Chapter 3
Parametrized Measure Models

3.1 The Space of Probability Measures and the Fisher Metric

This section has a more informal character. It introduces the basic concepts and problems of information geometry on—typically—infinite sample spaces and thereby sets the stage for the more formal considerations in the next section. The perspective here will be somewhat different from that developed in Chap. 2, as the constructions for the probability simplex presented there did not have to grapple with the measure theoretical complications that we shall encounter here. Nevertheless, the analogy with the finite-dimensional case will guide our intuition.

Let Ω be a set with a σ-algebra $\mathfrak{B}$ of subsets;[1] for example, Ω can be a topological space and $\mathfrak{B}$ the σ-algebra of Borel sets, i.e., the σ-algebra generated by the open sets. Later on, Ω will also have to carry a differentiable structure.

For a signed measure μ on Ω, we have the total variation

$$\|\mu\|_{TV} := \sup \sum_{i=1}^{n} |\mu(A_i)|, \qquad (3.1)$$

where the supremum is taken over all finite partitions $\Omega = A_1 \uplus \cdots \uplus A_n$ with disjoint sets $A_i \in \mathfrak{B}$. If $\|\mu\|_{TV} < \infty$, the signed measure μ is called finite. We consider the Banach space $\mathcal{S}(\Omega)$ of all signed finite measures on Ω with the total variation as Banach norm. The subsets of all finite *non-negative* measures and of probability measures on Ω will be denoted by $\mathcal{M}(\Omega)$ and $\mathcal{P}(\Omega)$, respectively.

The null sets of a measure μ are those subsets A of Ω with $\mu(A) = 0$. A finite non-negative measure μ_1 dominates another finite measure μ_2 if every null set of μ_1 is also a null set of μ_2. Two finite non-negative measures are called compatible if they dominate each other, i.e., if they have the same null sets. Spaces of such measures will be the basis of our subsequent constructions, and we shall therefore

---

[1] Ω will take over the role of the finite sample space I in Sect. 2.1.

now formalize these notions. We take some σ-finite non-negative measure μ_0. Then

$$\mathcal{S}(\Omega, \mu_0) := \left\{ \mu = \phi \, \mu_0 : \phi \in L^1(\Omega, \mu_0) \right\}$$

is the space of signed measures dominated by μ_0. This space can be identified in terms of the canonical map

$$i_{can} : \mathcal{S}(\Omega, \mu_0) \to L^1(\Omega, \mu_0)$$
$$\mu \qquad \mapsto \frac{d\mu}{d\mu_0}, \qquad (3.2)$$

where the latter is the Radon–Nikodym derivative of μ w.r.t. μ_0. Note that

$$\|\mu\|_{TV} = \|\phi\|_{L^1(\Omega,\mu_0)} = \left\| \frac{d\mu}{d\mu_0} \right\|_{L^1(\Omega,\mu_0)}.$$

As we see from this description that i_{can} is a Banach space isomorphism, we refer to the topology of $\mathcal{S}(\Omega, \mu_0)$ also as the L^1-*topology*.

If μ_1, μ_2 are compatible finite non-negative measures, they are absolutely continuous with respect to each other in the sense that there exists a non-negative function ϕ that is integrable with respect to either of them, such that

$$\mu_2 = \phi\mu_1, \quad \text{or, equivalently,} \quad \mu_1 = \phi^{-1}\mu_2. \qquad (3.3)$$

As noted before, ϕ is then the Radon–Nikodym derivative of μ_2 with respect to μ_1.

Being integrable, ϕ is finite almost everywhere (with respect to both μ_1 and μ_2) on Ω, and since the situation is symmetric between ϕ and ϕ^{-1}, ϕ is also positive almost everywhere. Thus, for any finite non-negative measure μ on Ω, we let

$$\mathcal{F}_+(\Omega, \mu) := \left\{ \phi \in L^1(\Omega, \mu), \, \phi > 0 \, \mu\text{-a.e.} \right\} \qquad (3.4)$$

be the space of integrable functions on Ω that are positive almost everywhere with respect to μ. (The reason for the notation will become apparent in Sect. 3.2 below.) In fact, in later sections, we shall find it more convenient to work with the space of measures

$$\mathcal{M}_+(\Omega, \mu) := \left\{ \phi\mu : \phi \in L^1(\Omega, \mu), \, \phi > 0 \, \mu\text{-a.e.} \right\} \qquad (3.5)$$

than with the space $\mathcal{F}_+(\Omega, \mu)$ of functions. Of course, these two spaces can easily be identified, as they simply differ by multiplication with μ.

In particular, the topology of $\mathcal{S}(\Omega, \mu_0)$ is independent of the particular choice of the reference measure μ_0 within its compatibility class, because if

$$\phi \in L^1(\Omega, \mu_0) \quad \text{and} \quad \psi \in L^1(\Omega, \phi\mu_0), \quad \text{then } \psi\phi \in L^1(\Omega, \mu_0). \qquad (3.6)$$

Compatibility is an equivalence relation on the space of finite non-negative measures on Ω, and that space is therefore partitioned into equivalence classes. The set of such equivalence classes is quite large. For instance, the Dirac measure at any

point of Ω generates its own such class. More generally, in Euclidean space, we can consider Hausdorff measures of subsets of possibly different Hausdorff dimensions. In the sequel, we shall usually work within a single such compatibility class with some base measure μ_0. The basic example that one may have in mind is of course the Lebesgue measure on a Euclidean or Riemannian domain, or else the Hausdorff measure on some fixed subset.

Any other finite non-negative measure μ that is compatible with μ_0 is thus of the form

$$\mu = \phi\mu_0 \quad \text{for some} \quad \phi \in \mathcal{F}_+(\Omega, \mu_0), \tag{3.7}$$

and therefore, by (3.3),

$$\mu_0 = \phi^{-1}\mu \quad \text{with} \quad \phi^{-1} \in \mathcal{F}_+(\Omega, \mu). \tag{3.8}$$

This yields the identifications

$$\begin{aligned}
\mathcal{F}_+(\Omega, \mu) &= \mathcal{F}_+(\Omega, \mu_0) = : \mathcal{F}_+, \\
\mathcal{M}_+(\Omega, \mu) &= \mathcal{M}_+(\Omega, \mu_0) = : \mathcal{M}_+,
\end{aligned} \tag{3.9}$$

where we use $\mathcal{F}_+$ and $\mathcal{M}_+$ if there is no ambiguity over which base measure μ_0 is used. Moreover, if

$$\mu_1 = \phi\mu_0, \quad \mu_2 = \psi\mu_1, \quad \text{with} \quad \phi \in \mathcal{F}_+(\Omega, \mu_0), \ \psi \in \mathcal{F}_+(\Omega, \mu_1)$$

then by (3.6)

$$\mu_2 = \psi\phi\mu_0 \quad \text{with} \quad \psi\phi \in \mathcal{F}_+(\Omega, \mu_0).$$

$\mathcal{F}_+(\Omega, \mu)$, however, is not a group under pointwise multiplication since for $\phi_1, \phi_2 \in L^1(\Omega, \mu_0)$, their product $\phi_1\phi_2$ need not be in $L^1(\Omega, \mu_0)$.

The question arises, however, whether $\mathcal{F}_+$ possesses an—ideally dense—subset that is a group, perhaps even a Lie group. That is, to what extent can we linearize the partial multiplicative group structure via an exponential map, in the same manner as $z \mapsto e^z$ maps the additive group $(\mathbb{R}, +)$ to the multiplicative group $(\mathbb{R}^+, \cdot)$? Formally, we might be inclined to identify the tangent space $T_\mu\mathcal{F}_+$ of $\mathcal{F}_+$ at any compatible μ with

$$B_\mu := \big\{f : \Omega \to \mathbb{R} \cup \{\pm\infty\}, \ e^{\pm f} \in L^1(\Omega, \mu)\big\}. \tag{3.10}$$

We should point out that here we attempt to revert the construction of Sects. 2.1, 2.2. However, we immediately run into the difficulty that the set B_μ is in general not a vector space, but only a convex subset of $L^1(\Omega, \mu)$. Thus, a reasonable definition of $T_\mu\mathcal{F}_+$ would be to define it as the vector space generated by B_μ.

We now switch to working with the space $\mathcal{M}_+$ of measures instead of the space $\mathcal{F}_+$ of functions, where, as mentioned, $\phi \in \mathcal{F}_+$ corresponds to $\phi\mu \in \mathcal{M}_+$. One approach to provide some structure on $\mathcal{M}_+$ has been pursued by Pistone and Sempi

who constructed a Banach norm on $T_\mu \mathcal{M}_+$ such that B_μ becomes a convex open subset. The Banach norm is—up to equivalence—independent of the choice of μ so that in this way $\mathcal{M}_+$ becomes a Banach manifold. The topology which is imposed on $\mathcal{M}_+$ in this way is called the e-topology. We shall describe this in detail in Sect. 3.3.

Let us return to our discussion of the analogy between the general situation and the construction of Sects. 2.1, 2.2. There the space of functions was the cotangent space of a measure m in the space of measures, and the tangent and the cotangent space were then identified through a scalar product (and the scalar product chosen then was the Fisher metric). Here, we take the duality between functions and measures

$$(\phi, m) \mapsto \int \phi \, dm$$

as a starting point and vary a measure via $m \mapsto Fm$ for some non-negative function F. This duality will then induce the scalar product

$$\langle f, F \rangle_m = \int f F \, dm$$

which we shall then use to define the Fisher metric.

The construction is tied together by the exponential map

$$\exp \colon T_\mu \mathcal{F}_+ \supseteq B_\mu \to \mathcal{F}_+$$
$$f \to e^f \tag{3.11}$$

that converts arbitrary functions into non-negative ones. In other words, we apply the exponential function $z \mapsto e^z$ to each value $f(x)$ of the function f.

Here, in fact, the underlying structure is even an affine one, in the following sense. When we take a measure $\mu' = \phi\mu$ in place of μ, then an exponential image $e^f \mu$ from $T_\mu \mathcal{F}_+$ is also an exponential image $e^g \mu' = e^g \phi\mu$ from $T_{\mu'} \mathcal{F}_+$, and the relationship is

$$f = g + \log \phi. \tag{3.12}$$

Thus, f and g are related by adding the function $\log \phi$ which is independent of both of them.

Also, of course,

$$\log \colon \mathcal{M}_+ \to B_\mu \subseteq T_\mu \mathcal{M}_+$$
$$\phi\mu \to \log \phi \tag{3.13}$$

is the inverse of the exponential map.[2]

---

[2]For reasons of integrability, this structure need not define an affine space in the sense of Sect. 2.8.1. We only have the structure of an affine manifold, in the sense of possessing affine coordinate changes. This issue will be clarified in Sect. 3.2 below.

There is also another, likewise ultimately futile, approach to a putative tangent space $T_\mu \mathcal{M}_+$ of $\mathcal{M}_+$ which is based on a duality relation between $\mathcal{M}_+$ and those f that are not only integrable (L^1) w.r.t. μ, but even of class $L^\infty(\Omega, \mu)$. This duality is given by

$$(f, \phi\mu) := \int f\phi \, d\mu, \qquad (3.14)$$

which exists since $\phi \in L^1(\Omega, \mu)$.

On this space, $T_\mu^\infty \mathcal{M}_+$, we then have a natural scalar product, namely the L^2-product

$$\langle f, g \rangle_\mu := \int fg \, d\mu \quad \text{for } f, g \in T_\mu^\infty \mathcal{M}_+. \qquad (3.15)$$

However, the completion of $T_\mu^\infty \mathcal{M}_+$ with respect to the norm $\|\cdot\|$ induced by $\langle \cdot, \cdot \rangle_\mu$ is the Hilbert space $T_\mu^2 \mathcal{M}_+$ of functions f of class L^2 with respect to μ, which is larger than those for which $e^{\pm f}$ are of class L^1.

Thus, by this construction, we do not quite succeed in making $\mathcal{M}_+$ into a Lie group. Nevertheless, (3.15) yields a natural Riemannian metric on $\mathcal{M}_+$ that will play an important role in the sequel. A Riemannian structure on a differentiable manifold X assigns to each tangent space $T_p X$ a scalar product, and this product has to depend differentiably on the base point. At this point, this is only formal, however. When Ω is not finite, spaces of measures or functions are infinite-dimensional in general because the value at every point is a degree of freedom. Only when Ω is a finite set do we obtain a finite-dimensional measure or function space. Thus, we need to deal with infinite-dimensional manifolds, e.g., compatibility classes of measures. In Appendix C, we discuss the appropriate concept, that of a Banach manifold. As mentioned, in the present discussion, we only have a weak such structure, however, as our space is not complete w.r.t. the L^2-structure of the metric. In other words, the topology induced by the metrics on the tangent spaces is weaker than the Banach space topology that we are working with. For the moment, we shall therefore proceed in a formal way. Below, in Sect. 3.2, we shall provide a rigorous construction that avoids these issues.

The natural identification between the spaces $T_\mu^2 \mathcal{M}_+$ and $T_{\mu'}^2 \mathcal{M}_+$ (which for the moment will take the role of tangent spaces—and we shall therefore omit the superscript 2) with $\mu' = \phi\mu$ is given by

$$f \to \frac{1}{\sqrt{\phi}} f; \qquad (3.16)$$

we then have

$$\left\langle \frac{1}{\sqrt{\phi}} f, \frac{1}{\sqrt{\phi}} g \right\rangle_{\mu'} = \int \frac{1}{\sqrt{\phi}} f \cdot \frac{1}{\sqrt{\phi}} g\phi \, d\mu = \int fg \, d\mu = \langle f, g \rangle_\mu.$$

In particular, if Ω is a differentiable manifold and

$$\kappa : \Omega \to \Omega$$

is a diffeomorphism, then μ is transformed into $\kappa_* \mu$, called the *push-forward of μ under κ*, with

$$\kappa_* \mu(V) := \mu\left(\kappa^{-1} V\right) \quad \text{for all } V \in \mathfrak{B}. \tag{3.17}$$

Since $\int_V \frac{1}{|\det d\kappa(x)|} \mu(d(\kappa(x))) = \int_{\kappa^{-1} V} \mu(dx)$ by the transformation formula, we have

$$\kappa_* \mu\left(\kappa(x)\right) \left|\det d\kappa(x)\right| = \mu(y) \quad \text{for } y = \kappa(x).$$

We thus have

$$\langle f, g \rangle_{\kappa_* \mu} = \left\langle \kappa^* f, \kappa^* g \right\rangle_\mu \tag{3.18}$$

with

$$\kappa^* f(x) = f\left(\kappa(x)\right).$$

In other words, if we employ this transformation rule for tangent vectors, then the diffeomorphism group of Ω acts by isometries on $\mathcal{M}_+$ with respect to the metric given by (3.15). Thus, our metric is invariant under diffeomorphisms of Ω. One might then wish to consider the quotient of $\mathcal{M}_+$ by the action of the diffeomorphism group $\mathcal{D}(\Omega)$. Of course, if Ω is finite, then $\mathcal{D}(\Omega)$ is simply the group of permutations of the elements of Ω. This group has a fixed point, namely the probability measure with

$$p(x_i) = \frac{1}{n} \quad \text{for } i = 1, \ldots, n$$

(assuming Ω consists of the n elements $x_1, \ldots, x_n$), and its scalar multiples. We therefore expect that the quotient $\mathcal{M}_+/\mathcal{D}(\Omega)$ will have singularities.

In the infinite case, the situation is somewhat different (although we still get singularities). Namely, if Ω is a compact oriented differentiable manifold, then a theorem of Moser [190] says that any two probability measures μ, ν that are volume forms, i.e., are smooth and positive on all open sets, (in local coordinates $(x^1, \ldots, x^n)$ on Ω, they are thus smooth positive multiples of $dx^1 \wedge \cdots \wedge dx^n$) are related by a diffeomorphism κ,

$$\nu = \kappa_* \mu,$$

or equivalently,

$$\mu = \left(\kappa^{-1}\right)_* \nu =: \kappa^* \nu.$$

Thus, the diffeomorphism group acts transitively on the space of volume forms of total measure 1, and the quotient by the diffeomorphism group of this space is therefore a single point.

We recall that a finite non-negative measure μ_1 dominates another finite non-negative measure μ_2 if every null set for μ_1 also is a null set for μ_2. Of course, two mutually dominant finite non-negative measures are compatible.

In the finite case, i.e., for $\Omega = \{x_1, \ldots, x_n\}$, μ_1 then dominates μ_2 if

$$\mu_1(x_i) = 0 \quad \text{implies } \mu_2(x_i) = 0 \quad \text{for any } i = 1, \ldots, n.$$

In particular, a measure μ with

$$\mu(x_i) > 0 \quad \text{for all } i$$

dominates every other measure on Ω.

This is different in the infinite case; for example, if Ω is a compact oriented differentiable manifold, then a volume form which is positive on every open set no longer dominates a Dirac measure supported at some point in Ω.

In information geometry, one wishes to study only probability measures, that is, measures that satisfy the normalization

$$\mu(\Omega) = \int_\Omega d\mu = 1. \tag{3.19}$$

It might seem straightforward to simply impose this condition upon measures and then study the space of those measures satisfying it as a subspace of the space of all measures. A somewhat different point of view, however, emerges from the following consideration. The normalization (3.19) can simply be achieved by rescaling a given measure μ, that is, by multiplying it by some appropriate $\lambda \in \mathbb{R}$. λ is simply obtained as $\mu(\Omega)^{-1}$. The freedom of rescaling a measure now expresses that we are not interested in absolute "sizes" $\mu(A)$ of subsets of Ω, but rather only in relative ones, like $\frac{\mu(A)}{\mu(\Omega)}$ or $\frac{\mu(A_1)}{\mu(A_2)}$. Therefore, we identify the space $\mathcal{P}(\Omega)$ of probability measures on Ω as the projective space

$$\mathbb{P}^1 \mathcal{M}(\Omega),$$

i.e., the space of all equivalence classes in $\mathcal{M}(\Omega)$ under multiplication by positive real numbers. Of course, elements of $\mathcal{P}(\Omega)$ can be considered as measures satisfying (3.19), but more appropriately as equivalence classes of measures giving the same relative sizes of subsets of Ω.

Our above metric then also induces a metric on $\mathcal{P}(\Omega)$ as a quotient of $\mathcal{M}(\Omega)$ which is different from the one obtained by identifying $\mathbb{P}^1 \mathcal{M}(\Omega)$ with the subspace of $\mathcal{M}(\Omega)$ consisting of metrics satisfying (3.19). Let us recall from Proposition 2.1 in Sect. 2.2 the case where Ω is finite, $\Omega = \{1, \ldots, n\}$. In that case, the probability measures on Ω are given by

$$\Sigma^{n-1} := \left\{ (p_1, \ldots, p_n) : p_i \geq 0 \text{ for } i = 1, \ldots, n, \text{ and } \sum_{i=1}^n p_i = 1 \right\}.$$

These form an $(n-1)$-dimensional simplex in the positive cone $\mathbb{R}_+^n$ of $\mathbb{R}^n$. The projective space

$$\mathbb{P}^1 \mathbb{R}_+^n,$$

however, is naturally identified with the corresponding spherical sector

$$S_+^{n-1} := \left\{ (z_1, \ldots, z_n) : z_i \geq 0 \text{ for } i = 1, \ldots, n, \ \sum_{i=1}^{n} z_i^2 = 1 \right\}.$$

There is a natural bijection

$$\Sigma^{n-1} \to S_+^{n-1}$$
$$(p_1, \ldots, p_n) \mapsto (\sqrt{p_1}, \ldots, \sqrt{p_n}). \tag{3.20}$$

Let us now try to carry this over to the infinite-dimensional case, under the assumption that Ω is a differentiable manifold of dimension n. In that case, we can consider each (Radon) measure μ as a density. This means that for each $x \in \Omega$, if we consider the space $Gl(n, \mathbb{R})$ as the space of all bases of $T_x \Omega$ (again, the identification is not canonical as we need to select one basis $V = (v^1, \ldots, v^n)$ that is identified with $\text{id} \in Gl(n, \mathbb{R})$),[3]

$$\mu(x)(XV) = |\det X| \mu(x)(V).$$

Likewise, we call ρ a half-density, if we have

$$\rho(x)(XV) = |\det X|^{1/2} \rho(x)(V)$$

for all $X \in Gl(n, \mathbb{R})$ and bases V. Below, in Definition 3.54, we shall give a precise definition of the space of half-densities (in fact, of any rth power of a measure with $0 < r \leq 1$).

In this interpretation, our above L^2-product on $\mathcal{M}_+(\Omega)$ becomes an L^2-product on the space of half-densities

$$\langle \rho, \sigma \rangle := \int_\Omega \rho\sigma,$$

where we no longer need a base measure μ. And the diffeomorphism group of Ω then acts by isometries on the Hilbert space of half-densities of class L^2.

If we now have a probability measure μ, then its square root $\sqrt{\mu}$ is a half-density that is contained in the unit sphere of that Hilbert space. Conversely, up to the issue of regularity, the part of that unit sphere that corresponds to non-negative half-densities can be identified with the probability measures on Ω. As mentioned, it carries a metric that is invariant under the action of the diffeomorphism group of Ω, i.e., under relabeling of the points of Ω.

---

[3] Again, we are employing a fundamental mathematical principle here: Instead of considering objects in isolation, we rather focus on the transformations between them. Thus, instead of an individual basis, we consider the transformation that generates it from some (arbitrarily chosen) standard basis. This automatically gives a very powerful structure, that of a group (of transformations).

The duality relation (3.14) for a probability measure $\mu' = \phi\mu$ then becomes

$$(f, \mu') = \mathbb{E}_{\mu'}(f), \tag{3.21}$$

the expectation value of f w.r.t. the probability measure μ'. This is a linear operation on the space of functions f and an affine operation on the space of probability measures μ'.

Let us clarify the relation with our previous construction of the metric: If f is a tangent vector to a probability measure μ, then it generates the curve

$$e^{tf}\mu, \quad t \in \mathbb{R},$$

through μ. By taking the square root as before, we obtain the curve

$$e^{\frac{1}{2}tf}\sqrt{\mu}, \quad t \in \mathbb{R},$$

in the space of half-densities that has the expansion in t

$$\sqrt{\mu} + \frac{1}{2}tf\sqrt{\mu} + O(t^2).$$

Thus, the tangent vector that corresponds to f in the space of half-densities is

$$\frac{1}{2}f\sqrt{\mu}.$$

The inner product of two such tangent vectors is

$$\int \frac{1}{2}f\sqrt{\mu} \cdot \frac{1}{2}g\sqrt{\mu} = \frac{1}{4}\int fg\mu.$$

Thus, up to the inessential factor $\frac{1}{4}$, we regain our original Riemannian metric on $\mathcal{M}_+(\Omega)$. In order to eliminate that factor, in Proposition 2.1, we had used the sphere with radius 2 instead of that of radius 1. Therefore, we should also modify (3.20) in the same manner if we want to have an isometry.

Let us now translate this into the usual statistical interpretation: We have a family $p(x; s)$ of probability measures depending on a parameter s, $-\varepsilon < s < \varepsilon$. Then the squared norm of the tangent vector to this family at $s = 0$ is (up to some factor 4)

$$4\int \frac{d}{ds}\sqrt{p(x; s)} \frac{d}{ds}\sqrt{p(x; s)}|_{s=0}\, dx$$

$$= \int \frac{d}{ds}\log p(x; s)\frac{d}{ds}\log p(x; s)p(x; 0)|_{s=0}\, dx$$

$$= \mathbb{E}_p\left(\left(\frac{d}{ds}\log p(x; s)\right)^2\Big|_{s=0}\right), \tag{3.22}$$

where $\mathbb{E}_p$ denotes the expectation with respect to the probability measure $p = p(\cdot; 0)$. By polarization, if $s = (s_1, \ldots, s_n)$ is now n-dimensional, we obtain the Fisher information metric

$$\mathbb{E}_p\left(\frac{\partial}{\partial s_\mu} \log p(x; s)\frac{\partial}{\partial s_\nu} \log p(x; s)\right). \tag{3.23}$$

We can also rewrite the above formula to obtain

$$\mathbb{E}_p\left(\frac{\partial}{\partial s_\mu} \log p(x; s)\frac{\partial}{\partial s_\nu} \log p(x; s)\right)$$

$$= \int \frac{\partial}{\partial s_\mu} \log p(x; s)\frac{\partial}{\partial s_\nu} \log p(x; s)p(x; 0)\, dx$$

$$= -\int \frac{\partial^2}{\partial s_\mu \partial s_\nu} \log p(x; s)p(x; 0)\, dx \tag{3.24}$$

since $\int \frac{\partial}{\partial s_\mu} \log p(x; s)p(x; 0)\, dx = \frac{\partial}{\partial s_\mu} \int \log p(x; s)p(x; 0)\, dx = \frac{\partial}{\partial s_\mu} \int p(x; s)\, dx = \frac{\partial}{\partial s_\mu} 1 = 0$, which implies

$$0 = \frac{\partial}{\partial s_\nu} \int \frac{\partial}{\partial s_\mu} \log p(x; s)\ p(x; s)\, dx = \int \frac{\partial^2}{\partial s_\mu \partial s_\nu} \log p(x; s)\ p(x; s)\, dx$$

$$+ \int \frac{\partial}{\partial s_\mu} \log p(x; s)\ \frac{\partial}{\partial s_\nu} \log p(x; s)\ p(x; s)\, dx.$$

This step also admits the following interpretation:

$$\frac{\partial}{\partial s_\nu} \log p(x; s) \tag{3.25}$$

which is called the *score* of the family with respect to the parameter s_ν. Our above computation then gives

$$\mathbb{E}_p\left(\frac{\partial}{\partial s_\nu} \log p(x; s)\right) = 0, \tag{3.26}$$

that is, the expectation value of the score vanishes. (This expresses the fact that the cross-entropy

$$-\int p(x) \log q(x)\, dx \tag{3.27}$$

is minimal w.r.t. q precisely for $q = p$.)

The Fisher metric (3.22) then expresses the covariance matrix of the score.

Returning to (3.24), (3.26) yields the formula

$$\mathbb{E}_p\left(\frac{\partial}{\partial s_\mu}\log p(x;s)\frac{\partial}{\partial s_\nu}\log p(x;s)\right)$$

$$= -\mathbb{E}_p\left(\frac{\partial^2}{\partial s_\mu \partial s_\nu}\log p(x;s)\right) \qquad (3.28)$$

as another representation of the Fisher metric (3.23).

We can also write our metric as

$$\int \frac{1}{p(x;0)}\frac{\partial}{\partial s_\mu}p(x;s)\frac{\partial}{\partial s_\nu}p(x;s)\,dx.$$

In the finite case, this becomes

$$\sum_{i=1}^n \frac{1}{p_i}\frac{\partial}{\partial s_\mu}p_i\frac{\partial}{\partial s_\nu}p_i.$$

Remark 3.1 This metric is called the Shashahani metric in mathematical biology, see Sect. 6.2.1.

As verified in Proposition 2.1, this is simply the metric obtained on the simplex Σ^{n-1} when identifying it with the spherical sector $S_{2,+}^{n-1}$ via the map $4p = q^2$, $q \in S_{2,+}^{n-1}$. If the second derivatives $\frac{\partial^2}{\partial s_\mu \partial s_\nu}p$ vanish, i.e., if $p(x;s)$ is linear in s, then

$$\sum_{i=1}^n \frac{1}{p_i}\frac{\partial}{\partial s_\mu}p_i\frac{\partial}{\partial s_\nu}p_i = \frac{\partial^2}{\partial s_\mu \partial s_\nu}\sum_{i=1}^n p_i\log p_i.$$

As will be discussed below, this means that the negative of the entropy is a potential for the metric. This will be applied in Theorem 6.4.

The Fisher metric then induces a metric on any smooth family of probability measures on Ω. To understand the Fisher metric, it is often useful to write a probability distribution in exponential form,

$$p(x;s) = \exp\big(-H(x,s)\big), \qquad (3.29)$$

where the normalization required for $\int p(x;s)\,dx = 1$ is supposed to be contained in H. The Fisher metric is then simply given by

$$\mathbb{E}_p\left(\frac{\partial}{\partial s_\mu}\log p(x;s)\frac{\partial}{\partial s_\nu}\log p(x;s)\right) = \mathbb{E}_p\left(\frac{\partial H}{\partial s_\mu}\frac{\partial H}{\partial s_\nu}\right) = \int \frac{\partial H}{\partial s_\mu}\frac{\partial H}{\partial s_\nu}p(x;s)\,dx.$$
$$(3.30)$$

Particularly important in this regard are the so-called exponential families (cf. Definition 2.10).

Definition 3.1 An *exponential family* is a family of probability distributions of the form

$$p(x; \vartheta) = \exp\left(\gamma(x) + \sum_{i=1}^{n} f_i(x)\vartheta^i - \psi(\vartheta)\right)\mu(x), \tag{3.31}$$

where $\vartheta = (\vartheta^1, \ldots, \vartheta^n)$ is an n-dimensional parameter, $\gamma(x)$ and $f_1(x), \ldots, f_n(x)$ are functions and $\mu(x)$ is a measure on Ω.

(Of course, γ could be absorbed into μ, but this would be inconvenient for our subsequent discussion of examples.) The function ψ simply serves to guarantee the normalization

$$\int_{\Omega} p(x; \vartheta) = 1;$$

namely

$$\psi(\vartheta) = \log \int \exp\left(\gamma(x) + \sum f_i(x)\vartheta^i\right)\mu(dx).$$

Here, the family is defined only for those ϑ for which

$$\int \exp\left(\gamma(x) + \sum f_i(x)\vartheta^i\right)\mu(dx) < \infty.$$

The set of those ϑ for which this is satisfied is convex, but can otherwise be quite complicated.

Exponential families will yield important examples of the parametrized measure models introduced in Sect. 3.2.4.

Example 3.1 The normal distribution $\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$ on $\mathbb{R}$, with Lebesgue measure dx, with parameters μ, σ can easily be written in this form by putting

$$\gamma(x) = 0, \quad f_1(x) = x, \quad f_2(x) = x^2, \quad \vartheta^1 = \frac{\mu}{\sigma^2}, \quad \vartheta^2 = -\frac{1}{2\sigma^2},$$

$$\psi(\vartheta) = \frac{\mu^2}{2\sigma^2} + \log\sqrt{2\pi}\sigma = -\frac{(\vartheta^1)^2}{4\vartheta^2} + \frac{1}{2}\log\left(-\frac{\pi}{\vartheta^2}\right),$$

and analogously for multivariate normal distributions $\mathcal{N}(y, \Lambda)$, i.e., Gaussian distributions on $\mathbb{R}^n$. See [169] for a systematic analysis.

For an exponential family, we have

$$\frac{\partial}{\partial\vartheta^i} \log p(x; \vartheta) = f_i(x) - \frac{\partial}{\partial\vartheta^i}\psi(\vartheta) \tag{3.32}$$

and

$$\frac{\partial^2}{\partial \vartheta^i \partial \vartheta^j} \log p(x; \vartheta) = -\frac{\partial^2}{\partial \vartheta^i \partial \vartheta^j} \psi(\vartheta). \tag{3.33}$$

This expression no longer depends on x, but only on the parameter θ. Therefore, the Fisher metric on such a family is given by

$$
\begin{aligned}
g_{ij}(p) &= -\mathbb{E}_p\left(\frac{\partial^2}{\partial \vartheta^i \partial \vartheta^j} \log p(x; \vartheta)\right) \\
&= \int \frac{\partial^2}{\partial \vartheta^i \partial \vartheta^j} \psi(\vartheta)\, p(x; \vartheta)\, dx \\
&= \frac{\partial^2}{\partial \vartheta^i \partial \vartheta^j} \psi(\vartheta) \quad \text{since } \int p(x; \vartheta)\, dx = 1. \tag{3.34}
\end{aligned}
$$

For the normal distribution, we compute the metric in terms of ϑ^1 and ϑ^2, using (3.33) and transform the result to the variables μ and σ with (B.16) and obtain at $\mu = 0$

$$g\left(\frac{\partial}{\partial \mu}, \frac{\partial}{\partial \mu}\right) = \frac{1}{\sigma^2}, \tag{3.35}$$

$$g\left(\frac{\partial}{\partial \mu}, \frac{\partial}{\partial \sigma}\right) = 0, \tag{3.36}$$

$$g\left(\frac{\partial}{\partial \sigma}, \frac{\partial}{\partial \sigma}\right) = \frac{2}{\sigma^2}. \tag{3.37}$$

As the Fisher metric is invariant under diffeomorphisms of $\Omega = \mathbb{R}$, and since $x \to x - \mu$ is such a diffeomorphism, it suffices to perform the computation at $\mu = 0$. The metric computed there, however, up to a simple scaling is the hyperbolic metric of the half-plane

$$H := \big\{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\big\},$$

and so, the Fisher metric on the family of normal distributions is the hyperbolic metric.

Let us summarize some points that will be important for the sequel. We have constructed the Fisher metric as the natural Riemannian metric in the space of lines of (finite non-negative) measures, i.e., on a projective space over a linear space. In the finite case, this projective space is simply a spherical sector. In particular, our metric is then the standard metric on the sphere, and it therefore has sectional curvature $\kappa \equiv 1$ (or, more precisely, $\frac{1}{4}$ if we utilize the sphere of radius 2, to get the normalizations right). This, in fact, carries over to the general case (see [99] for an explicit computation). Therefore, the Fisher metric is not Euclidean. By way of contrast, our space of probability measures can be viewed as a linear space in two different manners. On the one hand, as in the finite case, it can be represented as

a simplex in a vector space. Thus, any probability measure can be represented as a convex linear combination of certain extremal measures. More precisely, when Ω is a metrizable topological space, then the map $\Omega \to \mathcal{P}(\Omega)$, $x \mapsto \delta(x)$ that assigns to every $x \in \Omega$ the delta measure supported at x is an embedding. If Ω is also separable, then the image is a closed subspace of $\mathcal{P}(\Omega)$. Also, in this case, this image contains precisely the extreme points of the convex set $\mathcal{P}(\Omega)$. See [5], Sect. 15.2, for details.

We shall call this representation as a convex linear combination of extremal measures a *mixture representation*. On the other hand, our space of probability measures can be represented as the exponential image of a linear tangent space. This gives the so-called exponential representation. We shall see below that these two linear structures are dual to each other, in the sense that each of them is the underlying affine structure for some connection, and the two corresponding connections are dual with respect to the Fisher metric. Of course, neither of these connections can be the Levi-Civita connection of the Fisher metric as the latter does not have vanishing curvature.

The Fisher metric also allows the following construction: If Σ is a set with a measure $\sigma(u)$ on it, and if we have a mapping

$$h : \Sigma \to \mathcal{P}(\Omega),$$

i.e., if we have a family of measures on Ω parametrized by $u \in \Sigma$, we may then consider variations

$$h(u; s) : \Sigma \times (-\varepsilon, \varepsilon) \to \mathcal{P}(\Omega),$$

e.g.,

$$h(u; s) = \exp_{h(u;0)} s\varphi(u)$$

for some function φ.

If we have two such variations φ^1, φ^2, we can use the Fisher metric and the measure σ to form their L^2-product

$$\int_{\Sigma} \mathbb{E}_{h(u)}(\varphi^1, \varphi^2)\sigma(du)$$

$$= \int_{\Sigma} \left(\int_{\Omega} \varphi^1(u)(x)\, \varphi^2(u)(x)\, h(u)(dx) \right) \sigma(du).$$

In other words, we integrate the Fisher product with respect to the measure on our family of measures. The Fisher metric can formally be considered as a special case of this construction, namely when Σ is a singleton. The general construction allows us to average over a family of measures. For example, if we have a Markov process with transition probability $p(\cdot|y)$ for each $y \in \Omega$, we consider this as a family $h : \Omega \to \mathcal{P}(\Omega)$, with $h(y) = p(\cdot|y)$, and we may average the Fisher products taken with respect to the measures $p(\cdot|y)$ with respect to some initial probability distribution $p_0(y)$ for y. Thus, if we have families of such transition probabilities $p(\cdot|y; s)$,

we get the averaged Fisher product

$$g_{ij}(s) := \int \left(\int \frac{\partial}{\partial s^i} \log p(\cdot|y; s) \frac{\partial}{\partial s^j} \log p(\cdot|y; s) \, dx \right) p_0(dy). \qquad (3.38)$$

Under sufficient conditions, this metric is usually obtained from the distributions

$$p^{(n)}(x_0, x_1, \ldots, x_n; s) := p_0(x_0; s) p(x_1|x_0; s) \cdots p(x_n|x_{n-1}; s), \quad n = 0, 1, 2, \ldots$$

Denoting the Fisher metric for $s \mapsto p^{(n)}(\cdot; s)$ by $g_{ij}^{(n)}$, we can consider the limit

$$\overline{g}_{ij}(s) := \lim_{n \to \infty} \frac{1}{n} g_{ij}^{(n)}(s), \qquad (3.39)$$

if it exists. A sufficient condition for the existence of this limit is given by the stationarity of the process. In that case, the metric $\overline{\mathfrak{g}}$ defined by (3.39) reduces to the metric $\mathfrak{g}$ defined by (3.38). We shall take up this issue in Sect. 6.1.

3.2 Parametrized Measure Models

In this section, we shall give the formal definitions of parametrized measure models, providing solutions to some of the issues described in Sect. 3.1, improving upon [26]. First of all, the intuitive definition of a statistical model is to regard it as a family $\mathbf{p}(\xi)_{\xi \in M}$ of probability measures on some sample space Ω which varies in a differentiable fashion with $\xi \in M$. To make this formal, we need to provide some kind of differentiable structure on the space $\mathcal{P}(\Omega)$ of probability measures. This is done by noting that $\mathcal{P}(\Omega)$ is contained in the Banach space $\mathcal{S}(\Omega)$ of finite signed measures on Ω, provided with the Banach norm $\|\cdot\|_{TV}$ of total variation from (3.1). Therefore, we shall regard a statistical model as a C^1-map between Banach manifolds $\mathbf{p}: M \to \mathcal{S}(\Omega)$, as described in Appendix C, whose image is contained in $\mathcal{P}(\Omega)$.

Since $\mathcal{P}(\Omega)$ may also be regarded as the projectivization of the space of finite measures $\mathcal{M}(\Omega)$ via rescaling, any C^1-map $\mathbf{p}: M \to \mathcal{M}(\Omega)$ induces a statistical model $\mathbf{p}_0: M \to \mathcal{P}(\Omega)$ by

$$\mathbf{p}_0(\xi) = \frac{\mathbf{p}(\xi)}{\|\mathbf{p}(\xi)\|_{TV}}. \qquad (3.40)$$

It is often more convenient to study C^1-maps $\mathbf{p}: M \to \mathcal{M}(\Omega)$, called *parametrized measure models*, and then use (3.40) to obtain a statistical model $\mathbf{p}_0$. In case we consider C^1-maps $\mathbf{p}: M \to \mathcal{S}(\Omega)$, also allowing for non-positive measures, we call it a *signed parametrized measure model*.

Let us assume for simplicity that all measures $\mathbf{p}(\xi)$ are dominated by some fixed measure μ_0, even though later we shall show that this assumption is inessential. As

it turns out, a dominating measure μ_0 exists if M contains a countable dense subset which is the case, e.g., if M is a finite-dimensional manifold. In this case,

$$\mathbf{p}(\xi) = p(\cdot; \xi)\mu_0,$$

where $p : \Omega \times M \to \mathbb{R}$ is called the *density function of* $\mathbf{p}$ *w.r.t.* μ_0. Evidently, $p(\cdot; \xi) \in L^1(\Omega, \mu_0)$ for all ξ. The differential of $\mathbf{p}$ in the direction of a tangent vector $V \in T_\xi M$ is then given by

$$d_\xi \mathbf{p}(V) = \partial_V p(\cdot; \xi)\mu_0 \in L^1(\Omega, \mu_0),$$

where for simplicity we assume that the partial derivative of the density function exists, even though we shall show that this condition may be weakened as well.

Attempting to define an inner product on $T_\xi M$ analogous to the Fisher metric in (2.18), we have to regard $d_\xi \mathbf{p}(V)$ as an element of $T_{\mathbf{p}(\xi)} \mathcal{S}(\Omega) \cong L^1(\Omega, \mathbf{p}(\xi))$, which leads to

$$
\begin{aligned}
\mathfrak{g}_\xi(V, W) &= \big(d_\xi p(V), d_\xi p(W)\big)_{\mathbf{p}(\xi)} \\
&= \int_\Omega \frac{d\{d_\xi p(V)\}}{dp(\xi)} \frac{d\{d_\xi p(W)\}}{dp(\xi)} \, d\mathbf{p}(\xi) \qquad (3.41) \\
&= \int_\Omega \frac{\partial_V p(\cdot; \xi)}{p(\cdot; \xi)} \frac{\partial_W p(\cdot; \xi)}{p(\cdot; \xi)} \, d\mathbf{p}(\xi) \\
&= \int_\Omega \partial_V \log p(\cdot; \xi) \, \partial_W \log p(\cdot; \xi) \, d\mathbf{p}(\xi).
\end{aligned}
$$

We immediately encounter two problems. The first one is that—unlike in the case of a finite sample space Ω—the above integral may diverge, if we only assume that $\partial_V \log p(\cdot; \xi) \in L^1(\Omega, \mathbf{p}(\xi))$; rather, we should demand that $\partial_V \log p(\cdot; \xi) \in L^2(\Omega, \mathbf{p}(\xi))$ which in the case of an infinite sample space Ω is a proper subset.

The second problem is that the functions $\log p(\cdot; \xi)$ used in (3.41) to define the Fisher metric are not defined if we drop the assumption that $p > 0$, i.e., that all measures have the same null sets as μ_0. This is the reason why in most definitions of differentiable families of measures the equivalence of the measures $\mathbf{p}(\xi)$ in the family and hence the positivity of the density function p is required; cf., e.g., [9, 16, 25, 219]. For instance, if Ω is a finite sample space, then the description of the Fisher metric on $\mathcal{M}(\Omega)$ or on $\mathcal{P}(\Omega)$ in its canonical coordinates develops singularities outside the sets $\mathcal{M}_+(\Omega)$ and $\mathcal{P}_+(\Omega)$, respectively, cf. (2.13) and (2.19). However, if we use the coordinates $(\sqrt{p_i(\xi)})_{i \in I}$ instead, then this metric coincides—up to a constant factor—with the standard inner product on Euclidean space and hence extends to all of $\mathcal{M}(\Omega)$ and $\mathcal{P}(\Omega)$, respectively, cf. Proposition 2.1. That is, the seeming degeneracy of the Fisher metric near the boundary of $\mathcal{P}_+(\Omega, \mu_0)$ is only due to an inconvenient choice of coordinates, while with the right choice $(\sqrt{p_i(\xi)})_{i \in I}$ it becomes the Euclidean metric on the sphere. Formulating it slightly differently, the point is that not all the individual factors under the integral in (3.41) need to be well-defined, but it suffices that their product is.

Generalizing this approach, let us introduce *half-densities*, by which we mean formal square roots of measures. That is, we let

$$\sqrt{\mathbf{p}(\xi)} := \sqrt{p(\cdot;\xi)}\sqrt{\mu_0},$$

where for the moment we regard $\sqrt{\mu_0}$ merely as a formal symbol. Taking derivatives of this yields

$$d_\xi \sqrt{\mathbf{p}}(V) = \frac{1}{2}\frac{\partial_V p(\cdot;\xi)}{\sqrt{p(\cdot;\xi)}}\sqrt{\mu_0} = \frac{1}{2}\partial_V \log p(\cdot;\xi)\sqrt{\mathbf{p}(\xi)},$$

whence

$$d_\xi \sqrt{\mathbf{p}}(V) \cdot d_\xi \sqrt{\mathbf{p}}(W) = \frac{1}{4}\partial_V \log p(\cdot;\xi)\partial_W \log p(\cdot;\xi)p(\cdot;\xi)\mu_0,$$

and so

$$\mathfrak{g}_\xi(V,W) = 4\int_\Omega d\big(d_\xi \sqrt{\mathbf{p}}(V) \cdot d_\xi \sqrt{\mathbf{p}}(W)\big),$$

as in (3.22). Analogously, if we define $\sqrt[3]{\mathbf{p}(\xi)} := \sqrt[3]{p(\cdot;\xi)}\sqrt[3]{\mu_0}$, again regarding $\sqrt[3]{\mu_0}$ as a formal symbol, then

$$d_\xi \sqrt[3]{\mathbf{p}}(V) = \frac{1}{3}\frac{\partial_V p(\cdot;\xi)}{\sqrt[3]{p(\cdot;\xi)^2}}\sqrt[3]{\mu_0} = \frac{1}{3}\partial_V \log p(\cdot;\xi)\sqrt[3]{\mathbf{p}(\xi)},$$

so that the Amari–Chentsov tensor $\mathbf{T}$ from (2.51) can be written as

$$\mathbf{T}_\xi(V,W,U) = \int_\Omega \partial_V \log p(\cdot;\xi)\partial_W \log p(\cdot;\xi)\partial_U \log p(\cdot;\xi)\,d\mathbf{p}(\xi)$$

$$= 27\int_\Omega d\big(d_\xi \sqrt[3]{\mathbf{p}(\xi)}(V) \cdot d_\xi \sqrt[3]{\mathbf{p}(\xi)}(W) \cdot d_\xi \sqrt[3]{\mathbf{p}(\xi)}(U)\big). \quad (3.42)$$

This suggests that we should try to make formal sense of the objects $\sqrt{\mu_0}$, $\sqrt[3]{\mu_0}$ and hence of $\mathbf{p}(\xi)^{1/2} = \sqrt{\mathbf{p}(\xi)}$, $\mathbf{p}(\xi)^{1/3} = \sqrt[3]{\mathbf{p}(\xi)}$, etc. The idea of taking rth powers of a measure $\mathbf{p}(\xi)$ with $0 < r < 1$ has been introduced in a less rigorous way by Amari in [8, p. 66], where such powers are called α-*representations*.

We shall use a more formal approach and for $0 < r \leq 1$ construct the Banach spaces $\mathcal{S}^r(\Omega)$ in Sect. 3.2.3 by a certain direct limit construction, as well as the subsets

$$\mathcal{P}^r(\Omega) \subseteq \mathcal{M}^r(\Omega) \subseteq \mathcal{S}^r(\Omega).$$

The elements of these sets are called rth powers of probability measures (*finite measures, signed finite measures, respectively*), and they have the feature that one can formally take the $(1/r)$th power of them to obtain a probability measure (finite measure, signed finite measure, respectively), and raising to the $(1/r)$th power is a C^1-regular map between Banach spaces.

Once we have done this, we call a parametrized measure model *k-integrable* if the map

$$\mathbf{p}^{1/k} : M \longrightarrow \mathcal{M}^{1/k}(\Omega) \subseteq \mathcal{S}^{1/k}(\Omega), \qquad \xi \longmapsto \mathbf{p}(\xi)^{1/k} \tag{3.43}$$

is a C^1-map as defined in Appendix C. If $\mathbf{p}(\xi) = p(\cdot; \xi)\mu_0$ is dominated by μ_0, then the differential of this map is given by

$$\partial_V \mathbf{p}^{1/k} = d_\xi \mathbf{p}^{1/k}(V) := \frac{1}{k} \partial_V \log p(\cdot; \xi) \mathbf{p}^{1/k}(\xi),$$

and in order for this to be the *k*th root of a measure, we have to require that $\partial_V \log p(\cdot; \xi) \in L^k(\Omega, \mathbf{p}(\xi))$.

We call such a model *weakly k-integrable* if the map $\mathbf{p}^{1/k} : M \to \mathcal{M}^{1/k}(\Omega)$ is a weak C^1-map, cf. Appendix C for a definition. As we shall show in Theorem 3.2, the model is *k*-integrable if and only if the *k*-norm $V \mapsto \|\partial_V \mathbf{p}^{1/k}\|_k$ depends continuously on $V \in TM$, and it is weakly *k*-integrable if and only if the map $V \mapsto \partial_V \mathbf{p}^{1/k}$ is weakly continuous. Thus, our definition of *k*-integrability coincides with that given in [25, Definition 2.4].

In general, on an *n*-integrable parametrized measure model we may define the *canonical n-tensor*

$$\left(\tau_M^n\right)_\xi(V_1, \ldots, V_n) = \int_\Omega \partial_{V_1} \log p(\cdot; \xi) \cdots \partial_{V_n} \log p(\cdot; \xi) \, d\mathbf{p}(\xi), \tag{3.44}$$

which is a generalization of the Fisher metric $\mathfrak{g} = \tau_M^2$ (3.41) and the Amari–Chentsov tensor $\mathbf{T} = \tau_M^3$ (3.42). In fact, we show that τ_M^n is the pullback of a naturally defined covariant *n*-tensor L_Ω^n on $\mathcal{S}^{1/n}(\Omega)$ via the map $\mathbf{p}^{1/n} : M \to \mathcal{M}^{1/n}(\Omega) \subseteq \mathcal{S}^{1/n}(\Omega)$, where $k \geq n$. In particular, $\tau_M^n := (\mathbf{p}^{1/n})^* L_\Omega^n$ is well defined for a *k*-integrable parametrized measure model with $k \geq n$, even if p is not a positive function, in which case (3.44) has to be interpreted with care.

While for most applications it will suffice to consider statistical models which are dominated by some measure μ_0, our development of the theory will show that this is an inessential condition. Intuitively, it is plausible that the quantity $\partial_V p(\cdot; \xi)\mu_0$ which measures the change of measure w.r.t. the background measure μ_0 is not a significant quantity, but rather the rate of change of $\mathbf{p}(\xi)$ relative to the measure $\mathbf{p}(\xi)$ itself. That is, the relevant quantity to consider as a derivative is the *logarithmic derivative*

$$\partial_V \log p(\cdot; \xi) = \frac{d\{d_\xi \mathbf{p}(V)\}}{d\mathbf{p}(\xi)},$$

where the fraction stands for the Radon–Nikodym derivative. An important observation is that for any parametrized measure model, this Radon–Nikodym derivative always exists, so that the logarithmic derivative $\partial_V \log \mathbf{p}(\xi)$ may be defined also in the absence of a dominating measure μ_0. This is all that is needed to define the notions of *k*-integrability and the canonical tensors mentioned above.

3.2.1 The Structure of the Space of Measures

The aim of this section is to provide the formal set-up of parametrized measure models in order to make the discussion in the preceding section rigorous. As before, let Ω be a measurable space. We let

$$\mathcal{P}(\Omega) := \{\mu : \mu \text{ a probability measure on } \Omega\},$$

$$\mathcal{M}(\Omega) := \{\mu : \mu \text{ a finite measure on } \Omega\},$$

$$\mathcal{S}(\Omega) := \{\mu : \mu \text{ a signed finite measure on } \Omega\}, \tag{3.45}$$

$$\mathcal{S}_0(\Omega) := \left\{\mu \in \mathcal{S}(\Omega) : \int_\Omega d\mu = 0\right\}.$$

Clearly, $\mathcal{P}(\Omega) \subseteq \mathcal{M}(\Omega) \subseteq \mathcal{S}(\Omega)$, and $\mathcal{S}_0(\Omega), \mathcal{S}(\Omega)$ are real vector spaces. In fact, both $\mathcal{S}_0(\Omega)$ and $\mathcal{S}(\Omega)$ are Banach spaces whose norm is given by the total variation $\|\cdot\|_{TV}$ (3.1), whence any subset carries a canonical topology which is determined by saying that a sequence $(\nu_n)_{n \in \mathbb{N}}$ in (a subset of) $\mathcal{S}(\Omega)$ converges to ν_∞ if and only if

$$\lim_{n \to \infty} \|\nu_n - \nu_\infty\|_{TV} = 0.$$

With respect to this topology, the subsets

$$\mathcal{P}(\Omega) \subseteq \mathcal{M}(\Omega) \subseteq \mathcal{S}(\Omega)$$

are closed.

Remark 3.2 Evidently, for the applications we have in mind, we are interested mainly in statistical models. However, we can take the point of view that $\mathcal{P}(\Omega) = \mathbb{P}(\mathcal{M}(\Omega))$ is the projectivization of $\mathcal{M}(\Omega)$ via rescaling. Thus, given a parametrized measure model $(M, \Omega, \mathbf{p})$, normalization yields a statistical model $(M, \Omega, \mathbf{p}_0)$ defined by

$$\mathbf{p}_0(\xi) := \frac{\mathbf{p}(\xi)}{\|\mathbf{p}(\xi)\|_{TV}},$$

which is again a C^1-map. Indeed, the map $\mu \mapsto \|\mu\|_{TV}$ on $\mathcal{M}(\Omega)$ is a C^1-map, being the restriction of the linear (and hence differentiable) map $\mu \mapsto \int_\Omega d\mu$ on $\mathcal{S}(\Omega)$.

Observe that while $\mathcal{S}(\Omega)$ is a Banach space, the subsets $\mathcal{M}(\Omega)$ and $\mathcal{P}(\Omega)$ do not carry a canonical manifold structure.

By the *Jordan decomposition theorem*, each measure $\mu \in \mathcal{S}(\Omega)$ can be decomposed uniquely as

$$\mu = \mu_+ - \mu_- \quad \text{with } \mu_\pm \in \mathcal{M}(\Omega), \mu_+ \perp \mu_-. \tag{3.46}$$

The latter means that we have a disjoint union $\Omega = P \uplus N$ with $\mu_+(N) = \mu_-(P) = 0$. Thus, if we define

$$|\mu| := \mu_+ + \mu_- \in \mathcal{M}(\Omega),$$

then (3.46) implies

$$\left|\mu(A)\right| \leq |\mu|(A) \quad \text{for all } \mu \in \mathcal{S}(\Omega) \text{ and } A \in \Sigma, \tag{3.47}$$

so that

$$\|\mu\|_{TV} = \big\| \, |\mu| \, \big\|_{TV} = |\mu|(\Omega).$$

In particular,

$$\mathcal{P}(\Omega) = \big\{ \mu \in \mathcal{M}(\Omega) : \|\mu\|_{TV} = 1 \big\}.$$

Moreover, fixing a measure $\mu_0 \in \mathcal{M}(\Omega)$, we define

$$\begin{aligned}
\mathcal{M}(\Omega, \mu_0) &= \big\{ \mu = \phi \, \mu_0 : \phi \in L^1(\Omega, \mu_0), \ \phi \geq 0 \big\}, \\
\mathcal{M}_+(\Omega, \mu_0) &= \big\{ \mu = \phi \, \mu_0 : \phi \in L^1(\Omega, \mu_0), \ \phi > 0 \big\}, \\
\mathcal{P}(\Omega, \mu_0) &= \bigg\{ \mu \in \mathcal{M}(\Omega, \mu_0) : \int_\Omega d\mu = 1 \bigg\}, \\
\mathcal{P}_+(\Omega, \mu_0) &= \bigg\{ \mu \in \mathcal{M}_+(\Omega, \mu_0) : \int_\Omega d\mu = 1 \bigg\}, \\
\mathcal{S}(\Omega, \mu_0) &= \big\{ \mu = \phi \, \mu_0 : \phi \in L^1(\Omega, \mu_0) \big\}.
\end{aligned} \tag{3.48}$$

By the Radon–Nikodym theorem, $\mathcal{P}(\Omega, \mu_0) \subseteq \mathcal{M}(\Omega, \mu_0) \subseteq \mathcal{S}(\Omega, \mu_0)$ consist of those measures in $\mathcal{P}(\Omega) \subseteq \mathcal{M}(\Omega) \subseteq \mathcal{S}(\Omega)$ which are dominated by μ_0, and the canonical isomorphism, $i_{can} : \mathcal{S}(\Omega, \mu_0) \to L^1(\Omega, \mu_0)$ in (3.2) given by taking the Radon–Nikodym derivative w.r.t. μ_0 yields an isomorphism whose inverse is given by

$$\imath_{can}^{-1} : L^1(\Omega, \mu_0) \longrightarrow \mathcal{S}(\Omega, \mu_0), \qquad \phi \longmapsto \phi \, \mu_0.$$

Observe that $\imath_{can}$ is an isometry of Banach spaces, since evidently

$$\|\phi\|_{L^1(\Omega, \mu_0)} = \int_\Omega |\phi| \, d\mu_0 = \|\phi \, \mu_0\|_{TV}.$$

3.2.2 Tangent Fibration of Subsets of Banach Manifolds

In this section, we shall use the notion of differentiable maps between Banach spaces, as described in Appendix C. In particular, a *curve* in a Banach space

$(V; \|\cdot\|)$ is a differentiable map $c : I \to V$ with an (open) interval $I \subseteq \mathbb{R}$. That is, for each $t_0 \in I$ there exists the limit

$$\frac{d}{dt}\bigg|_{t=t_0} c(t) = \dot{c}(t_0) := \lim_{t \to t_0} \frac{c(t) - c(t_0)}{t - t_0}. \tag{3.49}$$

Definition 3.2 Let $(V; \|\cdot\|)$ be a Banach space, $X \subseteq V$ an arbitrary subset and $x_0 \in X$. Then $v \in V$ is called a *tangent vector of X at x_0*, if there is a curve $c : (-\varepsilon, \varepsilon) \to X \subseteq V$ such that $c(0) = x_0$ and $\dot{c}(0) = v$.

The set of all tangent vectors at x_0 is called the *tangent double cone of X at x_0* and is denoted by $T_{x_0} X$. We also define the *tangent fibration of X* as

$$TX := \biguplus_{x_0 \in X} T_{x_0} X \subseteq X \times V \subseteq V \times V,$$

equipped with the induced topology and with the canonical projection map $TX \to X$.

Remark 3.3 The reader should be aware that, unlike in some texts, we do not use *tangent fibration* as a synonym for the tangent bundle, since for general subsets $X \subseteq V$, $T_{x_0} X \subseteq V$ may fail to be a vector subspace, and for $x_0 \neq x_1$, the tangent cones $T_{x_0} X$ and $T_{x_1} X$ need not be homeomorphic.

For instance, let $X := \{(x, y)^\top \in \mathbb{R}^2 : xy = 0\} \subseteq \mathbb{R}^2$, so that X is the union of the two coordinate axes. Then $T_{(x,0)} X$ and $T_{(0,y)} X$ with $x, y \neq 0$ are the x-axis and the y-axis, respectively, and hence linear subspaces of $V = \mathbb{R}^2$, but $T_{(0,0)} X = X$ is not a subspace. This example also shows that $T_{x_0} X$ and $T_{x_1} X$ need not be homeomorphic if $x_0 \neq x_1$, whence the projection $TX \to X$ mapping $T_{x_0} X$ to x_0 is in general only a topological fibration, but not a vector bundle or a fiber bundle.

But at least, $T_{x_0} X$ is invariant under multiplication by (positive or negative) scalars and hence is a double cone. This is seen by replacing the curve c in Definition 3.2 by $\tilde{c}(t) := c(t_0 t) \in X$ for some $t_0 \in \mathbb{R}$, so that $\tilde{c}(0) = x_0$ and $\dot{\tilde{c}}(0) = t_0 v \in T_{x_0} X$.

If $X \subseteq V$ is a submanifold, however, then $T_{x_0} X$ and TX coincide with the standard notion of the tangent space at x_0 and the tangent bundle of X, respectively. Thus, in this case the tangent cone is a linear subspace, and the tangent fibration is the tangent bundle of X.

For instance, if $X = U \subseteq V$ is an open set, then $T_{x_0} U = V$ for all x_0 and hence, $TU = U \times V$. Indeed, the curve $c(t) = x_0 + tv \in U$ for small $|t|$ satisfies the properties required in the Definition 3.2. In this case, the tangent fibration $U \times V \to U$ is a (trivial) vector bundle.

If M is a Banach manifold and $F : M \to V$ is a C^1-map whose image is contained in $X \subseteq V$, then by the chain rule, for any $v \in T_{x_0} M$ and any curve $c : (-\varepsilon, \varepsilon) \to M$ with $c(0) = x_0$, $\dot{c}(0) = v$,

$$\frac{d}{dt}\bigg|_{t=0} F(c(t)) = d_{x_0} F(v),$$

and as $t \mapsto F(c(t))$ is a curve in X with $F(c(0)) = F(x_0)$, it follows that $d_{x_0}F(v) \in T_{F(x_0)}X$ for all $v \in TM$. That is, a C^1-map $F : M \to X \subseteq V$ induces a continuous map

$$dF : TM \longrightarrow TX, \qquad (x_0, v) \longmapsto d_{x_0}F(v).$$

Theorem 3.1 *Let $V = \mathcal{S}(\Omega)$ be the Banach space of finite signed measures on Ω. Then the tangent cones of $\mathcal{M}(\Omega)$ and $\mathcal{P}(\Omega)$ at μ are $T_\mu\mathcal{M}(\Omega) = \mathcal{S}(\Omega, \mu)$ and $T_\mu\mathcal{P}(\Omega) = \mathcal{S}_0(\Omega, \mu)$, respectively, so that*

$$T\mathcal{M}(\Omega) = \biguplus_{\mu \in \mathcal{M}(\Omega)} \mathcal{S}(\Omega, \mu) \subseteq \mathcal{M}(\Omega) \times \mathcal{S}(\Omega)$$

and

$$T\mathcal{P}(\Omega) = \biguplus_{\mu \in \mathcal{P}(\Omega)} \mathcal{S}_0(\Omega, \mu) \subseteq \mathcal{P}(\Omega) \times \mathcal{S}(\Omega).$$

Proof Let $v \in T_{\mu_0}\mathcal{M}(\Omega)$ and let $(\mu_t)_{t \in (-\varepsilon, \varepsilon)}$ be a curve in $\mathcal{M}(\Omega)$ with $\dot{\mu}_0 = v$. Let $A \subseteq \Omega$ be such that $\mu_0(A) = 0$. Then as $\mu_t(A) \geq 0$, the function $t \mapsto \mu_t(A)$ has a minimum at $t = 0$, whence

$$0 = \frac{d}{dt}\bigg|_{t=0} \mu_t(A) = \dot{\mu}_0(A) = v(A),$$

where the second equation is evident from (3.49). That is, $v(A) = 0$ whenever $\mu_0(A) = 0$, i.e., μ_0 dominates v, so that $v \in \mathcal{S}(\Omega, \mu_0)$. Thus, $T_{\mu_0}\mathcal{M}(\Omega) \subseteq \mathcal{S}(\Omega, \mu_0)$.
   Conversely, given $v = \phi\mu_0 \in \mathcal{S}(\Omega, \mu_0)$, define $\mu_t := p(\omega; t)\mu_0$ where

$$p(\omega; t) := \begin{cases} 1 + t\phi(\omega) & \text{if } t\phi(\omega) \geq 0, \\ \exp(t\phi(\omega)) & \text{if } t\phi(\omega) < 0. \end{cases}$$

As $p(\omega; t) \leq \max(1 + t\phi(\omega), 1)$, it follows that $\mu_t \in \mathcal{M}(\Omega)$, and as the derivative $\partial_t p(\omega; t)$ exists for all t and its absolute value is bounded by $|\phi| \in L^1(\Omega, \mu_0)$, it follows that $t \mapsto \mu_t$ is a C^1-curve in $\mathcal{M}(\Omega)$ with $\dot{\mu}_0 = \phi\mu_0 = v$, whence $v \in T_{\mu_0}\mathcal{M}(\Omega)$ as claimed.
   To show the statement for $\mathcal{P}(\Omega)$, let $(\mu_t)_{t \in (-\varepsilon, \varepsilon)}$ be a curve in $\mathcal{P}(\Omega)$ with $\dot{\mu}_0 = v$. Then as μ_t is a probability measure for all t, we conclude that

$$\left| \int_\Omega dv \right| = \left| \int_\Omega \frac{1}{t} d(\mu_t - \mu_0 - tv) \right| \leq \frac{\|\mu_t - \mu_0 - tv\|_{TV}}{|t|} \xrightarrow{t \to 0} 0,$$

so that $v \in \mathcal{S}_0(\Omega)$. Since $\mathcal{P}(\Omega) \subseteq \mathcal{M}(\Omega)$, it follows that $T_{\mu_0}\mathcal{P}(\Omega) \subseteq T_{\mu_0}\mathcal{M}(\Omega) \cap \mathcal{S}_0(\Omega) = \mathcal{S}_0(\Omega, \mu_0)$ for all $\mu_0 \in \mathcal{P}(\Omega)$.
   Conversely, given $v = \phi\mu_0 \in \mathcal{S}_0(\Omega, \mu_0)$, define the curve $\lambda_t := \mu_t\|\mu_t\|_{TV}^{-1} \in \mathcal{P}(\Omega)$ with μ_t from above, which is a C^1-curve in $\mathcal{P}(\Omega)$ as $\|\mu_t\|_{TV} > 0$, and it is straightforward that $\lambda_0 = \mu_0$ and $\dot{\lambda}_0 = \phi\mu_0 = v$. $\square$

Remark 3.4

(1) Even though the tangent cones of the subsets $\mathcal{P}(\Omega) \subseteq \mathcal{M}(\Omega) \subseteq \mathcal{S}(\Omega)$ at each point μ are closed vector subspaces of $\mathcal{S}(\Omega)$, these subsets are not Banach manifolds, and hence in particular not Banach submanifolds of $\mathcal{S}(\Omega)$. This is due to the fact that the spaces $\mathcal{S}_0(\Omega, \mu)$ need not be isomorphic to each other for different μ.

   This can already be seen for finite $\Omega = \{\omega_1, \ldots, \omega_k\}$. In this case, we may identify $\mathcal{S}(\Omega)$ with $\mathbb{R}^k$ by the map $\sum_{i=1}^{k} x_i \delta^{\omega_i} \cong (x_1, \ldots, x_k)$, and with this,

$$T\mathcal{M}(\Omega) \cong \left\{ (x_1, \ldots, x_k; y_1, \ldots, y_k) \in \mathbb{R}^k \times \mathbb{R}^k : \begin{array}{l} x_i \geq 0, \\ x_i = 0 \Rightarrow y_i = 0 \end{array} \right\},$$

   and this is evidently not a submanifold of $\mathbb{R}^{2k}$. Indeed, in this case the dimension of $T_\mu \mathcal{M}(\Omega) = \mathcal{S}(\Omega, \mu)$ equals $|\{\omega \in \Omega \mid \mu(\omega) > 0\}|$, which varies with μ. Of course, this simply reflects the geometric stratification of the closed probability simplex in terms of the faces of various dimensions. Theorem 3.1 then describes such a stratification also in infinite dimensions.

(2) Observe that the curves μ_t and λ_t, respectively, used in the proof of Theorem 3.1 are contained in $\mathcal{M}_+(\Omega, \mu_0)$ and $\mathcal{P}_+(\Omega, \mu_0)$, respectively, whence it also follows that

$$T_{\mu_0} \mathcal{M}_+(\Omega, \mu_0) = \mathcal{S}(\Omega, \mu_0) \quad \text{and} \quad T_{\mu_0} \mathcal{P}_+(\Omega, \mu_0) = \mathcal{S}_0(\Omega, \mu_0).$$

   But if $\mu_1 \in \mathcal{M}_+(\Omega, \mu_0)$, then μ_1 and μ_0 are compatible measures (i.e., they have the same null sets), whence in this case, $\mathcal{S}(\Omega, \mu_0) = \mathcal{S}(\Omega, \mu_1)$ and $\mathcal{S}_0(\Omega, \mu_0) = \mathcal{S}_0(\Omega, \mu_1)$. That is, the subset $\mathcal{M}_+(\Omega, \mu_0) \subseteq \mathcal{S}(\Omega, \mu_0)$ has at each point all of $\mathcal{S}(\Omega, \mu_0)$ as its tangent space, but in general, $\mathcal{M}_+(\Omega, \mu_0) \subseteq \mathcal{S}(\Omega, \mu_0)$ is not open.[4] This is a quite remarkable and unusual phenomenon in Differential Geometry.

   That is, neither on $\mathcal{M}(\Omega)$ nor on $\mathcal{M}_+(\Omega, \mu_0)$ is there a canonical manifold structure in general, and the same is true for $\mathcal{P}(\Omega)$ and $\mathcal{P}_+(\Omega, \mu_0)$, respectively. Nevertheless, Definition 3.2 and Theorem 3.1 allow us to speak of the tangent fibration of $\mathcal{M}(\Omega)$ and $\mathcal{P}(\Omega)$, respectively.

3.2.3 Powers of Measures

Let us now give the formal definition of roots of measures. On the set $\mathcal{M}(\Omega)$ we define the preordering $\mu_1 \leq \mu_2$ if μ_2 dominates μ_1. Then $(\mathcal{M}(\Omega), \leq)$ is a directed set, meaning that for any pair $\mu_1, \mu_2 \in \mathcal{M}(\Omega)$ there is a $\mu_0 \in \mathcal{M}(\Omega)$ dominating both of them (e.g., $\mu_0 := \mu_1 + \mu_2$).

---

[4]More precisely, $\mathcal{M}_+(\Omega, \mu_0)$ is not open unless Ω is the disjoint union of finitely many μ_0-atoms, where $A \subseteq \Omega$ is a μ_0-atom if for each $B \subseteq A$ either B or $A \backslash B$ is a μ_0-null set.

For fixed $r \in (0, 1]$ and measures $\mu_1 \leq \mu_2$ on Ω we define the linear embedding

$$\iota_{\mu_2}^{\mu_1} : L^{1/r}(\Omega, \mu_1) \longrightarrow L^{1/r}(\Omega, \mu_2), \qquad \phi \longmapsto \phi \left(\frac{d\mu_1}{d\mu_2} \right)^r.$$

Observe that

$$\left\| \iota_{\mu_2}^{\mu_1}(\phi) \right\|_{1/r} = \left(\int_\Omega \left| \iota_{\mu_2}^{\mu_1}(\phi) \right|^{1/r} d\mu_2 \right)^r = \left(\int_\Omega |\phi|^{1/r} \frac{d\mu_1}{d\mu_2} \, d\mu_2 \right)^r$$

$$= \left(\int_\Omega |\phi|^{1/r} \, d\mu_1 \right)^r = \|\phi\|_{1/r}, \tag{3.50}$$

so that $\iota_{\mu_2}^{\mu_1}$ is an isometry. Moreover, $\iota_{\mu_2}^{\mu_1} \iota_{\mu_3}^{\mu_2} = \iota_{\mu_3}^{\mu_1}$ whenever $\mu_1 \leq \mu_2 \leq \mu_3$. Then we define the *space of rth roots of measures on Ω* to be the directed limit over the directed set $(\mathcal{M}(\Omega), \leq)$

$$\mathcal{S}^r(\Omega) := \varinjlim L^{1/r}(\Omega, \mu). \tag{3.51}$$

Let us give a more concrete definition of $\mathcal{S}^r(\Omega)$. On the disjoint union of the spaces $L^{1/r}(\Omega, \mu)$ for $\mu \in \mathcal{M}(\Omega)$ we define the equivalence relation

$$L^{1/r}(\Omega, \mu_1) \ni \phi \sim \psi \in L^{1/r}(\Omega, \mu_2) \quad \Longleftrightarrow \quad \iota_{\mu_0}^{\mu_1}(\phi) = \iota_{\mu_0}^{\mu_2}(\psi)$$

$$\Longleftrightarrow \quad \phi \left(\frac{d\mu_1}{d\mu_0} \right)^r = \psi \left(\frac{d\mu_2}{d\mu_0} \right)^r$$

for some $\mu_0 \geq \mu_1, \mu_2$. Then $\mathcal{S}^r(\Omega)$ is the set of all equivalence classes of this relation, cf. Fig. 3.1.

Let us denote the equivalence class of $\phi \in L^{1/r}(\Omega, \mu)$ by $\phi\mu^r$, so that $\mu^r \in \mathcal{S}^r(\Omega)$ is the equivalence class represented by $1 \in L^{1/r}(\Omega, \mu)$. Then the equivalence relation yields

$$\mu_1^r = \left(\frac{d\mu_1}{d\mu_2} \right)^r \mu_2^r \quad \text{as elements of } \mathcal{S}^r(\Omega) \tag{3.52}$$

whenever $\mu_1 \leq \mu_2$, justifying this notation. In fact, from this description in the case $r = 1$ we see that

$$\mathcal{S}^1(\Omega) = \mathcal{S}(\Omega).$$

Observe that by (3.50) $\|\phi\|_{1/r}$ is constant on equivalence classes, whence there is a norm on $\mathcal{S}^r(\Omega)$, also denoted by $\| \cdot \|_{1/r}$, for which the inclusions

$$\mathcal{S}^r(\Omega, \mu) \hookrightarrow \mathcal{S}^r(\Omega) \quad \text{and} \quad L^{1/r}(\Omega, \mu) \longrightarrow \mathcal{S}^r(\Omega), \quad \phi \longmapsto \phi\mu^r \tag{3.53}$$

are isometries. For $r = 1$, we have $\| \cdot \|_1 = \| \cdot \|_{TV}$. Thus,

$$\left\| \phi\mu^r \right\|_{1/r} = \|\phi\|_{1/r} = \left(\int_\Omega |\phi|^{1/r} \, d\mu \right)^r \quad \text{for } 0 < r \leq 1. \tag{3.54}$$
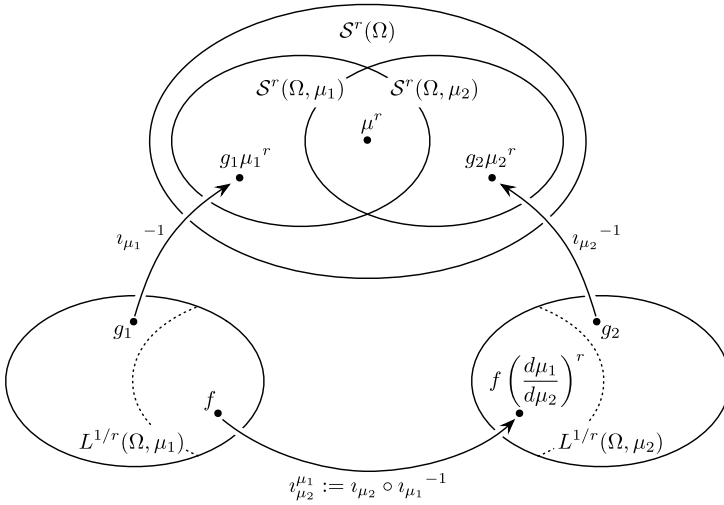
Fig. 3.1 Natural description of $\mathcal{S}^r(\Omega)$ in terms of powers of the Radon–Nikodym derivative

Note that the equivalence relation also preserves non-negativity of functions, whence we may define the subsets

$$
\begin{aligned}
\mathcal{M}^r(\Omega) &:= \big\{\phi\mu^r : \mu \in \mathcal{M}(\Omega), \phi \ge 0\big\}, \\
\mathcal{P}^r(\Omega) &:= \big\{\phi\mu^r : \mu \in \mathcal{P}(\Omega), \phi \ge 0, \|\phi\|_{1/r} = 1\big\}.
\end{aligned}
\tag{3.55}
$$

In analogy to (3.48) we define for a fixed measure $\mu_0 \in \mathcal{M}(\Omega)$ and $r \in (0, 1]$ the spaces

$$
\begin{aligned}
\mathcal{S}^r(\Omega, \mu_0) &:= \big\{\phi\,\mu_0^r : \phi \in L^{1/r}(\Omega, \mu_0)\big\}, \\
\mathcal{M}^r(\Omega, \mu_0) &:= \big\{\phi\,\mu_0^r : \phi \in L^{1/r}(\Omega, \mu_0), \phi \ge 0\big\}, \\
\mathcal{P}^r(\Omega, \mu_0) &:= \big\{\phi\,\mu_0^r : \phi \in L^{1/r}(\Omega, \mu_0), \phi \ge 0, \|\phi\|_{1/r} = 1\big\}, \\
\mathcal{S}_0^r(\Omega, \mu_0) &:= \Big\{\phi\mu_0^r : \phi \in L^{1/r}(\Omega, \mu_0), \int_\Omega \phi\,d\mu = 0\Big\}.
\end{aligned}
$$

The elements of $\mathcal{P}^r(\Omega, \mu_0), \mathcal{M}^r(\Omega, \mu_0), \mathcal{S}^r(\Omega, \mu_0)$ are said to be *dominated* by μ_0^r.

Remark 3.5 The concept of rth roots of measures has been indicated in [200, Ex. IV.1.4]. Moreover, if Ω is a manifold and $r = 1/2$, then $\mathcal{S}^{1/2}(\Omega)$ is even a Hilbert space which has been considered in [192, 6.9.1]. This Hilbert space is also related to the Hilbert manifold of finite-entropy probability measures defined in [201].

The product of powers of measures can now be defined for all $r, s \in (0, 1)$ with $r + s \leq 1$ and for measures $\phi \mu^r \in \mathcal{S}^r(\Omega, \mu)$ and $\psi \mu^s \in \mathcal{S}^s(\Omega, \mu)$:

$$\left(\phi \mu^r\right) \cdot \left(\psi \mu^s\right) := \phi \psi \mu^{r+s}.$$

By definition $\phi \in L^{1/r}(\Omega, \mu)$ and $\psi \in L^{1/s}(\Omega, \mu)$, whence Hölder's inequality implies that $\|\phi \psi\|_{1/(r+s)} \leq \|\phi\|_{1/r} \|\psi\|_{1/s} < \infty$, so that $\phi \psi \in L^{1/(r+s)}(\Omega, \mu)$ and hence, $\phi \psi \mu^{r+s} \in \mathcal{S}^{r+s}(\Omega, \mu)$. Since by (3.52) this definition of the product is independent of the choice of representative μ, it follows that it induces a bilinear product

$$\cdot : \mathcal{S}^r(\Omega) \times \mathcal{S}^s(\Omega) \longrightarrow \mathcal{S}^{r+s}(\Omega), \quad \text{where } r, s, r + s \in (0, 1], \tag{3.56}$$

satisfying the Hölder inequality

$$\|\nu_r \cdot \nu_s\|_{1/(r+s)} \leq \|\nu_r\|_{1/r} \|\nu_s\|_{1/s}, \tag{3.57}$$

so that the product in (3.56) is a bounded bilinear map.

Definition 3.3 (Canonical pairing) For $r \in (0, 1)$ we define the pairing

$$(\cdot; \cdot) : \mathcal{S}^r(\Omega) \times \mathcal{S}^{1-r}(\Omega) \longrightarrow \mathbb{R}, \qquad (\nu_1; \nu_2) := \int_\Omega d(\nu_1 \cdot \nu_2). \tag{3.58}$$

It is straightforward to verify that this pairing is non-degenerate in the sense that

$$(\nu_r; \cdot) = 0 \Longleftrightarrow \nu_r = 0. \tag{3.59}$$

Lemma 3.1 (Cf. [200, Ex. IV.1.3]) *Let $\{\nu_n : n \in \mathbb{N}\} \subseteq \mathcal{S}(\Omega)$ be a countable family of (signed) measures. Then there is a measure $\mu_0 \in \mathcal{M}(\Omega)$ dominating ν_n for all n.*

Proof We assume w.l.o.g. that $\nu_n \neq 0$ for all n and define

$$\mu_0 := \sum_{n=1}^{\infty} \frac{1}{2^n \|\nu_n\|_{TV}} |\nu_n|.$$

Since $\|\nu_n\|_{TV} = |\nu_n|(\Omega)$, it follows that this sum converges, so that $\mu_0 \in \mathcal{M}(\Omega)$ is well defined. Moreover, if $\mu_0(A) = 0$, then $|\nu_n|(A) = 0$ for all n, showing that μ_0 dominates all ν_n as claimed. $\qquad\square$

From Lemma 3.1, we can now conclude the following statement:

Any sequence in $\mathcal{S}^r(\Omega)$ is contained in $\mathcal{S}^r(\Omega, \mu_0)$ for some $\mu_0 \in \mathcal{M}(\Omega)$.

In particular, any Cauchy sequence in $\mathcal{S}^r(\Omega)$ is a Cauchy sequence in $\mathcal{S}^r(\Omega, \mu_0) \cong L^{1/r}(\Omega, \mu_0)$ for some μ_0 and hence convergent. Thus,

$(\mathcal{S}^r(\Omega), \|\cdot\|_{1/r})$ is a Banach space. It also follows that $\mathcal{S}^r(\Omega, \mu_0)$ is a closed subspace of $\mathcal{S}(\Omega)$ for all $\mu_0 \in \mathcal{M}(\Omega)$.

In analogy to Theorem 3.1, we can also determine the tangent cones of the subsets $\mathcal{P}^r(\Omega) \subseteq \mathcal{M}^r(\Omega) \subseteq \mathcal{S}^r(\Omega)$.

Proposition 3.1   *For each $\mu \in \mathcal{M}(\Omega)$ ($\mu \in \mathcal{P}(\Omega)$, respectively), the tangent cone of $\mathcal{P}^r(\Omega) \subseteq \mathcal{M}^r(\Omega) \subseteq \mathcal{S}^r(\Omega)$ at μ^r are $T_{\mu^r} \mathcal{M}^r(\Omega) = \mathcal{S}^r(\Omega, \mu)$ and $T_{\mu^r} \mathcal{P}^r(\Omega) = \mathcal{S}_0^r(\Omega, \mu)$, respectively, so that the tangent fibrations are given as*

$$T\mathcal{M}^r(\Omega) = \biguplus_{\mu^r \in \mathcal{M}^r(\Omega)} \mathcal{S}^r(\Omega, \mu) \subseteq \mathcal{M}^r(\Omega) \times \mathcal{S}^r(\Omega)$$

and

$$T\mathcal{P}^r(\Omega) = \biguplus_{\mu^r \in \mathcal{P}^r(\Omega)} \mathcal{S}_0^r(\Omega, \mu) \subseteq \mathcal{P}^r(\Omega) \times \mathcal{S}^r(\Omega).$$

Proof   We have to adapt the proof of Theorem 3.1. The proof of the statements $\mathcal{S}^r(\Omega, \mu) \subseteq T_{\mu^r} \mathcal{M}^r(\Omega)$ and $\mathcal{S}_0^r(\Omega, \mu) \subseteq T_{\mu^r} \mathcal{P}^r(\Omega)$ is identical to that of the corresponding statement in Theorem 3.1; just as in that case, one shows that for $\phi \in L^{1/r}(\Omega, \mu_0)$ the curves $\mu_t^r := p(\omega; t)\mu_0^r$ with $p(\omega; t) := 1 + t\phi(\omega)$ if $t\phi(\omega) \geq 0$ and $p(\omega; \xi) = \exp(t\phi(\omega))$ if $t\phi(\omega) < 0$ is a differentiable curve in $\mathcal{M}^r(\Omega)$, and $\lambda_t^r := \mu_t^r / \|\mu_t^r\|_{1/r}$ is a differentiable curve in $\mathcal{P}^r(\Omega)$, and their derivative is $\phi\mu_0^r$ at $t = 0$.

In order to show the other direction, let $(\mu_t^r)_{t \in (-\varepsilon, \varepsilon)}$ be a curve in $\mathcal{M}^r(\Omega)$. Then $\mathbb{Q} \cap (-\varepsilon, \varepsilon)$ is countable, whence by Lemma 3.1 there is a measure $\hat{\mu}$ such that $(\mu_t^r) \in \mathcal{M}^r(\Omega, \hat{\mu})$ for all $t \in \mathbb{Q}$, and since $\mathcal{S}^r(\Omega, \hat{\mu}) \subseteq \mathcal{S}^r(\Omega)$ is closed, it follows that $\mu_t^r \in \mathcal{M}(\Omega, \hat{\mu})$ for all t. Now we can apply the argument from Theorem 3.1 to the curve $t \mapsto (\mu_t^r \cdot \hat{\mu}^{1-r})(A)$ for $A \subseteq \Omega$. $\square$

Remark 3.6   Just as in the case of $r = 1$, we may take two points of view on the relation of $\mathcal{M}^r(\Omega)$ and $\mathcal{P}^r(\Omega)$. The one is that $\mathcal{P}^r(\Omega)$ may be regarded as a subset of $\mathcal{M}^r(\Omega)$, but also, the normalization map

$$\mathcal{M}^r(\Omega) \longrightarrow \mathcal{P}^r(\Omega), \qquad \mu_r \longmapsto \frac{\mu_r}{\|\mu_r\|_{1/r}}$$

allows us to regard $\mathcal{P}^r(\Omega)$ as the projectivization $\mathbb{P}(\mathcal{M}^r(\Omega))$. It will depend on the context which point of view is better adapted.

Besides multiplying roots of measures, we also wish to take their powers. Here, we have two possibilities for dealing with signs. For $0 < k \leq r^{-1}$ and $\nu_r = \phi\mu^r \in \mathcal{S}^r(\Omega)$ we define

$$|\nu_r|^k := |\phi|^k \mu^{rk} \quad \text{and} \quad \tilde{\nu}_r^k := \text{sign}(\phi)|\phi|^k \mu^{rk}. \tag{3.60}$$

Since $\phi \in L^{1/r}(\Omega, \mu)$, it follows that $|\phi|^k \in L^{1/kr}(\Omega, \mu)$, so that $|\nu_r|^k, \tilde{\nu}_r^k \in \mathcal{S}^{rk}(\Omega)$. By (3.52) these powers are well defined, independent of the choice of the measure μ, and, moreover,

$$\left\| \, |\nu_r|^k \, \right\|_{1/(rk)} = \left\| \tilde{\nu}_r^k \right\|_{1/(rk)} = \|\nu_r\|_{1/r}^k. \tag{3.61}$$

Proposition 3.2 *Let $r \in (0, 1]$ and $0 < k \leq 1/r$, and consider the maps*

$$\pi^k, \tilde{\pi}^k : \mathcal{S}^r(\Omega) \longrightarrow \mathcal{S}^{rk}(\Omega), \qquad \begin{array}{l} \pi^k(\nu) := |\nu|^k, \\ \tilde{\pi}^k(\nu) := \tilde{\nu}^k. \end{array}$$

Then $\pi^k, \tilde{\pi}^k$ are continuous maps. Moreover, for $1 < k \leq 1/r$ they are C^1-maps between Banach spaces, and their derivatives are given as

$$d_{\nu_r}\tilde{\pi}^k(\rho_r) = k \, |\nu_r|^{k-1} \cdot \rho_r \quad and \quad d_{\nu_r}\pi^k(\rho_r) = k \, \tilde{\nu}_r^{k-1} \cdot \rho_r. \tag{3.62}$$

Observe that for $k = 1$, $\pi^1(\nu_r) = |\nu_r|$ fails to be C^1, whereas $\tilde{\pi}^1(\nu_r) = \nu_r$, so that $\tilde{\pi}^1$ is the identity and hence a C^1-map.

Proof Let us first assume that $0 < k \leq 1$. We assert that for all $x, y \in \mathbb{R}$ we have the estimates

$$\begin{aligned} \left| |x + y|^k - |x|^k \right| &\leq |y|^k \quad and \\ \left| \mathrm{sign}(x + y)|x + y|^k - \mathrm{sign}(x)|x|^k \right| &\leq 2^{1-k}|y|^k. \end{aligned} \tag{3.63}$$

For $k = 1$, (3.63) is obvious. If $0 < k < 1$, then by homogeneity it suffices to show these for $y = 1$. Note that the functions

$$x \longmapsto |x + 1|^k - |x|^k \quad and \quad x \longmapsto \mathrm{sign}(x + 1)|x + 1|^k - \mathrm{sign}(x)|x|^k$$

are continuous and tend to 0 for $x \to \pm\infty$, and then (3.63) follows by elementary calculus.

Let $\nu_1, \nu_2 \in \mathcal{S}^r(\Omega)$, and choose $\mu_0 \in \mathcal{M}(\Omega)$ such that $\nu_1, \nu_2 \in \mathcal{S}^r(\Omega, \mu_0)$, i.e., $\nu_i = \phi_i \mu_0^r$ with $\phi_i \in L^{1/r}(\Omega, \mu_0)$. Then

$$\begin{aligned} \left\| \pi^k(\nu_1 + \nu_2) - \pi^k(\nu_1) \right\|_{1/rk} &= \left\| |\phi_1 + \phi_2|^k - |\phi_1|^k \right\|_{1/rk} \\ &\leq \left\| \, |\phi_2|^k \right\|_{1/rk} \quad \text{by (3.63)} \\ &= \|\nu_2\|_{1/r}^k \quad \text{by (3.61)}, \end{aligned}$$

so that $\lim_{\|\nu_2\|_{1/r} \to 0} \|\pi^k(\nu_1 + \nu_2) - \pi^k(\nu_1)\|_{1/rk} = 0$, showing the continuity of π^k for $0 < k \leq 1$. The continuity of $\tilde{\pi}^k$ follows analogously.

Now let us assume that $1 < k \leq 1/r$. In this case, the functions

$$x \longmapsto |x|^k \quad and \quad x \longmapsto \mathrm{sign}(x)|x|^k$$

with $x \in \mathbb{R}$ are C^1-maps with respective derivatives

$$x \longmapsto k \ \text{sign}(x)|x|^{k-1} \quad \text{and} \quad x \longmapsto k|x|^{k-1}.$$

Thus, if we pick $v_i = \phi_i \mu_0^r$ as above, then by the mean value theorem we have

$$\pi^k(v_1 + v_2) - \pi^k(v_1) = \left(|\phi_1 + \phi_2|^k - |\phi_1|^k\right)\mu_0^{rk}$$
$$= k \ \text{sign}(\phi_1 + \eta\phi_2)|\phi_1 + \eta\phi_2|^{k-1}\phi_2\mu_0^{rk}$$
$$= k \ \text{sign}(\phi_1 + \eta\phi_2)|\phi_1 + \eta\phi_2|^{k-1}\mu_0^{r(k-1)} \cdot v_2$$

for some function $\eta : \Omega \to (0, 1)$. If we let $v_\eta := \eta\phi_2\mu_0^r$, then $\|v_\eta\|_{1/r} \leq \|v_2\|_{1/r}$, and we get

$$\pi^k(v_1 + v_2) - \pi^k(v_1) = k\tilde{\pi}^{k-1}(v_1 + v_\eta) \cdot v_2.$$

With the definition of $d_{v_1}\tilde{\pi}^k$ from (3.62) we have

$$\left\|\pi^k(v_1 + v_2) - \pi^k(v_1) - d_{v_1}\pi^k(v_2)\right\|_{1/(rk)}$$
$$= \left\|k\left(\tilde{\pi}^{k-1}(v_1 + v_\eta) - \tilde{\pi}^{k-1}(v_1)\right) \cdot v_2\right\|_{1/(rk)}$$
$$\leq k\left\|\tilde{\pi}^{k-1}(v_1 + v_\eta) - \tilde{\pi}^{k-1}(v_1)\right\|_{1/(r(k-1))}\|v_2\|_{1/r}$$

and hence,

$$\frac{\left\|\pi^k(v_1 + v_2) - \pi^k(v_1) - d_{v_1}\pi^k(v_2)\right\|_{\frac{1}{rk}}}{\|v_2\|_{\frac{1}{r}}} \leq k\left\|\tilde{\pi}^{k-1}(v_1 + v_\eta) - \tilde{\pi}^{k-1}(v_1)\right\|_{\frac{1}{r(k-1)}}.$$

Thus, the differentiability of π^k will follow if

$$\left\|\tilde{\pi}^{k-1}(v_1 + v_\eta) - \tilde{\pi}^{k-1}(v_1)\right\|_{1/(r(k-1))} \xrightarrow{\|v_2\|_{1/r} \to 0} 0,$$

and because of $\|v_\eta\|_{1/r} \leq \|v_2\|_{1/r}$, this is the case if $\tilde{\pi}^{k-1}$ is continuous.

Analogously, one shows that $\tilde{\pi}^k$ is differentiable if π^{k-1} is continuous.

Since we already know continuity of π^k and $\tilde{\pi}^k$ for $0 < k \leq 1$, and since C^1-maps are continuous, the claim now follows by induction on $\lceil k \rceil$. $\qquad\square$

Thus, (3.62) implies that the differentials of π^k and $\tilde{\pi}^k$ (which coincide on $\mathcal{P}^r(\Omega)$ and $\mathcal{M}^r(\Omega)$) yield continuous maps

$$d\pi^k = d\tilde{\pi}^k : \begin{array}{ccc} T\mathcal{P}^r(\Omega) & \longrightarrow & T\mathcal{P}^{rk}(\Omega) \\ T\mathcal{M}^r(\Omega) & \longrightarrow & T\mathcal{M}^{rk}(\Omega), \end{array} \qquad (\mu, \rho) \longmapsto k\mu^{rk-r} \cdot \rho.$$

3.2.4 Parametrized Measure Models and k-Integrability

In this section, we shall now present our notion of a parametrized measure model.

Definition 3.4 (Parametrized measure model) Let Ω be a measurable space.

(1) A *parametrized measure model* is a triple $(M, \Omega, \mathbf{p})$ where M is a (finite or infinite-dimensional) Banach manifold and $\mathbf{p} : M \to \mathcal{M}(\Omega) \subseteq \mathcal{S}(\Omega)$ is a C^1-map in the sense of Definition C.3.
(2) The triple $(M, \Omega, \mathbf{p})$ is called a *statistical model* if it consists only of probability measures, i.e., such that the image of $\mathbf{p}$ is contained in $\mathcal{P}(\Omega)$.
(3) We call such a model *dominated by* μ_0 if the image of $\mathbf{p}$ is contained in $\mathcal{M}(\Omega, \mu_0)$. In this case, we use the notation $(M, \Omega, \mu_0, \mathbf{p})$ for this model.

If a parametrized measure model $(M, \Omega, \mu_0, \mathbf{p})$ is dominated by μ_0, then there is a *density function* $p : \Omega \times M \to \mathbb{R}$ such that

$$\mathbf{p}(\xi) = p(\cdot; \xi)\mu_0. \tag{3.64}$$

Evidently, we must have $p(\cdot; \xi) \in L^1(\Omega, \mu_0)$ for all ξ. In particular, for fixed ξ, $p(\cdot; \xi)$ is defined only up to changes on a μ_0-null set. The existence of a dominating measure μ_0 is not a strong restriction, as the following shows.

Proposition 3.3 *Let $(M, \Omega, \mathbf{p})$ be a parametrized measure model. If M contains a countable dense subset, e.g., if M is a finite-dimensional manifold, then there is a measure $\mu_0 \in \mathcal{M}(\Omega)$ dominating the model.*

Proof Let $(\xi_n)_{n \in \mathbb{N}} \subseteq M$ be a dense countable subset. By Lemma 3.1, there is a measure μ_0 dominating all measures $\mathbf{p}(\xi_n)$ for $n \in \mathbb{N}$, i.e., $\mathbf{p}(\xi_n) \in \mathcal{M}(\Omega, \mu_0)$. If $\xi_{n_k} \to \xi$, so that $\mathbf{p}(\xi_{n_k}) \to \mathbf{p}(\xi)$, then as the inclusion $\mathcal{S}(\Omega, \mu_0) \hookrightarrow \mathcal{S}(\Omega)$ is an isometry by (3.53), it follows that $(\mathbf{p}(\xi_{n_k}))_{k \in \mathbb{N}}$ is a Cauchy sequence in $\mathcal{S}(\Omega, \mu_0)$, and as the latter is complete, it follows that $\mathbf{p}(\xi) \in \mathcal{S}(\Omega, \mu_0) \cap \mathcal{M}(\Omega) = \mathcal{M}(\Omega, \mu_0)$. $\qquad\square$

Definition 3.5 (Regular density function) Let $(M, \Omega, \mu_0, \mathbf{p})$ be a parametrized measure model dominated by μ_0. We say that this model has a *regular density function* if the density function $p : \Omega \times M \to \mathbb{R}$ satisfying (3.64) can be chosen such that for all $V \in T_\xi M$ the partial derivative $\partial_V p(\cdot; \xi)$ exists and lies in $L^1(\Omega, \mu_0)$.

Remark 3.7 The standard notion of a statistical model always assumes that it is dominated by some measure and has a positive regular density function (e.g., [9, §2 , p. 25], [16, §2.1], [219], [25, Definition 2.4]). In fact, the definition of a parametrized measure model or statistical model in [25, Definition 2.4] is equivalent to a parametrized measure model or statistical model with a positive regular density function in the sense of Definition 3.5. In contrast, in [26], the assumption

of regularity and, more importantly, of the positivity of the density function p is dropped.

It is worth pointing out that the density function p of a parametrized measure model $(M, \Omega, \mu_0, \mathbf{p})$ does not need to be regular, so that the present notion is indeed more general. The formal definition of differentiability of $\mathbf{p}$ implies that for each C^1-path $\xi(t) \in M$ with $\xi(0) = \xi$, $\dot{\xi}(0) =: V \in T_\xi M$, the curve $t \mapsto p(\cdot; \xi(t)) \in L^1(\Omega, \mu_0)$ is differentiable. This implies that there is a $d_\xi \mathbf{p}(V) \in L^1(\Omega, \mu_0)$ such that

$$\left\| \frac{p(\cdot; \xi(t)) - p(\cdot; \xi)}{t} - d_\xi \mathbf{p}(V)(\cdot) \right\|_1 \xrightarrow{\ t \to 0\ } 0.$$

If this is a *pointwise* convergence, then $d_\xi \mathbf{p}(V) = \partial_V p(\cdot; \xi)$ is the partial derivative and whence, $\partial_V p(\cdot; \xi)$ lies in $L^1(\Omega, \mu_0)$, so that the density function is regular.

However, in general convergence in $L^1(\Omega, \mu_0)$ does *not* imply pointwise convergence, whence there are parametrized measure models in the sense of Definition 3.4 without a regular density function, cf. Example 3.2.3 below. Nevertheless, we shall use the following notations interchangeably,

$$d_\xi \mathbf{p}(V) = \partial_V \mathbf{p} = \partial_V p(\cdot; \xi) \, \mu_0, \tag{3.65}$$

even if $\mathbf{p}$ does not have a regular density function and the derivative $\partial_V p(\cdot; \xi)$ does not exist.

Example 3.2

(1) The family of normal distributions on $\mathbb{R}$

$$p(\mu, \sigma) := \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2} \right) dx$$

is a statistical model with regular density function on the upper half-plane $H = \{(\mu, \sigma) : \mu, \sigma \in \mathbb{R}, \sigma > 0\}$.

(2) To see that there are parametrized measure models without a regular density function, consider the family $(\mathbf{p}(\xi))_{\xi > -1}$ of measures on $\Omega = (0, \pi)$

$$\mathbf{p}(\xi) := \begin{cases} (1 + \xi \, (\sin^2(t - 1/\xi))^{1/\xi^2}) \, dt & \text{for } \xi \neq 0, \\ dt & \text{for } \xi = 0. \end{cases}$$

This model is dominated by the Lebesgue measure dt, with density function $p(t; \xi) = 1 + \xi \, (\sin^2(t - 1/\xi))^{1/\xi^2}$ for $\xi \neq 0$, $p(t; 0) = 1$. Thus, the partial derivative $\partial_\xi p$ does not exist at $\xi = 0$, whence the density function is not regular.

On the other hand, $\mathbf{p} : (-1, \infty) \to \mathcal{M}(\Omega, dt)$ is differentiable at $\xi = 0$ with $d_0 \mathbf{p}(\partial_\xi) = 0$, so that $(M, \Omega, \mathbf{p})$ is a parametrized measure model in the sense of

Definition 3.4. To see this, we calculate

$$\left\| \frac{\mathbf{p}(\xi) - \mathbf{p}(0)}{\xi} \right\|_{TV} = \left\| \left(\sin^2(t - 1/\xi) \right)^{1/\xi^2} dt \right\|_1$$

$$= \int_0^\pi \left(\sin^2(t - 1/\xi) \right)^{1/\xi^2} dt$$

$$= \int_0^\pi \left(\sin^2 t \right)^{1/\xi^2} dt \xrightarrow{\xi \to 0} 0,$$

which shows the claim. Here, we used the π-periodicity of the integrand for fixed ξ and dominated convergence in the last step.

Since for a parametrized measure model $(M, \Omega, \mathbf{p})$ the map $\mathbf{p}$ is C^1, it follows that its derivative yields a continuous map between the tangent fibrations

$$d\mathbf{p} : TM \longrightarrow T\mathcal{M}(\Omega) = \biguplus_{\mu \in \mathcal{M}(\Omega)} \mathcal{S}(\Omega, \mu).$$

That is, for each tangent vector $V \in T_\xi M$, its differential $d_\xi \mathbf{p}(V)$ is contained in $\mathcal{S}(\Omega, \mathbf{p}(\xi))$ and hence dominated by $\mathbf{p}(\xi)$. Therefore, we can take the Radon–Nikodym derivative of $d_\xi \mathbf{p}(V)$ w.r.t. $\mathbf{p}(\xi)$.

Definition 3.6 Let $(M, \Omega, \mathbf{p})$ be a parametrized measure model. Then for each tangent vector $V \in T_\xi M$ of M, we define

$$\partial_V \log \mathbf{p}(\xi) := \frac{d\{d_\xi \mathbf{p}(V)\}}{d\mathbf{p}(\xi)} \in L^1\big(\Omega, p(\xi)\big) \tag{3.66}$$

and call this the *logarithmic derivative of* $\mathbf{p}$ *at* ξ *in the direction* V.

If such a model is dominated by μ_0 and has a positive regular density function p for which (3.64) holds, then we calculate the Radon–Nikodym derivative as

$$\frac{d\{d_\xi \mathbf{p}(V)\}}{d\mathbf{p}(\xi)} = \frac{d\{d_\xi \mathbf{p}(V)\}}{d\mu_0} \cdot \left(\frac{d\mathbf{p}(\xi)}{d\mu_0} \right)^{-1}$$

$$= \partial_V p(\cdot; \xi)\big(p(\cdot; \xi)\big)^{-1} = \partial_V \log p(\cdot; \xi).$$

This justifies the notation in (3.66) even for models without a positive regular density function.

For a parametrized measure model $(M, \Omega, \mathbf{p})$ and $k > 1$ we consider the map

$$\mathbf{p}^{1/k} := \pi^{1/k} \mathbf{p} : M \longrightarrow \mathcal{M}^{1/k}(\Omega) \subseteq \mathcal{S}^{1/k}(\Omega), \qquad \xi \longmapsto \mathbf{p}(\xi)^{1/k}.$$

Since $\pi^{1/k}$ is continuous by Proposition 3.2, it follows that $\mathbf{p}^{1/k}$ is continuous as well. We define the following notions of k-integrability.

Definition 3.7 (*k*-Integrable parametrized measure model)   A parametrized measure model $(M, \Omega, \mathbf{p})$ (statistical model, respectively) is called *k-integrable* if the map

$$\mathbf{p}^{1/k} : M \longrightarrow \mathcal{M}^{1/k}(\Omega) \subseteq \mathcal{S}^{1/k}(\Omega)$$

is a C^1-map in the sense of Definition C.1. It is called *weakly k-integrable* if this map is a weak C^1-map, again in the sense of Definition C.1.

Furthermore, we call the model *(weakly) ∞-integrable* if it is (weakly) *k*-integrable for all $k \geq 1$.

Evidently, every parametrized measure model is 1-integrable by definition. Moreover, since for $1 \leq l < k$ we have $\mathbf{p}^{1/l} = \pi^{k/l} \mathbf{p}^{1/k}$ and $\pi^{k/l}$ is a C^1-map by Proposition 3.2, it follows that (weak) *k*-integrability implies (weak) *l*-integrability for $1 \leq l < k$.

Example 3.3   An exponential family, as defined by (3.31), generalizes the family of normal distributions and represents an extremely important statistical model. It will play a central role throughout this book.

Adapting (3.31) to the notation in this context, we can write it for $\xi = (\xi^1, \ldots, \xi^n) \in U \subseteq \mathbb{R}^n$ as

$$\mathbf{p}(\xi) = \exp\left(\gamma(\omega) + \sum_{i=1}^n f_i(\omega)\xi^i - \psi(\xi)\right)\mu(\omega), \qquad (3.67)$$

for suitable functions f_i, γ on Ω and ψ on U. Therefore, for $V = (v_1, \ldots, v_n)$,

$$\partial_V \mathbf{p}^{1/k}(\xi) = \frac{1}{k}\left(\sum_{i=1}^n v_i f_i(\omega) + \partial_V \psi(\xi)\right)$$

$$\times \exp\left(\gamma(\omega)/k + \sum_{i=1}^n f_i(\omega)\xi^i/k - \psi(\xi)/k\right)\mu(\omega)^{1/k},$$

and the *k*-integrability of this model for any *k* is easily verified from there. Therefore, exponential families provide a class of examples of ∞-integrable parametrized measure models. See also [216, p. 1559].

Remark 3.8

(1) In Sect. 2.8.1, we have introduced exponential families for finite sets as affine spaces. Let us comment on that structure for the setting of arbitrary measurable spaces as described in [192]. Consider the affine action $(\mu, f) \mapsto e^f \mu$, defined by (2.130). Clearly, multiplication of a finite measure $\mu \in \mathcal{M}_+(\Omega)$ with e^f will in general lead to a positive measure on Ω that is not finite but σ-finite. As we restrict attention to finite measures, and thereby obtain the Banach space structure of $\mathcal{S}(\Omega)$, it is not possible to extend the affine structure of Sect. 2.8.1 to the

general setting of measurable spaces. However, we can still define exponential families as geometric objects that correspond to a "section" of an affine space, leading to the expression (3.31).

(2) Fukumizu [101] proposed the notion of a *kernel exponential family*, based on the theory of reproducing kernel Hilbert spaces. In this approach, one considers a Hilbert space $\mathcal{H}$ of functions $\Omega \to \mathbb{R}$ that is defined in terms of a so-called *reproducing kernel* $k : \Omega \times \Omega \to \mathbb{R}$. In particular, the inner product on $\mathcal{H}$ is given as

$$\langle f, g \rangle_{\mathcal{H}} = \int \int f(\omega) \, g(\omega') \, k(\omega, \omega') \, \mu_0(d\omega) \, \mu_0(d\omega').$$

We will revisit this product again in Sect. 6.4 (see (6.196)). Within this framework, the sum $\sum_{i=1}^{n} f_i(\omega) \, \xi^i$ in (3.67) is then replaced by an integral $\int k(\omega, \omega') \, \xi(\omega') \, \mu_0(d\omega')$, leading to the definition of a kernel exponential family.

Proposition 3.4 *Let $(M, \Omega, \mathbf{p})$ be a (weakly) k-integrable parametrized measure model. Then its (weak) derivative is given as*

$$\partial_V \mathbf{p}^{1/k}(\xi) := \frac{1}{k} \partial_V \log \mathbf{p}(\xi) \, \mathbf{p}^{1/k}(\xi) \in \mathcal{S}^{1/k}\big(\Omega, \mathbf{p}(\xi)\big), \qquad (3.68)$$

and for any functional $\alpha \in \mathcal{S}^{1/k}(\Omega)'$ we have

$$\partial_V \alpha\big(\mathbf{p}(\xi)^{1/k}\big) = \alpha\big(\partial_V \mathbf{p}^{1/k}(\xi)\big). \qquad (3.69)$$

Observe that if $\mathbf{p}(\xi) = p(\cdot; \xi)\mu_0$ with a regular density function p, the derivative $\partial_V p^{1/k}(\omega; \xi)$ is indeed the partial derivative of the function $p(\omega; \xi)^{1/k}$.

Proof Suppose that the model is weakly k-integrable, i.e., $\mathbf{p}^{1/k}$ is weakly differentiable, and let $V \in T_\xi M$ and $\alpha \in \mathcal{S}(\Omega)'$. Then

$$\alpha\big(\partial_V \log \mathbf{p}(\xi)\mathbf{p}(\xi)\big) \;=\; \alpha(\partial_V \mathbf{p}) \overset{(3.66)}{=} \alpha\big(\partial_V \big(\pi^k \mathbf{p}^{1/k}\big)\big)$$

$$\overset{(3.62)}{=} \alpha\big(k\mathbf{p}(\xi)^{1-1/k} \cdot \partial_V \mathbf{p}^{1/k}\big),$$

whence

$$\alpha\big(\big(k\partial_V \mathbf{p}^{1/k} - \partial_V \log \mathbf{p}(\xi)\mathbf{p}^{1/k}\big) \cdot \mathbf{p}(\xi)^{1-1/k}\big) = 0$$

for all $\alpha \in \mathcal{S}(\Omega)'$. On the other hand, $\partial_V \log \mathbf{p} \in T_{\mathbf{p}^{1/k}(\xi)}\mathcal{M}^{1/k}(\Omega) = \mathcal{S}^{1/k}(\Omega, \mathbf{p}(\xi))$ according to Proposition 3.1, and from this, (3.68) follows.

The identity (3.69) is simply the definition of $\partial_V \mathbf{p}^{1/k}(\xi)$ being the weak Gâteaux-derivative of $\mathbf{p}^{1/k}$, cf. Proposition C.2. $\qquad\qquad \square$

The following now gives a description of (weak) integrability in terms of the (weak) derivative.

Theorem 3.2 *Let $(M, \Omega, \mathbf{p})$ be a parametrized measure model. Then the following hold*:

(1) *The model is k-integrable if and only if the map*

$$V \longmapsto \left\| \partial_V \mathbf{p}^{1/k} \right\|_k \tag{3.70}$$

defined on TM is continuous.

(2) *The model is weakly k-integrable if and only if the weak derivative of $\mathbf{p}^{1/k}$ is weakly continuous, i.e., if for all $V_0 \in TM$*

$$\partial_V \mathbf{p}^{1/k} \rightharpoonup \partial_{V_0} \mathbf{p}^{1/k} \quad as \quad V \to V_0.$$

Remark 3.9 In [25, Definition 2.4], k-integrability (in case of a positive regular density function) was defined by the continuity of the norm function in (3.70), whence it coincides with Definition 3.7 by Theorem 3.2.

Our motivation for also introducing the more general definition of weak k-integrability is that it is the weakest condition that ensures that integration and differentiation of kth roots can be interchanged, as explained in the following.

If $(M, \Omega, \mu_0, \mathbf{p})$ has the density function $p : \Omega \times M \to \mathbb{R}$ given by (3.64), then for $\xi \in M$ and $V \in T_\xi M$ we have

$$\mathbf{p}^{1/k}(\xi) = p(\cdot; \xi)^{1/k} \mu_0^{1/k} \quad and \quad \partial_V \mathbf{p}^{1/k} = \partial_V p(\cdot; \xi)^{1/k} \mu_0^{1/k},$$

where

$$\partial_V p(\cdot; \xi)^{1/k} := \frac{1}{k} \frac{\partial_V p(\cdot; \xi)}{p(\cdot; \xi)^{1-1/k}} \in L^k(\Omega, \mu_0). \tag{3.71}$$

Thus, if we let $\alpha(\cdot) := (\cdot; \mu_0^{1-1/k})$ with the canonical pairing from (3.58), then (3.69) takes the form

$$\partial_V \int_A p(\omega; \xi)^{1/k} \, d\mu_0(\omega) = \int_A \partial_V p(\omega; \xi)^{1/k} \, d\mu_0(\omega). \tag{3.72}$$

Evidently, if p is a regular density function, then the weak partial derivative is indeed the partial derivative of p, in which case (3.72) is obvious, as integration and differentiation may be interchanged under these regularity conditions.

Example 3.4 For arbitrary $k > 1$, the following is an example of a parametrized measure model which is l-integrable for all $1 \leq l < k$, weakly k-integrable, but not k-integrable.

Let $\Omega = (-1, 1)$ with the Lebesgue measure dt, and let $f : [0, \infty) \longrightarrow \mathbb{R}$ be a smooth function such that

$$f(u) > 0, f'(u) < 0 \quad \text{for } u \in [0, 1), \qquad f(u) \equiv 0 \quad \text{for } u \geq 1.$$

For $\xi \in \mathbb{R}$, define the measure $\mathbf{p}(\xi) = p(\xi; t) \, dt$, where

$$p(\xi; t) := \begin{cases} 1 & \text{if } t \leq 0 \text{ and } \xi \in \mathbb{R} \text{ arbitrary,} \\ |\xi|^{k-1} f(t|\xi|^{-1})^k \, dt & \text{if } \xi \neq 0 \text{ and } t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Since on its restrictions to $(-1, 0]$ and $(0, 1)$ the density function p as well as p^r with $0 < r < 1$ are positive with bounded derivative for $\xi \neq 0$, it follows that $\mathbf{p}$ is ∞-integrable for $\xi \neq 0$. For $\xi = 0$, we have

$$\left\| \mathbf{p}(\xi) - \mathbf{p}(0) \right\|_1 = |\xi|^{k-1} \int_0^{|\xi|} f\big(t|\xi|^{-1}\big)^k \, dt = |\xi|^k \int_0^1 f(u)^k \, du,$$

whence, as $k > 1$,

$$\lim_{\xi \to 0} \frac{\|\mathbf{p}(\xi) - \mathbf{p}(0)\|_1}{|\xi|} = 0,$$

showing that $\mathbf{p}$ is also a C^1-map at $\xi = 0$ with differential $\partial_\xi \mathbf{p}(0) = 0$. That is, $(\mathbb{R}, \Omega, \mathbf{p})$ is a parametrized measure model with $d_0\mathbf{p} = 0$.

For $1 \leq l \leq k$ and $\xi \neq 0$ we calculate

$$\partial_\xi \mathbf{p}^{1/l} = \chi_{(0,1)}(t) \, \partial_\xi \left(|\xi|^{(k-1)/l} f\big(t|\xi|^{-1}\big)^{k/l} \right) dt^{1/l}$$

$$= \chi_{(0,1)}(t) \, \text{sign}(\xi) |\xi|^{(k-l-1)/l} \left(\frac{k-1}{l} f(u)^{k/l} - u\big(f^{k/l}\big)'(u) \right)\bigg|_{u=t|\xi|^{-1}} dt^{1/l}.$$

Thus, it follows that

$$\left\| \partial_\xi \mathbf{p}^{1/l}(\xi) \right\|_l^l = |\xi|^{k-l} \int_0^1 \left(\frac{k-1}{l} f(u)^{k/l} - u\big(f^{k/l}\big)'(u) \right)^l du,$$

so that

$$\lim_{\xi \to 0} \left\| \partial_\xi \mathbf{p}^{1/l}(\xi) \right\|_l = 0 = \left\| \partial_\xi \mathbf{p}^{1/l}(0) \right\|_l \qquad \text{for } 1 \leq l < k,$$

$$\lim_{\xi \to 0} \left\| \partial_\xi \mathbf{p}^{1/k}(\xi) \right\|_k > 0 = \left\| \partial_\xi \mathbf{p}^{1/k}(0) \right\|_k.$$

That is, by Theorem 3.2 the model is l-integrable for $1 \leq l < k$, but it fails to be k-integrable. On the other hand,

$$\left\| \partial_\xi \mathbf{p}^{1/k}(\xi) \cdot dt^{1-1/k} \right\|_1 = |\xi|^{1-1/k} \int_0^1 \left(\left(1 - \frac{1}{k}\right) f(u) - u f'(u) \right) du,$$

so that $\lim_{\xi \to 0} \| \partial_\xi \mathbf{p}^{1/k}(\xi) \cdot dt^{1-1/k} \|_1 = 0$. As we shall show in Lemma 3.3 below, this implies that $\partial_\xi \mathbf{p}^{1/k}(\xi) \rightharpoonup 0 = \partial_\xi \mathbf{p}^{1/k}(0)$ as $\xi \to 0$. Thus, the model is weakly k-integrable by Theorem 3.2.

The rest of this section will be devoted to the proof of Theorem 3.2 which is somewhat technical and therefore will be divided into several lemmas.

Before starting the proof, let us give a brief outline of its structure.

We begin by proving the second statement of Theorem 3.2. Note that for a weak C^1-map the differential is weakly continuous by definition, so one direction of the proof is trivial. The reverse implication is the content of Lemmata 3.2 through 3.6.

We give a decomposition of the dual space $\mathcal{S}^{1/k}(\Omega)'$ in Lemma 3.2 and a sufficient criterion for the weak convergence of sequences in $\mathcal{S}^{1/k}(\Omega, \mu_0)$ in Lemma 3.3 as well as a criterion for weak k-integrability in terms of interchanging differentiation and integration along curves in M in Lemma 3.4.

Unfortunately, we are not able to verify this criterion directly for an arbitrary model $\mathbf{p}$. The technical obstacle is that the measures of the family $\mathbf{p}(\xi)$ need not be equivalent. We overcome this difficulty by modifying the model $\mathbf{p}$ to a model $\mathbf{p}_\varepsilon(\xi) := \mathbf{p}(\xi) + \varepsilon \mu_0$, where $\varepsilon > 0$ and $\mu_0 \in \mathcal{M}(\Omega)$ is a suitable measure, so that $\mathbf{p}_\varepsilon$ has a positive density function. Then we show in Lemma 3.5 that the differential of $\mathbf{p}_\varepsilon$ remains weakly continuous, and finally in Lemma 3.6 we show that $\mathbf{p}_\varepsilon$ is weakly k-integrable as it satisfies the criterion given in Lemma 3.4; furthermore it is shown that taking the limit $\varepsilon \to 0$ implies the weak k-integrability of $\mathbf{p}$ as well, proving the second part of Theorem 3.2.

The first statement of Theorem 3.2 is proven in Lemmata 3.7 and 3.8. Again, one direction is trivial: if the model is k-integrable, then its differential is continuous by definition, whence so is its norm (3.70). That is, we have to show the converse.

In Lemma 3.7 we show that the continuity of the map (3.70) implies the weak continuity of the differential of $\mathbf{p}^{1/k}$. This implies, on the one hand, that the model is weakly k-integrable by the second part of Theorem 3.2 which was already shown, and on the other hand, the Radon–Riesz theorem (cf. Theorem C.3) implies that the differential of $\mathbf{p}^{1/k}$ is even norm-continuous. Then in Lemma 3.8 we give the standard argument that a weak C^1-map with a norm-continuous differential must be a C^1-map, and this will complete the proof.

Lemma 3.2 *For each $\mu_0 \in \mathcal{M}(\Omega)$, there is an isomorphism*

$$\left(\mathcal{S}^{1/k}(\Omega)\right)' \cong \left(\mathcal{S}^{1/k}(\Omega, \mu_0)\right)^\perp \oplus \mathcal{S}^{k/(k-1)}(\Omega, \mu_0), \tag{3.73}$$

where $(\mathcal{S}^{1/k}(\Omega, \mu_0))^\perp$ denotes the annihilator of $\mathcal{S}^{1/k}(\Omega, \mu_0) \subseteq \mathcal{S}^{1/k}(\Omega)$ and where $\mathcal{S}^{k/(k-1)}(\Omega, \mu_0)$ is embedded into the dual via the canonical pairing (3.58). That is, we can write any $\alpha \in (\mathcal{S}^{1/k}(\Omega))'$ uniquely as

$$\alpha(\cdot) = \left(\cdot; \phi_\alpha \mu_0^{1-1/k}\right) + \beta^{\mu_0}(\cdot), \tag{3.74}$$

where $\phi_\alpha \in L^{k/(k-1)}(\Omega, \mu_0)$ and $\beta^{\mu_0}(\mathcal{S}^{1/k}(\Omega, \mu_0)) = 0$.

Proof The restriction of α to $\mathcal{S}^{1/k}(\Omega, \mu_0)$ yields a functional on $L^k(\Omega, \mu_0)$ given as $\psi \mapsto \alpha(\psi \mu_0^{1/k})$. Since the dual of $L^k(\Omega, \mu_0)$ is $L^{k/(k-1)}(\Omega, \mu_0)$, there is a ϕ_α

such that

$$\alpha\big(\psi\mu_0^{1/k}\big) = \int_\Omega \psi\phi_\alpha \, d\mu_0 = \big(\psi\mu_0^{1/k}; \phi_\alpha\mu_0^{1-1/k}\big),$$

and then (3.74) follows by letting $\beta^{\mu_0}(\cdot) := \alpha(\cdot) - (\cdot; \phi_\alpha\mu_0^{1-1/k})$.   □

Lemma 3.3 *Let* $v_n^{1/k} = \psi_n\mu_0^{1/k}$ *be a bounded sequence in* $\mathcal{S}^{1/k}(\Omega, \mu_0)$, *i.e.,* $\limsup \|v_n^{1/k}\|_k < \infty$. *If* $\lim \int_\Omega |\psi_n| \, d\mu_0 = 0$, *then*

$$v_n^{1/k} \rightharpoonup 0,$$

i.e., $\lim \alpha(v_n^{1/k}) = 0$ *for all* $\alpha \in (\mathcal{S}^{1/k}(\Omega))'$.

Proof Suppose that $\int_\Omega |\psi_n| \, d\mu_0 \to 0$ and let $\phi \in L^{k/(k-1)}(\Omega, \mu_0)$ and $\tau \in L^\infty(\Omega)$. Then

$$\limsup \big|(v_n^{1/k}; \phi\mu_0^{1-1/k})\big| \leq \limsup\bigg(\int_\Omega |\psi_n||\phi - \tau| \, d\mu_0 + \|\tau\|_\infty \int_\Omega |\psi_n| \, d\mu_0\bigg)$$

$$\leq \limsup \|\psi_n\|_k \|\phi - \tau\|_{k/(k-1)},$$

using Hölder's inequality in the last estimate. Since $\|v_n^{1/k}\|_k = \|\psi_n\|_k$ and hence, $\limsup \|\psi_n\|_k < \infty$, the bound on the right can be made arbitrarily small as $L^\infty(\Omega) \subseteq L^{k/(k-1)}(\Omega, \mu_0)$ is a dense subspace. Therefore, $\lim(v_n^{1/k}; \phi\mu_0^{1-1/k}) = 0$ for all $\phi \in L^{k/(k-1)}$, and since $\beta(v_n^{1/k}) = 0$ for all $\beta \in (\mathcal{S}^{1/k}(\Omega, \mu_0))^\perp$, the assertion follows from (3.74).   □

Before we go on, let us introduce some notation. Let $(M, \Omega, \mathbf{p})$ be a parametrized measure model such that $\partial_V \log \mathbf{p}(\xi) \in L^k(\Omega, \mathbf{p}(\xi))$ for all $V \in T_\xi M$, and let $(\xi_t)_{t \in I}$ be a curve in M. By Proposition 3.3, there is a measure $\mu_0 \in \mathcal{M}(\Omega)$ dominating all $\mathbf{p}(\xi_t)$. For $t, t_0 \in I$ and $1 \leq l \leq k$ we define the *remainder term* as

$$\mathbf{r}_l(t, t_0) := \mathbf{p}^{1/l}(\xi_{t+t_0}) - \mathbf{p}^{1/l}(\xi_{t_0}) - t\partial_{\xi_{t_0}'}\mathbf{p}^{1/l} \in \mathcal{S}^{1/l}(\Omega), \qquad (3.75)$$

and we define the functions $p_t, q_t \in L^1(\Omega, \mu_0)$ with $p_t \geq 0$ and $q_{t;l}, r_{t,t_0;l} \in L^l(\Omega, \mu_0)$ such that

$$\mathbf{p}(\xi_t) = p_t\mu_0 \qquad \text{and} \qquad \partial_{\xi_t'}\mathbf{p} = q_t\mu_0,$$

$$q_{t;l} := \frac{q_t}{lp_t^{1-1/l}} \qquad \text{so that} \quad \partial_{\xi_t'}\mathbf{p}^{1/l} = q_{t;l}\mu_0^{1/l}, \qquad (3.76)$$

$$r_{t,t_0;l} := p_{t+t_0}^{1/l} - p_{t_0}^{1/l} - tq_{t_0;l} \quad \text{so that} \quad \mathbf{r}_l(t, t_0) = r_{t,t_0;l}\mu_0^{1/l}.$$

Lemma 3.4 *Let* $(M, \Omega, \mathbf{p})$ *be a parametrized measure model such that* $\partial_V \log \mathbf{p}(\xi) \in L^k(\Omega, \mathbf{p}(\xi))$ *for all* $V \in T_\xi M$ *and the function* $V \mapsto \partial_V \mathbf{p}^{1/k} \in \mathcal{S}^{1/k}(\Omega)$ *is weakly continuous.*

Then the model is weakly k-integrable if for any curve $(\xi_t)_{t \in I}$ in M and the measure $\mu_0 \in \mathcal{M}(\Omega)$ and the functions $p_t, q_t, q_{t;l}$ defined in (3.76) and any $A \subseteq \Omega$ and $t_0 \in I$

$$\frac{d}{dt}\bigg|_{t=t_0} \int_A p_t^{1/k} \, d\mu_0 = \int_A q_{t_0;k} \, d\mu_0 \tag{3.77}$$

or, equivalently, for all $a, b \in I$

$$\int_A \left(p_b^{1/k} - p_a^{1/k}\right) d\mu_0 = \int_a^b \int_A q_{t;k} \, d\mu_0 \, dt. \tag{3.78}$$

Proof Note that the right-hand side of (3.77) can be written as

$$\int_A q_{t_0;k} \, d\mu_0 = \left(\partial_{\xi'_{t_0}} \mathbf{p}^{1/k}; \chi_A \mu_0^{1-1/k}\right)$$

with the pairing from (3.58), whence it depends continuously on t_0 by the weak continuity of $V \mapsto \partial_V \mathbf{p}^{1/k}$. Thus, the equivalence of (3.77) and (3.78) follows from the fundamental theorem of calculus.

Now (3.78) can be rewritten as

$$\left(\mathbf{p}^{1/k}(\xi_b) - \mathbf{p}^{1/k}(\xi_a); \phi\mu_0^{1-1/k}\right) = \int_a^b \left(\partial_{\xi'_t} \mathbf{p}^{1/k}; \phi\mu_0^{1-1/k}\right) dt \tag{3.79}$$

for $\phi = \chi_A$, and hence, (3.79) holds whenever $\phi = \tau$ is a step function. But now, if $\phi \in L^{k/(k-1)}$ is given, then there is a sequence of step functions (τ_n) such that $\text{sign}(\phi)\tau_n \nearrow |\phi|$, and since (3.79) holds for all step functions, it also holds for ϕ by dominated convergence.

If $\beta \in \mathcal{S}^{1/k}(\Omega, \mu_0)^\perp$, then clearly, $\beta(\mathbf{p}^{1/k}(\xi_t)) = \beta(\partial_{\xi'_t} \mathbf{p}^{1/k}) = 0$, whence by (3.74) we have for all $\alpha \in (\mathcal{S}^{1/k}(\Omega, \mu_0))'$

$$\alpha\left(\mathbf{p}^{1/k}(\xi_b) - \mathbf{p}^{1/k}(\xi_a)\right) = \int_a^b \alpha\left(\partial_{\xi'_t} \mathbf{p}^{1/k}\right) dt,$$

and since the function $t \mapsto \alpha(\partial_{\xi'_t} \mathbf{p}^{1/k})$ is continuous by the assumed weak continuity of $V \mapsto \partial_V \mathbf{p}^{1/k}$, differentiation and the fundamental theorem of calculus yield (3.69) for $V = \xi'_t$, and as the curve (ξ_t) was arbitrary, (3.69) holds for arbitrary $V \in TM$.

But (3.69) is equivalent to saying that $\partial_V \mathbf{p}^{1/k}(\xi)$ is the weak Gâteaux-differential of $\mathbf{p}^{1/k}$ (cf. Definition C.1), and since this map is assumed to be weakly continuous, it follows that $\mathbf{p}^{1/k}$ is a weak C^1-map, whence $(M, \Omega, \mathbf{p})$ is weakly k-integrable. $\square$

Lemma 3.5 *Let $(M, \Omega, \mathbf{p})$ be a parametrized measure model for which the map $V \mapsto \partial_V \mathbf{p}^{1/k}$ is weakly continuous. Let $(\xi_t)_{t \in I}$ be a curve in M, and let $\mu_0 \in \mathcal{M}(\Omega)$ be a measure dominating $\mathbf{p}(\xi_t)$ for all t. For $\varepsilon > 0$, define the parametrized measure model $\mathbf{p}_\varepsilon$ as*

$$\mathbf{p}_\varepsilon(\xi) := \mathbf{p}(\xi) + \varepsilon\mu_0. \tag{3.80}$$

Then the map $t \mapsto \partial_{\xi'_t} \mathbf{p}_\varepsilon^{1/k} \in \mathcal{S}^{1/k}(\Omega)$ is weakly continuous, and for all $t_0 \in I$,

$$\frac{1}{t} \mathbf{r}_k^\varepsilon(t, t_0) \rightharpoonup 0 \quad as\ t \to 0,$$

where $\mathbf{r}_k^\varepsilon(t, t_0)$ is defined analogously to (3.75).

Proof We define the functions p_t^ε, q_t^ε, $q_{t;l}^\varepsilon$ and $r_{t,t_0;l}^\varepsilon$ satisfying (3.76) for the parametrized measure model $\mathbf{p}_\varepsilon$, so that $p_t^\varepsilon = p_t + \varepsilon$ and $q_t^\varepsilon = q_t$. For $t, t_0 \in I$ we have

$$
\begin{aligned}
\left| q_{t;k}^\varepsilon - q_{t_0;k}^\varepsilon \right| &= \left| \frac{q_t}{k(p_t^\varepsilon)^{1-1/k}} - \frac{q_{t_0}}{k(p_{t_0}^\varepsilon)^{1-1/k}} \right| \\
&\leq \frac{1}{k(p_{t_0}^\varepsilon)^{1-1/k}} |q_t - q_{t_0}| + \left| \frac{1}{k(p_t^\varepsilon)^{1-1/k}} - \frac{1}{k(p_{t_0}^\varepsilon)^{1-1/k}} \right| |q_t| \\
&\leq \frac{1}{k(p_{t_0}^\varepsilon)^{1-1/k}} \left(|q_t - q_{t_0}| + k \left| (p_{t_0}^\varepsilon)^{1-1/k} - (p_t^\varepsilon)^{1-1/k} \right| |q_{t;k}^\varepsilon| \right) \\
&\overset{(3.63)}{\leq} \frac{1}{k\varepsilon^{1-1/k}} \left(|q_t - q_{t_0}| + k|p_{t_0} - p_t|^{1-1/k} |q_{t;k}| \right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\int_\Omega \left| q_{t;k}^\varepsilon - q_{t_0;k}^\varepsilon \right| d\mu_0 &\leq \frac{1}{k\varepsilon^{1-1/k}} \int_\Omega \left(|q_t - q_{t_0}| + k|p_{t_0} - p_t|^{1-1/k} |q_{t;k}| \right) d\mu_0 \\
&\leq \frac{1}{k\varepsilon^{1-1/k}} \left(\|q_t - q_{t_0}\|_1 + k\|p_{t_0} - p_t\|_1^{1-1/k} \|q_{t;k}\|_k \right)
\end{aligned}
$$

and since $\|q_t - q_{t_0}\|_1 = \|d_{\xi'_t}\mathbf{p} - d_{\xi'_{t_0}}\mathbf{p}\|_1$ and $\|p_{t_0} - p_t\|_1 = \|\mathbf{p}(\xi_t) - \mathbf{p}(t_0)\|_1$ tend to 0 for $t \to t_0$ as $\mathbf{p}$ is a C^1-map by the definition of parametrized measure model, and $\|q_{t;k}\|_k = \|\partial_{\xi'_t}\mathbf{p}^{1/k}\|_k$ is bounded for $t \to t_0$ by (C.1) as $\partial_{\xi'_t}\mathbf{p}^{1/k} \rightharpoonup \partial_{\xi'_{t_0}}\mathbf{p}^{1/k}$, it follows that the integral tends to 0 and hence, Lemma 3.3 implies that $\partial_{\xi'_t}\mathbf{p}_\varepsilon^{1/k} \rightharpoonup \partial_{\xi'_{t_0}}\mathbf{p}_\varepsilon^{1/k}$ as $t \to t_0$, showing the weak continuity of $t \mapsto \partial_{\xi'_t}\mathbf{p}_\varepsilon^{1/k}$.

For the second claim, note that by the mean value theorem there is an η_t between $p_{t+t_0}^\varepsilon$ and $p_{t_0}^\varepsilon$ (and hence, $\eta_t \geq \varepsilon$) for which

$$
\begin{aligned}
\left| r_{t,t_0;k}^\varepsilon \right| &= \left| (p_{t+t_0}^\varepsilon)^{1/k} - (p_{t_0}^\varepsilon)^{1/k} - t\frac{q_{t_0}}{k(p_{t_0}^\varepsilon)^{1-1/k}} \right| = \left| \frac{p_{t+t_0} - p_{t_0}}{k\eta_t^{1-1/k}} - t\frac{q_{t_0}}{k(p_{t_0}^\varepsilon)^{1-1/k}} \right| \\
&\leq \frac{|r_{t,t_0;1}|}{k\eta_t^{1-1/k}} + |t| \left| \frac{q_{t_0}}{k\eta_t^{1-1/k}} - \frac{q_{t_0}}{k(p_{t_0}^\varepsilon)^{1-1/k}} \right| \\
&= \frac{|r_{t,t_0;1}|}{k\eta_t^{1-1/k}} + \frac{|t||q_{t_0}| |(p_{t_0}^\varepsilon)^{1-1/k} - \eta_t^{1-1/k}|}{k(p_{t_0}^\varepsilon)^{1-1/k}\eta_t^{1-1/k}} \\
&\overset{(3.63)}{\leq} C\left(|r_{t,t_0;1}| + |t||q_{t_0;k}||p_{t+t_0} - p_{t_0}|^{1-1/k} \right)
\end{aligned}
$$

for $C := 1/k\varepsilon^{1-1/k} > 0$ depending only on k and ε. Thus,

$$\int_\Omega |r^\varepsilon_{t,t_0;k}|\, d\mu_0 \le C\left(\int_\Omega |r_{t,t_0;1}|\, d\mu_0 + |t|\int_\Omega |q_{t_0;k}||p_{t+t_0} - p_{t_0}|^{1-1/k}\, d\mu_0\right)$$

$$\le C\left(\|r_{t,t_0;1}\|_1 + |t|\|q_{t_0;k}\|_k\|p_{t+t_0} - p_{t_0}\|_1^{1-1/k}\right)$$

$$= C\left(\|\mathbf{r}_1(t,t_0)\|_1 + |t|\|\partial_{\xi_t'}\mathbf{p}^{1/k}\|_k\|\mathbf{p}(t+t_0) - \mathbf{p}(t_0)\|_1^{1-1/k}\right),$$

using Hölder's inequality in the second line. Since $\mathbf{p}$ is a C^1-map, $\|\mathbf{r}_1(t,t_0)\|_1/|t|$ and $\|\mathbf{p}(t+t_0) - \mathbf{p}(t_0)\|_1$ tend to 0, whereas $\|\partial_{\xi_t'}\mathbf{p}^{1/k}\|_k$ is bounded close to t_0 by (C.1) since $t \mapsto \partial_{\xi_t'}\mathbf{p}^{1/k}$ is weakly continuous, so that

$$\frac{1}{|t|}\int_\Omega |r^\varepsilon_{t,t_0;k}|\, d\mu_0 \xrightarrow{t\to 0} 0,$$

which by Lemma 3.3 implies the second assertion. $\qquad\square$

Since by definition the derivative of $V \mapsto \partial_V\mathbf{p}^{1/k}$ is weakly continuous for any k-integrable model, the second assertion of Theorem 3.2 will follow from the following.

Lemma 3.6 *Let $(M, \Omega, \mathbf{p})$ be a parametrized measure model for which the map $V \mapsto \partial_V\mathbf{p}^{1/k}$ is weakly continuous. Then $(M, \Omega, \mathbf{p})$ is weakly k-integrable.*

Proof Let $(\xi_t)_{t\in I}$ be a curve in M, let $\mu_0 \in \mathcal{M}(\Omega)$ be a measure dominating $\mathbf{p}(\xi_t)$ for all t, and define the parametrized measure model $\mathbf{p}_\varepsilon(\xi)$ and $\mathbf{r}_k(t,t_0)$ as in (3.80). By Lemma 3.5, we have for any $A \subseteq \Omega$ and $t_0 \in I$ and the pairing $(\cdot;\cdot)$ from (3.58)

$$0 = \lim_{t\to 0}\frac{1}{t}\left(\mathbf{r}_k(t,t_0); \chi_A\mu_0^{1-1/k}\right) = \lim_{t\to 0}\frac{1}{t}\int_A\left((p^\varepsilon_{t+t_0})^{1/k} - (p^\varepsilon_{t_0})^{1/k} - tq^\varepsilon_{t_0;k}\right) d\mu_0$$

$$= \frac{d}{dt}\bigg|_{t=t_0}\int_A (p^\varepsilon_t)^{1/k}\, d\mu_0 - \int_A q^\varepsilon_{t_0;k}\, d\mu_0,$$

showing that

$$\frac{d}{dt}\bigg|_{t=t_0}\int_A (p^\varepsilon_t)^{1/k}\, d\mu_0 = \int_A q^\varepsilon_{t_0;k}\, d\mu_0 \tag{3.81}$$

for all $t_0 \in I$. As we observed in the proof of Lemma 3.4, the weak continuity of the map $t \mapsto \partial_{\xi_t'}\mathbf{p}_\varepsilon^{1/k}$ implies that the integral on the right-hand side of (3.81) depends continuously on $t_0 \in I$, whence integration implies that for all $a, b \in I$ we have

$$\int_A \left((p^\varepsilon_b)^{1/k} - (p^\varepsilon_a)^{1/k}\right) d\mu_0 = \int_a^b \int_A q^\varepsilon_{t;k}\, d\mu_0\, dt. \tag{3.82}$$

Now $|q_{t;k}^{\varepsilon}| \leq |q_{t;k}|$ and $|(p_b^{\varepsilon})^{1/k} - (p_a^{\varepsilon})^{1/k}\| \leq |p_b - p_a|^{1/k}$ by (3.63), whence we can use dominant convergence for the limit as $\varepsilon \searrow 0$ in (3.82) to conclude that (3.78) holds, so that Lemma 3.4 implies the weak k-integrability of the model. $\square$

We now have to prove the first assertion of Theorem 3.2. We begin by showing the weak continuity of the differential of $\mathbf{p}^{1/k}$.

Lemma 3.7 *Let* $(M, \Omega, \mathbf{p})$ *be a parametrized measure model with* $\partial_V \log \mathbf{p}(\xi) \in L^k(\Omega, \mathbf{p}(\xi))$ *for all* ξ, *and suppose that the function* (3.70) *is continuous. Then the map* $V \mapsto \partial_V \mathbf{p}^{1/k}$ *is weakly continuous, so that the model is weakly k-integrable.*

Proof Let $(V_n)_{n \in \mathbb{N}}$ be a sequence, $V_n \in T_{\xi_n} M$ with $V_n \to V_0 \in T_{\xi_0} M$, and let $\mu_0 \in \mathcal{M}(\Omega)$ be a measure dominating all $\mathbf{p}(\xi_n)$. In fact, we may assume that there is a decomposition $\Omega = \Omega_0 \uplus \Omega_1$ such that

$$\mathbf{p}(\xi_0) = \chi_{\Omega_0} \mu_0.$$

We adapt the notation from (3.76) and define $p_n, q_n \in L^1(\Omega, \mu_0)$ and $q_{n;l} \in L^l(\Omega, \mu_0)$, replacing $t \in I$ by $n \in \mathbb{N}_0$ in (3.76). In particular, $p_0 = \chi_{\Omega_0}$. Then on Ω_0 we have

$$
\begin{aligned}
|q_{n;k} - q_{n;0}| &= \left| \frac{q_n}{k p_n^{1-1/k}} - \frac{q_0}{k} \right| \leq \frac{1}{k} |q_n - q_0| + \frac{|q_n|}{k} \left| \frac{1}{p_n^{1-1/k}} - 1 \right| \\
&= \frac{1}{k} |q_n - q_0| + |q_{n;k}| \left| 1 - p_n^{1-1/k} \right| \\
&\stackrel{(3.63)}{\leq} \frac{1}{k} |q_n - q_0| + |q_{n;k}| |1 - p_n|^{1-1/k} \\
&= \frac{1}{k} |q_n - q_0| + |q_{n;k}| |p_n - p_0|^{1-1/k}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\int_{\Omega_0} |q_{n;k} - q_{n;0}| \, d\mu_0 &\leq \frac{1}{k} \int_{\Omega_0} |q_n - q_0| \, d\mu_0 + \int_{\Omega_0} |q_{n;k}| |p_n - p_0|^{1-1/k} \, d\mu_0 \\
&\leq \frac{1}{k} \|q_n - q_0\|_1 + \|q_{n;k}\|_k \|p_n - p_0\|_1^{1-1/k} \\
&= \frac{1}{k} \|\partial_{V_n} \mathbf{p} - \partial_{V_0} \mathbf{p}\|_1 + \left\| \partial_{V_n} \mathbf{p}^{1/k} \right\|_k \left\| \mathbf{p}(\xi_n) - \mathbf{p}(\xi_0) \right\|_1^{1-1/k}.
\end{aligned}
$$

Since $\mathbf{p}$ is a C^1-map, both $\|\partial_{V_n} \mathbf{p} - \partial_{V_0} \mathbf{p}\|_1$ and $\|\mathbf{p}(\xi_n) - \mathbf{p}(\xi_0)\|_1$ tend to 0, whereas $\|\partial_{V_n} \mathbf{p}^{1/k}\|_k$ tends to $\|\partial_{V_0} \mathbf{p}^{1/k}\|_k$ by the continuity of (3.70). Moreover, $\partial_{V_0} \mathbf{p}$ is dominated by $\mathbf{p}(\xi_0) = \chi_{\Omega_0} \mu_0$, whence q_0 and $q_{0;k}$ vanish on Ω_1. Thus, we conclude that

$$0 = \lim \int_{\Omega_0} |q_{n;k} - q_{n;0}| \, d\mu_0 = \lim \int_{\Omega} |\chi_{\Omega_0} q_{n;k} - q_{n;0}| \, d\mu_0. \qquad (3.83)$$

By Lemma 3.3, this implies that $\chi_{\Omega_0} \partial_{V_n} \mathbf{p}^{1/k} \rightharpoonup \partial_{V_0} \mathbf{p}^{1/k}$, whence by (C.1) we have

$$\|\partial_{V_0} \mathbf{p}\|_k \leq \liminf \|\chi_{\Omega_0} \partial_{V_n} \mathbf{p}\|_k \leq \limsup \|\chi_{\Omega_0} \partial_{V_n} \mathbf{p}\|_k$$
$$\leq \limsup \|\partial_{V_n} \mathbf{p}\|_k = \|\partial_{V_0} \mathbf{p}\|_k,$$

using again the continuity of (3.70). Thus, we have equality in these estimates, i.e.,

$$\|\partial_{V_0} \mathbf{p}\|_k = \lim \|\chi_{\Omega_0} \partial_{V_n} \mathbf{p}\|_k = \lim \|\partial_{V_n} \mathbf{p}\|_k,$$

and since $\|\partial_{V_n} \mathbf{p}\|_k^k = \|\chi_{\Omega_0} \partial_{V_n} \mathbf{p}\|_k^k + \|\chi_{\Omega_1} \partial_{V_n} \mathbf{p}\|_k^k$, this implies that

$$\lim \|\chi_{\Omega_1} \partial_{V_n} \mathbf{p}\|_k = 0.$$

Thus,

$$\lim \int_{\Omega_1} |q_{n;k} - q_{n;0}| \, d\mu_0 = \lim \int_{\Omega_1} |q_{n;k}| \, d\mu_0 = \lim \|\chi_{\Omega_1} \partial_{V_n} \mathbf{p}\|_1 = 0$$

as $\|\chi_{\Omega_1} \partial_{V_n} \mathbf{p}\|_k \to 0$, so that together with (3.83) we conclude that

$$\lim \int_{\Omega} |q_{n;k} - q_{n;0}| \, d\mu_0 = 0,$$

and now, Lemma 3.3 implies that $\partial_{V_n} \mathbf{p}^{1/k} \rightharpoonup \partial_{V_0} \mathbf{p}^{1/k}$ for an arbitrary convergent sequence $(V_n) \in TM$, showing the weak continuity, and the last assertion now follows from Lemma 3.6. □

Lemma 3.8 *The first assertion of Theorem* 3.2 *holds.*

Proof By the definition of k-integrability, the continuity of the map $\mathbf{p}^{1/k}$ for a k-integrable parametrized measure model is evident from the definition, so that the continuity of (3.70) follows.

Thus, we have to show the converse and assume that the map (3.70) is continuous. By Lemma 3.7, this implies that the map $V \mapsto \partial_V \mathbf{p}^{1/k}$ is weakly continuous and hence, the model is weakly k-integrable by Lemma 3.6. In particular, (3.69) holds by Proposition 3.4.

Together with the continuity of the norm, it follows from the Radon–Riesz theorem (cf. Theorem C.3) that the map $V \mapsto \partial_V \mathbf{p}^{1/k}$ is continuous even in the norm of $\mathcal{S}^{1/k}(\Omega)$.

Let $(\xi_t)_{t \in I}$ be a curve in M and let $V := \xi_0' \in T_{\xi_0} M$, and recall the definition of the remainder term $\mathbf{r}_k(t, t_0)$ from (3.75). Thus, what we have to show is that

$$\frac{1}{|t|} \|\mathbf{r}_k(t, t_0)\|_k \xrightarrow{t \to 0} 0. \tag{3.84}$$

By the Hahn–Banach theorem (cf. Theorem C.1), we may for each pair $t, t_0 \in I$ choose an $\alpha \in \mathcal{S}^{1/k}(\Omega)'$ with $\alpha(\mathbf{r}_k(t, t_0)) = \|\mathbf{r}_k(t, t_0)\|_k$ and $\|\alpha\|_k = 1$. Then we have

$$
\begin{aligned}
\|\mathbf{r}_k(t, t_0)\|_k &= \left| \alpha\big(\mathbf{r}_k(t, t_0)\big) \right| = \left| \alpha\big(\mathbf{p}^{1/k}(\xi_{t+t_0}) - \mathbf{p}^{1/k}(\xi_{t_0})\big) - t \partial_{\xi'_{t_0}} \mathbf{p}^{1/k}) \right| \\
&\overset{(3.69)}{=} \left| \int_{t_0}^{t+t_0} \alpha\big(\partial_{\xi'_s} \mathbf{p}^{1/k} - \partial_{\xi'_{t_0}} \mathbf{p}^{1/k}\big) \, ds \right| \\
&\leq \int_{t_0}^{t+t_0} \|\alpha\|_k \left\| \partial_{\xi'_s} \mathbf{p}^{1/k} - \partial_{\xi'_{t_0}} \mathbf{p}^{1/k} \right\|_k \, ds \\
&\leq |t| \max_{|s - t_0| \leq t} \left\| \partial_{\xi'_s} \mathbf{p}^{1/k} - \partial_{\xi'_{t_0}} \mathbf{p}^{1/k} \right\|_k
\end{aligned}
$$

Thus,

$$
\frac{\|\mathbf{r}_k(t, t_0)\|_k}{|t|} \leq \max_{|s - t_0| \leq t} \left\| |\partial_{\xi'_s} \mathbf{p}^{1/k} - \partial_{\xi'_{t_0}} \mathbf{p}^{1/k} \right\|_k,
$$

and by the continuity of the map $V \mapsto \partial_V \mathbf{p}^{1/k}$ in the norm of $\mathcal{S}^{1/k}(\Omega)$, the right-hand side tends to 0 as $t \to 0$, showing (3.84) and hence the claim. $\qquad\square$

3.2.5 Canonical n-Tensors of an n-Integrable Model

We begin this section with the formal definition of an n-tensor on a vector space.

Definition 3.8 Let $(V, \|\cdot\|)$ be a normed vector space (e.g., a Banach space). A *covariant n-tensor on V* is a multilinear map $\Theta : \bigtimes^n V \to \mathbb{R}$ which is continuous w.r.t. the product topology.

We can characterize covariant n-tensors by the following proposition.

Proposition 3.5 *Let $(V, \|\cdot\|)$ be a Banach space and $\Theta : \bigtimes^n V \to \mathbb{R}$ a be multilinear map. Then the following are equivalent.*

(1) *Θ is a covariant n-tensor on V, i.e., continuous w.r.t. the product topology.*
(2) *There is a $C > 0$ such that for $V_1, \ldots, V_n \in V$*

$$
\left| \Theta(V_1, \ldots, V_n) \right| \leq C \|V_1\| \cdots \|V_n\|. \tag{3.85}
$$

Proof To see that the first condition implies (3.85), we proceed by induction on n. For $n = 1$ this is clear as a continuous linear map $\Theta : V \to \mathbb{R}$ is bounded. Suppose that (3.85) holds for all n-tensors and let $\Theta^{n+1} : \bigtimes^{n+1} V \to \mathbb{R}$ be a covariant $(n+1)$-tensor. For fixed $V_1, \ldots, V_n \in V$, the map $\Theta^{n+1}(\cdot, V_1, \ldots, V_n)$ is continuous and hence bounded linear.

On the other hand, for fixed V_0, the map $\Theta^{n+1}(V_0, V_1, \ldots, V_n)$ is a covariant n-tensor and hence by induction hypothesis,

$$\left|\Theta^{n+1}(V_0, V_1, \ldots, V_n)\right| \leq C_{V_0} \quad \text{if } \|V_i\| = 1 \text{ for all } i > 0. \qquad (3.86)$$

The uniform boundedness principle (cf. Theorem C.2) now shows that the constant C_{V_0} in (3.86) can be chosen to be $C\|V_0\|$ for some fixed $C \in \mathbb{R}$, so that (3.85) holds for Θ^{n+1}, completing the induction.

Next, to see that (3.85) implies the continuity of Θ^n, let $(V_i^{(k)})_{k \in \mathbb{N}} \in V$, $i = 1, \ldots, n$ be sequences converging to V_i^0. Then

$$\left|\Theta\left(V_1^{(k)}, \ldots, V_n^{(k)}\right) - \Theta\left(V_1^0, \ldots, V_n^0\right)\right| = \left|\sum_{i=1}^{n} \Theta(V_1^{(k)}, \ldots, V_i^{(k)} - V_i^0, \ldots, V_n^0)\right|$$

$$\overset{(3.85)}{\leq} \sum_{i=1}^{n} C\|V_1^{(k)}\| \cdots \|V_i^{(k)} - V_i^0\| \cdots \|V_n^0\|,$$

and this tends to 0 as $\|V_i^{(k)}\| \xrightarrow{k \to \infty} \|V_i^0\|$ and $\|V_i^{(k)} - V_i^0\| \xrightarrow{k \to \infty} 0$. Thus, Θ is continuous in the product topology. □

Definition 3.9 (Covariant n-tensors on a manifold) Let M be a C^1-manifold. A *covariant n-tensor field on M* is a family $(\Theta_\xi)_{\xi \in M}$ of covariant n-tensor fields on $T_\xi M$ which are weakly continuous, i.e., such that for continuous vector fields $V_1, \ldots, V_n$ on M the function $\Theta(V_1, \ldots, V_n)$ is continuous on M.

An important example of such a tensor is given by the following

Definition 3.10 (Canonical n-tensor) For $n \in \mathbb{N}$, the *canonical n-tensor* is the covariant n-tensor on $\mathcal{S}^{1/n}(\Omega)$, given by

$$L_\Omega^n(v_1, \ldots, v_n) = n^n \int_\Omega d(v_1 \cdots v_n), \quad \text{where } v_i \in \mathcal{S}^{1/n}(\Omega). \qquad (3.87)$$

Moreover, for $0 < r \leq 1/n$ the *canonical n-tensor* $\tau_{\Omega;r}^n$ on $\mathcal{S}^r(\Omega)$ is defined as

$$\left(\tau_{\Omega;r}^n\right)_{\mu_r}(v_1, \ldots, v_n) := \begin{cases} \frac{1}{r^n} \int_\Omega d(v_1 \cdots v_n \cdot |\mu_r|^{1/r-n}) & \text{if } r < 1/n, \\ L_\Omega^n(v_1, \ldots, v_n) & \text{if } r = 1/n, \end{cases} \qquad (3.88)$$

where $v_i \in \mathcal{S}^r(\Omega) = T_{\mu_r}\mathcal{S}^r(\Omega)$.

Observe that for a finite set $\Omega = I$, the definition of L_I^n coincides with the covariant n-tensor given in (2.53).

For $n = 2$, the pairing $(\cdot; \cdot) : \mathcal{S}^{1/2}(\Omega) \times \mathcal{S}^{1/2}(\Omega) \to \mathbb{R}$ from (3.58) satisfies

$$(v_1; v_2) = \frac{1}{4} L_\Omega^2(v_1, v_2).$$

Since $(\nu; \nu) = \|\nu\|_2^2$ by (3.54), it follows:[5]

$$\left(\mathcal{S}^{1/2}(\Omega), \frac{1}{4} L_\Omega^2\right) \text{ is a Hilbert space with norm } \|\cdot\|_2. \tag{3.89}$$

For a C^1-map $\Phi : M_1 \to M_2$ and a covariant n-tensor field Θ on M_2, the pull-back $\Phi^*\Theta$ given by

$$\Phi^*\Theta(V_1, \ldots, V_n) := \Theta\big(d\Phi(V_1), \ldots, d\Phi(V_n)\big) \tag{3.90}$$

is a covariant n-tensor field on M_1. If $\Phi : M_1 \to M_2$ and $\Psi : M_2 \to M_3$ are differentiable, then this immediately implies for the composition $\Psi\Phi : M_1 \to M_3$:

$$(\Psi\Phi)^*\Theta = \Phi^*\Psi^*\Theta. \tag{3.91}$$

Proposition 3.6 *Let $n \in \mathbb{N}$ and $0 < s \leq r \leq 1/n$. Then*

$$\tau_{\Omega;r}^n = \big(\tilde{\pi}^{1/rn}\big)^* L_\Omega^n \tag{3.92}$$

and

$$\big(\tilde{\pi}^{r/s}\big)^* \tau_{\Omega;r}^n = \tau_{\Omega;s}^n, \tag{3.93}$$

with the C^1-maps $\tilde{\pi}^{1/rn}$ and $\tilde{\pi}^{r/s}$ from Proposition 3.2, respectively.

Proof Unwinding the definition we obtain for $k := 1/rn \geq 1$:

$$
\begin{aligned}
\big(\big(\tilde{\pi}^k\big)^* L_\Omega^n\big)_{\mu_r}\big(v_r^1, \ldots, v_r^n\big) &= L_\Omega^n\big(d_{\mu_r}\tilde{\pi}^k\big(v_r^1\big), \ldots, d_{\mu_r}\tilde{\pi}^k\big(v_r^n\big)\big) \\
&\overset{(3.62)}{=} L_\Omega^n\big(k|\mu_r|^{k-1} \cdot v_r^1, \ldots, k|\mu_r|^{k-1} \cdot v_r^n\big) \\
&= k^n n^n \int_\Omega d\big(\big(|\mu_r|^{k-1} \cdot v_r^1\big) \cdots \big(|\mu_r|^{k-1} \cdot v_r^n\big)\big) \\
&= \frac{1}{r^n} \int_\Omega d\big(v_r^1 \cdots v_r^n \cdot |\mu_r|^{1/r-n}\big) \\
&= \big(\tau_{\Omega;r}^n\big)_{\mu_r}\big(v_r^1, \ldots, v_r^n\big),
\end{aligned}
$$

showing (3.92). Thus,

$$
\begin{aligned}
\big(\tilde{\pi}^{r/s}\big)^* \tau_{\Omega;r}^n &= \big(\tilde{\pi}^{r/s}\big)^*\big(\tilde{\pi}^{1/rn}\big)^* L_\Omega^n = \big(\tilde{\pi}^{1/rn}\tilde{\pi}^{r/s}\big)^* L_\Omega^n = \big(\tilde{\pi}^{1/sn}\big)^* L_\Omega^n \\
&= \tau_{\Omega;s}^n,
\end{aligned}
$$

showing (3.93). $\square$

---

[5]Observe that the factor $1/4$ in (3.89) by which the canonical form differs from the Hilbert inner product is responsible for having to use the sphere of radius 2 rather than the unit sphere in Proposition 2.1.

Remark 3.10 If $\Omega = I$ is a finite measurable space, then $\mathcal{M}_+^r(I)$ is an open subset of $\mathcal{S}^r(I)$ and hence a manifold. Moreover, the restrictions $\tilde{\pi}^k : \mathcal{M}_+^r(I) \to \mathcal{M}_+^s(I)$ is a C^1-map even for $k \leq 1$, so that we may use (3.92) to define $\tau_{\Omega;r}^n$ for *all* $r \in (0, 1]$ on $\mathcal{M}_+(I)$.

That is, if $\mu = \sum_{i \in I} m_i \delta^i \in \mathcal{M}_+(I)$ and $V_k = \sum_{i \in I} v_{k;i} \delta^i \in \mathcal{S}(I)$, then

$$\tau_{I;r}^n(V_1, \ldots, V_n) = \frac{1}{r^n} \int_I v_{1;i} \cdots v_{n;i} m_i^{1/r-n} \delta^i = \sum_{i \in I} \frac{1}{r^n m_i^{n-1/r}} v_{1;i} \cdots v_{n;i}. \tag{3.94}$$

For $r = 1$, the tensor $\tau_{I;1}^n$ coincides with the tensor τ_I^n in (2.76).

This explains on a more conceptual level why $\tau_I^n = \tau_{I;1}^n$ cannot be extended from $\mathcal{M}_+(I)$ to $\mathcal{S}(I)$, whereas $L_I^2 = \tilde{\pi}^{1/2} \tau_I^n$ *can* be extended to a tensor on $\mathcal{S}^2(I)$, cf. (2.52).

If $(M, \Omega, \mathbf{p})$ is an n-integrable parametrized measure model, then we define the *canonical n-tensor of $(M, \Omega, \mathbf{p})$* as

$$\tau_M^n = \tau_{(M,\Omega,\mathbf{p})}^n := (\mathbf{p}^{1/n})^* L_\Omega^n, \tag{3.95}$$

so that by (3.68)

$$\tau_M^n(V_1, \ldots, V_n) := L_\Omega^n \big(d_\xi \mathbf{p}^{1/n}(V_1), \ldots, d_\xi \mathbf{p}^{1/n}(V_n) \big)$$

$$= \int_\Omega \partial_{V_1} \log \mathbf{p}(\xi) \cdots \partial_{V_n} \log \mathbf{p}(\xi) \, d\mathbf{p}(\xi), \tag{3.96}$$

where the second line follows immediately from (3.68) and (3.87). That is, τ_M^n coincides with the tensor given in (3.44).

Observe that τ_M^n is indeed continuous, since $\tau_M^n(V, \ldots, V) = n^n \|d_\xi \mathbf{p}^{1/n}(V)\|_n^n$ is continuous for all V.

Example 3.5

(1) For $n = 1$, the canonical 1-form is given as

$$\tau_M^1(V) := \int_\Omega \partial_V \log \mathbf{p}(\xi) \, d\mathbf{p}(\xi) = \partial_V \big\| \mathbf{p}(\xi) \big\|. \tag{3.97}$$

Thus, it vanishes if and only if $\|\mathbf{p}(\xi)\|$ is locally constant, e.g., if $(M, \Omega, \mathbf{p})$ is a statistical model.

(2) For $n = 2$, τ_M^2 coincides with the *Fisher metric* $\mathfrak{g}$ from (3.41).

(3) For $n = 3$, τ_M^3 coincides with the *Amari–Chentsov 3-symmetric tensor* **T** from (3.42).

Remark 3.11 While the above examples show the statistical significance of τ_M^n for $n = 1, 2, 3$, we shall show later that the tautological forms of even degree τ_M^{2n} can

be used to measure the information loss of statistics and Markov kernels, cf. Theorem 5.5. Moreover, in [160, p. 212] the question is posed if there are other significant tensors on statistical manifolds, and the canonical n-tensors may be considered as natural candidates.

3.2.6 Signed Parametrized Measure Models

In this section, we wish to comment on the generalization of Definition 3.4 to families of finite *signed* measures $(\mathbf{p}(\xi))_{\xi \in M}$, i.e., dropping the assumption that $\mathbf{p}(\xi)$ is a non-negative measure. That is, we may simply consider C^1-maps $\mathbf{p} : M \to \mathcal{S}(\Omega)$.

However, with this approach there is no notion of k-integrability or of canonical tensors, as the term $\partial_V \log \mathbf{p}(\xi)$ from (3.66), which is necessary to define these notions, cannot be given sense without further assumptions.

For instance, if $\Omega = \{\omega\}$ is a singleton and $\mathbf{p}(\xi) := \xi \delta^\omega$ for $\xi \in \mathbb{R}$, then $\mathbf{p} : \mathbb{R} \to \mathcal{S}(\Omega)$ is certainly a C^1-map, but $\log \mathbf{p}(\xi) = \log \xi$ cannot be continuously extended at $\xi = 0$.

Therefore, we shall make the following definition.

Definition 3.11   Let Ω be a measurable space and M be a (Banach-)manifold.

(1) A *signed parametrized measure model* is a triple $(M, \Omega, \mathbf{p})$, where M is a (finite or infinite-dimensional) Banach manifold and $\mathbf{p} : M \to \mathcal{S}(\Omega)$ is a C^1-map in the sense of Sect. 3.2.2.
(2) A signed parametrized measure model $(M, \Omega, \mathbf{p})$ is said to have a *logarithmic derivative* if for each $\xi \in M$ and $V \in T_\xi M$ the derivative $d_\xi \mathbf{p}(V) \in \mathcal{S}(\Omega)$ is dominated by $|\mathbf{p}(\xi)|$. In this case, we define analogously to (3.66)

$$\partial_V \log |\mathbf{p}(\xi)| := \frac{d\{d_\xi \mathbf{p}(V)\}}{d\mathbf{p}(\xi)}. \tag{3.98}$$

Here, for signed measures $\mu, \nu \in \mathcal{S}(\Omega)$ such that $|\mu|$ dominates ν, the Radon–Nikodym derivative $\dfrac{d\nu}{d\mu}$ is the unique function in $L^1(\Omega, |\mu|)$ such that

$$\nu = \frac{d\nu}{d\mu}\mu.$$

Just as in the non-signed case, we can now consider the $(1/k)$th power of $\mathbf{p}$. Here, we shall use the signed power in order not to lose information on $\mathbf{p}(\xi)$. That is, we consider the (continuous) map

$$\tilde{\mathbf{p}}^{1/k} : M \longrightarrow \mathcal{S}^{1/k}(\Omega), \qquad \xi \longmapsto \tilde{\pi}^{1/k} \mathbf{p}(\xi) =: \tilde{\mathbf{p}}(\xi)^{1/k}.$$

Assuming that this map is differentiable and differentiating the equation $\mathbf{p} = \tilde{\pi}^k \tilde{\mathbf{p}}^{1/k}$ just as in the non-signed case, we obtain in analogy to (3.68)

$$d_\xi \tilde{\mathbf{p}}^{1/k}(V) := \frac{1}{k} \, \partial_V \log|\mathbf{p}(\xi)| \, \tilde{\mathbf{p}}^{1/k}(\xi) \in \mathcal{S}^{1/k}\big(\Omega, |\mathbf{p}(\xi)|\big) \qquad (3.99)$$

so that, in particular, $\partial_V \log|\mathbf{p}(\xi)| \in L^k(\Omega, |\mathbf{p}(\xi)|)$, and in analogy to Definition 3.7 we define the following.

Definition 3.12 (k-Integrable signed parametrized measure model)   A signed parametrized measure model $(M, \Omega, \mathbf{p})$ is called *k-integrable* for $k \geq 1$ if it has a logarithmic derivative and the map $\tilde{\mathbf{p}}^{1/k} : M \to \mathcal{S}^{1/k}(\Omega)$ is a C^1-map. Furthermore, we call the model *∞-integrable* if it is k-integrable for all $k \geq 1$.

Just as in the non-signed case, this allows us to define the canonical n-tensor for an n-integrable signed parametrized measure model. Namely, in analogy to (3.95) and (3.96) we define for an n-integrable signed parametrized measure model $(M, \Omega, \mathbf{p})$ the *canonical n-tensor of $(M, \Omega, \mathbf{p})$* as

$$\tau_M^n = \tau_{(M, \Omega, \mathbf{p})}^n := \big(\tilde{\mathbf{p}}^{1/n}\big)^* L_\Omega^n, \qquad (3.100)$$

so that by (3.99) and (3.87)

$$\tau_M^n(V_1, \ldots, V_n) := L_\Omega^n\big(d_\xi \tilde{\mathbf{p}}^{1/n}(V_1), \ldots, d_\xi \tilde{\mathbf{p}}^{1/n}(V_n)\big)$$

$$= \int_\Omega \partial_{V_1} \log|\mathbf{p}(\xi)| \cdots \partial_{V_n} \log|\mathbf{p}(\xi)| \, d\mathbf{p}(\xi).$$

Example 3.6

(1) For $n = 1$, the canonical 1-form is given as

$$\tau_M^1(V) := \int_\Omega \partial_V \log|\mathbf{p}(\xi)| \, dp(\xi) = \partial_V\big(\mathbf{p}(\xi)(\Omega)\big). \qquad (3.101)$$

Thus, it vanishes if and only if $\mathbf{p}(\xi)(\Omega)$ is locally constant, but of course, as $\mathbf{p}(\xi)$ is a signed measure, in general this quantity does not equal $\|\mathbf{p}(\xi)\|_{TV} = |\mathbf{p}(\xi)|(\Omega)$.

(2) For $n = 2$ and $n = 3$, τ_M^2 and τ_M^3 coincide with the Fisher metric and the Amari–Chentsov 3-symmetric tensor $\mathbf{T}$ from (3.41) and (3.42), respectively. That is, for signed parametrized measure models which are k-integrable for $k \geq 2$ or $k \geq 3$, respectively, the tensors $\mathfrak{g}_M$ and $\mathbf{T}$, respectively, still can be defined. Note, however, that $\mathfrak{g}_M$ may fail to be positive definite. In fact, it may even be degenerate, so that $\mathfrak{g}_M$ does not necessarily yield a Riemannian (and possibly not even a pseudo-Riemannian) metric on M for signed parametrized measure models.

3.3 The Pistone–Sempi Structure

The definition of (k-integrable) parametrized measure models (statistical models, respectively) discussed in Sect. 3.2 is strongly inspired by the definition of Amari [9]. There is, however, another beautiful concept of geometrizing the space of finite measures (probability measures, respectively) which was first suggested by Pistone and Sempi in [216].

This approach is, on the one hand, more restrictive, since instead of geometrizing all of $\mathcal{M}(\Omega)$, it only geometrizes the subset $\mathcal{M}_+ = \mathcal{M}_+(\Omega, \mu_0)$ of finite measures compatible with a fixed measure μ_0, where two measures are called compatible if they have the same null sets. On the other hand, it defines on $\mathcal{M}_+$ the structure of a Banach manifold, whence a much stronger geometric structure than that of $\mathcal{M}(\Omega)$ defined in Sect. 3.2. We shall refer to this as the *Pistone–Sempi structure on* $\mathcal{M}_+(\Omega, \mu_0)$.

The starting point is to define a topology on $\mathcal{M}_+$, called the *topology of e-convergence* or *e-topology*, cf. Definition 3.13 below. For this topology, the inclusion $\mathcal{M}_+ = \mathcal{M}_+(\Omega, \mu_0) \hookrightarrow \mathcal{M}(\Omega)$ is continuous, where $\mathcal{M}(\Omega)$ is equipped with the L^1-topology induced by the inclusion $\mathcal{M}(\Omega, \mu_0) \hookrightarrow \mathcal{S}(\Omega)$.

Since $\mathcal{M}_+$ is a Banach manifold, it is natural to consider models which are given by C^1-maps $\mathbf{p} : M \to \mathcal{M}_+$ from a (Banach-)manifold M to the Banach manifold $\mathcal{M}_+$. As it turns out, such a map, when regarded as a parametrized measure model $\mathbf{p} : M \to \mathcal{M}_+ \hookrightarrow \mathcal{M}(\Omega)$ is ∞-integrable in the sense of Definition 3.7. The converse, however, is far from being true in general, see Example 3.8 below.

In fact, we shall show that in the *e*-topology, the set $\mathcal{M}_+$ decomposes into several connected components each of which can be canonically identified with a convex open set in a Banach space, and this induces the Banach manifold structure established in [216].

We shall begin our discussion by introducing the notion of the *e*-topology, which is defined using the notion of convergence of sequences. We shall then recall some basic facts about Orlicz spaces and show that to each $\mu \in \mathcal{M}_+$ there is an associated *exponential tangent space* $T_\mu \mathcal{M}_+$ which is an Orlicz space. It then turns out that the image of the injective map

$$\log_{\mu_0} : e^f \mu_0 \longmapsto f$$

maps the connected component of μ in $\mathcal{M}_+$ into an open convex subset of $T_\mu \mathcal{M}_+$. We also refer to the results in [230] where this division of $\mathcal{M}_+$ into open cells is established as well.

3.3.1 e-Convergence

Definition 3.13 (Cf. [216]) Let $(g_n)_{n \in \mathbb{N}}$ be a sequence of measurable functions in the finite measure space (Ω, μ), and let $g \in L^1(\Omega, \mu)$ be such that $g_n, g > 0$ μ-a.e. We say that $(g_n)_{n \in \mathbb{N}}$ *is e-convergent to* g if the sequences (g_n/g) and (g/g_n)

converge to 1 in $L^p(\Omega, g\mu)$ for all $p \geq 1$. In this case, we also say that the sequence of measures $(g_n\mu)_{n\in\mathbb{N}}$ is e-convergent to the measure $g\mu$.

It is evident from this definition that the measure μ may be replaced by a compatible measure $\mu' \in \mathcal{M}_+(\Omega, \mu)$ without changing the notion of e-convergence. There are equivalent reformulations of e-convergence given as follows.

Proposition 3.7 (Cf. [216]) *Let $(g_n)_{n\in\mathbb{N}}$ and g be as above. Then the following are equivalent*:

(1) *$(g_n)_{n\in\mathbb{N}}$ is e-convergent to g.*
(2) *For all $p \geq 1$, we have*

$$\lim_{n\to\infty} \int_\Omega \left| \left(\frac{g_n}{g}\right)^p - 1 \right| g\, d\mu = \lim_{n\to\infty} \int_\Omega \left| \left(\frac{g}{g_n}\right)^p - 1 \right| g\, d\mu = 0.$$

(3) *The following conditions hold*:

    (a) *(g_n) converges to g in $L^1(\Omega, \mu)$,*
    (b) *for all $p \geq 1$ we have*

$$\limsup_{n\to\infty} \int_\Omega \left(\frac{g_n}{g}\right)^p g\, d\mu < \infty \quad \text{and} \quad \limsup_{n\to\infty} \int_\Omega \left(\frac{g}{g_n}\right)^p g\, d\mu < \infty.$$

For the proof, we shall use the following simple

Lemma 3.9 *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions in (Ω, μ) such that*

(1) *$\lim_{n\to\infty} \|f_n\|_1 = 0$,*
(2) *$\limsup_{n\to\infty} \|f_n\|_p < \infty$ for all $p > 1$.*

Then $\lim_{n\to\infty} \|f_n\|_p = 0$ for all $p \geq 1$, i.e., (f_n) converges to 0 in $L^p(\Omega, \mu)$ for all $p \geq 1$.

Proof For $p \geq 1$, we have by Hölder's inequality

$$\|f_n\|_p^p = \int_\Omega |f_n|^p d\mu = \int_\Omega |f_n|^{1/2} |f_n|^{p-1/2}\, d\mu$$

$$\leq \left\| f_n^{1/2} \right\|_2 \left\| f_n^{p-1/2} \right\|_2 = \|f_n\|_1^{1/2} \|f_n\|_{2p-1}^{(2p-1)/2}.$$

Since $\|f_n\|_1 \to 0$ and $\|f_n\|_{2p-1}$ is bounded, $\|f_n\|_p \to 0$ follows. $\square$

Proof of Proposition 3.7 (1) $\Rightarrow$ (2) Note that the expressions $|(g_n/g)^p - 1|$ and $|(g/g_n)^p - 1|$ are increasing in p, hence we may assume w.l.o.g. that $p \in \mathbb{N}$. Let $f_n := g_n/g - 1$, so that by hypothesis $\lim_{n\to\infty} \int_\Omega |f_n|^p g\, d\mu = 0$ for all $p \geq 1$.

Then

$$\int_{\Omega} \left| \left(\frac{g_n}{g} \right)^p - 1 \right| g \, d\mu = \int_{\Omega} \left| (1 + f_n)^p - 1 \right| g \, d\mu$$

$$\leq \sum_{k=1}^{p} \binom{p}{k} \int_{\Omega} |f_n|^k g \, d\mu \longrightarrow 0,$$

and the other assertion follows analogously.

(2) $\Rightarrow$ (3) The first equation in (2) for $p = 1$ reads

$$0 = \lim_{n \to \infty} \int_{\Omega} \left| \frac{g_n}{g} - 1 \right| g \, d\mu = \lim_{n \to \infty} \int_{\Omega} |g_n - g| \, d\mu,$$

so that (3)(a) is satisfied. Moreover, it is evident that the convergence conditions in (2) imply the boundedness of (g_n/g) and (g/g_n) in $L^p(\Omega, g\mu)$.

(3) $\Rightarrow$ (1) Again, let $f_n := g_n/g - 1$. Then (3)(a) implies that (f_n) converges to 0 in $L^1(\Omega, g\mu)$, and the first condition in (3)(b) implies that (f_n) is bounded in $L^p(\Omega, g\mu)$ for all $p > 1$. This together with Lemma 3.9 implies that (f_n) converges to 0 and hence (g_n/g) to 1 in $L^p(\Omega, g\mu)$ for all $p \geq 1$.

Note that $g/g_n - 1 = -f_n \cdot (g/g_n)$. By the above, (f_n) tends to 0 in $L^2(\Omega, g\mu)$, and (g/g_n) is bounded in that space by (3)(b). Thus, by Hölder's inequality, $(g/g_n - 1) = -f_n \cdot (g/g_n)$ tends to 0 in $L^1(\Omega, g\mu)$ and, moreover, this sequence is bounded in $L^p(\Omega, g\mu)$ for all $p \geq 1$ by the second condition in (3)(b). Thus, (g/g_n) tends to 1 in $L^p(\Omega, g\mu)$ for all $p \geq 1$ by Lemma 3.9. $\square$

3.3.2 Orlicz Spaces

In this section, we recall the theory of Orlicz spaces which is needed in Sect. 3.3.3 for the description of the geometric structure on $\mathcal{M}(\Omega)$. Most of the results can be found, e.g., in [153].

A function $\phi : \mathbb{R} \to \mathbb{R}$ is called a *Young function* if $\phi(0) = 0$, ϕ is even, convex, strictly increasing on $[0, \infty)$ and $\lim_{t \to \infty} t^{-1} \phi(t) = \infty$. Given a finite measure space (Ω, μ) and a Young function ϕ, we define the *Orlicz space*

$$L^{\phi}(\mu) := \left\{ f : \Omega \to \mathbb{R} \, \middle| \, \int_{\Omega} \phi(\varepsilon f) \, d\mu < \infty \text{ for some } \varepsilon > 0 \right\}.$$

The elements of $L^{\phi}(\mu)$ are called *Orlicz functions* of (ϕ, μ). Convexity of ϕ and $\phi(0) = 0$ implies

$$\phi(cx) \leq c\phi(x) \quad \text{and} \quad \phi(c^{-1}x) \geq c^{-1}\phi(x) \quad \text{for all } c \in (0, 1), x \in \mathbb{R}. \quad (3.102)$$

We define the *Orlicz norm* on $L^\phi(\mu)$ as

$$\|f\|_{\phi,\mu} := \inf\left\{a > 0 \,\Big|\, \int_\Omega \phi\left(\frac{f}{a}\right) d\mu \le 1\right\}.$$

If f is an Orlicz function and $\varepsilon > 0$ and $K \ge \int_\Omega \phi(\frac{f}{a})\,d\mu \ge 1$, then

$$\int_\Omega \phi(\varepsilon f)\,d\mu \le K \Rightarrow \int_\Omega \phi(K^{-1}\varepsilon f)\,d\mu \overset{(3.102)}{\le} K^{-1}\int_\Omega \phi(\varepsilon f)\,d\mu = 1,$$

whence

$$\int_\Omega \phi(\varepsilon f)\,d\mu \le K \Rightarrow \|f\|_{\phi,\mu} \le \frac{K}{\varepsilon}, \tag{3.103}$$

as long as $K \ge 1$. In particular, every Orlicz function has finite norm.

Observe that the infimum in the definition of the norm is indeed attained, as for a sequence $(a_n)_{n\in\mathbb{N}}$ descending to $\|f\|_{\phi,\mu}$ the sequence $g_n := \phi(f/a_n)$ is monotonically increasing as ϕ is even and increasing on $[0, \infty)$. Thus, by the monotone convergence theorem,

$$\int_\Omega \phi\left(\frac{f}{\|f\|_{\phi,\mu}}\right) d\mu = \lim_{n\to\infty} \int_\Omega \phi\left(\frac{f}{a_n}\right) d\mu \le 1.$$

We assert that $\|\cdot\|_{\phi,\mu}$ is indeed a norm on $L^\phi(\mu)$. For the positive definiteness, suppose that $\|f\|_{\phi,\mu} = 0$. Then $\int_\Omega \phi(nf)\,d\mu \le 1$ for all $n \in \mathbb{N}$, and again by the monotone convergence theorem,

$$1 \ge \lim_{n\to\infty} \int_\Omega \phi(nf)\,d\mu = \int_\Omega \lim_{n\to\infty} \phi(nf)\,d\mu,$$

so that, in particular, $\lim_{n\to\infty} \phi(nf(\omega)) < \infty$ for a.e. $\omega \in \Omega$. But since $\lim_{t\to\infty} \phi(t) = \infty$, this implies that $f(\omega) = 0$ for a.e. $\omega \in \Omega$, as asserted.

The homogeneity $\|cf\|_{\phi,\mu} = |c|\,\|f\|_{\phi,\mu}$ is immediate from the definition. Finally, for the triangle inequality let $f_1, f_2 \in L^\phi(\Omega, \mu)$ and let $c_i := \|f_i\|_{\phi,\mu}$. Then the convexity of ϕ implies

$$\int_\Omega \phi\left(\frac{f_1 + f_2}{c_1 + c_2}\right) d\mu = \int_\Omega \phi\left(\frac{c_1}{c_1 + c_2}\frac{f_1}{c_1} + \frac{c_2}{c_1 + c_2}\frac{f_2}{c_2}\right) d\mu$$

$$\le \frac{c_1}{c_1 + c_2}\int_\Omega \phi\left(\frac{f_1}{c_1}\right) d\mu + \frac{c_2}{c_1 + c_2}\int_\Omega \phi\left(\frac{f_2}{c_2}\right) d\mu$$

$$\le 1.$$

Thus, $\|f_1 + f_2\|_{\phi,\mu} \le c_1 + c_2 = \|f_1\|_{\phi,\mu} + \|f_2\|_{\phi,\mu}$ so that the triangle inequality holds.

Example 3.7 For $p > 1$, let $\phi(t) := |t|^p$. This is then a Young function, and $L^\phi(\mu) = L^p(\Omega, \mu)$ as normed spaces.

In this example, we could also consider the case $p = 1$, even though $\phi(t) = |t|$ is *not* a Young function as it fails to meet the condition $\lim_{t \to \infty} t^{-1}\phi(t) = \infty$. However, this property of Young functions was not used in the verification of the norm properties of $\| \cdot \|_{\phi,\mu}$.

Proposition 3.8   *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $L^\phi(\mu)$. Then the following are equivalent:*

(1) *$\lim_{n\to\infty} f_n = 0$ in the Orlicz norm.*
(2) *There is a $K > 0$ such that for all $c > 0$, $\limsup_{n\to\infty} \int_\Omega \phi(cf_n)\,d\mu \leq K$.*
(3) *For all $c > 0$, $\lim_{n\to\infty} \int_\Omega \phi(cf_n)\,d\mu = 0$.*

Proof Suppose that (1) holds. Then for any $c > 0$ and $\varepsilon \in (0, 1)$ we have

$$
\varepsilon \geq \varepsilon \limsup_{n\to\infty} \underbrace{\int_\Omega \phi\big(c\varepsilon^{-1}f_n\big)\,d\mu}_{\leq 1 \text{ if } \|f_n\|_{\phi,\mu} \leq \varepsilon/c} \overset{(3.102)}{\geq} \limsup_{n\to\infty} \int_\Omega \phi(cf_n)\,d\mu,
$$

and since $\varepsilon \in (0, 1)$ is arbitrary, (3) follows. Obviously, (3) implies (2), and if (2) holds for some K, then assuming w.l.o.g. that $K \geq 1$, we conclude from (3.103) that $\limsup_{n\to\infty} \|f_n\|_{\phi,\mu} \leq Kc^{-1}$ for all $c > 0$, whence $\lim_{n\to\infty} \|f_n\|_{\phi,\mu} = 0$, which shows (1). $\square$

Now let us investigate how the Orlicz spaces behave under a change of the Young function ϕ.

Proposition 3.9   *Let (Ω, μ) be a finite measure space, and let $\phi_1, \phi_2 : \mathbb{R} \to \mathbb{R}$ be two Young functions. If*

$$
\limsup_{t\to\infty} \frac{\phi_1(t)}{\phi_2(t)} < \infty,
$$

then $L^{\phi_2}(\mu) \subseteq L^{\phi_1}(\mu)$, and the inclusion is continuous, i.e., $\|f\|_{\phi_1,\mu} \leq c\, \|f\|_{\phi_2,\mu}$ for some $c > 0$ and all $f \in L^{\phi_2}(\mu)$. In particular, if

$$
0 < \liminf_{t\to\infty} \frac{\phi_1(t)}{\phi_2(t)} \leq \limsup_{t\to\infty} \frac{\phi_1(t)}{\phi_2(t)} < \infty,
$$

then $L^{\phi_1}(\mu) = L^{\phi_2}(\mu)$, and the Orlicz norms $\| \cdot \|_{\phi_1,\mu}$ and $\| \cdot \|_{\phi_2,\mu}$ are equivalent.

Proof By our hypothesis, $\phi_1(t) \leq K\phi_2(t)$ for some $K \geq 1$ and all $t \geq t_0$. Let $f \in L^{\phi_2}(\mu)$ and $a := \|f\|_{\phi_2,\mu}$. Moreover, decompose

$$
\Omega := \Omega_1 \uplus \Omega_2 \quad \text{with} \quad \Omega_1 := \big\{\omega \in \Omega \,\big|\, \big|f(\omega)\big| \geq at_0\big\}.
$$

Then

$$K \geq K \int_\Omega \phi_2\left(\frac{|f|}{a}\right) d\mu \geq \int_{\Omega_1} K\phi_2\left(\frac{|f|}{a}\right) d\mu$$

$$\geq \int_{\Omega_1} \phi_1\left(\frac{|f|}{a}\right) d\mu \quad \text{as } \frac{|f|}{a} \geq t_0 \text{ on } \Omega_1$$

$$= \int_\Omega \phi_1\left(\frac{|f|}{a}\right) d\mu - \int_{\Omega_2} \phi_1\left(\frac{|f|}{a}\right) d\mu$$

$$\geq \int_\Omega \phi_1\left(\frac{|f|}{a}\right) d\mu - \int_{\Omega_2} \phi_1(t_0)\, d\mu \quad \text{as } \frac{|f|}{a} < t_0 \text{ on } \Omega_2$$

$$\geq \int_\Omega \phi_1\left(\frac{|f|}{a}\right) d\mu - \phi_1(t_0)\mu(\Omega).$$

Thus, $\int_\Omega \phi_1(\frac{|f|}{a}) \leq K + \phi_1(t_0)\mu(\Omega) =: c$, hence $f \in L^{\phi_1}(\mu)$. As $c \geq 1$, (3.103) this implies that $\|f\|_{\phi_1,\mu} \leq ca = c\|f\|_{\phi_2,\mu}$, and this proves the claim. $\square$

The following lemma is a straightforward consequence of the definitions and we omit the proof.

Lemma 3.10 *Let (Ω, μ) be a finite measure space, let $\phi : \mathbb{R} \to \mathbb{R}$ be a Young function, and let $\tilde{\phi}(t) := \phi(\lambda t)$ for some constant $\lambda > 0$.*

Then $\tilde{\phi}$ is also a Young function. Moreover, $L^\phi(\mu) = L^{\tilde{\phi}}(\mu)$ and $\|\cdot\|_{\tilde{\phi},\mu} = \lambda\|\cdot\|_{\phi,\mu}$, so that these norms are equivalent.

Furthermore, we investigate how the Orlicz spaces relate when changing the measure μ to another measure $\mu' \in \mathcal{M}(\Omega, \mu)$.

Proposition 3.10 *Let $0 \neq \mu' \in \mathcal{M}(\Omega, \mu)$ be a measure such that $d\mu'/d\mu \in L^p(\Omega, \mu)$ for some $p > 1$, and let $q > 1$ be the dual index, i.e., $p^{-1} + q^{-1} = 1$. Then for any Young function ϕ we have*

$$L^{\phi^q}(\mu) \subseteq L^\phi(\mu'),$$

and this embedding is continuous.

Proof Let $h := d\mu'/d\mu \in L^p(\Omega, \mu)$ and $c := \|h\|_p > 0$. If $f \in L^{\phi^q}(\mu)$ and $a := \|f\|_{\phi^q,\mu}$, then by Hölder's inequality we have

$$\int_\Omega \phi\left(\frac{|f|}{a}\right) d\mu' = \int_\Omega \phi\left(\frac{|f|}{a}\right) h\, d\mu \leq c\left\|\phi\left(\frac{|f|}{a}\right)\right\|_q = c \underbrace{\left\|\phi^q\left(\frac{|f|}{a}\right)\right\|_1^{1/q}}_{\leq 1} \leq c.$$

Thus, $f \in L^\phi(\mu')$, and (3.103) implies that $\|f\|_{\phi,\mu'} \leq ca = c\|f\|_{\phi^q,\mu}$, which shows the claim. $\square$

Finally, we show that the Orlicz norms are complete.

Theorem 3.3 *Let ϕ be a Young function. Then $(L^\phi(\mu), \|\cdot\|_{\phi,\mu})$ is a Banach space.*

Proof Since $\lim_{t\to\infty}\phi(t)/t = 0$, Proposition 3.9 implies that we have a continuous inclusion $(L^\phi(\mu), \|\cdot\|_{\phi,\mu}) \hookrightarrow L^1(\Omega,\mu)$. In particular, any Cauchy sequence $(f_n)_{n\in\mathbb{N}}$ in $L^\phi(\mu)$ is a Cauchy sequence in $L^1(\Omega,\mu)$, and since the latter space is complete, it L^1-converges to a limit function $f \in L^1(\Omega,\mu)$. Therefore, after passing to a subsequence, we may assume that $f_n \to f$ pointwise almost everywhere.

Let $\varepsilon > 0$. Then there is an $N(\varepsilon)$ such that for all $n,m \geq N(\varepsilon)$ we have $\|f_n - f_m\|_{\phi,\mu} < \varepsilon$, that is, for all $n,m \geq N(\varepsilon)$ we have

$$\int_\Omega \phi\big(\varepsilon^{-1}(f_m - f_n)\big)\, d\mu \leq 1.$$

Taking the pointwise limit $n \to \infty$, Fatou's lemma yields

$$\int_\Omega \phi\big(\varepsilon^{-1}(f_m - f)\big)\, d\mu \leq \liminf_{n\to\infty} \int_\Omega \phi\big(\varepsilon^{-1}(f_m - f_n)\big)\, d\mu \leq 1,$$

which implies that $f_m - f \in L^\phi(\mu)$ and $\|f_m - f\|_{\phi,\mu} \leq \varepsilon$ for all $m \geq N(\varepsilon)$. Therefore, $f \in L^\phi(\mu)$ and $\lim_{m\to\infty} \|f_m - f\|_{\phi,\mu} = 0$. $\qquad\square$

3.3.3 Exponential Tangent Spaces

For an arbitrary $\mu \in \mathcal{M}_+$, we define the set

$$\hat{B}_\mu(\Omega) := \big\{ f : \Omega \to [-\infty, +\infty], |f| < \infty \ \mu\text{-a.e.} : e^f \in L^1(\Omega,\mu) \big\},$$

which by Hölder's inequality is a convex subset of the space of measurable functions $\Omega \to [-\infty, +\infty]$. For μ_0, there is a bijection

$$\log_{\mu_0} : \mathcal{M}_+ \longrightarrow \hat{B}_{\mu_0}(\Omega), \qquad \phi\mu_0 \longmapsto \log(\phi).$$

That is, $\log_{\mu_0}$ canonically identifies $\mathcal{M}_+$ with a convex set. Replacing μ_0 by a measure $\mu_0' \in \mathcal{M}_+$, we have $\log_{\mu_0'} = \log_{\mu_0} -u$, where $u := \log_{\mu_0'} \mu_0$. Moreover, we let

$$B_\mu(\Omega) := \hat{B}_\mu(\Omega) \cap \big(-\hat{B}_\mu(\Omega)\big)$$

$$= \big\{ f : \Omega \to [-\infty, \infty] \,\big|\, e^{\pm f} \in L^1(\Omega,\mu) \big\}$$

$$= \big\{ f : \Omega \to [-\infty, \infty] \,\big|\, e^{|f|} \in L^1(\Omega,\mu) \big\}$$

and

$$B_\mu^0(\Omega) := \big\{ f \in B_\mu(\Omega) \,\big|\, (1+s)f \in B_\mu(\Omega) \text{ for some } s > 0 \big\}.$$

The points of $B_\mu^0(\Omega)$ are called *inner points* of $B_\mu(\Omega)$.

Definition 3.14   Let $\mu \in \mathcal{M}_+$. Then

$$T_\mu \mathcal{M}_+ := \left\{ f : \Omega \to [-\infty, \infty] \,\middle|\, tf \in B_\mu(\Omega) \text{ for some } t \neq 0 \right\}$$

is called the *exponential tangent space of* $\mathcal{M}_+$ *at* μ.

Evidently, $f \in T_\mu \mathcal{M}_+$ iff $\int_\Omega (\exp|tf| - 1)\, d\mu < \infty$ for some $t > 0$. Thus,

$$T_\mu \mathcal{M}_+ = L^{\exp|t|-1}(\mu),$$

and hence is an Orlicz space, i.e., it has a Banach norm.[6] If $\|f\|_{L^{\exp|t|-1}(\mu)} < 1$, then

$$\int_\Omega (\exp|f| - 1)\, d\mu \leq 1 \quad \Longrightarrow \quad \int_\Omega e^{|f|}\, d\mu < \infty \Longrightarrow f \in B_\mu(\Omega).$$

That is, $B_\mu(\Omega) \subseteq T_\mu \mathcal{M}_+$ contains the unit ball w.r.t. the Orlicz norm and hence is a neighborhood of the origin. Furthermore, $\lim_{t \to \infty} t^p / (\exp|t| - 1) = 0$ for all $p \geq 1$, so that Proposition 3.9 implies that

$$L^\infty(\Omega, \mu) \subseteq T_\mu \mathcal{M}_+ \subseteq \bigcap_{p \geq 1} L^p(\Omega, \mu), \tag{3.104}$$

where all inclusions are continuous.

Observe that the inclusions in (3.104) are proper, in general. As an example, let $\Omega = (0, 1)$ be the unit interval, and let $\mu = dt$ be the Lebesgue measure.

Let $f(t) := (\log t)^2$. Since $\int_0^1 |\log t|^n dt = n!$ for all $n \in \mathbb{N}$, it follows that $f \in L^p((0, 1), dt)$ for all p. However, for $x > 0$ we have

$$\int_0^1 \exp\bigl(x f(t)\bigr)\, dt = \sum_{n=0}^\infty \frac{1}{n!} x^n \int_0^1 \log(t)^{2n}\, dt = \sum_{n=0}^\infty \frac{(2n)!}{n!} x^n.$$

But this power series diverges for all $x \neq 0$, hence $(\log t)^2 \notin T_\mu(\Omega, \mu)$.

For the first inclusion, observe that $|\log t| \in T_\mu(\Omega, \mu)$ as $\exp(\alpha|\log t|) = t^{-\alpha}$, which is integrable for $\alpha < 1$. However, $|\log t|$ is unbounded and hence not in $L^\infty((0, 1), dt)$.

Remark 3.12   In [106, Definition 6], $T_\mu \mathcal{M}_+$ is called the *Cramer class* of μ. Moreover, in [106, Proposition 7] (see also [216, Definition 2.2]), the *centered Cramer class* is defined as the functions $u \in T_\mu \mathcal{M}_+$ with $\int_\Omega u\, d\mu = 0$. Thus, the centered Cramer class is a closed subspace of codimension one.

---

[6]In [216] the Young function $\cosh t - 1$ was used instead of $\exp|t| - 1$. However, these produce equivalent Orlicz spaces by Proposition 3.9.

In order to understand the topological structure of $\mathcal{M}_+$ with respect to the e-topology, it is useful to introduce the following preorder on $\mathcal{M}_+$:

$$\mu' \preceq \mu \quad \text{if and only if} \quad \mu' = \phi\mu \text{ with } \phi \in L^p(\Omega, \mu) \text{ for some } p > 1. \quad (3.105)$$

In order to see that $\preceq$ is indeed a preorder, we have to show transitivity, as the reflexivity of $\preceq$ is obvious. Thus, let $\mu'' \preceq \mu'$ and $\mu' \preceq \mu$, so that $\mu' = \phi\mu$ and $\mu'' = \psi\mu'$ with $\phi \in L^p(\Omega, \mu)$ and $\psi \in L^{p'}(\Omega, \mu')$, then $\phi^p, \psi^{p'}\phi \in L^1(\Omega, \mu)$ for some $p, p' > 1$. Let $\lambda := (p' - 1)/(p + p' - 1) \in (0, 1)$. Then by Hölder's inequality, we have

$$L^1(\Omega, \mu_1) \ni \left(\psi^{p'}\phi\right)^{1-\lambda}\left(\phi^p\right)^\lambda = \psi^{p'(1-\lambda)}\phi^{1+\lambda(p-1)} = (\psi\phi)^{p''},$$

where $p'' = pp'/(p + p' - 1) > 1$, so that $\psi\phi \in L^{p''}(\Omega, \mu)$, and hence, $\mu'' \preceq \mu$ as $\mu'' = \psi\phi\mu$.

From the preorder $\preceq$ we define the equivalence relation on $\mathcal{M}_+$ by

$$\mu' \sim \mu \quad \text{if and only if} \quad \mu' \preceq \mu \text{ and } \mu \preceq \mu', \quad (3.106)$$

in which case we call μ and μ' *similar*, and hence we obtain a partial ordering on the set of equivalence classes $\mathcal{M}_+/\sim$

$$[\mu'] \preceq [\mu] \quad \text{if and only if} \quad \mu' \preceq \mu.$$

Proposition 3.11 *Let $\mu' \preceq \mu$. Then $T_\mu\mathcal{M}_+ \subseteq T_{\mu'}\mathcal{M}_+$ is continuously embedded w.r.t. the Orlicz norms on these tangent spaces.*

In particular, if $\mu \sim \mu'$, then $T_\mu\mathcal{M}_+ = T_{\mu'}\mathcal{M}_+$ with equivalent Banach norms. If we denote the isomorphism class of these spaces as $T_{[\mu]}\mathcal{M}_+$, then there are continuous inclusions

$$T_{[\mu]}\mathcal{M}_+ \hookrightarrow T_{[\mu']}\mathcal{M}_+ \quad \text{if} \quad [\mu'] \preceq [\mu]. \quad (3.107)$$

Proof Let $\mu' = \phi\mu$ with $\phi \in L^p(\Omega, \mu)$, $p > 1$, and let $q > 1$ be the dual index, i.e., $p^{-1} + q^{-1} = 1$. Then by Hölder's inequality,

$$\int_\Omega \left(\exp|tf| - 1\right) d\mu' = \int_\Omega \left(\exp|tf| - 1\right)\phi \, d\mu$$

$$\leq \|\phi\|_p \left(\int_\Omega \left(\exp|tf| - 1\right)^q d\mu\right)^{1/q}. \quad (3.108)$$

Let $\psi : \mathbb{R} \to \mathbb{R}$, $\psi(t) := \|\phi\|_p^q(\exp|t| - 1)^q$, which is a Young function. Let $f \in L^\psi(\mu)$ and $a := \|f\|_{L^\psi(\mu)}$. Then the right-hand side of (3.108) with $t := a^{-1}$ is bounded by 1, whence so is the left-hand side, so that $f \in L^{\exp|t|-1}(\mu') = T_{\mu'}\mathcal{M}_+$ and $\|f\|_{L^\psi(\mu)} = a \geq \|f\|_{L^{\exp|x|-1}(\mu')}$. Thus, there is a continuous inclusion

$$L^\psi(\mu) \hookrightarrow L^{\exp|x|-1}(\mu') = T_{\mu'}\mathcal{M}_+(\Omega).$$

But now, as $\lim_{t\to\infty}\frac{\psi(t)}{\exp|qt|-1} = \|\phi\|_p^q \in (0,\infty)$, Proposition 3.9 implies that as Banach spaces, $L^\psi(\mu) \cong L^{\exp|qt|-1}(\mu)$ and furthermore, $L^{\exp|qt|-1}(\mu) \cong L^{\exp|t|-1}(\mu) = T_{\mu'}\mathcal{M}_+(\Omega)$ by Lemma 3.10. $\qquad\square$

Proposition 3.12 *The subspace in* (3.107) *is neither closed nor dense, unless* $\mu \sim \mu'$, *in which case these subspaces are equal. In fact,* $f \in T_{[\mu']}\mathcal{M}_+$ *lies in the closure of* $T_{[\mu]}\mathcal{M}_+$ *if and only if*

$$\big(|f| + \varepsilon \log(d\mu'/d\mu)\big)_+ \in T_{[\mu]}\mathcal{M}_+ \quad \text{for all } \varepsilon > 0. \tag{3.109}$$

Here, the subscript refers to the decomposition of a function into its non-negative and non-positive part, i.e., to the decomposition

$$\psi = \psi_+ - \psi_- \quad \text{with} \quad \psi_\pm \geq 0, \psi_+ \perp \psi_-.$$

Proof Let $p > 1$ be such that $\phi := d\mu'/d\mu \in L^p(\Omega, \mu)$, and assume w.l.o.g. that $p \leq 2$. Furthermore, let $u := \log\phi$. Then

$$K := \int_\Omega \big(\exp((p-1)|u|) - 1\big)\, d\mu' = e^{-1}\int_\Omega \max(\phi^{p-1}, \phi^{1-p})\, d\mu'$$

$$= e^{-1}\int_\Omega \max(\phi^p, \phi^{2-p})\, d\mu \leq e^{-1}\int_\Omega \max(\phi^p, 1)\, d\mu < \infty.$$

If we let $\psi(t) := K^{-1}(\exp|t| - 1)$, then ψ is a Young function, and by Proposition 3.9, $L^\psi(\mu') = L^{\exp|t|-1}(\mu') = T_{[\mu']}\mathcal{M}_+$. For $f \in T_{[\mu']}\mathcal{M}_+$ we have

$$\big||f| - (|f| + \varepsilon u)_+\big| \leq \varepsilon|u|$$

and therefore

$$\int_\Omega \psi\left(\frac{||f| - (|f| + \varepsilon u)_+|}{\varepsilon(p-1)^{-1}}\right) d\mu' \leq \int_\Omega \psi\big((p-1)|u|\big)\, d\mu'$$

$$= K^{-1}\int_\Omega \big(\exp(p-1)|u| - 1\big)\, d\mu' = 1,$$

so that $\||f| - (|f| + \varepsilon u)_+\|_{L^\psi(\mu')} \leq \varepsilon(p-1)^{-1}$ by the definition of the Orlicz norm. That is, $\lim_{\varepsilon\to0}(|f| + \varepsilon u)_+ = |f|$ in $L^\psi(\mu') = L^{\exp|t|-1}(\mu') = T_{[\mu']}\mathcal{M}_+$.

In particular, if (3.109) holds, then $|f|$ lies in the closure of $T_{[\mu']}\mathcal{M}_+$. Observe that $(f_\pm + \varepsilon u)_+ \leq (|f| + \varepsilon u)_+$, whence (3.109) implies that $(f_\pm + \varepsilon u)_+ \in T_{[\mu']}\mathcal{M}_+$, whence $f_\pm$ lie in the closure of $T_{[\mu']}\mathcal{M}_+$, and so does $f = f_+ - f_-$.

On the other hand, if $(f_n)_{n\in\mathbb{N}}$ is a sequence in $T_{[\mu]}\mathcal{M}_+$ converging to f, then $\min(|f|, |f_n|)$ is also in $T_{[\mu]}\mathcal{M}_+$ converging to $|f|$, whence we may assume w.l.o.g. that $0 \leq f_n \leq f$. Let $\varepsilon > 0$ and choose n such that $\|f - f_n\| < \varepsilon$ in $T_{[\mu']}\mathcal{M}_+$. Then

by the definition of the Orlicz space we have

$$1 \geq \int_\Omega \left(\exp\left(\varepsilon^{-1}(f - f_n)\right) - 1\right) d\mu' = \int_\Omega \exp\left(\varepsilon^{-1}(f - f_n + \varepsilon u)\right) d\mu - \mu'(\Omega)$$

$$\geq \int_\Omega \exp\left(\varepsilon^{-1}(f - f_n + \varepsilon u)_+\right) d\mu - \mu'(\Omega),$$

so that $\int_\Omega (\exp(\varepsilon^{-1}(f - f_n + \varepsilon u)_+) - 1)\, d\mu < \infty$, whence $(f - f_n + \varepsilon u)_+ \in T_{[\mu]}\mathcal{M}_+$. On the other hand,

$$(f + \varepsilon u)_+ \leq (f + \varepsilon u - f_n)_+ + f_n,$$

and since the summands on the right are contained in $T_{[\mu]}\mathcal{M}_+$, so is $(f + \varepsilon u)_+$ as asserted.

Thus, we have shown that f lies in the closure of $T_{[\mu']}\mathcal{M}_+$ iff (3.109) holds.

Let us assume from now on that $\mu' \not\sim \mu$. It remains to show that in this case, $T_{[\mu]}\mathcal{M}_+ \subseteq T_{[\mu']}\mathcal{M}_+$ is neither closed nor dense. In order to see this, let $\Omega_+ := \{\omega \in \Omega \mid u(\omega) \geq 0\}$ and $\Omega_- := \Omega \setminus \Omega_+$. Observe that

$$\int_\Omega \exp(p u_+)\, d\mu = \mu(\Omega_-) + \int_{\Omega_+} \phi^p\, d\mu < \infty,$$

as $\phi \in L^p(\Omega, \mu)$, whence $u_+ \in T_{[\mu]}\mathcal{M}_+$.

We assert that $|u|^a \in T_{[\mu']}\mathcal{M}_+$, but $\notin T_{[\mu]}\mathcal{M}_+$ for all $a > 0$. Namely, pick $t > 0$ such that $ta < \min(1, p - 1)$ and calculate

$$\int_\Omega \exp\left(t|u|^a\right) d\mu' = \int_{\Omega_+} \phi^{ta}\, d\mu' + \int_{\Omega_-} \phi^{-ta}\, d\mu'$$

$$\leq \int_{\Omega_+} \phi^{1+ta}\, d\mu + \int_{\Omega_-} \underbrace{\phi^{1-ta}}_{\leq 1}\, d\mu \leq \int_\Omega \max\{\phi^p, 1\}\, d\mu < \infty.$$

Thus, $|u|^a \in T_{[\mu']}\mathcal{M}_+$ for all $a > 0$. On the other hand, if $t > 0$, then

$$\int_\Omega \exp\left(t|u|^a\right) d\mu = \int_{\Omega_+} \phi^{ta}\, d\mu + \int_{\Omega_-} \phi^{-ta}\, d\mu$$

$$\geq \int_{\Omega_+} 1\, d\mu + \int_{\Omega_-} \phi^{-(1+ta)}\, d\mu' \geq \int_\Omega \phi^{-(1+ta)}\, d\mu',$$

and the last integral is finite if and only if $\phi^{-1} \in L^{1+ta}(\Omega, \mu')$ for some $t > 0$, if and only if $\mu \preceq \mu'$ and hence $\mu \sim \mu'$. Since this case was excluded, it follows that $\int_\Omega \exp(t|u|^a) d\mu = \infty$ for all $t > 0$, whence $|u|^a \notin T_{[\mu]}\mathcal{M}_+$ as asserted.

Thus, our assertion follows if we can show that $|u|^a$ is contained in the closure of $T_{[\mu]}\mathcal{M}_+$ for $0 < a < 1$, but it is *not* in the closure of $T_{[\mu]}\mathcal{M}_+$ for $a = 1$.

For $a = 1$ and $\varepsilon \in (0, 1)$, $(|u| + \varepsilon u)_+ = (1 + \varepsilon)u_+ + (1 - \varepsilon)u_- = 2\varepsilon u_+ + (1 - \varepsilon)|u|$. Since $u_+ \in T_{[\mu]}\mathcal{M}_+$ and $|u| \notin T_{[\mu]}\mathcal{M}_+$, it follows that $(|u| + \varepsilon u)_+ \notin$

$T_{[\mu]}\mathcal{M}_+$, which shows that (3.109) is violated for $f = |u| \in T_{[\mu']}\mathcal{M}_+$, whence $|u|$ does not lie in the closure of $T_{[\mu]}\mathcal{M}_+$, so that the latter is not a dense subspace.

If $0 < a < 1$, then $(|u|^a + \varepsilon u)_+ = u_+^a + \varepsilon u_+ + (u_-^a - \varepsilon u_-)_+$. Now $u_+^a + \varepsilon u_+ \leq (a + \varepsilon)u_+ + 1$, whereas $u_-^a - \varepsilon u_- \geq 0$ implies that $u_- \leq \varepsilon^{1/(a-1)}$, so that $(u_-^a - \varepsilon u_-)_+ \leq C_\varepsilon$, where $C_\varepsilon := \max\{t^a - \varepsilon t \mid 0 \leq t \leq \varepsilon^{1/(a-1)}\}$.

Thus, $(|u|^a + \varepsilon u)_+ \leq (a + \varepsilon)u_+ + 1 + C_\varepsilon$, and since $u_+ \in T_{[\mu]}\mathcal{M}_+$ this implies that $(|u|^a + \varepsilon u)_+ \in T_{[\mu]}\mathcal{M}_+$, so that $|u|^a$ lies in the closure of $T_{[\mu]}\mathcal{M}_+$, but not in $T_{[\mu]}\mathcal{M}_+$ for $0 < a < 1$. Thus, $T_{[\mu]}\mathcal{M}_+$ is not a closed subspace of $T_{[\mu']}\mathcal{M}_+$.   □

The following now is a reformulation of Propositions 3.4 and 3.5 in [216].

Proposition 3.13 *A sequence* $(g_n\mu_0)_{n\in\mathbb{N}} \in \mathcal{M}(\Omega, \mu_0)$ *is e-convergent to* $g_0\mu_0 \in \mathcal{M}(\Omega, \mu_0)$ *if and only if* $g_n\mu_0 \sim g_0\mu_0$ *for large n, and* $u_n := \log g_n \in T_{g_0\mu_0}\mathcal{M}_+$ *converges to* $u_0 := \log g_0 \in T_{g\mu_0}\mathcal{M}_+$ *in the Banach norm on* $T_{g\mu_0}\mathcal{M}_+$ *described above.*

Proof If $(g_n\mu_0)_{n\in\mathbb{N}} \in \mathcal{M}(\Omega, \mu_0)$ *e*-converges to $g_0\mu_0 \in \mathcal{M}(\Omega, \mu_0)$, then for large n, (g_n/g_0) and (g_0/g_n) are contained in $L^p(\Omega, g_0\mu_0)$, so that $g_n\mu_0 \sim g_0\mu_0$ and hence, $u_n := \log |g_n| \in T_{g_0\mu_0}\mathcal{M}_+$.

Moreover, by Proposition 3.7 $(g_n\mu_0)_{n\in\mathbb{N}} \in \mathcal{M}(\Omega, \mu_0)$ *e*-converges to $g_0\mu_0 \in \mathcal{M}(\Omega, \mu_0)$ if and only if for all $p \geq 1$

$$0 = \lim_{n\to\infty} \int_\Omega \left\{ \left| \left(\frac{g_n}{g_0}\right)^p - 1 \right| + \left| \left(\frac{g_0}{g_n}\right)^p - 1 \right| \right\} g_0 \, d\mu_0$$

$$= \lim_{n\to\infty} \int_\Omega \left\{ \left| e^{p(u_n - u_0)} - 1 \right| + \left| e^{p(u_0 - u_n)} - 1 \right| \right\} g_0 \, d\mu_0$$

$$= \lim_{n\to\infty} \int_\Omega 2 \sinh\left(p|u_n - u_0|\right) g_0 \, d\mu_0.$$

By Proposition 3.8, this is equivalent to saying that $(u_n)_{n\in\mathbb{N}}$ converges to u_0 in the Orlicz space $L^{\sinh|t|}(g_0\mu_0)$.

However, $L^{\sinh|t|}(g_0\mu_0) = L^{\exp|t|-1}(g_0\mu_0) = T_{g_0\mu_0}\mathcal{M}_+$ by Proposition 3.9, since $\lim_{t\to\infty} \frac{\sinh|t|}{\exp|t|-1} = 1/2 \in (0, \infty)$.   □

By virtue of this proposition, we shall refer to the topology on $T_\mu\mathcal{M}_+$ obtained above as the *topology of e-convergence* or the *e-topology*. Note that the first statement in Proposition 3.13 implies that the equivalence classes of $\sim$ are open and closed in the *e*-topology.

Theorem 3.4 *Let* $K \subseteq \mathcal{M}_+$ *be an equivalence class w.r.t.* $\sim$, *and let* $T := T_{[\mu]}\mathcal{M}_+$ *for* $\mu \in K$ *be the common exponential tangent space, equipped with the e-topology. Then for all* $\mu \in K$,

$$A_\mu := \log_\mu(K) \subseteq T$$

is open convex, and $\log_\mu : K \to A_\mu$ *is a homeomorphism where K is equipped with the e-topology. In particular, the identification* $\log_\mu : A_\mu \to K$ *allows us to canonically identify K with an open convex subset of the affine space associated to T.*

Remark 3.13 This theorem shows that the equivalence classes w.r.t. $\sim$ are the connected components of the *e*-topology on $\mathcal{M}(\Omega, \mu_0)$, and since each such component is canonically identified as a subset of an affine space whose underlying vector space is equipped with a family of equivalent Banach norms, it follows that $\mathcal{M}(\Omega, \mu_0)$ is a Banach manifold. This is the affine Banach manifold structure on $\mathcal{M}(\Omega, \mu_0)$ described in [216], therefore we refer to it as the *Pistone–Sempi structure*.

The subdivision of $\mathcal{M}_+(\Omega)$ into disjoint open connected subsets was also noted in [230].

Proof of Theorem 3.4 If $f \in A_\mu$, then, by definition, $(1 + s)f, -sf \in \hat{B}_\mu(\Omega)$ for some $s > 0$. In particular, $sf \in B_\mu(\Omega)$, so that $f \in T$ and hence, $A_\mu \subseteq T$. Moreover, if $f \in A_\mu$ then $\lambda f \in A_\mu$ for $\lambda \in [0, 1]$.

Next, if $g \in A_\mu$, then $\mu' := e^g \mu \in K$. Therefore, $f \in A_{\mu'}$ if and only if $K \ni e^f \mu' = e^{f+g}\mu$ if and only if $f + g \in A_\mu$, so that $A_{\mu'} = g + A_\mu$ for a fixed $g \in T$. From this, the convexity of A_μ follows.

Therefore, in order to show that $A_\mu \subseteq T$ is open, it suffices to show that $0 \in A_{\mu'}$ is an inner point for all $\mu' \in K$. For this, observe that for $f \in B^0_{\mu'}(\Omega)$ we have $(1 + s)f \in B_{\mu'}(\Omega)$ and hence $e^{\pm(1+s)f} \in L^1(\Omega, \mu')$, so that $e^f \in L^{1+s}(\Omega, \mu')$ and $e^{-f} \in L^{2+s}(\Omega, e^f \mu')$, whence $e^f \mu' \sim \mu' \sim \mu$, so that $e^f \mu' \in K$ and hence, $f \in A_{\mu'}$. Thus, $0 \in B^0_{\mu'}(\Omega) \subseteq A_{\mu'}$, and since $B^0_{\mu'}(\Omega)$ contains the unit ball of the Orlicz norm, the claim follows. □

In the terminology which we developed, we can formulate the relation of the Pistone–Sempi structure on $\mathcal{M}_+$ with k-integrability as follows.

Proposition 3.14 *The parametrized measure model $(\mathcal{M}_+(\Omega, \mu), \Omega, i)$ is ∞-integrable, where $i : \mathcal{M}_+(\Omega, \mu) \hookrightarrow \mathcal{M}(\Omega)$ is the inclusion map and $\mathcal{M}_+(\Omega, \mu)$ carries the Banach manifold structure from the e-topology.*

Proof The measures in $\mathcal{M}_+(\Omega, \mu)$ are dominated by μ, and for the inclusion map we have

$$\partial_V \log i\big(\exp(f)\mu\big) = \partial_V i$$

and the inclusion $i : T_\mu \mathcal{M}_+(\Omega) \hookrightarrow L^k(\Omega, \mu)$ is continuous for all $k \geq 1$ by (3.104). Thus, $(\mathcal{M}_+(\Omega, \mu), \Omega, i)$ is k integrable for all such k. □

Example 3.8 Let $\Omega := (0, 1)$, and consider the 1-parameter family of finite measures

$$\mathbf{p}(\xi) = p(\xi, t)\, dt := \exp\big(-\xi^2 (\log t)^2\big)\, dt \in \mathcal{M}_+\big((0, 1), dt\big), \quad x \in \mathbb{R}.$$

Since $\partial_\xi \log p(\xi, t) = -2\xi (\log t)^2$ and

$$\int_0^1 \left| \partial_\xi \log p(\xi, t) \right|^k dp(\xi) = 2^k |\xi|^k \int_0^1 (\log t)^{2k} \exp\left(-\xi^2 (\log t)^2\right) dt < \infty$$

and this expression depends continuously on ξ for all k, it follows that this parametrized measure model is ∞-integrable.

However, $\mathbf{p}$ is not even continuous w.r.t. the e-topology. Namely, in this topology $\mathbf{p}(\xi) \to \mathbf{p}(0)$ as $\xi \to 0$ would imply that

$$\int_0^1 \left(\frac{p(0, t)}{p(\xi, t)} - 1 \right) dp(0) \xrightarrow{\xi \to 0} 0.$$

Since $p(0) = dt$, this is equivalent to

$$\int_0^1 \exp\left(\xi^2 (\log t)^2\right) dt \xrightarrow{\xi \to 0} 1.$$

However,

$$\int_0^1 \exp\left(\xi^2 (\log t)^2\right) dt = \sum_{k=0}^\infty \frac{1}{n!} \xi^{2n} \int_0^1 (\log t)^{2n} dt = \sum_{k=0}^\infty \frac{(2n)!}{n!} \xi^{2n} = \infty$$

for all $\xi \neq 0$, so that this expression does not converge.

We end this section with the following result which illustrates how the ordering $\preceq$ provides a stratification of $\hat{B}_{\mu_0}(\Omega)$.

Proposition 3.15 Let $\mu_0', \mu_1' \in \mathcal{M}_+$ with $f_i := \log_{\mu_0}(\mu_i') \in \hat{B}_{\mu_0}(\Omega)$, and let $\mu_\lambda' := \exp(f_0 + \lambda(f_1 - f_0))\mu_0$ for $\lambda \in [0, 1]$ be the segment joining μ_0' and μ_1'. Then the following hold:

(1) The measures μ_λ' are similar for $\lambda \in (0, 1)$.
(2) $\mu_\lambda' \preceq \mu_0'$ and $\mu_\lambda' \preceq \mu_1'$ for $\lambda \in (0, 1)$.
(3) $T_{\mu_\lambda'}\mathcal{M}_+ = T_{\mu_0'}\mathcal{M}_+ + T_{\mu_1'}\mathcal{M}_+$ for $\lambda \in (0, 1)$.

Proof Let $\delta := f_1 - f_0$ and $\phi := \exp(\delta)$. Then for all $\lambda_1, \lambda_2 \in [0, 1]$, we have

$$\mu_{\lambda_1}' = \phi^{\lambda_1 - \lambda_2} \mu_{\lambda_2}'. \tag{3.110}$$

For $\lambda_1 \in (0, 1)$ and $\lambda_2 \in [0, 1]$, we pick $p > 1$ such that $\lambda_2 + p(\lambda_1 - \lambda_2) \in (0, 1)$. Then by (3.110) we have

$$\phi^{p(\lambda_1 - \lambda_2)} \mu_{\lambda_2}' = \mu_{\lambda_2 + p(\lambda_1 - \lambda_2)}' \in \mathcal{M}_+,$$

so that $\phi^{p(\lambda_1 - \lambda_2)} \in L^1(\Omega, \mu_{\lambda_2}')$ or $\phi^{\lambda_1 - \lambda_2} \in L^p(\Omega, \mu_{\lambda_2})$ for small $p - 1 > 0$. Therefore, $\mu_{\lambda_1}' \preceq \mu_{\lambda_2}'$ for all $\lambda_1 \in (0, 1)$ and $\lambda_2 \in [0, 1]$, which implies the first and second statement.

This implies that $T_{\mu'_i}\mathcal{M}_+ \subseteq T_{\mu'_\lambda}\mathcal{M}_+ = T_{\mu'_{1/2}}\mathcal{M}_+$ for $i = 0, 1$ and all $\lambda \in (0, 1)$ which shows one inclusion in the third statement.

In order to complete the proof, observe that

$$T_{\mu'_{1/2}}\mathcal{M}_+ = T_{\mu'_{1/2}}\mathcal{M}_+(\Omega_+, \mu_0) \oplus T_{\mu'_{1/2}}\mathcal{M}_+(\Omega_-, \mu_0),$$

where $\Omega_+ := \{\omega \in \Omega \mid \delta(\omega) > 0\}$ and $\Omega_- := \{\omega \in \Omega \mid \delta(\omega) \leq 0\}$. If $g \in T_{\mu'_{1/2}}\mathcal{M}_+(\Omega_+, \mu_0)$, then for some $t \neq 0$

$$\int_\Omega \exp(|tg|)\,d\mu'_0 \leq \int_{\Omega_+} \exp\left(|tg| + \frac{1}{2}\delta\right)d\mu'_0 + \int_{\Omega_-} d\mu'_0$$

$$= \int_{\Omega_+} \exp(|tg|)\,d\mu'_{1/2} + \int_{\Omega_-} d\mu'_0 < \infty,$$

so that $g \in T_{\mu'_0}(\Omega, \mu_0)$ and hence, $T_{\mu'_{1/2}}\mathcal{M}_+(\Omega_+, \mu_0) \subseteq T_{\mu'_0}(\Omega, \mu_0)$. Analogously, one shows that $T_{\mu'_{1/2}}\mathcal{M}_+(\Omega_-, \mu_0) \subseteq T_{\mu'_0}(\Omega, \mu_1)$, which completes the proof. $\square$