \xi \, d\mu \qquad u, v \in T_e\mathfrak{D}(M),\ \xi \in \mathfrak{D}(M), \tag{52}$$

whose geometry is relatively easy to visualize: a curve $t \mapsto \xi(t)$ is a geodesic in $\mathfrak{D}(M)$ if and only if $t \mapsto \xi(t)(x)$ is a geodesic in M for each x. Observe that in general this metric is neither left- nor right-invariant. It is right-invariant when restricted to $\mathfrak{D}_\mu(M)$ and it becomes left-invariant only when restricted to the subgroup of isometries.

In order to see that Π also defines a Riemannian submersion note that the horizontal vectors with respect to (52) have the form $\nabla f \circ \xi$ for some $f : M \to \mathbb{R}$ and that the metric descends to a (weak) Riemannian metric on the base

$$\langle \alpha, \beta \rangle_\rho = \int_M \langle \nabla f, \nabla g \rangle \, \rho \, d\mu, \tag{53}$$

where f and g solve the equations $\mathrm{div}\,(\rho \nabla f) = -\alpha$ and $\mathrm{div}\,(\rho \nabla g) = -\beta$ and the mean-zero functions α and β are tangent vectors to $\mathfrak{Dens}(M)$ at ρ.

The Riemannian distance of (53) between two measures ν and λ on the space of densities $\mathfrak{Dens}(M)$ (viewed as the space of left cosets) has a very appealing interpretation as the L^2 cost of transporting one density to the other

$$\mathrm{dist}_W^2(\nu, \lambda) = \inf_\xi \int_M \mathrm{dist}_M^2(x, \xi(x)) \, dx \tag{54}$$

where the infimum is over all diffeomorphisms ξ such that $\xi^*\lambda = \nu$ and dist_M is the Riemannian distance on M. In this setting dist_W is often referred to as the L^2-*Wasserstein* or (perhaps more appropriately) the *Kantorovich-Rubinstein distance* between densities ν and λ. It is of fundamental importance in Optimal Transport problems. We refer to the papers [10], [45] or the comprehensive monograph [53] for more details.

References

1. Amari, S.: Differential geometry of curved exponential families - curvature and information loss. Ann. Stat. **10**, 357–385 (1982)
2. Amari, S.: Information Geometry and its Applications. Springer Japan (2016)
3. Amari, S., Barndorff-Nielsen, O., Kass, R., Lauritzen, L., Rao, C.: Differential Geometry in Statistical Inference. Lecture Notes, Institute of Mathematical Statistics (1987)

4. Amari, S., Nagaoka, H.: Methods of Information Geometry. Translations of Mathematical Monographs, American Mathematical Society, Providence (2000)
5. Arnold, V.: Sur la geometrie differentielle des groupes de Lie de dimension infinie et ses application a l'hydrodynamique des fluides parfaits. Ann. Inst. Fourier **16**, 319–361 (1966)
6. Arnold, V., Khesin, B.: Topological Methods in Hydrodynamics. Springer, New York (1998)
7. Ay, N., Jost, J., Le, H., Schwachhofer, L.: Information Geometry. Springer International (2017)
8. Banyaga, A.: The Structure of Classical Diffeomorphism Groups. Kluwer, Dortrecht (1997).
9. Bauer, M., Bruveris, M., Michor, P.: Uniqueness of the Fisher-Rao metric on the space of densities. Bull. Lond. Math. Soc. **48**, 499–506 (2016)
10. Benamou, J., Brenier, Y.: A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numer. Math. **84**, 375–393 (2001)
11. Bhattacharyya, A.: On a measure of divergence between two statistical populations defined by their probability distributions. Bull. Calcutta Math. Soc. **35**, 99–109 (1943)
12. Birkhoff, G.: Analytical groups. Trans. Amer. Math. Soc. **43**, 61–101 (1938)
13. Bourbaki, N.: Lie Groups and Lie Algebras. Hermann, Paris (1975)
14. Cheeger, J., Ebin, D.: Comparison Theorems in Riemannian Geometry. North Holland, New York (1975)
15. Chentsov, N.: Statistical Decision Rules and Optimal Inference. (In Russian). Nauka, Moscow (1972); english translation in Translations of Mathematical Monographs, American Mathematical Society, Providence (1982)
16. Chenstov, N., Morozova E.: Natural geometry of families of probabilistic laws. (In russian). Current Problems of Mathematics, Fundamental Directions, Akad, Nauk SSSR **83**, 133–265 (1991)
17. David, A.: Further comments on some comments on a paper by Bradley Efron. Ann. Stat. **6**, 1249 (1977)
18. Dieudonne, J.: Foundations of Modern Analysis. Treatise on Analysis. Academic Press, New York (1969)
19. Dunford, N., Schwartz, J.: Linear Operators I; II. Interscience Publishers, New York (1958-1963)
20. Ebin, D., Marsden, J.: Groups of diffeomorphisms and the motion of an incompressible fluid. Ann. Math. **92**, 102–163 (1970)
21. Eells, J.: On the geometry of function spaces. Symp. Int. de Topologia Algebraica, Mexico City, 303–307 (1958)
22. Eells, J.: A setting for global analysis. Bull. Amer. Math. Soc. **72**, 751–807 (1966)
23. Efron, B.: Defining the curvature of a statistical problem (with applications to second order efficiency). Ann. Stat. **3**, 1189–1242 (1975)
24. Friedrich, T.: The Fisher Information and symplectic structures. (In german). Math. Nachr. **153**, 273–296 (1991)
25. Gibilisco, P., Pistone, G.: Connections on non-parametric statistical models by Orlicz space geometry. Infinite Dimensional Analysis, Quantum Probability and Related Topics, 325–347 (1998)
26. Grabowski, J.: Free subgroups of diffeomorphism groups. Fund. Math. **313**, 103–121 (1988)
27. Hamilton, R.: The inverse function theorem of Nash and Moser. Bull. Amer. Math. Soc. **7**, 65–222 (1982)
28. Hormander, L.: The boundary problems of physical geodesy. Arch. Ration. Mech. Anal. **62**, 1–52 (1976)
29. Khesin, B., Misiołek, G.: Euler equations on homogeneous spaces and Virasoro orbits. Adv. Math. **176**, 116–144 (2003)
30. Khesin, B., Lenells, J., Misiołek, G., Preston, S.: Geometry of diffeomorphism groups, complete integrability and geometric statistics. Geom. funct. anal. 23, 334–366 (2013)
31. Köthe, G.: Topological Vector Spaces I. Springer, New York (1969)
32. Kriegl, A., Michor, P.: The Convenient Setting of Global Analysis. Mathematical Surveys and Monographs, American Mathematical Society, Providence (1997)
33. Ladyzhenskaya, O.: The Mathematical Theory of Viscous Incompressible Flow. Gordon and Breach, New York (1969).

34. Lang, S.: Fundamentals of Differential Geometry. Springer, New York (1999)
35. Lenells, J.: The Hunter-Saxton equation describes the geodesic flow on a sphere. J. Geom. Phys. **57**, 2049–2064 (2007)
36. Lenells, J., Misiołek, G.: Amari-Chetnsov connections and their geodesics on homogeneous spaces of diffeomorphism groups. Zapis. Nauch. Sem. POMI **411**, 49–62 (2013).
37. Łojasiewicz, S., Zehnder, E.: An inverse function theorem in Fréchet spaces. J. Funct. Anal. **33**, 165–174 (1979)
38. Milnor, J.: Remarks on infinite-dimensional Lie groups. in "Relativite, groupes et topologie II" (Les Houches 1983), North Holland, Amsterdam (1984)
39. Modin, K.: Generalized Hunter-Saxton equations, optimal information transport and factorization of diffeomorphisms. J. Geom. Anal. **24**, 1306–1334 (2015)
40. Morrey, C.: Multiple Integrals in the Calculus of Variations. Springer, New York (1966)
41. Moser, J.: A new technique for the construction of solutions of nonlinear partial differential equations. Proc. Natl. Acad. Sci USA **47**, 1824–1831 (1961)
42. Moser, J.: On the volume elements of a manifold. Trans. Amer.Math.Soc. **120**, 286–294 (1965)
43. Nash, J.: The embedding problem for Riemannian manifolds. Ann. Math. **63**, 20–63 (1956)
44. Omori, H.: Infinite-dimensional Lie Groups. Translations of Mathematical Monographs, American Mathematical Society, Providence (1997)
45. Otto, F.: The geometry of dissipative evolution equations: the porous media equation. Comm. Partial Differential Eq. **26**, 101–174 (2001)
46. Pistone, G., Sempi, C.: An infinite-dimensional geometric structure on the space of all probability measures equivalent to a given one. Ann. Stat. **23**, 1543–1561 (1995)
47. Rao, C.R.: Information and accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc. **37**, 81–91 (1945)
48. Rudin, W.: Functional Analysis. McGraw-Hill, New York (1973)
49. Schaefer, H.H.: Topological Vector Spaces. Macmillan, New York (1966)
50. Schmid, R.: Infinite Dimensional Hamiltonian Systems. Monographs and Textbooks in Physical Science, Lecture Notes **3**, Bibliopolis, Naples (1987)
51. Sergeraert, F.: Un theoreme de fonctions implicites sur certains espaces de Fréchet et quelques applications. Ann. Scient. Ecole Norm. Sup. **5**, 599–660 (1972)
52. Steenrod, N.: The Topology of Fibre Bundles. Princeton Univeristy Press, Princeton (1951).
53. Villani, C.: Optumal Transport: Old and New. Springer, New York 2009
54. Vizman, C.: Geodesic Equations on Diffeomorphism Groups. Orizonturi Univeristare, Timisoara (2008)
55. Warner, F.: Foundations of Differentiable Manifolds and Lie Groups. Scott, Foresman & Co., Glenview Ill. (1971)