

# Designing a general library for convolutions

Floris van Doorn

## Abstract

We will discuss our experiences and design decisions obtained from building a formal library for the convolution of two functions. Convolution is a fundamental concept with applications throughout mathematics. We will focus on the design decisions we made to make the convolution general and easy to use, and the incorporation of this development in Lean's mathematical library `mathlib`.

## 1 Introduction

Convolutions are used throughout mathematics. They are used in probability theory to compute the probability density function for the sum of two independent random variables, in image processing to blur images and detect edges, in differential topology to approximate functions with smooth functions, and it is used to define distributions as generalized functions.

In the most basic form, the convolution of two functions  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  is defined to be

$$(f * g)x = \int f(t)g(x - t)dt.$$

However, there are many variants and generalizations of convolutions:

- Discrete convolution of functions defined on  $\mathbb{Z}$ ;
- Convolution for functions defined on  $\mathbb{R}^n$
- More generally, convolution for functions defined on a group  $G$ ; equipped with a measure  $\mu$ ;
- Convolution of a scalar function with a vector-valued function.

It is tricky to design a general library for convolution capturing all these variations, especially if you realize that the various instances of convolution satisfy

different properties. For example, the convolution operator for functions defined on  $\mathbb{R}^n$  is commutative, but this no longer holds on (nonabelian) groups  $G$ . And while the convolution of functions on  $\mathbb{R}^n$  is smooth under suitable hypotheses, this statement is nonsensical for the convolution of functions on  $\mathbb{Z}$ .

In this paper we discuss the design of a convolution library we designed for the Lean 3 theorem prover [dMKA<sup>+</sup>15]. Lean is developed principally by Leonardo de Moura at Microsoft Research and implements a version of the calculus of inductive constructions [CP88] with quotient types, non-cumulative universes, and proof irrelevance. We use and incorporate our library into its mathematical library `mathlib` [Com20], which has more than 250 contributors and almost one million lines of code. We give a single definition of convolution that captures all aforementioned cases, and we prove all properties about the convolution using the minimal assumptions needed for that property. This flexibility in assumptions for each theorem is greatly aided by the flexible type-class system implemented in Lean.

In fact, we generalize the convolution enough in order to compute the total derivative of the convolution of two multivariate functions, even in the case where one of them takes values in any Banach space. To do this we have to define a convolution in the case where both functions take values in a normed space, and we use a bilinear map to combine the functions. We could not find the computation of the total derivative of the convolution anywhere in the literature; all books that the author of this paper consulted compute only the partial derivatives of the convolution [Hir76, Yos95].

This formalization is interesting since the convolution has a lot of counterintuitive properties. For example, the convolution of two smooth functions need not exist. And even if the convolution exists everywhere, it need not be continuous. This makes the formalization of these results particularly useful, since one can formally check which results are true and what assumptions are needed for each result.

To get well-behaved properties of the convolution, one needs to impose conditions on one or both functions stating that they decay sufficiently rapidly at  $\infty$ . The strongest such condition, on which we will focus, is the requirement that one of the functions has compact support. This is a condition that will ensure that the convolution well-behaved, under a very weak condition of the other function (local integrability).

Other conditions could be imposed on the functions to obtain a well-behaved convolution. For example, one can assume that both functions in the convolution are rapidly decreasing. However, such results are typically not more general than the results in this paper, since a relaxation of the behavior of one function has to be counteracted by a stricter condition on the other


function.

We prove the following properties about convolution:

- Algebraic properties: commutativity, associativity and bilinearity;
- Integrability properties for the convolution of integrable functions;
- Smoothness properties: continuity, differentiability and being  $C^n$ ;
- Approximation properties: the convolution of  $f$  with a small bump function is close to  $f$ .

The last property shows the power of convolutions: they give a very general framework to approximate functions with smooth functions, even if the original function has very weak properties (being locally integrable). For the smoothness properties we assume that one of the functions has compact support.

When formalizing mathematics, there is often a difference in formalizing a specific theorem and writing a reusable library that is nice and convenient to use. In the former one often takes shortcuts when setting up the development in a manner that is sufficient for proving the relevant theorem. When designing a reusable library, one wants to focus more on giving very general definitions, and proving the properties in a great generality, so that the library can be reused in many areas. One often also spends more time on reliable naming conventions, documentation and code readability. These differences are one of the reasons that the reviewing process for `mathlib` contributions is lengthy, especially for new users. The written code has to be tidied, pass the linter suite of `mathlib` [vDEL20] and satisfy `mathlib`'s standards for code quality.

This paper is not a formalization of a single big theorem, but instead describes the design of a reusable library of a tool, to be used in other developments. Unless specifically mentioned, all results in this paper have been incorporated into `mathlib`. We will link to specific results in `mathlib` using the icon  using static links to the current version of `mathlib`.

There are many other proof assistants that have developed a substantial analysis library, most notably Isabelle [HIH13], HOL Light [Har13] and Coq [BLM15]. However, to my knowledge there are no proof assistants that have formalized the properties of the convolution of smooth functions.

## 2 Preliminaries

### 2.1 Type Classes

`mathlib` uses partially bundled *type-classes* [WB89, SvdW11] throughout the library. Almost all results in `mathlib` depend on one or more type-classes, and this paper we will see a couple of them. For example, the type-class `[nontrivially_normed_field ℚ]` states that  $\mathbb{Q}$  is a field with a norm  $\|\cdot\| : \mathbb{Q} \rightarrow \mathbb{R}$  satisfying

- $d(x, y) := \|x - y\|$  defines a metric on  $\mathbb{Q}$ ;
- $\|xy\| = \|x\| \cdot \|y\|$ ;
- the norm on  $\mathbb{Q}$  is nontrivial: there is an element  $x \in \mathbb{Q}$  with  $\|x\| > 1$ .

Outside code-snippets, we will use the term *normed field* to refer to nontrivially normed fields. When applying a lemma that depends on a type-class, Lean will automatically find these type-class arguments during *type-class synthesis*, using the data provided in a theorem and globally-declared instances. An example of a globally declared instance is `[nontrivially_normed_field ℝ]`.<sup>1</sup>

If  $\mathbb{K}$  is a normed field, then we can consider all normed spaces over  $\mathbb{K}$ . For technical reasons we split this condition in two parts.

The use the class `[normed_add_comm_group E]` or “ $E$  is a normed group” to state that  $(E, +, 0)$  is an abelian group with a norm, such that  $d(x, y) := \|x - y\|$  defines a metric on  $E$ . A seminormed group has a similar definition, with as only difference that  $\|x\| = 0$  need not imply that  $x = 0$ .

If  $\mathbb{K}$  is a normed field and  $E$  is a normed group, then we have an additional type-class `[normed_space ℚ E]` stating that  $E$  is a vector space over  $\mathbb{K}$  with  $\|k \cdot v\| = \|k\| \cdot \|v\|$  for  $k \in \mathbb{K}$  and  $v \in E$ . The type-class `normed_space` is a *mixin*, which means that it takes other type-classes as arguments. In this case the arguments are `[normed_field ℚ]` and `[seminormed_add_comm_group E]`.

The reason for this separation is that the class `[seminormed_add_comm_group E]` contains all the properties that *only* mention  $E$ . So for example, if we want to use a lemma about the addition on  $E$ , we only have to search for the instance `[seminormed_add_comm_group E]` and not for an instance of `[normed_space ℚ E]`. If we had a single type-class `[normed_space ℚ E]` with all the properties, then we would have to

---

<sup>1</sup>To be precise, this is a consequence of  $\mathbb{R}$  being a densely-ordered field, but this derivation happens automatically by type-class synthesis.

search for an instance `[normed_space ℚ E]` without knowing what `ℚ` is (since addition on `E` does not refer to `ℚ`). With Lean’s algorithm for type-class synthesis, it can be problematic to search for instances without knowing all the arguments [Baa22]. In the definition given in `mathlib`, only properties of normed spaces that refer to both `ℚ` and `E` are in the `[normed_space ℚ E]` class. This means that when we use one of these properties, both `ℚ` and `E` are known once Lean has to synthesize an instance for this type-class.

Many particular properties of objects are also recorded in type-classes. For example, the properties `[sigma_finite μ]`, `[is_add_left_invariant μ]` and `[is_neg_invariant μ]` (that we will define below) are all properties on measures that are marked as type-class arguments. This means that we only have to prove these properties if they cannot be inferred from the globally-declared instances, which does not come up often. An example of such a global instance for these classes states that  $\mu \times \nu$  is  $\sigma$ -finite if both  $\mu$  and  $\nu$  are. [↗](#)

```
instance [sigma_finite μ] [sigma_finite ν] :
  sigma_finite (μ.prod ν)
```

## 2.2 Filters

`mathlib` has a general library for filters, which are used throughout the topology, analysis and measure theory libraries. Filters are used to make precise arguments that involve phrases such as “for all  $x$  close to  $x_0$ ,” “for large enough  $N$ ” and “for almost every  $x \in \mathbb{R}$ .” In this way, filters can be used as a generalized version of bounded quantification. The filter library is heavily inspired by the one in Isabelle [HH13].

Formally, a filter  $F$  on  $X$  is a collection of subsets of  $X$  satisfying the following properties

- $X \in F$ ;
- $F$  is upwards closed: if  $X \in F$  and  $Y \supseteq X$  then  $Y \in F$ ;
- $F$  is closed under binary intersections: if  $X, Y \in F$  then  $X \cap Y \in F$ .

Given a filter  $F$ , we can introduce the notion  $\forall^f x \in F, P(x)$  stating that  $P$  holds eventually for  $x \in F$ . It is defined to mean that  $\{x \mid P(x)\} \in F$ .

Now the aforementioned examples become special cases of this notions. “For all  $x$  close to  $x_0$ ,  $P(x)$  holds” means  $\forall^f x \in \mathcal{N}_{x_0}, P(x)$ , where the *neighborhood filter*  $\mathcal{N}_x$  is defined as the collection of all neighborhoods of a point  $x$  in a topological space:

$$\mathcal{N}_x = \{N \mid \text{there exists an open set } U \subseteq N \text{ with } x \in U\}.$$

Similarly, if  $\mu$  is a measure on  $X$  then “for almost every  $x \in X$ ,  $P(x)$  holds” means  $\forall^f x \in \text{AE}_\mu, P(x)$ , where the *almost everywhere* filter of co-null sets is defined to be

$$\text{AE}_\mu = \{A \mid \mu(X \setminus A) = 0\}.$$

Note that the phrase “ $x$  is close to  $x_0$ ” by itself is not a precisely defined notion. Filters only give a precise notion of “for all  $x$  close to  $x_0$ ,  $P(x)$  holds.”

We will see the definition of another filter in Section 7.

The filters  $\mathcal{F}_X$  on  $X$  are ordered by reverse inclusion:

$$F \leq G \iff G \subseteq F.$$

If  $f : X \rightarrow Y$  then it induces a forward map  $f_* : \mathcal{F}_X \rightarrow \mathcal{F}_Y$  defined by

$$A \in f_*(F) \iff f^{-1}(A) \in F.$$

We can use filters to define limits and convergence. If  $f : X \rightarrow Y$ ,  $F$  is a filter on  $X$  and  $G$  is a filter on  $Y$ , we say that  $f(x)$  *tends to  $G$  as  $x$  tends to  $F$*  to mean that  $f_*F \leq G$ . One can derive that if  $F$  and  $G$  are neighborhood filters, this exactly corresponds to the usual definition in a topological space. If moreover  $X$  and  $Y$  are metric spaces, it corresponds to the  $\varepsilon$ - $\delta$  definition of limits.

## 2.3 Differentiation

Lean has a general theory of derivatives, formulated in terms of the Fréchet derivative and developed mainly by [redacted]<sup>2</sup>.

If  $E$  and  $F$  are normed spaces over a normed field  $\mathbb{k}$ , we denote the space of continuous  $\mathbb{k}$ -linear maps from  $E$  to  $F$  as  $L(E, F) = L_{\mathbb{k}}(E, F)$ . The space  $L(E, F)$  is itself a normed space, where the norm  $\|L\|$  is the least  $c$  such that  $\|Lx\| \leq c\|x\|$  for all  $x \in E$ . Suppose  $f : E \rightarrow F$  is a map,  $f' : L_{\mathbb{k}}(E, F)$  is a continuous linear map and  $x \in E$ . Then we write `has_fderiv_at f f' x` for the statement that  $f$  has derivative  $f'$  at  $x$ . It is defined using the little- $o$  notation as  $f(x') - f(x) - f'(x' - x) = o(x' - x)$  as  $x'$  tends to  $x$ . If a derivative exists at  $x$ , we say that  $f$  is differentiable at  $x$ . We define the derivative `fderiv  $\mathbb{k}$  f x` or  $Df$  of  $f$  at  $x$  to be the derivative of  $f$  at  $x$  if it exists. It is defined to be 0 if  $f$  has no derivative at  $x$ .

Furthermore, if  $n \in \mathbb{N} \cup \{\infty\}$  then we write `cont_diff  $\mathbb{k}$  n f` or “ $f$  is  $C^n$ ” to state that  $f$  is  $n$ -times continuously differentiable. It is defined in terms of the existence of a Taylor series, and it has the following defining properties.

<sup>2</sup>In the anonymous version of this paper we avoid mentioning `mathlib` contributors by name.

```

cont_diff ℓ 0 f ↔ continuous f
cont_diff ℓ (n + 1) f ↔ differentiable ℓ f ∧
  cont_diff ℓ n (fderiv ℓ f)
cont_diff ℓ ∞ f ↔ ∀ (n : ℕ), cont_diff ℓ n f

```

## 2.4 Integration

Lean has a large library for measure theory and integration, including the Bochner integral, Fubini’s theorem, Differentiation under an integral and the Haar measure [vD21].<sup>3</sup>

Lean’s Bochner integral is defined for maps from a measure space into a Banach space (complete normed space) over  $\mathbb{R}$ . Previously we required the codomain to be second countable, but after a Herculean refactor by [redacted] this assumption is not necessary anymore.

A *measurable space* is a space equipped with a designated  $\sigma$ -algebra of measurable sets, and a *measure space* is a measurable space with a specified measure. A *simple function* is a function with finitely many values where all preimages are measurable. If  $f : X \rightarrow Y$  is a map from a measure space  $(X, \mu)$  to a topological space  $Y$ , then  $f$  is  *$\mu$ -a.e. strongly measurable* if it is  $\mu$ -almost everywhere equal to the limit of a sequence of simple functions. If the codomain of  $f$  is a normed group, then  $f$  is  *$\mu$ -integrable* if it is  $\mu$ -a.e. strongly measurable and the Lower Lebesgue integral of  $\|f\|$  is finite.  $f$  is *integrable on  $A \subseteq X$*  if  $f$  is integrable w.r.t. to the measure  $\mu|_A$  defined by  $\mu|_A(B) = \mu(A \cap B)$ . If  $X$  is also a topological space, we say that  $f$  is *locally integrable* if  $f$  is integrable on all compact subsets of  $X$ .

One can define the Bochner integral by first defining it for simple  $L^1$  functions (as a simple linear combination), and then extend it to all  $L^1$  functions by noticing that the simple  $L^1$ -functions are dense in all  $L^1$ -functions. Then the *Bochner integral*  $\int_X f(x) d\mu(x)$  of a function  $f$  from a measure space  $(X, \mu)$  to a Banach space  $E$  is defined to be 0 if  $f$  is not integrable, and otherwise it is defined to be the integral of  $f$  viewed as an element of  $L^1$ .

We say that a  $\mu$  is  *$\sigma$ -finite* if there is a countable collection of sets with finite measure that span  $X$ . If  $(X, \mu)$  and  $(Y, \nu)$  are  $\sigma$ -finite measure spaces, then there is a *product measure*  $\mu \times \nu$  on  $X \times Y$  satisfying

$$(\mu \times \nu)(A \times B) = \mu(A)\nu(B).$$

The Bochner integral satisfies Fubini’s theorem [vD21], which comes in two parts.

---

<sup>3</sup>Results in `mathlib` that we will not use in this paper include the fundamental theorems of calculus, change of variables for higher-dimensional integrals [Gou22], and the divergence theorem [Kud22].

Fubini's theorem neatly expresses the integrability condition for functions  $f : X \times Y \rightarrow E$ . If  $f$  is  $(\mu \times \nu)$ -a.e. strongly measurable then  $f$  is  $(\mu \times \nu)$ -integrable iff  $y \mapsto f(x, y)$  is  $\nu$ -integrable for  $\mu$ -almost all  $x \in X$  and  $x \mapsto \int_Y \|f(x, y)\| d\nu(y)$  is  $\mu$ -integrable. There is of course also a symmetric version of this condition.

Furthermore, Fubini's theorem states how to compute the integral. If  $f : X \times Y \rightarrow E$  is  $(\mu \times \nu)$ -integrable, then

$$\begin{aligned} \int_{X \times Y} f(z) d(\mu \times \nu)(z) &= \int_X \int_Y f(x, y) d\nu(y) d\mu(x) \\ &= \int_Y \int_X f(x, y) d\mu(x) d\nu(y). \end{aligned}$$

Using the dominated convergence theorem, we can prove under what conditions a parametric integral is continuous.

**Proposition 2.1.** *Let  $Z$  be a first-countable topological space,  $F : Z \times X \rightarrow E$  be a map and  $z_0 \in Z$  such that*

- $F(z, -)$  is  $\mu$ -a.e. strongly measurable for  $z$  near  $z_0$ ;
- $F(-, x)$  is continuous for  $\mu$ -a.e.  $x \in X$ ;
- There is an integrable bounding function  $B : X \rightarrow \mathbb{R}$  such that  $\|F(z, x)\| \leq B(x)$  for  $z$  near  $z_0$  and  $\mu$ -almost all  $x$ .

Then  $z \mapsto \int F(z, x) d\mu(x)$  is continuous at  $z_0$ .

[redacted] has proven very general results in Lean that show that we can differentiate under a parametric integral. The result we will use is the following, but it is not the most general version that is provided.

**Proposition 2.2.** *Let  $\mathbb{k} = \mathbb{R}$  or  $\mathbb{k} = \mathbb{C}$ , let  $H$  be a normed space and  $E$  be a Banach space, both over  $\mathbb{k}$ . Let  $F : H \times X \rightarrow E$ ,  $F' : H \times X \rightarrow L_{\mathbb{k}}(H, E)$ , and  $z_0 \in Z$  such that*

- $F(z, -)$  is  $\mu$ -a.e. strongly measurable for  $z$  near  $z_0$ ;
- $F(z_0, -)$  is  $\mu$ -integrable;
- $F'(z_0, -)$  is  $\mu$ -a.e. strongly measurable;
- $F(-, x)$  has derivative  $F'(z, x)$  at  $z$  for  $\mu$ -a.e.  $x \in X$  and all  $z$  near  $z_0$ ;
- There is an integrable bounding function  $B : X \rightarrow \mathbb{R}$  such that  $\|F'(z, x)\| \leq B(x)$  for all  $z$  and  $\mu$ -almost all  $x$ .



Then  $z \mapsto \int F(z, x)d\mu(x)$  has derivative  $\int F'(z_0, x)d\mu(x)$  at  $z_0$ .<sup>☞</sup>

Let  $(G, +, 0)$  be a measurable group. This means that  $G$  is a group equipped with a  $\sigma$ -algebra such that the addition  $+ : G \times G \rightarrow G$  and negation  $- : G \rightarrow G$  are measurable. If  $\mu$  a measure on  $G$ , we say that  $\mu$  is *invariant under left addition* or *left-invariant*<sup>☞</sup> if  $\mu(g + A) = \mu(A)$  for all  $g \in G$  and measurable  $A \subseteq G$ . In this case, we also have that for any function  $f : G \rightarrow E$  that<sup>☞</sup>

$$\int_G f(g + x)d\mu(x) = \int_G f(x)d\mu(x).$$

Note that this equality also holds if  $f$  is not integrable and even if  $f$  is not (a.e.) measurable. The reason is that  $f$  is integrable if and only if  $x \mapsto f(g + x)$  is integrable, and if this is false, both sides of the equation equal 0.

Similarly, we say that  $\mu$  is *invariant under negation* or *negation-invariant*<sup>☞</sup> if  $\mu(-A) = \mu(A)$  for all measurable  $A \subseteq G$ .

If  $G$  is a second-countable locally compact Hausdorff topological group, then there is a left-invariant measure  $\mu$  that is finite on compact sets, and nonzero on non-empty open sets, called the *Haar measure*.<sup>☞</sup> Moreover, this measure is unique up to a scalar factor [vD21].<sup>☞</sup>

### 3 Definition of the convolution

As mentioned in the introduction, we want to define a general definition of the convolution, so that it is applicable in many situations. For the convolution  $f * g$  we don't want to require that the domains of  $f$  and  $g$  is  $\mathbb{R}$ , or even any Euclidean space. We can be more general by requiring that the domain  $f$  and  $g$  is an arbitrary group  $(G, +, 0)$  equipped with a  $\sigma$ -algebra and a measure  $\mu$ . In many of the properties we prove we will assume that addition and negation are measurable operations, and that the measure is invariant under addition and negation. This is fine, since in most applications  $G$  will be a locally compact group with Haar measure  $\mu$ . In fact, for the bare definition of convolution we don't even assume that  $G$  is a group, but just a type equipped with a binary subtraction operation  $- : G \times G \rightarrow G$ .

For the codomains of  $f$  and  $g$  we also want to have a general choice. One of the applications we want to provide is when one starts with a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  that is merely continuous or even locally integrable. One can take the convolution with a small bump function  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  around the origin to obtain a smooth function  $\varphi * f$  that is close to the original function  $f$ .

This suggests the definition where we define the convolution  $\varphi *_{\mu} f$  for  $\varphi : G \rightarrow \mathbb{k}$  and  $f : G \rightarrow E$  where  $\mathbb{k}$  is a normed field and  $E$  is a normed vector space over  $\mathbb{k}$ , which we can define as

$$(\varphi *_{\mu} f)(x) = \int_G \varphi(t)f(x-t)d\mu(t).$$

This definition has two disadvantages. First, it is asymmetric: the first function is scalar-valued and the second function can be vector-valued. This is not a problem per se, but it makes it awkward to work with. We cannot state commutativity of the convolution in general, but only in the special case where the second function is also scalar-valued. Since we cannot state the commutativity of the convolution in general, we cannot use it to prove symmetric versions lemmas. Suppose we have proven that  $\varphi *_{\mu} f$  is continuous when  $\varphi$  is continuous with compact support and  $f$  is locally integrable. We cannot conclude the symmetric version from this, which states that  $\varphi *_{\mu} f$  is also continuous when  $f$  is continuous with compact support and  $\varphi$  is locally integrable.

Second, there is a problem with computing the total derivative of the convolution. To work with derivatives, we have to assume that  $G$  is also a normed vector space over  $\mathbb{k}$ . If  $f$  is  $C^1$  with compact support and  $\varphi$  is locally integrable, then we can prove that  $D_x(\varphi *_{\mu} f) = (\varphi *_{\mu} Df)(x)$  for  $x \in G$ . Note that on the right hand-side  $Df : G \rightarrow L(G, E)$  is a family of continuous linear maps, and so we take the convolution in the space  $L(G, E)$  of continuous  $\mathbb{k}$ -linear maps from  $G$  to  $E$ .

However, more commonly we will be dealing with the symmetric case, where  $\varphi$  is  $C^1$  with compact support and  $f$  is locally integrable. This case occur for example when we want to smoothen a vector-valued function by taking the convolution with a smooth scalar-valued function. In this case stating any formula for  $D_x(\varphi *_{\mu} f)$  is difficult. We want to say that it is the convolution of  $D\varphi$  with  $f$ , evaluated at  $x$ , but this does not fit in our current definition, since  $D\varphi$  has type  $L(G, \mathbb{k})$  and is not a scalar-valued function anymore. In this particular case, it is not particularly hard to write down what the derivative should be, namely for  $x, x' \in G$

$$D_x(\varphi *_{\mu} f)(x') = \int_G D_t\varphi(x')f(x-t)d\mu(t). \quad (1)$$

The right-hand side is almost a convolution, except that we have replaced scalar multiplication inside the integral with a different bilinear map. Indeed, we can define a bilinear map  $A : L(G, \mathbb{k}) \times E \rightarrow L(G, E)$  by  $A(v, y)(x) = v(x)y$ . Now we can rewrite (1) as

$$D_x(\varphi *_{\mu} f) = \int_G A(D_t\varphi, f(x-t))d\mu(t).$$

This suggests a more general definition of convolution that depends on a bilinear map. This is the definition of convolution that we have formalized.

For the remainder of the paper, we assume that  $E, E'$  are normed spaces and  $F$  is a Banach space over a normed field  $\mathbb{k}$ . Furthermore, we assume that  $(G, \mu)$  is a measure space with a designated subtraction operation,  $L : E \times E' \rightarrow F$  is a continuous bilinear map, and that  $f : G \rightarrow E$  and  $g : G \rightarrow E'$  are functions.

**Definition 3.1.** The *convolution of  $f$  and  $g$  with respect to  $L$  and  $\mu$*  is defined by

$$(f *_{L,\mu} g)(x) := \int_G L(f(t), g(x-t)) d\mu(t). \quad (2)$$

We say that *the convolution  $f *_{L,\mu} g$  exists at  $x$*  if the integrand of (2) is integrable.

Recall that an integral is defined to be 0 if the integrand is not integrable. Of course, for most lemmas about convolution we will need to assume that the convolution exists, or a stronger condition that implies that the convolution exists.

If  $G$  is nonabelian, there are multiple definitions of convolution. We could have also defined the convolution  $(f *_{L,\mu} g)(x)$  to be

$$\int_G L(f(-t+x), g(t)) d\mu(t),$$

which is different unless  $G$  is unimodular. While we do not go into the theory of unimodular groups and the modular function in the formalization, we generally avoid assuming that the group  $G$  is abelian. In this paper, we only assume that  $G$  is abelian in the commutativity of the convolution, and when we assume that  $G$  is a (semi)normed group in parts of Sections 6 and 7.

## 4 Algebraic Properties

**Proposition 4.1.** *The convolution satisfies the following algebraic properties for  $f, f' : G \rightarrow E, g, g' : G \rightarrow E'$  and  $c \in \mathbb{k}$*

1.  $(cf) *_{L,\mu} g = f *_{L,\mu} (cg) = c(f *_{L,\mu} g)$ ;
2. if  $f *_{L,\mu} g$  and  $f *_{L,\mu} g'$  exist, then

$$f *_{L,\mu} (g + g') = f *_{L,\mu} g + f *_{L,\mu} g';$$

3. if  $f *_{L,\mu} g$  and  $f' *_{L,\mu} g$  exist, then  $\square$

$$(f + f') *_{L,\mu} g = f *_{L,\mu} g + f' *_{L,\mu} g;$$

4. Let  $L^t$  be the transpose of  $L$ , defined by  $L^t(x, y) = L(y, x)$ . If  $G$  is an abelian measurable group and  $\mu$  is invariant under left addition and negation, then  $f *_{L,\mu} g = g *_{L^t,\mu} f$ .  $\square$

*Proof.* The first 3 items are immediate. The last equality follows from

$$\begin{aligned} & \int L(f(t), g(x-t)) d\mu(t) \\ &= \int L(f(x-s), g(x-(x-s))) d\mu(s) \\ &= \int L^t(g(s), f(x-s)) d\mu(s), \end{aligned}$$

where we used the invariance of  $\mu$  in the first equality.  $\square$

Stating associativity is trickier in this general setting.

**Proposition 4.2.** *Assume that  $\mathbb{k}$  is either  $\mathbb{R}$  or  $\mathbb{C}$  and suppose that we are given a group  $G$  with measurable addition,  $x_0 \in G$ , a measure  $\mu$  and a right-invariant measure  $\nu$  on  $G$ , both of which are  $\sigma$ -finite, normed spaces  $E_i$  and Banach spaces  $F_i$  over  $\mathbb{k}$  ( $i \in \{1, 2, 3\}$ ), functions  $f_i : G \rightarrow E_i$  and continuous bilinear maps*

$$\begin{aligned} L_1 : E_1 \times E_2 &\rightarrow F_1 & L_3 : E_1 \times F_2 &\rightarrow F_3 \\ L_2 : F_1 \times E_3 &\rightarrow F_3 & L_4 : E_2 \times E_3 &\rightarrow F_2. \end{aligned}$$

Suppose that

- For all  $x_i \in E_i$  we have

$$L_2(L_1(x_1, x_2), x_3) = L_3(x_1, L_4(x_2, x_3)); \quad (3)$$

- $f_1 *_{L_4,\mu} f_2$  and  $f_2 *_{L_1,\nu} f_3$  exist;
- the map  $(t, s) \mapsto L_3(f_1(s), L_4(f_2(t-s), f_3(x_0-t)))$  is  $(\nu \times \mu)$ -integrable,

then  $\square$

$$(f_1 *_{L_1,\mu} f_2) *_{L_2,\nu} f_3 = f_1 *_{L_3,\mu} (f_2 *_{L_4,\nu} f_3).$$

Note that the conditions are immediately satisfied if all the  $L_j$  are (scalar) multiplication, and all the  $f_i$  are integrable.

*Proof.* We compute

$$\begin{aligned}
& \int L_2(\int L_1(f_1(s), f_2(t-s))d\mu(s), f_3(x_0-t))d\nu(t) \\
&= \iint L_2(L_1(f_1(s), f_2(t-s)), f_3(x_0-t))d\mu(s)d\nu(t) \\
&= \iint L_3(f_1(s), L_4(f_2(t-s), f_3(x_0-t)))d\mu(s)d\nu(t) \\
&= \iint L_3(f_1(s), L_4(f_2(t-s), f_3(x_0-t)))d\nu(t)d\mu(s) \\
&= \int L_3(f_1(s), \int L_4(f_2(t-s), f_3(x_0-t))d\nu(t))d\mu(s) \\
&= \int L_3(f_1(s), \int L_4(f_2(t), f_3(x_0-s-t))d\nu(t))d\mu(s).
\end{aligned}$$

In the first and fourth equality we use linearity (of  $L_2$  resp.  $L_3$ ), which requires that the appropriate convolutions exist. In the second equality we use (3), in the third equality we use Fubini's theorem (using the integrability condition), and in the fifth equality we use that  $\nu$  is right-invariant.  $\square$

## 5 Integrability

If  $f$  and  $g$  are both integrable, we cannot conclude that the convolution  $f *_{L,\mu} g$  exists everywhere. However, we can get close.

**Proposition 5.1.** *If  $G$  is a measurable group,  $\mu$  is  $\sigma$ -finite and right-invariant, and both  $f$  and  $g$  are integrable, then*

- *The convolution  $f *_{L,\mu} g$  exists at  $\mu$ -almost all  $x$  in  $G$ ,*
- *$f *_{L,\mu} g$  is integrable.*
- *Furthermore, if  $\mathbb{k}$  is either  $\mathbb{R}$  or  $\mathbb{C}$  and both  $E$  and  $E'$  are Banach spaces, then*

$$\int (f *_{L,\mu} g)(x)d\mu(x) = L(\int f(x)d\mu(x), \int g(x)d\mu(x)).$$

*Proof.* We first prove that the map  $(x, t) \mapsto L(f(t), g(x-t))$  is  $(\mu \times \mu)$ -integrable. First note that this map is  $(\mu \times \mu)$ -a.e. strongly measurable. By Fubini's theorem, it is sufficient to show that

$$x \mapsto L(f(t), g(x-t))$$

is  $\mu$ -integrable for  $\mu$ -almost all  $t$  and that

$$t \mapsto \int L(f(t), g(x-t)) d\mu(x)$$

is  $\mu$ -integrable.

The first claim follows immediately from the integrability of  $g$ . For the second claim, notice that

$$\begin{aligned} & \left\| \int L(f(t), g(x-t)) d\mu(x) \right\| \\ & \leq \int \|L(f(t), g(x-t))\| d\mu(x) \\ & \leq \|L\| \cdot \|f(t)\| \cdot \int \|g(x-t)\| d\mu(x) \\ & = \|L\| \cdot \|f(t)\| \cdot \int \|g(x)\| d\mu(x). \end{aligned}$$

We use the right-invariance of  $\mu$  in the last equality. This last expression is  $\mu$ -integrable in  $t$  since  $f$  is integrable, and therefore  $t \mapsto \int L(f(t), g(x-t)) d\mu(x)$  is also  $\mu$ -integrable.

We can now apply the symmetric version of Fubini's theorem to conclude that

$$t \mapsto L(f(t), g(x-t))$$

is  $\mu$ -integrable for  $\mu$ -almost all  $x$  and that

$$x \mapsto \int L(f(t), g(x-t)) d\mu(t)$$

is  $\mu$ -integrable.

Finally, we can use Fubini's theorem one more time and use the bilinearity of  $L$  twice to calculate

$$\begin{aligned} & \iint L(f(t), g(x-t)) d\mu(t) d\mu(x) \\ & = \iint L(f(t), g(x-t)) d\mu(x) d\mu(t) \\ & = \int L(f(t), \int g(x-t) d\mu(x)) d\mu(t) \\ & = \int L(f(t), \int g(x) d\mu(x)) d\mu(t) \\ & = L(\int f(t) d\mu(t), \int g(x) d\mu(x)). \end{aligned} \quad \square$$

## 6 Smoothness

One main property of the convolution is that it can smoothen a function. That is to say, if one takes the convolution with a smooth function, the convolution is smooth than the original function, even if the original function is only locally integrable. We will explore this behavior in this section. In this section we will assume that  $g$  is continuous/differentiable/ $C^n$ . Using the commutativity of convolution one can show the same for the case where  $f$  is smooth (assuming that  $G$  is abelian and  $\mu$  left-invariant and negation invariant).

### 6.1 Continuity and Differentiability

In this paper, we denote the *support*  $\square$  of a function  $f$  as

$$\text{supp}(f) := \{x \mid f(x) \neq 0\}.$$

This notion should be contrasted with the *topological support*  $\square$  of a function  $f$ , which is the closure of  $\text{supp}(f)$ :

$$\text{tsupp}(f) := \overline{\text{supp}(f)}.$$

We say that a function *has compact support*  $\square$  if its topological support is compact.

**Proposition 6.1.** *Suppose that  $G$  is a second-countable locally compact Hausdorff topological group. If  $f$  is locally  $\mu$ -integrable and  $g$  is continuous with compact support then  $f *_{L,\mu} g$  is continuous.  $\square$*

*Proof.* To show that  $f *_{L,\mu} g$  is continuous at  $x_0$ , we use Theorem 2.1. Clearly,  $x \mapsto L(f(t), g(x-t))$  is continuous.

To show that the map  $t \mapsto L(f(t), g(x-t))$  is  $\mu$ -a.e. strongly measurable, note that  $f$  is  $\mu$ -a.e. strongly measurable, since we can write  $G$  as the union of a countable collection of compact sets, and  $f$  is  $\mu$ -a.e. strongly measurable on each of those sets, hence on all of  $G$ . Furthermore,  $L$  and  $t \mapsto g(x-t)$  are continuous, hence  $\mu$ -a.e. strongly measurable, so the composition is also  $\mu$ -a.e. strongly measurable.

To provide the integrable bound, let  $K$  be a compact neighborhood of  $x_0$  and let  $K' = -\text{tsupp}(g) + K$ , which is also compact. For  $x \in K$  and any

$t \in G$  we have

$$\begin{aligned} & \|L(f(t), g(x-t))\| \\ & \leq \|L\| \cdot \|f(t)\| \cdot \|g(x-t)\| \\ & \leq \begin{cases} \|L\| \cdot \|f(t)\| \cdot \sup \|g\| & \text{if } t \in K' \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Since  $f$  is integrable on  $K'$ , this last function is integrable.  $\square$

We have to be careful with the above statement: our definition of integration includes the integral is defined to be 0 if the integrand is not integrable. So we have to separately prove that the convolution really exists. In this case, that is indeed true, since the integrand is a  $\mu$ -a.e. strongly measurable function that is bounded by an integrable function, so it is itself integrable.

We also prove a variant of Theorem 6.1 that with a stronger condition on  $f$ , namely that that  $f$  is  $\mu$ -integrable, but weaker conditions on  $g$ , namely that  $g$  is continuous and  $\|g\|$  is bounded.  $\square$

For the remainder of this section, assume that  $\mathbb{k} = \mathbb{R}$  or  $\mathbb{k} = \mathbb{C}$ , that  $G$  is a finite-dimensional normed space over  $\mathbb{k}$  and that  $\mu$  is  $\sigma$ -finite and left-invariant.

As in the discussion in Section 3, we have to define a new bilinear map to state what the derivative of a convolution is. In this case, we define the continuous linear map

$$L^G : E \times L(G, E') \rightarrow L(G, F)$$

as

$$L^G(x, f) = L(x, -) \circ f.$$

We can now write down the derivative of a convolution elegantly.

**Proposition 6.2.** *If  $f$  is locally  $\mu$ -integrable and  $g$  is  $C^1$  with compact support then  $f *_{L, \mu} g$  has derivative  $f *_{L^G, \mu} Dg$ .  $\square$*

*Proof.* We use Theorem 2.2 with  $F(x, t) := L(f(t), g(x-t))$  and  $F'(x, t) := L^G(f(t), D_{x-t}g)$ . We already showed that  $F(x, -)$  is integrable and hence  $\mu$ -a.e. strongly measurable. Since  $Dg$  is continuous,  $F'(x, -)$  is also  $\mu$ -a.e. strongly measurable. The reason that the derivative is dominated by an integrable function is the same as in the proof of Theorem 6.1. Finally, to compute the derivative of  $F$  we use the chain rule:

$$\begin{aligned} D_x F(-, t) &= D_{g(x-t)} L(f(t), -) \circ D_x(x \mapsto g(x-t)) \\ &= L(f(t), -) \circ D_{x-t} g \\ &= L^G(f(t), D_{x-t} g) = F'(x, t). \end{aligned} \quad \square$$



From this, we get the appropriate smoothness condition for convolutions.

**Proposition 6.3.** *If  $f$  is locally  $\mu$ -integrable and  $g$  is  $C^n$  with compact support (for  $n \in \mathbb{N} \cup \{\infty\}$ ) then  $f *_{L,\mu} g$  is  $C^n$ . [↗](#)*

## 6.2 A Digression on Universes

Before we delve into the proof of Theorem 6.3, we need to discuss a foundational obstacle in Lean when doing inductive proofs of this form. The idea of this proof is by performing induction on  $n$ . In the successor case, one can use the equivalence

```
cont_diff ℓ (n + 1) f ↔ differentiable ℓ f ∧
  cont_diff ℓ n (fderiv ℓ f)
```

To show that  $f *_{L,\mu} g$  is differentiable we use Theorem 6.2 and we use the induction principle to show that  $D(f *_{L,\mu} g)$  is  $C^n$ . The problem is that  $f *_{L,\mu} g$  has type  $G \rightarrow F$  and  $D(f *_{L,\mu} g)$  has type  $G \rightarrow L(G, F)$ . The types  $G$  and  $F$  live in certain universes, say  $\mathcal{U}_a$  resp.  $\mathcal{U}_b$  (this would be denoted **Type a** and **Type b** in Lean). In `mathlib`, we want to be maximally general, so we want to allow for the possibility that  $G$  and  $F$  live in different universes, so we don't want to assume that  $a$  and  $b$  are the same. Now the catch is that  $L(G, F)$  has type  $\mathcal{U}_{\max(a,b)}$ . Since Lean does not have cumulative universes [vD18], we cannot get  $E$  to have type  $\mathcal{U}_{\max(a,b)}$  automatically. This is a problem for the induction: the universe levels in the base case have to be the same as the universe levels in the induction hypothesis. A similar problem occurs with the derivative of  $g : G \rightarrow E'$ , since  $Dg$  has type  $G \rightarrow L(G, E')$ , which lives in a different universe than  $E'$ .

To fix the induction argument, one has to assume that the universe levels of  $E'$  and  $F$  are larger than the universe level of  $G$ . One can do that by assuming that  $G : \mathcal{U}_a$ ,  $E : \mathcal{U}_b$ ,  $E' : \mathcal{U}_{\max(a,c)}$  and  $F : \mathcal{U}_{\max(a,d)}$ . In this case, the type of  $E'$  and  $L(G, E)$  are the same, namely  $\mathcal{U}_{\max(a,c)}$ . Similarly, the types of  $F$  and  $L(G, F)$  are the same, and the induction goes through. To get the fully general version of the lemma without universe restrictions, one can use *universe lifting* `ulift : \mathcal{U}_a \rightarrow \mathcal{U}_{\max(a,b)}`. `ulift(F)` is a copy of type  $F$  in a higher universe level, which comes with a bijection `up : F \rightarrow ulift F`. We can apply the version of the lemma with restricted universes to  $f *_{\text{up} \circ L, \mu}$  (`up \circ g`) to derive it for  $f *_{L, \mu} g$ . However, this is unpleasant to do, since it requires a lot of lemmas stating that the map `up` commutes with various operations and preserves various properties (for example, to write the aforementioned convolution, we need to know that `up` is a continuous linear map).

This proof works, but it is not convenient to give, especially since these induction proofs are very common to prove that some operation respects  $C^n$

function. Note that with universe cumulativity this problem disappears: if  $F : \mathcal{U}_a$  then  $F$  also has type  $\mathcal{U}_{\max(a,b)}$ . Usually the lack of universe cumulativity is completely invisible to the user, and occasionally it is a very minor inconvenience (for example, when one has to write down universe levels explicitly). This is the only issue that the author of this paper has encountered where the lack of universe cumulativity is annoying.<sup>4</sup>

However, for our current purpose there is a much simpler proof that avoids all of these universe issues.

We have proven a different version of the characterization of being  $C^{n+1}$  (in collaboration with [redacted] and [redacted]), namely<sup>↗</sup>

```
cont_diff k (n + 1) f ↔ differentiable k f ∧
  ∀ v, cont_diff k n (λ x, fderiv k f x v)
```

However, we have proven this only in the case that the domain of  $f$  is finite dimensional. The difference with the previous formulation is that instead of stating that  $Df$  is  $C^n$ , viewed as a map into the space of continuous linear maps, we look at all directional derivatives, and require that  $x \mapsto D_x f(v)$  (denoted  $D_{(-)}f(v)$ ) is smooth for all vectors  $v$ .

The reason that this avoids the universe issue mentioned above is that the map  $D_{(-)}f(v)$  has the same domain and codomain as  $f$ , and therefore the induction works without any universe restrictions.

The proof of this characterization of being  $C^{n+1}$  is not very difficult if the domain is finite dimensional, since we can reconstruct a linear map from its values on set of vectors forming a basis, and this reconstructed linear map is  $C^n$  iff its values on the basis are  $C^n$ . However, this characterization of being  $C^{n+1}$  is false when the domain of  $f$  is not finite-dimensional, so we cannot use this property for all proofs. In particular, the proof that the composition of  $C^n$  functions is  $C^n$  (done by [redacted]) still uses a cumbersome proof that has to work with universe lifts.

### 6.3 Proving Smoothness

In Theorem 6.3, we are assuming that the domain  $G$  of all functions involved is finite dimensional, so we can use the the characterization of being  $C^{n+1}$  described in Section 6.2.

*Proof of Theorem 6.3.* We perform induction on  $n$ . For  $n = 0$  it follows directly from Theorem 6.1. If  $n = k + 1$ , then we know that  $g$  is differentiable,

---

<sup>4</sup>But it should be noted that there are also issues with universe cumulativity [Luo12].

so  $f *_{L,\mu} g$  is differentiable by Theorem 6.2. We can also compute

$$\begin{aligned} D_x(f *_{L,\mu} g)(v) &= (f *_{L^G,\mu} Dg)(x)(v) \\ &= \left( \int L^G(f(t), D_{x-t}g) d\mu(t) \right) (v) \\ &= \int L(f(t), D_{x-t}g(v)) \circ d\mu(t) \\ &= (f *_{L,\mu} D_{(-)}g(v))(x). \end{aligned}$$

Note that  $D_{(-)}g(v)$  has compact support and is  $C^n$ , so we can apply the induction hypothesis to  $f$  and  $D_{(-)}g(v)$ . We conclude that  $f *_{L,\mu} g$  is  $C^{n+1}$ .

We automatically get the case  $n = \infty$ , since being  $C^\infty$  is equivalent to being  $C^n$  for all  $n$ .  $\square$

## 7 Approximations of Functions

In this section we will discuss how the convolution can be used to approximate functions. If  $x$  is a point in a metric space, we write

$$B_R(x) = \{y \mid d(x, y) < R\}$$

for the open ball around  $x$  with radius  $R$ .

In this section, let  $G$  be a second-countable seminormed group,  $\mu$  a left-invariant  $\sigma$ -finite measure,  $\mathbb{k} = \mathbb{R}$  and  $E$  be a Banach space.

**Lemma 7.1.** *Let  $x_0 \in G$  and assume that  $f$  is  $\mu$ -integrable with  $\text{supp}(f) \subseteq B_R(0)$ . Suppose that  $g$  is strongly  $\mu$ -a.e. measurable such that for all  $x \in B_R(x_0)$  we have  $d(g(x), g(x_0)) \leq \varepsilon$  for some  $\varepsilon \geq 0$ . Then the distance between  $(f *_{L,\mu} g)(x_0)$  and  $\int_G L(f(t), g(x_0)) d\mu(t)$  is at most*

$$\varepsilon \|L\| \cdot \int_G \|f(x)\| d\mu(x). \quad \square$$

*Proof.* First note that the convolution  $f *_{L,\mu} g$  exists at  $x_0$ , since  $f$  is integrable on  $B_R(0)$  and  $g$  is bounded on  $B_R(x_0)$  (we omit the details). Note that for all  $t \in G$  we have

$$d(L(f(t), g(x_0 - t)), L(f(t), g(x_0))) \leq \varepsilon \|L(f(t), -)\|. \quad (4)$$

Indeed, if  $t \in \text{supp}(f) \subseteq B_R(0)$  then  $d(g(x), g(x_0)) \leq \varepsilon$ , hence (4) holds. Otherwise the inequality is also valid, since both sides are equal to 0. Therefore,

$$\begin{aligned} & d((f *_{L,\mu} g)(x_0), \int_G L(f(t), g(x_0)) d\mu(t)) \\ & \leq \int_G \varepsilon \|L(f(t), -)\| d\mu(t) \\ & \leq \int_G \varepsilon \|L\| \cdot \|f(t)\| d\mu(t) \\ & = \varepsilon \|L\| \cdot \int_G \|f(t)\| d\mu(t) \end{aligned}$$

as desired.  $\square$

Write  $\varphi *_{\mu} g$  for the convolution when the bilinear map is scalar multiplication.

Given a filter  $M$  on  $X$  we can define the filter “small sets w.r.t.  $M$ ” on the power set  $\mathcal{P}(X)$ , defined as the largest filter containing  $\mathcal{P}(A)$  for  $A \in M$ .  $\square$

The defining property is that if  $f : Y \rightarrow \mathcal{P}(X)$  is a family of sets and  $N$  is a filter on  $Y$ , then  $f(x)$  tends to the small sets w.r.t.  $M$  as  $x$  tends to  $N$  if and only if for all  $A \in M$  we have  $\square$

$$\forall^f y \in N, f(x) \subseteq A.$$

**Proposition 7.2.** *Let  $M$  be a filter on some indexing type  $I$  and let  $(\varphi_i)_{i \in I}$  be a sequence of nonnegative functions  $G \rightarrow \mathbb{R}$  with integral 1. Suppose that  $\text{supp}(\varphi_i)$  tends to small sets w.r.t.  $\mathcal{N}_0$  as  $i$  tends to  $M$ . If  $g$  is a strongly  $\mu$ -a.e. measurable function continuous at  $x_0 \in G$ , then  $i \mapsto (\varphi_i *_{\mu} g)(x_0)$  tends to  $\mathcal{N}_{g(x_0)}$  as  $i$  tends to  $M$ .  $\square$*

*Proof.* We have to show for all  $\varepsilon > 0$  that

$$\forall^f i \in M, d((\varphi_i *_{\mu} g)(x_0), g(x_0)) < \varepsilon.$$

Since  $g$  is continuous at  $x_0$ , we find a  $\delta > 0$  such that  $d(g(x), g(x_0)) < \frac{\varepsilon}{2}$  for  $x \in B_{\delta}(x_0)$ . By the condition on  $\text{supp}(\varphi_i)$  we know that

$$\forall^f i \in M, \text{supp}(\varphi_i) \subseteq B_{\delta}(0).$$

If  $i \in I$  such that  $\text{supp}(\varphi_i) \subseteq B_{\delta}(0)$  then it suffices to show that  $d((\varphi_i *_{\mu} g)(x_0), g(x_0)) < \varepsilon$ . We apply Theorem 7.1 (noting that  $\varphi_i$  is integrable, since its integral is nonzero) to conclude that

$$d((\varphi_i *_{\mu} g)(x_0), \int_G \varphi_i(t) g(x_0) d\mu(t)) \leq \frac{\varepsilon}{2} < \varepsilon,$$

using that the norm of scalar multiplication is 1. We obtain the desired result by noting that

$$\int_G \varphi_i(t)g(x_0)d\mu(t) = g(x_0) \int_G \varphi_i(t)d\mu(t) = g(x_0). \quad \square$$

In the formalization the of the proof of Theorem 7.2 is a combination of two lemmas. The combined length of the proofs of these lemmas is 14 lines long, showing that `mathlib` provides a useful library for topology and calculus, where we can efficiently prove tricky arguments.

Theorem 7.2 is particularly useful when combined with [redacted]'s work on defining smooth bump functions  $E \rightarrow \mathbb{R}$ ,<sup>☞</sup> where  $E$  is a finite dimensional normed vector space over  $\mathbb{R}$ . Suppose  $\varphi_R$  is such a smooth bump function on  $G$  with  $\text{tsupp}(\varphi_R) = B_R(0)$ , normed so that  $\int \varphi_R(x)d\mu(x) = 1$ . Using these functions, we get a much simpler corollary from Theorem 7.2.

**Corollary 7.3.** *If  $g$  is a strongly  $\mu$ -a.e. measurable function continuous at  $x_0 \in G$ , then  $R \mapsto (\varphi_R *_{\mu} g)(x_0)$  tends to  $\mathcal{N}_{g(x_0)}$  as  $R$  tends to 0.*<sup>☞</sup>

## 8 Concluding Thoughts

This formalization was added to `mathlib` in a sequence of 12 pull requests. These pull requests added roughly 400 lines of code to existing files throughout the filter, analysis and measure theory libraries and 900 lines of code to a new file about convolutions.

Therefore, this is not a particularly long formalization, despite the various results we have formalized. The shortness is a virtue: it shows that one can do real mathematics and do complex arguments in analysis with relatively little effort.

We notice that the type-classes Lean uses reduce the cognitive overhead of the formalizer. Lean regularly gives an error, complaining that there is a missing type-class assumption in a proposition statement. Furthermore, `mathlib` has a suite of linters that check if lemmas unused type-class assumptions [vDEL20]. It can also automatically check whether a weaker type-class suffices for a particular lemma [Bes21]. This means that the end-user can give a quick approximation of the needed type-classes, and Lean will give feedback if more or fewer are needed.

This convolution library has already been used in the Sphere Eversion Project<sup>☞</sup> to show that a particular sequence of loops approximate a given loop using Theorem 7.3.<sup>☞</sup>

It would be interesting to write more theories on top of this library for convolutions. In particular it would be interesting to define distributions.

The inclusion map from locally integrable function to distributions can be defined in terms of a convolution of two functions, and the properties of this map should be easily derivable from the properties in this paper.

In Sections 6 and 7 we focused on the convolution when one of the functions is compactly supported. There are other conditions one can impose on the functions to still get a nice theory of convolutions. For example, one can relax the condition that one of the function is compactly supported, by requiring that both functions are rapidly decreasing. It would be interesting to investigate the convolution with other requirements on the functions.

## Acknowledgements

I would like to thanks to Patrick Massot for the mentoring and collaboration during my postdoc, and for proofreading the first version of this article. I would like to thank everyone in the `mathlib` community for developing a usable and convenient library. Finally, I would like to thank Fondation Mathématique Jacques Hadamard for the financial support for this project.

## References

- [Baa22] Anne Baanen, *Use and Abuse of Instance Parameters in the Lean Mathematical Library*, 13th International Conference on Interactive Theorem Proving (ITP 2022) (Dagstuhl, Germany), Leibniz International Proceedings in Informatics (LIPIcs), vol. 237, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, pp. 4:1–4:20, doi:10.4230/LIPIcs.ITP.2022.4, <https://drops.dagstuhl.de/opus/volltexte/2022/16713>.
- [Bes21] Alexander Best, *Automatically generalizing theorems using typeclasses*, 2021, <https://easychair.org/publications/preprint/KLfT>.
- [BLM15] Sylvie Boldo, Catherine Lelay, and Guillaume Melquiond, *Coquelicot: A user-friendly library of real analysis for coq*, Mathematics in Computer Science **9** (2015), no. 1, 41–62.
- [Com20] The `mathlib` Community, *The Lean mathematical library*, Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (New York, NY, USA), CPP 2020, Association for Computing Machinery, 2020, p. 367–381, doi:10.1145/3372885.3373824.

- [CP88] Thierry Coquand and Christine Paulin, *Inductively defined types*, COLOG-88, International Conference on Computer Logic, Tallinn, USSR, December 1988, Proceedings, 1988, pp. 50–66.
- [dMKA<sup>+</sup>15] Leonardo de Moura, Soonho Kong, Jeremy Avigad, Floris van Doorn, and Jakob von Raumer, *The Lean Theorem Prover (system description)*, 378–388.
- [Gou22] Sébastien Gouëzel, *A formalization of the change of variables formula for integrals in mathlib*, Intelligent Computer Mathematics (Cham) (Kevin Buzzard and Temur Kutsia, eds.), Springer International Publishing, 2022, pp. 3–18.
- [Har13] John Harrison, *The HOL Light theory of Euclidean space*, Journal of Automated Reasoning **50** (2013), no. 2, 173–190, doi:10.1007/s10817-012-9250-9.
- [HIH13] Johannes Hölzl, Fabian Immler, and Brian Huffman, *Type Classes and Filters for Mathematical Analysis in Isabelle/HOL*, Interactive Theorem Proving - 4th International Conference, ITP 2013, Rennes, France, July 22-26, 2013. Proceedings, 2013, pp. 279–294.
- [Hir76] Morris Hirsch, *Differential topology*, 1976, doi:10.1007/978-1-4684-9449-5.
- [Kud22] Yury Kudryashov, *Formalizing the Divergence Theorem and the Cauchy Integral Formula in Lean*, 13th International Conference on Interactive Theorem Proving (ITP 2022) (Dagstuhl, Germany), Leibniz International Proceedings in Informatics (LIPIcs), vol. 237, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, pp. 23:1–23:19, doi:10.4230/LIPIcs.ITP.2022.23, <https://drops.dagstuhl.de/opus/volltexte/2022/16732>.
- [Luo12] Zhaohui Luo, *Notes on universes in type theory*, preprint, 2012, <http://www.cs.rhul.ac.uk/home/zhaohui/universes.pdf>.
- [SvdW11] B.A.W. Spitters and E.E. van der Weegen, *Type classes for mathematics in type theory*, Mathematical Structures in Computer Science **21** (2011), 795–825.
- [vD18] Floris van Doorn, *On the formalization of higher inductive types and synthetic homotopy theory*, Ph.D. thesis, Carnegie Mellon University, 2018, <https://arxiv.org/abs/1808.10690>.

- [vD21] ———, *Formalized Haar Measure*, 12th International Conference on Interactive Theorem Proving (ITP 2021), Leibniz International Proceedings in Informatics (LIPIcs), vol. 193, 2021, pp. 18:1–18:17, [doi:10.4230/LIPIcs.ITP.2021.18](https://doi.org/10.4230/LIPIcs.ITP.2021.18).
- [vDEL20] Floris van Doorn, Gabriel Ebner, and Robert Y. Lewis, *Maintaining a library of formal mathematics*, Intelligent Computer Mathematics, Springer International Publishing, 2020, pp. 251–267.
- [WB89] P. Wadler and S. Blott, *How to make ad-hoc polymorphism less ad hoc*, Proceedings of the 16th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (New York, NY, USA), POPL '89, Association for Computing Machinery, 1989, p. 60–76, [doi:10.1145/75277.75283](https://doi.org/10.1145/75277.75283), <https://doi.org/10.1145/75277.75283>.
- [Yos95] Kôsaku Yosida, *Functional analysis*, 1995, [doi:10.1007/978-3-642-61859-8](https://doi.org/10.1007/978-3-642-61859-8).