Formalisation of the Calderón Transference Principle in Ergodic Theory

Felix Pernegger

Born 1st September, 2004 in Ried im Innkreis, Austria

1st September, 2025

Bachelor's Thesis Mathematics

Advisor: Prof. Dr. Floris van Doorn

Second Advisor: Prof. Dr. Asgar Jamneshan

Mathematisches Institut

Mathematisch-Naturwissenschaftliche Fakultät der

Rheinischen Friedrich-Wilhelms-Universität Bonn

Abstract

This Bachelor's thesis deals with the formalisation of an instance of the Calderón transference principle via the proof assistant Lean and its mathematical library mathlib. In particular, we reduce norm-variational estimations of ergodic averages to a result purely in harmonic analysis. Doing so, our formalisation gives a "first step" of a potential formalised proof of these estimations.

Furthermore, we discuss the r-variation of a function in a general setting and present a formalisation thereof.
Finally, we also briefly introduce and formalise a combinatorial space together with a technical lemma, filling a small gap in the original proof of the transference principle in a certain case.

While we only present selected parts of the Lean code, all formalised results can be found in `https://www.github.com/felixpernegger/ErgodicAverages`.

Zusammenfassung in deutscher Sprache

Diese Bachelorarbeit befasst sich mit der Formalisierung einer Instanz des Calderón Übertragungsprinzips mithilfe des Beweisassistenten Lean und seiner mathematischen library mathlib. Insbesondere reduzieren wir normvariationale Schätzungen ergodischer Mittelwerte auf ein Ergebnis in harmonischer Analysis. Damit stellt unsere Formalisierung einen "ersten Schritt" für einen möglichen formalisierten Beweis dieser Schätzungen dar.

Darüber hinaus diskutieren wir die r-Variation einer Funktion in einem allgemeinen Kontext und präsentieren eine entsprechende Formalisierung.
Ebenso führen wir kurz einen kombinatorischen Raum ein und formalisieren ihn zusammen mit einem technischen Lemma. Damit schließen wir eine kleine Lücke im ursprünglichen Beweis des Transformationsprinzips in einem bestimmten Fall.

Wir präsentieren zwar nur ausgewählte Teile des Lean-Codes; alle formalisierten Ergebnisse sind zu finden in `https://www.github.com/felixpernegger/ErgodicAverages`.

Contents

1 Introduction

1.1 Ergodic averages and rate of convergence

A classical theorem in Ergodic theory is the mean-ergodic theorem as proved by von Neumann.

Theorem 1.1. *Given a measure space* $(X, \mathcal{F}, \mu)$ *and measure-preserving* $S : X \to X$*, i.e.* S *measurable and for all* $E \in \mathcal{F}$*,* $\mu(S^{-1}E) = E$*, and* $f \in \mathrm{L}^2(X)$ *we define the* n-th ergodic average *of* f *as*

$$M_n f(x) := \frac{1}{n} \sum_{k=0}^{n-1} f(S^k x).$$

Then $M_n f$ *converges in the* $\mathrm{L}^2(X)$ *norm for* $n \to \infty$*.*

One of several possible generalisations of this result introduces *multilinear ergodic averages.* Namely, for $S_1, \dots, S_p : X \to X$ and $f_1, \dots, f_l : X \to \mathbb{R}$, we define

$$M_n(f_1, \dots, f_l)(x) := \frac{1}{n} \sum_{k=0}^{n-1} \prod_{j=1}^{l} f_j(S_j^k x).$$

Typically, we will assume that $S_1, \dots, S_l$ are measure-preserving and pairwise commuting. Analogous statements to Theorem 1.1 in the case $l = 2$ were proven by Conze and Lesigne in [5] and for $l = 3$ by Frantzikinakis and Kra [11].
Finally, Tao [15] proved a version for general $1 \leq l$ via a combinatorial approach.

Theorem 1.2. *Assume that* $S_1, \dots, S_l : X \to X$ *are measure-preserving, pairwise commuting and invertible. Then for* $f_1, \dots, f_p \in \mathrm{L}^\infty(X)$*,* $M_n(f_1, \dots, f_l)$ *converges in the* L^2*-norm as* $n \to \infty$*.*

Typically, classical proofs of these results do not give us estimates about the rate of convergence. We are interested in quantitative results about the rate of convergence that may be stated in terms of the variation norm, as introduced and discussed in detail in chapter 4.
For the case $l = 1$, such an estimate, utilizing the spectral theorem for unitary operators, was proven by Jones, Ostrovskii and Rosenblatt [12], showing the existence of a finite constant C, independent of $(X, \mathcal{F}, \mu)$, $f \in \mathrm{L}^2(X)$ and $S : X \to X$, such that

$$\sum_{j=1}^{m} \left\| M_{n_j} f - M_{n_{j-1}} f \right\|_{\mathrm{L}^2(X)}^2 \leq C \left\| f \right\|_{\mathrm{L}^2(X)}^2 \tag{1.1}$$

for all $0 < n_0 < \dots < n_m$.
Via the variation norm, this may be briefly stated as

$$\left\| M_n f \right\|_{\mathrm{V}_n^2(\mathbb{N} - \{0\}, \mathrm{L}^2(X))} \leq C^{\frac{1}{2}} \left\| f \right\|_{\mathrm{L}^2(X)}.$$

For $l = 2$, Durcik, Kovač, Škreb and Thiele [10] proved a similar result.

Theorem 1.3. *There is a finite constant C such that for any σ-finite measure space $(X, \mathcal{F}, \mu)$ and commuting measure-preserving transformations $S_1, S_2 : X \to X$ and $f_1, f_2 \in \mathrm{L}^4(X)$ we have*

$$\sum_{j=1}^{m} \left\| M_{n_j}(f_1, f_2) - M_{n_{j-1}}(f_1, f_2) \right\|_{\mathrm{L}^2(X)}^2 \leq C \, \|f_1\|_{\mathrm{L}^4(X)}^2 \, \|f_2\|_{\mathrm{L}^4(X)}^2$$

And for $l = 3$ by Durcik, Slavíková and Thiele [9], albeit in a slightly different manner.

Theorem 1.4. *For all $r > 4$, there exists a constant $C_r > 0$ such that the following holds. Let $(X, \mathcal{F}, \mu)$ be a σ-finite measure space, $S_1, S_2, S_3 : X \to X$ pairwise commuting measure-preserving transformations, $n_0 < \cdots < n_m$ positive integers. For any $f_1, f_2 \in \mathrm{L}^8(X)$ and $f_3 \in \mathrm{L}^4(X)$, we have*

$$\sum_{j=1}^{m} \left\| M_{n_j}(f_1, f_2, f_3) - M_{n_{j-1}}(f_1, f_2, f_3) \right\|_{\mathrm{L}^2(X)}^r \leq C_r \, \|f_1\|_{\mathrm{L}^8(X)}^r \, \|f_2\|_{\mathrm{L}^8(X)}^r \, \|f_3\|_{\mathrm{L}^4(X)}^r .$$

Theorem 1.4 can be generalised to other tuples than $(8, 8, 4)$ via interpolation of permutations of $(8, 8, 4)$, which for example gives an analogue statement for $(6, 6, 6)$.

We can write both theorems briefly via the variation norm from chapter 4 as

$$\|M_n(f_1, f_2)\|_{\mathrm{V}_n^2(\mathbb{N}, \mathrm{L}^2(X))} \leq C^{\frac{1}{2}} \, \|f_1\|_{\mathrm{L}^4(X)} \, \|f_2\|_{\mathrm{L}^4(X)}$$

and

$$\|M_n(f_1, f_2, f_3)\|_{\mathrm{V}_n^r(\mathbb{N}, \mathrm{L}^2(X))} \leq C_r^{\frac{1}{r}} \, \|f_1\|_{\mathrm{L}^8(X)} \, \|f_2\|_{\mathrm{L}^8(X)} \, \|f_3\|_{\mathrm{L}^4(X)}$$

respectively.

Note that all of the estimates about rate of convergence do in fact imply convergence in the L^2-norm.

1.2 Calderón's transference principle

While the proof of Theorem 1.1 is relatively short and direct, both Theorem 1.3 and Theorem 1.4 reduce the statement first to a statement in harmonic analysis. This is done via a version of the well-known *Calderón transference principle.* The principle is named after Alberto Pedro Calderón, who first proved a version thereof in 1968 [3].
For Theorem 1.3 the transferred version as used in [10] is

Theorem 1.5. *There exists a finite constant C, such that for all $F_1, F_2 \in \mathrm{L}^4(\mathbb{R}^2)$ we have*

$$\sum_{j=1}^{m} \left\| A_{t_j}^{\chi_{[0,1)}}(F_1, F_2) - A_{t_{j-1}}^{\chi_{[0,1)}}(F_1, F_2) \right\|_{\mathrm{L}^2(\mathbb{R}^2)} \leq C \, \|F_1\|_{\mathrm{L}^4(\mathbb{R}^2)} \, \|F_2\|_{\mathrm{L}^4(\mathbb{R}^2)}$$

for all positive real numbers $t_0 < \cdots < t_m$, where for $\varphi : \mathbb{R} \to \mathbb{R}$

$$A_t^{\varphi}(F_1, F_2)(x, y) := \int_{\mathbb{R}} F_1(x + s, y) F_2(x, y + s) t^{-1} \varphi(t^{-1} s) \, ds.$$

Similarly, Theorem 1.4 in [9] is transferred from

Theorem 1.6. *For all $r > 4$, there exists a finite constant C_r, such that for all $F_1, F_2 \in \mathrm{L}^8(\mathbb{R}^3)$ and $F_3 \in \mathrm{L}^4(\mathbb{R}^3)$ we have*

$$\sum_{j=1}^{m} \left\| A_{t_j}^{\chi_{[0,1)}}(F_1, F_2, F_3) - A_{t_{j-1}}^{\chi_{[0,1)}}(F_1, F_2, F_3) \right\|_{\mathrm{L}^2(\mathbb{R}^2)}^r \leq C \left\|F_1\right\|_{\mathrm{L}^8(\mathbb{R}^3)}^r \left\|F_2\right\|_{\mathrm{L}^8(\mathbb{R}^3)}^r \left\|F_3\right\|_{\mathrm{L}^4(\mathbb{R}^3)}^r$$

for all positive real numbers $t_0 < \cdots < t_m$, where A_t^{φ} is defined analogously as in Theorem 1.5.

Note that in both cases, the transferred version entirely avoids measure-preserving transformations and only deals with function on $\mathbb{R}^n$. The appropriate version of Calderón's transference principle is described in [10] and content of chapter 2.
On a high level, the proof idea is to first apply Theorem 1.3 to certain simple functions, which after some estimations gives us the analogue statement for $\widetilde{F}_1, \widetilde{F}_2 \in \ell^4(\mathbb{Z}^2)$ and

$$\widetilde{A}(\widetilde{F}_1, \widetilde{F}_2)(k, l) := \frac{1}{n} \sum_{i=0}^{n-1} \widetilde{F}_1(k + i, l)\widetilde{F}_2(k, l + i).$$

Then for $f_1, f_2 : X \to \mathbb{R}$ this discrete version is applied to push-forward functions of the form

$$\widetilde{F}_{i,x,N}(k, l) := \begin{cases} f_i(S_1^k S_2^l x) & \text{if } 0 \leq k, l \leq 2N - 1 \\ 0 & \text{otherwise} \end{cases}$$

For some fixed $x \in X$ and $N \in \mathbb{N}$, which again after some estimations gives Theorem 1.3.

The aim of this thesis is to formalise this version of the Calderón Transference principle, generalised for l functions, in the proof assistant Lean and therefore to provide a "first step" of a possible formalisation of Theorem 1.3 and Theorem 1.4.
Formalising a statement in Lean typically requires a very detailed written proof, called "blueprint" to work with. The blueprint of the transference principle as described above is the content of chapter 2 and thus our core argument.

Coming back to the $l = 3$ case, in [9] the transfer of Theorem 1.6 to Theorem 1.4 is omitted and instead just referred to the $l = 2$ case from before as described in [10].
However, we did not actually manage to construct an analogue proof for $l = 3$. The reason for this is the exponent $r > 4$ in Theorem 1.6. Namely, the transfer in the $l = 2$ often (i.e. in Fact 6 in the blueprint) switches finite sums and integrals of the form

$$\sum_{i=1}^{m} \left\|f_i\right\|_{\mathrm{L}^2(X)}^r = \left\| \sum_{i=1}^{m} f_i \right\|_{\mathrm{L}^2(X)}^r.$$

For nonnegative f_i's and $r = 2$, this is trivially the case.
However, for $r > 4$ the equality evidently does not hold in general.
While there might be a work-around to this issue i.e. using Minkowski's inequality, this motivated us to take a slightly different approach for $l = 3$.

In [9], Theorem 1.6 itself is proven by using an endpoint estimate, whose details are again not discussed in the paper, for $r = 2$ of the form

$$\sum_{j=1}^{J} \left\| A_{t_j}(F_1, F_2, F_3) - A_{t_{j-1}}(F_1, F_2, F_3) \right\|_2^r \leq C J^{\frac{1}{2}} \left\| F_1 \right\|_8^r \left\| F_2 \right\|_8^r \left\| F_3 \right\|_4^r, \tag{1.2}$$

with C again independent of the F_i's and $0 < t_0 < \cdots < t_J$.

Instead of using the endpoint estimate, we can instead first apply Calderón's transference principle as described in chapter 3 to (1.2) giving us an estimation of the form

$$\sum_{j=1}^{J} \left\| M_{n_j}(F_1, F_2, F_3) - M_{n_{j-1}}(F_1, F_2, F_3) \right\|_2^r \leq C J^{\frac{1}{2}} \left\| F_1 \right\|_8^r \left\| F_2 \right\|_8^r \left\| F_3 \right\|_4^r, \tag{1.3}$$

on which we can then apply the endpoint estimate, giving us Theorem 1.4.
The discrete version of Theorem 1.4 can be proved via the same strategy.
The details of the endpoint estimate, presented in a general manner, and formalisation thereof are the content of chapter 3. The main theorem of that chapter (and therefore of this thesis) will be Theorem 2.5.

Many of our statements can be written briefly using the r-variation of a function, which we discuss in detail and formalise in chapter 4.

Finally, in our proof of the transference principle, of the endpoint estimate and formalisations thereof, we never utilize σ-finiteness of X.[1] As again, the harmonic analysis statements entirely avoid talking about abstract measure spaces, we can therefore improve Theorem 1.3 and Theorem 1.4 by not requiring σ-finiteness.

1.3 Lean and mathlib

To formalise our result, we use the proof assistant Lean. Lean, which can also be used as a functional programming language, was originally developed by Leonardo de Moura in 2013 and, as of writing, is currently in its 4th main version.

The main mathematical library for Lean is called mathlib (again, in its current version mathlib4), as of writing, consisting of about 2 million lines of code and over 200.000 formalised mathematical statements with hundreds of contributors. It is being maintained and expanded by an active community.
Very roughly speaking, it consists of about an undergraduate education. When formalising a mathematical statement, one should typically rely on Lean and already formalised mathematical theorems within it as much as possible. Likewise, our formalisation of the Calderón transference principle relies heavily on mathlib. There are also a number of systems in and related to Lean, which help finding appropriate theorems from mathlib, which in practice

[1] As does our Lean formalisation.

reduces the time needed to prove statements dramatically.

The main mathematical theory behind Lean is *dependent type theory.* Roughly speaking, every object in Lean is given some unique type, which can be thought of similar to data types in programming. If some object `a` has type `b`, we write `a : b`.
For example, π (typically) has the type `Real`, the type of real numbers. Depending on the context 0 can have type `Real`, `Nat` (natural numbers), `NNReal` (nonnegative real numbers) and many more, importantly `0 : Nat` and `0 : Real` are formally still different objects. Types themselves also have type, giving us an infinite Type hierarchy.

Mathematical statements, like $\forall p.(\forall q.(p \wedge q \Leftrightarrow q \wedge p))$, $1 + 1 = 2$ or $0 = 1$ have type `Prop` (meaning Proposition). A *proof* of a statement `p : Prop` is an object `q :p`. Lean supports *proof irrelevance*, so if `q : p` and `q' : p` where `p : Prop`, `q` and `q'` are treated as the same. Proving a statement now comes down to constructing a member of a statement via proofs of different statements and rules how to connect them.
For example, `@Eq.refl Nat` has type `∀ a : Nat : a = a` and `And.intro` has type `∀ a b : Prop. a → b → a ∧ b`, meaning if we have a proof of `a` and a proof of `b`, it gives us a proof of `a ∧ b`.
Then, with `0 : Nat` and `1 : Nat`, we therefore have `And.intro (@Eq.refl Nat 0) (@Eq.refl Nat 1) : 0 = 0 ∧ 1=1`. In Lean this looks like:

```
theorem easy_example : 0 = 0 ∧ 1 = 1 :=
  And.intro (@Eq.refl Nat 0) (@Eq.refl Nat 1)
```

Now that we have a proof of `0 =0∧ 1 = 1` we can in turn use it to prove other statements:

```
theorem another_example (n : Nat) : n = n ∧ 0 = 0 ∧ 1 = 1 :=
    And.intro (Eq.refl n) easy_example
```

The `(n : Nat)` is the local *context* of the statement `another_example`. Also, we could have used `@Eq.refl Nat n` instead of `Eq.refl n`, however Lean automatically detects from the context, that `Nat` is the type of equality to be used, thus `Nat` was inserted as an *implicit variable*.
It is easy to see, that for more complicated statements, a proof of this form can get extremely complicated, which is why Lean also has a *tactic* mode which allows to write proofs "backwards" and is much closer to writing a proof on paper, it can be entered by writing "by" after the := at the beginning of a proof.

Arrow types `f: A → B` where `A` and `B` are types, can be thought of as functions. For example @sq Nat has type `Nat → Nat`. For an arrow type `f: A → B` and `a : A`, we can apply `f` to `a`, giving us `f a : B`.
`Nat.add`, addition on natural numbers, has type `Nat → Nat → Nat` (if no brackets are used, this always means `Nat →` (`Nat → Nat`), so for `a : Nat` , we can apply `a` partly to `Nat.add`, via `Nat.add a` or `Nat.add _ a`, which in both cases have type `Nat → Nat`, while `Nat.add a a` has type `Nat`. For readability reasons, unless explicitly stated, for the rest of this thesis we write $f(a, b)$ instead of `f a b` as usual.

Definitions are similar to theorems, minus the proof irrelevance. For example, we can define the natural number zero_add_one as:

```
def zero_add_n : Nat → Nat :=
    fun n ↦ Nat.add 0 n
```

Now proving zero_add_n = 1 can be done as

```
example (n : Nat): zero_add_n n = n := by
  unfold zero_add_one
  exact zero_add n
```

where `zero_add` is a proof that for any monoid `M` and `a : M`, `0 + a = a`. `example` is the same thing as `theorem`, except we cannot name it and therefore not use it for other proofs as well. Note, that definitions in Lean can be similar to definitions of functions in programming languages.
The close correspondence between theorem proving in type theory and programming languages is called the *Curry-Howard isomorphism*. A version thereof was initially observed by Haskell Curry in 1934 [6].

Lean also supports ways to define your own types, most notably inductive types and the special case `structure` we will encounter in chapter 3, but will not elaborate further at this point.

As noted before, mathematical objects that are informally typically treated as the same thing when writing down proofs on paper, like `0 : NNReal` and `0 : Nat` are formally not the same object in Lean. Therefore *coercions* are often needed, which are formally just functions between types, but can be thought of as "embedding" one type in another. For example, there is a coercion from `Nat` to `NNReal` and `NNReal` to `Real`. In many cases, Lean will automatically put coercions on objects, for example `(0 : Nat + (1.5 : Real)` will be automatically interpreted as `(↑(0 : Nat) : Real) + (1.5 : Real)`.
Nevertheless, issues with types of this kind, pose a big challenge when formalising statements in Lean. For example, for much of our formalisations we worked with numbers in ENNReal, the extended nonnegative numbers, where something like $a - b + c = a + c - b$ will not hold in general.

Another challenge one encounters in formalisation, especially in analysis, is needing to prove statements and properties which are typically not even thought of. For example, we sometimes had to prove manually that a given function is measurable. An interesting instance of such a problem, which on paper seems trivial but requires a lot of thought in Lean is described in chapter 5.

Obviously, this short introduction is far short of a proper introduction to Lean and formalised mathematics. A playful introduction to Lean is the Natural Numbers Game.
A beginner-friendly book to working with Lean and mathlib for mathematics is *Mathematics in Lean* [2]. Another introductory book focusing more on the technical aspects of Lean and formalisation is *Theorem Proving in Lean 4* [1].

2 Blueprint of the transference principle

In the following, let $g : \mathbb{N} \to \mathbb{R}_{\geq 0}$.
The following is a detailed generalisation of the version of the Calderón transference principle presented in [10] for p functions.
With g constant and $p = 2, q_1 = q_2 = \frac{1}{4}$ this gives the deduction of Theorem 1 from Theorem 2 in [10]. For the $p = 3$ version as in [9], we need an additional argument for a suitable deduction, which will be discussed in the next chapter 3.
We adapt the proof structure and indexing from [10], namely "Fact 2" refers to "(5.2)" in [10] and theorem `fact5_2` in the formalisation.

Definition 2.1 (Ergodic average)**.** *Let $(X, \mathcal{F}, \mu)$ be a measure space. For $f_1, \dots, f_p : X \to \mathbb{R}$ and $S_1, \dots, S_p : X \to X$ measure-preserving, let the ergodic average be defined as:*

$$M_n(f_1, \dots, f_p)(x) := \frac{1}{n} \sum_{i=0}^{n-1} \prod_{j=1}^{p} f_j(S_j^i x)$$

Definition 2.2 (Multilinear averages)**.** *Let $\varphi : \mathbb{R} \to \mathbb{R}$ integrable, $t > 0$, and $F_1, \dots, F_p : \mathbb{R}^p \to \mathbb{R}$ (such that the expression below is well-defined), then let*

$$A_t^{\varphi}(F_1, \dots, F_p)(x_1, \dots, x_p) := \int_{\mathbb{R}} \left(\prod_{j=1}^{p} F_j(x_1, \dots, x_{j-1}, x_j + s, x_{j+1}, \dots, x_p) \right) \cdot t^{-1} \varphi(t^{-1} s) \, ds.$$

In particular let

$$\begin{aligned} A_t(F_1, \dots, F_p) &:= A_t^{\chi_{[0,1)}}(F_1, \dots, F_p) \\ &= \frac{1}{t} \int_{[0,t)} \left(\prod_{j=1}^{p} F_j(x + s \cdot e_j) \right) ds \\ &= \frac{1}{t} \int_{[\sum_j x_j, t + \sum_j x_j)} \left(\prod_{j=1}^{p} F_j\left(x - e_j \cdot \left(-u + \sum_{j=1}^{p} x_j \right)\right) \right) du. \end{aligned}$$

(This is technically Fact 1.)

Definition 2.3 (Curly discrete averages)**.** *Given $\widetilde{F}_1, \dots, \widetilde{F}_p : \mathbb{Z}^p \to \mathbb{R}$ and $n \in \mathbb{N}$, define*

$$\widetilde{A}_n \left(\widetilde{F}_1, \dots, \widetilde{F}_p \right) (k_1, \dots, k_p) = \frac{1}{n} \sum_{i=0}^{n-1} \prod_{j=1}^{p} \widetilde{F}_j(k_1, \dots, k_{j-1}, k_j + i, k_{j+1}, \dots, k_p).$$

Definition 2.4. *Let $g : \mathbb{N} \to \mathbb{R}_{\geq 0}$ and $q_1, \dots, q_p$ positive real numbers (where p is some natural number). We call the tuple (g, q)* good *if $\sum_{i=1}^{p} \frac{1}{q_i} = \frac{1}{2}$, $1 \leq g(m)$ for all $1 \leq m$ and*[1] *if there is a finite constant C such that for any $m \in \mathbb{N}, \sigma$-finite measure space $(X, \mathcal{F}, \mu)$, any commuting measure-preserving $S_1, \dots, S_p : X \to X$ and all functions $[f_i] \in \mathrm{L}^{q_i}(\mathbb{R}^p)$ for $i \in \{1, \dots, p\}$, we have*

$$\sum_{j=1}^{m} \left\| A_{t_j}(f_1, \dots, f_p) - A_{t_{j-1}}(f_1, \dots, f_p) \right\|_{\mathrm{L}^2(\mathbb{R}^p)}^2 \leq C\, g(m) \prod_{i=1}^{p} \|f_i\|_{\mathrm{L}^{q_i}(\mathbb{R}^2)}^2 ,$$

for any positive real $t_0 < \cdots < t_m$.

For the rest of the thesis, we leave the q_i's implicit and call, by slight abuse of notation, a function $g : \mathbb{N} \to \mathbb{R}_{\geq 0}$ good.

The main theorem of this thesis, which will correspond to the `og_goal` Lean statement later, now is:

Theorem 2.5. *For any good g, there is a finite constant C such that for any $m \in \mathbb{N}$, measure space $(X, \mathcal{F}, \mu)$, any commuting measure-preserving $S_1, \dots, S_p : X \to X$ and all functions $[f_i] \in \mathrm{L}^{q_i}(X)$ for $i \in \{1, \dots, p\}$, we have*

$$\sum_{j=1}^{m} \left\| M_{n_j}(f_1, \dots, f_p) - M_{n_{j-1}}(f_1, \dots, f_p) \right\|_{\mathrm{L}^2(X)}^2 \leq C\, g(m) \prod_{i=1}^{p} \|f_i\|_{\mathrm{L}^{q_i}(X)}^2$$

for all positive integers $n_0 < \cdots < n_m$.

Which will be proved via the intermediate result:

Theorem 2.6 (Discrete version of assumption)**.** *For any good g, there exists a finite constant C such that for any $[\widetilde{F}_i] \in \ell^{q_i}(\mathbb{Z}^p)$ for $i \in \{1, \dots, p\}$, we have*

$$\sum_{j=1}^{m} \left\| \widetilde{A}_{n_j}(\widetilde{F}_1, \dots, \widetilde{F}_p) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_1, \dots, \widetilde{F}_p) \right\|_{\ell^2(\mathbb{Z}^p)}^2 \leq C\, g(m) \prod_{i=1}^{p} \left\| \widetilde{F}_i \right\|_{\ell^{q_i}(\mathbb{Z}^p)}^2$$

for all positive integers $n_0 < \cdots < n_m$

2.1 Proof of Theorem 2.6

Proof. Let $g : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be good. It suffices to show the theorem when $\left\| \widetilde{F}_i \right\|_{\ell^{q_i}(\mathbb{Z}^p)} = 1$ for all i as (directly from definition) for all $a \in \mathbb{R}$:

$$\sum_{j=1}^{m} \left\| \widetilde{A}_{n_j}(a \cdot \widetilde{F}_1, \dots, \widetilde{F}_p) - \widetilde{A}_{n_{j-1}}(a \cdot \widetilde{F}_1, \dots, \widetilde{F}_p) \right\|_{L^2(\mathbb{R}^p)}^2$$
$$= |a|^2 \sum_{j=1}^{m} \left\| \widetilde{A}_{n_j}(\widetilde{F}_1, \dots, \widetilde{F}_p) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_1, \dots, \widetilde{F}_p) \right\|_{L^2(\mathbb{R}^p)}^2$$

[1] More generally, g bounded below by a nonzero number on $1 \leq m$ is enough for Theorem 2.5 to work.

and

$$\left\|\widetilde{F}_i\right\|_{\ell^{q_i}(\mathbb{Z}^p)} = 0 \Rightarrow \widetilde{F}_i = 0$$
$$\Rightarrow \widetilde{A}_n(\widetilde{F}_1, \dots, \widetilde{F}_p) = 0$$
$$\Rightarrow \left\|\widetilde{A}_{n_j}(\widetilde{F}_1, \dots, \widetilde{F}_p) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_1, \dots, \widetilde{F}_p)\right\|^2_{L^2(\mathbb{R}^p)} = 0$$

.

Introduce $m \in \mathbb{N}$ and integers $0 < n_0 < \cdots < n_m$.
Fact 2. *There exists a constant $C \in \mathbb{R}$, such that for all normalised $[F_i] \in \mathrm{L}^{q_i}(\mathbb{R}^p)$ for $i \in \{1, \dots, p\}$ with norms 1, such that*

$$\sum_{j=1}^{m} \left\|A_{n_j}(F_1, \dots, F_p) - A_{n_{j-1}}(F_1, \dots, F_p)\right\|^2_{\mathrm{L}^2(\mathbb{R}^p)} \leq C\, g(m).$$

Proof. Follows from g being good applied to $F_1, \dots, F_p$ and $t_j = n_j$. □

Fact 3.

$$\widetilde{A}_n(\widetilde{F}_1, \dots, \widetilde{F}_p)(k_1, \dots, k_p) = \frac{1}{n} \sum_{\substack{k_1 + \cdots + k_p \leq i \wedge \\ i \leq k_1 + \cdots + k_p - 1}} \widetilde{F}_1\left(k_1 + i - \sum_{i=1}^{p} k_i, \dots, k_p\right) \dots \widetilde{F}_p\left(k_1, \dots, k_p + i - \sum_{i=1}^{p} k_i\right)$$

Proof. Trivial. □

Introduce normalised $[\widetilde{F}_i] \in \ell^{q_i}(\mathbb{Z}^p)$ for $i \in \{1, \dots, p\}$. Define $F_i : \mathbb{R}^p \to \mathbb{R}$ as

$$F_i(x_1, \dots, x_n) := \sum_{j_1, \dots, j_n \in \mathbb{Z}} \widetilde{F}_i\left(j + e_i \cdot \left(j_i - \sum_{r=1}^{p} j_p\right)\right) \chi_{[j_1, j_1+1)}(x_1) \dots \chi_{[j_i, j_i+1)}\left(\sum_{r=1}^{p} x_r\right) \dots \chi_{[j_p, j_p+1)}(x_p).$$

Now let $\alpha_1, \dots, \alpha_p \in [0, 1)$.
Fact 4.

$$A_n(F_1, \dots, F_p)(k + \alpha) = \frac{1}{n} \sum_{i \in \mathbb{Z}} a_i \prod_{j=1}^{p} \widetilde{F}_j\left(k_1, \dots, k_{j-1}, i + k_j - \sum_{r=1}^{p} k_r, \dots, k_p\right)$$

where

$$a_i = \left|[i, i+1) \cap \left[\sum_{r=1}^{p} k_r + \sum_{r=1}^{p} \alpha_r, \sum_{r=1}^{p} k_r + \sum_{r=1}^{p} \alpha_r + n\right)\right|$$

(Lebesgue 1 measure to be exact)

Proof. First observe $\|F_i\|_{\mathrm{L}^{q_i}(\mathbb{R}^p)} = 1$ for all i. It suffices to show $\|F_i\|^{q_i}_{\mathrm{L}^{q_i}(\mathbb{R}^p)} = 1$, as $\|F\|_{L^{q_i}(\mathbb{R}^p)} \in \mathbb{R}_{\geq 0}$. Note that

$$[i, i+1) \cap [k, k+1) = \varnothing$$

for $i, k \in \mathbb{Z}$ and $i \neq k$. Therefore (by induction over s),

$$F_i(x_1, \dots, x_p)^s = \sum_{j_1,\dots,j_p \in \mathbb{Z}} \widetilde{F}_i \left(j + e_i \cdot \left(j_i - \sum_{r=1}^{p} j_p \right) \right)^s \chi_{[j_1, j_1+1)}(x_1) \dots \chi_{[j_i, j_i+1)} \left(\sum_{r=1}^{p} x_r \right) \dots \chi_{[j_1, j_1+1)}(x_p)$$

for $s \in \mathbb{N}_{>0}$. Likewise we have

$$\begin{aligned}
\|F_i\|^{q_i}_{L^{q_i}(\mathbb{R}^p)} &= \int_{\mathbb{R}^p} |F_i|^{q_i} \, d\mathcal{L}^p \\
&= \int_{\mathbb{R}^p} \left(\sum_{j_1 \dots, j_p \in \mathbb{Z}} \widetilde{F}_i^{q_i}(\dots)\chi \dots \right) d\mathcal{L}^p \quad \text{by above observation} \\
&= \sum_{a_1,\dots,a_p \in \mathbb{Z}} \left(\int_{[a_1, a_1+1) \times \dots \times [a_p, a_p+1)} \left(\sum_{j_1,\dots,j_p \in \mathbb{Z}} \widetilde{F}_i^{q_i}(\dots)\chi \dots \right) d\mathcal{L}^p \right) \\
&= \sum_{a_1,\dots,a_p \in \mathbb{Z}} \left(\int_{[a_1, a_1+1) \times \dots \times [a_p, a_p+1)} \widetilde{F}_i(a_1, a_2, \dots, 2a_i - \sum_{r=1}^{p} a_r, \dots, a_p)^{q_i} \, d\mathcal{L}^p \right) \\
&= \sum_{a_1,\dots,a_p \in \mathbb{Z}} \widetilde{F}_i(a_1, a_2, \dots, 2a_i - \sum_{r=1}^{p} a_r, \dots, a_p)^{q_i} \;\; \text{(constant over measure 1)} \\
&= \left\| \widetilde{F}_i \right\|^{q_i}_{\ell^{q_i}(\mathbb{Z}^p)} \\
&= 1.
\end{aligned}$$

So with $I := [\sum_r k_r + \sum_r \alpha_r, \sum_r k_r + \sum_r \alpha_r + n)$:

$$
\begin{aligned}
&A_n(F_1,\dots,F_p)(k_1+\alpha_1,\dots,k_p+\alpha_p)\\
&\overset{\text{Fact 1}}{=} \frac{1}{n}\int_I \left(\prod_{j=1}^{p} F_j(k+\alpha-e_j\cdot\left(-u+\sum_{r=1}^{p}k_r+\alpha_r\right)\right)\,du\\
&=\frac{1}{n}\int_I \left(\prod_{i=1}^{p}\sum_{j\in\mathbb{Z}^p}\widetilde{F}_i\left(j+e_i\cdot\left(j_i-\sum_{r=1}^{p}j_r\right)\right)\chi\dots\right)\,du\\
&=\frac{1}{n}\int_I \left(\prod_{i=1}^{p}\sum_{a\in\mathbb{Z}}\widetilde{F}_i\left(k+e_i\cdot\left(a-\sum_{r=1}^{p}k_r\right)\right)\chi_{[a,a+1)}(u)\right)\,du\\
&=\frac{1}{n}\int_I \left(\sum_{a\in\mathbb{Z}}\prod_{i=1}^{p}\widetilde{F}_i\left(k+e_i\cdot\left(a-\sum_{r=1}^{p}k_r\right)\right)\chi_{[a,a+1)}(u)\right)\,du\\
&=\frac{1}{n}\sum_{a\in\mathbb{Z}}\int_I \left(\prod_{i=1}^{p}\widetilde{F}_i\left(k+e_i\cdot\left(a-\sum_{r=1}^{p}k_r\right)\right)\chi_{[a,a+1)}(u)\right)\,du\\
&=\frac{1}{n}\sum_{i\in\mathbb{Z}}a_i\prod_{j=1}^{p}\widetilde{F}_j\left(k_1,k_2\dots,k_{j-1},i+k_j-\sum_{r=1}^{p}k_r,\dots,k_p\right)
\end{aligned}
$$

Where the last few steps can be argued for as follows:
As $0 \leq \alpha_i < 1$, necessarily $a_j = k_j$ for all $j \neq i$ as for all other choices, the characteristic maps will vanish.
Similarly, $a := a_i$ is constant as (if we were to multiply out) one of the two characteristic maps of a summand would be zero. Instead of switching integral and sum, we can also split the area we are integrating over in intervals of $[a, a+1)$. This avoids potential convergence issues.

□

Combining Fact 4 and Fact 3, with $r_i := \prod_{j=1}^{p} \widetilde{F}_j\left(k_1, \ldots, i + k_j - \sum_{r=1}^{p} k_r, \ldots, k_p\right)$ and $u := \sum_{j=1}^{p} k_j$:

$$\begin{aligned}
&|A_n(F_0, \ldots, F_p)(k_1 + \alpha_1, \ldots, k_p + \alpha_p) - \widetilde{A}_n(\widetilde{F}_1, \ldots, \widetilde{F}_p)(k_1, \ldots, k_p)| \\
&\quad = \left| \left(\frac{1}{n} \sum_{i \in \mathbb{Z}} a_i \, r_i \right) - \left(\frac{1}{n} \sum_{u \leq i \leq u+n-1} r_i \right) \right| \\
&\quad = \frac{1}{n} \left| \left(\sum_{\substack{i \leq u-1 \, \vee \\ i \geq u+n+p}} a_i \, r_i \right) + \left(\sum_{\substack{u+p \leq i \, \wedge \\ i \leq u+n-1}} a_i \, r_i \right) - \left(\sum_{\substack{u \leq i \, \wedge \\ i \leq u+n-1}} r_i \right) \right. \\
&\qquad \left. + \sum_{\substack{u \leq i < u+p \, \vee \\ u+n \leq i < u+n+1}} a_i r_i \right| \\
&\quad = \frac{1}{n} \left| \left(\sum_{\substack{i \leq u-1 \, \vee \\ i \geq u+n+p}} 0 \right) + \left(\sum_{\substack{u+p \leq i \, \wedge \\ i \leq u+n-1}} r_i \right) - \left(\sum_{\substack{u \leq i \, \wedge \\ i \leq u+n-1}} r_i \right) \right. \\
&\qquad \left. + \sum_{\substack{u \leq i < u+p \, \vee \\ u+n \leq i < u+n+1}} a_i r_i \right| \\
&\quad = \frac{1}{n} \left| \sum_{\substack{u \leq i < u+p \, \vee \\ u+n \leq i < u+n+1}} a_i r_i \right| \\
&\quad \leq \frac{1}{n} \sum_{\substack{u \leq i < u+p \, \vee \\ u+n \leq i < u+n+1}} |a_i r_i| \\
&\quad \leq \frac{1}{n} \sum_{\substack{u \leq i < u+p \, \vee \\ u+n \leq i < u+n+1}} |r_i| \\
&\quad = \frac{1}{n} \sum_{\substack{u \leq i < u+p \, \vee \\ u+n \leq i < u+n+1}} \left| \prod_{j=1}^{p} \widetilde{F}_j(k + e_j \cdot i) \right|
\end{aligned}$$

Therefore we get regarding the ℓ^2-norm (note it's monotone):

$$
\begin{aligned}
&\left\| A_n(F_1,\ldots,F_p)(k+\alpha) - \widetilde{A}_n(\widetilde{F}_1,\ldots,\widetilde{F}_p) \right\|^2_{\ell^2_k(\mathbb{Z}^p)} \\
&\leq \left\| \frac{1}{n} \sum_{\substack{u\leq i<u+p \,\vee \\ u+n\leq i<u+n+1}} \left| \prod_{j=1}^p \widetilde{F}_j(k+e_j\cdot i) \right| \right\|^2_{\ell^2_k(\mathbb{Z}^p)} \\
&= \frac{1}{n^2} \left\| \sum_{\substack{u\leq i<u+p \,\vee \\ u+n\leq i<u+n+1}} \left| \prod_{j=1}^p \widetilde{F}_j(k+e_j\cdot i) \right| \right\|^2_{\ell^2_k(\mathbb{Z}^p)} \\
&= \frac{1}{n^2} \sum_{k\in\mathbb{Z}^p} \left(\sum_{\substack{u\leq i<u+p \,\vee \\ u+n\leq i<u+n+1}} \left| \prod_{j=1}^p \widetilde{F}_j(k+e_j\cdot i) \right| \right)^2 \\
&\leq \frac{4}{n^2} \sum_{k\in\mathbb{Z}^p} \sum_{\substack{u\leq i<u+p \,\vee \\ u+n\leq i<u+n+1}} \left(\prod_{j=1}^p \widetilde{F}_j(k+e_j\cdot i) \right)^2 \\
&\leq \frac{4}{n^2} \sum_{k\in\mathbb{Z}^p} \sum_{\substack{u\leq i<u+p \,\vee \\ u+n\leq i<u+n+1}} \sum_{j=1}^p \frac{2}{q_i} \cdot \widetilde{F}_j(k+e_j\cdot i)^{q_i} \\
&\leq \frac{8}{n^2} \cdot \max\{q_i^{-1}\} \sum_{k\in\mathbb{Z}^p} \sum_{\substack{u\leq i<u+p \,\vee \\ u+n\leq i<u+n+1}} \sum_{j=1}^p \widetilde{F}_j(k+e_j\cdot i)^{q_i} \\
&\leq \frac{16p}{n^2} \cdot \max\{q_i^{-1}\} \sum_{j=1}^p \left\| \widetilde{F}_j \right\|^{q_i}_{\ell^{q_i}(\mathbb{Z}^p)} \\
&= \frac{16p}{n^2} \cdot \max\{q_i^{-1}\} \sum_{j=1}^p 1 \\
&= \frac{16p^2}{n^2} \cdot \max\{q_i^{-1}\} \\
&\leq \frac{8p^2}{n^2}
\end{aligned}
$$

Implying

$$
\left\| A_n(f_1,\ldots,f_p)(k+\alpha) - \widetilde{A}_n(\widetilde{F}_1,\ldots,\widetilde{F}_p) \right\|_{\ell^2_{(\sum_j k_j)}(\mathbb{Z}^p)} \leq \frac{K}{n}
$$

where $K := \sqrt{8}\cdot p$. Note that for any normed space

$$
|\,\|a-b\| - \|c-d\|\,| \leq \|a-c-b+d\| \leq \|a-c\| + \|b-d\|\,.
$$

Using this fact with $a := A_{n_j}(F_1,\ldots,F_p)(k+\alpha), b := A_{n_{j-1}}(F_1,\ldots,F_p)(k+\alpha), c := \widetilde{A}_{n_j}(\widetilde{F}_1,\ldots,\widetilde{F}_p), d := \widetilde{A}_{n_{j-1}}(\widetilde{F}_1,\ldots,\widetilde{F}_p)$ (where a and b are functions in k,l) with the $\ell^2(\mathbb{Z}^p)$

norm, $u_j(x) := A_{n_j}(F_1,\dots,F_p)(x_1,\dots,x_p) - A_{n_{j-1}}(F_1,\dots,F_p)(x_1,\dots,x_p)$ and the above inequality this gets us:

$$\left| \|u_j(k+\alpha)\|_{\ell^2_k(\mathbb{Z}^p)} - \left\|\widetilde{A}_{n_j}(\widetilde{F}_1,\dots,\widetilde{F}_p) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_1,\dots,\widetilde{F}_p)\right\|_{\ell^2(\mathbb{Z}^p)} \right| \leq \frac{K}{n_j} + \frac{K}{n_{j-1}} \leq \frac{2K}{n_{j-1}}$$

Note that if $f : \mathbb{R}^p \to \mathbb{R}$ is a $\mathrm{L}^2(\mathbb{R}^p)$ function, then:

$$\begin{aligned}
\left\| \|f(x_1,\dots,x_p)\|_{\ell^2(\mathbb{Z}^p)} \right\|_{\mathrm{L}^2_{[0,1)^p}(x_1,\dots,x_p)} &= \left(\int_{[0,1)^p} \sum_{k\in\mathbb{Z}^p} f(k+x)^2 \, d\mathcal{L}^p(x) \right)^{\frac{1}{2}} \\
&= \left(\sum_{k,l,r\in\mathbb{Z}} \int_{[0,1)^p} f(k+x)^2 \, d\mathcal{L}^p(x) \right)^{\frac{1}{2}} \\
&= \left(\int_{\mathbb{R}^p} f \, d\mathcal{L}^2 \right)^{\frac{1}{2}} \\
&= \|f\|_{\mathrm{L}^2(\mathbb{R}^p)}
\end{aligned}$$

In particular, if f is constant on $[0,1)^p$ cubes, taking the L^2 norm on top doesn't change the value.
Therefore applying the L^2-norm to above inequality with $\alpha_1,\dots,\alpha_p$ gives:

$$\begin{aligned}
\frac{2K}{n_j} &= \left\| \frac{2K}{n_j} \right\|_{\mathrm{L}^2([0,1)^p)} \\
&\geq \left\| \|u_j(k+\alpha)\|_{\ell^2(\mathbb{Z}^p)} - \left\|\widetilde{A}_{n_j}(\widetilde{F}_1,\dots,\widetilde{F}_p) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_1,\dots,\widetilde{F}_p)\right\|_{\ell^2(\mathbb{Z}^p)} \right\|_{\mathrm{L}^2([0,1)^p)} \\
&\geq \left| \left\| \|u_j(k+\alpha)\|_{\ell^2(\mathbb{Z}^p)} \right\|_{\mathrm{L}^2([0,1)^\mathrm{p})} - \left\| \left\|\widetilde{A}_{n_j}(\widetilde{F}_1,\dots,\widetilde{F}_p) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_1,\dots,\widetilde{F}_p)\right\|_{\ell^2(\mathbb{Z}^p)} \right\|_{\mathrm{L}^2([0,1)^p)} \right| \\
&= \left| \|u_j\|_{\mathrm{L}^2(\mathbb{R}^p)} - \left\|\widetilde{A}_{n_j}(\widetilde{F}_1,\dots,\widetilde{F}_p) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_1,\dots,\widetilde{F}_p)\right\|_{\ell^2(\mathbb{Z}^p)} \right|
\end{aligned}$$

Meaning:

$$\begin{aligned}
\frac{2K}{n_j} + \|u_j\|_{\mathrm{L}^2(\mathbb{R}^p)} &\geq \left| \|u_j\|_{\mathrm{L}^2(\mathbb{R}^p)} - \left\|\widetilde{A}_{n_j}(\widetilde{F}_1,\dots,\widetilde{F}_p) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_1,\dots,\widetilde{F}_p)\right\|_{\ell^2(\mathbb{Z}^p)} \right| + \|u_j\|_{\mathrm{L}^2(\mathbb{R}^p)} \\
&\geq \left\|\widetilde{A}_{n_j}(\widetilde{F}_1,\dots,\widetilde{F}_p) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_1,\dots,\widetilde{F}_p)\right\|_{\ell^2(\mathbb{Z}^p)}
\end{aligned}$$

Now let C be the constant from Fact 2. Then summing up gives:

$$\begin{aligned}\sum_{j=1}^{m}\left\|\widetilde{A}_{n_j}(\widetilde{F}_1,\dots,\widetilde{F}_p))-\widetilde{A}_{n_{j-1}}(\widetilde{F}_1,\dots,\widetilde{F}_p))\right\|^2_{\ell^2(\mathbb{Z}^p)} &\leq \sum_{j=1}^{m}\left(\frac{2K}{n_j}+\|u_j\|_{\mathrm{L}^2(\mathbb{R}^p)}\right)^2\\ &\leq 2\cdot\sum_{j=1}^{m}\left(\frac{4K^2}{n_j^2}+\|u_j\|^2_{\mathrm{L}^2(\mathbb{R}^p)}\right)\\ &\leq 8K^2\left(\sum_{n=1}^{\infty}\frac{1}{n^2}\right)+2\cdot\sum_{j=1}^{m}\|u_j\|^2_{\mathrm{L}^2(\mathbb{R}^p)}\\ &=\frac{4K^2\pi^2}{3}+2\cdot\sum_{j=1}^{m}\|u_j\|^2_{\mathrm{L}^2(\mathbb{R}^p)}\\ &\leq\frac{4K^2\pi^2}{3}+2C\,g(m)\\ &=\left(\frac{4K^2\pi^2}{3g(m)}+2C\right)g(m)\\ &\leq^2\left(\frac{4K^2\pi^2}{3}+2C\right)g(m)\end{aligned}$$

$\left(\frac{4K^2\pi^2}{3}+2C\right)$ is independent of the $\widetilde{F}_1,\dots,\widetilde{F}_p$ chosen, so we can choose it as the constant. (Taking the supremum over all variations gives the variation norm if $g(m)$ is constant.) □

2.2 Proof of Theorem 2.5 using Theorem 2.6

Proof. Let $g:\mathbb{N}\to\mathbb{R}_{\geq 0}$ be good. Once again it suffices to show the statement for normalised $[f_i]\in\mathrm{L}^{q_i}(X)$, as $M_n(c\cdot f_1,\dots,f_p)=c\cdot M_n(f_1,\dots,f_p)$ and if $\|f_i\|_{\mathrm{L}^{q_i}(X)}=0$ every integral over $M_n(f_1,\dots,f_p)$ will be zero almost everywhere, as f_i is zero almost everywhere, thus

$$\left\|M_{n_j}(f_1,\dots,f_p)-M_{n_{j-1}}(f_1,\dots,f_p)\right\|_{\mathrm{L}^2(X)}=0.$$

Due to $M_{n_j}(f_1,\dots,f_p)-M_{n_{j-1}}(f_1,\dots,f_p)$ being zero almost everywhere as well.

So introduce $[f_i]\in\mathrm{L}^{q_i}(X)$ for all $i\in\{1,\dots,p\}$ and $x\in X$ (fixed for now) and a positive integer $N\geq n_m$.

We define the following functions:

$$\widetilde{F}_{i,x,N}(k_1,\dots,k_p):=\begin{cases}f_i(S_1^{k_1}\dots S_p^{k_p}x) & \text{if } 0\leq k_1,\dots,k_p\leq 2N-1\\ 0 & \text{else}\end{cases}$$

Fact 5.

$$M_n(f_1,\dots,f_p)(S_1^{k_1}\dots S_p^{k_p}x)=\widetilde{A}_n(\widetilde{F}_{1,x,N},\dots,\widetilde{F}_{p,x,N})(k_1,\dots,k_p)$$

Proof.

$$\begin{aligned}M_n(f_1,\dots,f_p)(S_1^{k_1}\dots S_p^{k_p}x) &= \frac{1}{n}\sum_{i=0}^{n-1}\prod_{j=1}^{p} f_j(S_j^i(S_1^{k_1}\dots S_p^{k_p}x))\\ &= \frac{1}{n}\sum_{i=0}^{n-1}\prod_{j=1}^{p} f_j(S_1^{k_1}\dots S_j^{k_j+i}\dots S_p^{k_p}x)\\ &= \frac{1}{n}\sum_{i=0}^{n-1}\prod_{j=1}^{p} \widetilde{F}_{j,x,N}(k_1,\dots,k_j+i,\dots,k_p)\\ &= \widetilde{A}_n(\widetilde{F}_{1,x,N},\dots,\widetilde{F}_{p,x,N})(k_1,\dots,k_p)\end{aligned}$$

□

Fact 6.

$$\begin{aligned}&\left\|M_{n_j}(f_1,\dots,f_p)-M_{n_{j-1}}(f_1,\dots,f_p)\right\|_{\mathrm{L}^2(X)}^2\\ &\leq \frac{1}{N^p}\int_X \left\|\widetilde{A}_{n_j}(\widetilde{F}_{1,x,N},\dots,\widetilde{F}_{p,x,N})-\widetilde{A}_{n_{j-1}}(\widetilde{F}_{1,x,N},\dots,\widetilde{F}_{p,x,N})\right\|_{\ell^2(\mathbb{Z}^p)}^2 dx\end{aligned}$$

Proof.

$$\begin{aligned}&\left\|M_{n_j}(f_1,\dots,f_p)-M_{n_{j-1}}(f_1,\dots,f_p)\right\|_{\mathrm{L}^2(X)}^2\\ &= \int_X \left|M_{n_j}(f_1,\dots,f_p)(x)-M_{n_{j-1}}(f_1,\dots,f_p)(x)\right|^2 d\mu(x)\\ &= \frac{1}{N^p}\sum_{k_1,\dots,k_p=0}^{N-1}\int_X \left|M_{n_j}(f_1,\dots,f_p)(x)-M_{n_{j-1}}(f_1,\dots,f_p)(x)\right|^2 d\mu(x)\\ &= \frac{1}{N^p}\sum_{k_1,\dots,k_p=0}^{N-1}\int_X \left|M_{n_j}(f_1,\dots,f_p)(S_1^{k_1}\dots S_p^{k_p}x)-M_{n_{j-1}}(f_1,\dots,f_p)(S_1^{k_1}\dots S_p^{k_p}x)\right|^2 d\mu(x)\\ &= \frac{1}{N^p}\int_X\sum_{k_1,\dots,k_p=0}^{N-1} \left|M_{n_j}(f_1,\dots,f_p)(S_1^{k_1}\dots S_p^{k_p}x)-M_{n_{j-1}}(f_1,\dots,f_p)(S_1^{k_1}\dots S_p^{k_p}x)\right|^2 d\mu(x)\\ &= \frac{1}{N^p}\int_X\sum_{k_1,\dots,k_p=0}^{N-1} \left|\widetilde{A}_{n_j}(\widetilde{F}_{1,x,N},\dots,\widetilde{F}_{p,x,N})-\widetilde{A}_{n_{j-1}}(\widetilde{F}_{1,x,N},\dots,\widetilde{F}_{p,x,N})\right|^2 d\mu(x)\quad\text{(Fact 5)}\\ &\leq \frac{1}{N^p}\int_X \left\|\widetilde{A}_{n_j}(\widetilde{F}_{1,x,N},\dots,\widetilde{F}_{p,x,N})-\widetilde{A}_{n_{j-1}}(\widetilde{F}_{1,x,N},\dots,\widetilde{F}_{p,x,N})\right\|_{\ell^2(\mathbb{Z}^p)}^2 d\mu(x)\end{aligned}$$

Where the third equality follows from the fact of $S_1^{k_1}\dots S_p^{k_p}$ being measure preserving, thus $\mu(X)-\mu((S_1^{k_1}\dots S_p^{k_p})^{-1}(X))=0$ and the theorem about integrals and measure preserving transformations.

□

Fact 7. *For* $i\in\{1,\dots,p\}$, *we have*

$$\int_X \left\|\widetilde{F}_{i,x,N}\right\|_{\ell^{q_i}(\mathbb{Z}^p)}^{q_i} d\mu(x) = 2^pN^p.$$

Proof.

$$\begin{aligned}
\int_X \left\| \widetilde{F}_{i,x,N} \right\|_{\ell^{q_i}(\mathbb{Z}^p)}^{q_i} d\mu(x) &= \int_X \sum_{k_1,\ldots,k_p=0}^{2N-1} |f_i(S_1^{k_1} \ldots S_p^{k_p} x)|^{q_i} \, d\mu(x) \\
&= \sum_{k_1,\ldots,k_p=0}^{2N-1} \int_X |f_i(S_1^{k_1} \ldots S_p^{k_p} x)|^{q_i} \, d\mu(x) \\
&= \sum_{k_1,\ldots,k_p=0}^{2N-1} \int_X |f_i(x)|^{q_i} \, d\mu(x) \\
&= 2^p N^p \cdot \|f_i\|_{\mathrm{L}^{q_i}(X)}^{q_i} \\
&= 2^p N^p
\end{aligned}$$

Where we again used the theorem about integrals and measure preserving transformations in the second equality.

□

Fact 8. We have

$$\begin{aligned}
&\sum_{j=1}^{m} \left\| \widetilde{A}_{n_j}(\widetilde{F}_{1,x,N}, \ldots, \widetilde{F}_{p,x,N}) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_{1,x,N}, \ldots, \widetilde{F}_{p,x,N}) \right\|_{\ell^2(\mathbb{Z}^p)}^2 \\
&\qquad \leq C\, g(m) \cdot \sum_{i=1}^{p} \left\| \widetilde{F}_{i,x,N} \right\|_{\ell^{q_i}(\mathbb{Z}^p)}^{q_i}
\end{aligned}$$

for a constant C not depending on $f_1, \ldots, f_p$.

Proof. Applying Theorem 2.6 to $\widetilde{F}_i := \widetilde{F}_{i,x,N}$ gives us

$$\begin{aligned}
&\sum_{j=1}^{m} \left\| \widetilde{A}_{n_j}(\widetilde{F}_{1,x,N}, \ldots, \widetilde{F}_{p,x,N}) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_{1,x,N}, \ldots, \widetilde{F}_{p,x,N}) \right\|_{\ell^2(\mathbb{Z}^p)}^2 \\
&\leq C \cdot \prod_{i=1}^{p} \left\| \widetilde{F}_{i,x,N} \right\|_{\ell^{q_i}(\mathbb{Z}^p)}^2 \\
&\leq C \sum_{i=1}^{p} \frac{2}{q_i} \left\| \widetilde{F}_{i,x,N} \right\|_{\ell^{q_i}(\mathbb{Z}^p)}^{q_i} \\
&\leq 2C \cdot \max\{q_i^{-1}\} \sum_{i=1}^{p} \left\| \widetilde{F}_{i,x,N} \right\|_{\ell^{q_i}(\mathbb{Z}^p)}^{q_i} \\
&\leq C \sum_{i=1}^{p} \left\| \widetilde{F}_{i,x,N} \right\|_{\ell^{q_i}(\mathbb{Z}^p)}^{q_i}
\end{aligned}$$

Where C is the constant from Theorem 2.6. □

Now we are ready to finish off:

$$\sum_{j=1}^{m} \left\|M_{n_j}(f_1,\dots,f_p) - M_{n_{j-1}}(f_1,\dots,f_p)\right\|_{\mathrm{L}^2(X)}^2$$

$$\overset{\text{Fact 6}}{\leq} \sum_{i=1}^{m} \frac{1}{N^p} \int_X \left\|\widetilde{A}_{n_j}(\widetilde{F}_{1,x,N},\dots,\widetilde{F}_{p,x,N}) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_{1,x,N},\dots,\widetilde{F}_{p,x,N})\right\|_{\ell^2(\mathbb{Z}^p)}^2 d\mu(x)$$

$$= \frac{1}{N^p} \int_X \sum_{i=1}^{m} \left\|\widetilde{A}_{n_j}(\widetilde{F}_{1,x,N},\dots,\widetilde{F}_{p,x,N}) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_{1,x,N},\dots,\widetilde{F}_{p,x,N})\right\|_{\ell^2(\mathbb{Z}^p)}^2 d\mu(x)$$

$$\overset{\text{Fact 8}}{\leq} \frac{1}{N^p} \int_X C\, g(m) \left(\sum_{i=1}^{p} \left\|\widetilde{F}_{i,x,N}\right\|_{\ell^2(\mathbb{Z}^p)}^{q_i} \right) d\mu(x)$$

$$= \frac{C\, g(m)}{N^p} \cdot \sum_{i=1}^{p} \int_X \left\|\widetilde{F}_{i,x,N}\right\|_{\ell^{q_i}(\mathbb{Z}^p)}^{q_i} d\mu(x)$$

$$\overset{\text{Fact 7}}{=} \frac{C\, g(m)}{N^p} \sum_{i=1}^{p} 2^p N^p$$

$$= C \cdot g(m) \cdot p 2^p$$

Where $C \cdot g(m) \cdot p2^p$ does not depend on $f_1, \dots, f_p$.

□

2.3 Formalisation

Our main results, Definition 2.4, Theorem 2.6 and Theorem 2.5, we formalised as

```
def good (g : ℕ → NNReal) (q : ι → ℝ) : Prop :=
  (∀ (i : ℕ), 1 ≤ i → 1 ≤ g i) ∧
  ∃ C : NNReal, ∀ (m : ℕ) (idx : Fin (m + 1) → NNReal)
  (mon : Monotone idx) (hidx : Injective idx) (hidx' : 0 < idx 0)
  (F : ι → (ι → ℝ) → ℝ) (hF : ∀ (i : ι), MeasureTheory.MemLp (F i) (q i).
    toNNReal),
  ∑ i : Fin m, (∫⁻ (a : ι → ℝ), ‖(multilinear_avg_spec F (idx i.succ) a
  - multilinear_avg_spec F (idx (Fin.castSucc i)) a)‖ₑ ^ 2)
  ≤ C * g m * Π (i : ι), (∫⁻ (a : ι → ℝ), ‖(F i) a‖ₑ ^ (↑(q i) : ℝ)) ^ (1 / (↑(q
    i) : ℝ))

theorem discrete_ver {g : ℕ → NNReal} {q : ι → ℝ} (hg: good g q)
    (hq : (∀(i : ι), 0 < q i) ∧ ∑ (i : ι), 1 / (q i) = 1 / 2) [unι: Nontrivial ι]
    (m : ℕ) (idx : Fin (m + 1) → ℕ) (mon : Monotone idx)
    (hidx : Injective idx) (hidx': 0 < idx 0)
    (F : ι → (ι → ℕ) → ℝ) (hF: ∀ (j : ι), Memℓp (F j) (q j).toNNReal):
    ∑ i : Fin m, ∑' a : ι → ℕ,
    (discrete_avg (idx i.succ) F a - discrete_avg (idx (Fin.castSucc i)) F a)^2
     ≤ (((256 * (Fintype.card ι) ^ 3 * π ^ 2) / 3) + 2 * (good_const hg))
        * g m * Π (j : ι), (∑' a : ι → ℕ, |(F j a)| ^ (q j)) ^ (2 / (q j)) :=
```

```
theorem og_goal
    {ι : Type*} [Fintype ι] [Nonempty ι]
    {g : ℕ → NNReal} {q : ι → ℝ} (hg : good g q)
    (hq : (∀ (i : ι), 0 < q i) ∧ ∑ (i : ι), 1 / (q i) = 1 / 2) :
    ∃ C : NNReal, ∀ {X : Type*} [MeasurableSpace X] {μ : Measure X}
    {f : ι → X → ℝ} {S : ι → X → X}
    (m : ℕ) (idx : Fin (m + 1) → ℕ) (_ : Monotone idx) (_ : Injective idx) (_ :
    0 < idx 0)
    (_ : MMeasurePreserving μ S) (_ : pairwise_commuting S) (_ : ∀ (i : ι),
    Measurable (f i)),
    ∑ i : Fin m, ∫⁻ (x : X),
    ‖(ergodic_avg (idx i.succ) S f x - ergodic_avg (idx (Fin.castSucc i)) S f x)‖ₑ
     ^ 2 ∂μ ≤
    (↑C : ENNReal) * (↑(g m) : ENNReal)
      * Π(i : ι), (∫⁻ (x : X), ‖f i x‖ₑ ^ (q i) ∂μ) ^ (2 / (q i)) := by
  use (↑(Fintype.card ι) : NNReal) *
    2 ^ (Fintype.card ι) * (((256 *
    (↑(Fintype.card ι) : ℝ) ^ 3 * π ^ 2) / 3).toNNReal + 2 * (good_const hg))
  intro X _ μ f S m idx mon hidx hidx' hS hS' hf
  exact goal f hg μ hq m idx mon hidx hidx' hS hS' hf
```

Here, `ergodic_avg` refers to M_n, `multilinear_avg_spec` to A_n, `discrete_avg` to $\widetilde{A}_n$, `idx` to the n_j and `good_const hg` as the constant from Definition 2.4. Meaning we prove the formalisation of Theorem 2.5 with the constant $\frac{256p^4 2^p}{3}C$ where C is the constant from g good. One can improve this constant with both a more careful formalisation and refining the proof itself, though the factor 2^p seems hard to avoid.

As stated in section 1.3, we extensively used the measure theory portion of mathlib in our formalisation. Additionally, for some measurability arguments, we also used results from mathlib's library about topology and for $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$, the corresponding identity in mathlib we used is located in the number theory section.

In its current version, the proof is very long, about 7000 lines of code. Much of this comes from issues with types, as briefly described in section 1.3. Staying more consistent with types than we did, together with improved code quality, which currently still is quite poor, could therefore shorten the formalisation a lot.
Furthermore, most of the proof comes down to proving technical estimations about particular integrals and sums, which are typically not very interesting on their own.
Therefore we will mostly not elaborate on the formalised proof steps and instead refer to `https://www.github.com/felixpernegger/ErgodicAverages`.

However, we want to mention that for Fact 4, we used that the measure of some parallelotope (n-dimensional parallelogram) is 1, which is easy to see on paper. In Lean, we formalised this via an inductive approach and by using mathlib's marginal construction, a tool for iterated integration, which is described in [8].
A similar instance of a part of the formalisation which is "trivial on paper" but interesting in Lean, related to the construction of $\widetilde{F}_{i,x,N}$ in subsection $\widetilde{F}_{i,x,N}$, is the topic of chapter 5.

3 Finclosed spaces

As mentioned in section 1.2 and the beginning of chapter 2, the version of the Calderón transference principle we proved in this thesis as in chapter 2 is not strong enough to reduce the main result of [9], Theorem 1.4, to a result purely in harmonic analysis directly. As a workaround we can use the following definition and lemma.

Definition 3.1. *Let X be a set. We denote $Q(X)$ as the set of finite subsets of X.*
We say a subset $F \subseteq Q(X)$ is finclosed *if $\varnothing \in F$ and if there is a k, such that for any $U \in F$ and $V \subseteq U$, there is a $V' \in F$ with $V \subseteq V'$ and $|V'| \leq k \cdot |V|$.*
We denote such a number k as $d(F)$ and call the members of F closed *subsets of X. We call the tuple (X, F) with $F \subseteq Q(X)$ and F finclosed, a* finclosed space.

The following lemma is then central.

Lemma 3.2. *Let $b \in [0, 1)$, X be a set and an evaluating map $e : X \to \mathbb{R}_{\geq 0}$ and $F \subseteq Q(X)$ finclosed.*
Assume there is a constant C such that for all $U \in F$

$$\sum_{u \in U} e(u) \leq C \cdot |U|^b.$$

Then for all $r > \frac{1}{1-b}$ there is a constant C_r, such that for all $U \in F$:

$$\sum_{u \in U} e(u)^r \leq C_r$$

Proof. By abuse of notation, we ignore the evaluating function in the following and apply it only implicitly. Let $U \in F$. Define a bijection $f : \{1, \ldots, |U|\} \to U$. For $e \in \mathbb{R}$ we define

$$A_e := \{u \in U \mid u > e\}$$

From F being finclosed, we get an A'_t with $A'_t \in F, A_t \subseteq A'_t$ and $|A'_t| \leq d(F) \cdot |A_t|$. But then:

$$\begin{aligned} t \cdot |A_t| &\leq \sum_{u \in A_t} u \\ &\leq \sum_{u \in A'_t} u \\ &\leq C \cdot |A'_t|^b \\ &\leq C \cdot (d(F) \cdot |A_t|)^b \end{aligned}$$

Implying

$$|A_t| \leq C^{\frac{1}{1-b}} d(F)^{\frac{b}{1-b}} \cdot t^{-\frac{1}{1-b}}$$

Note this implies the existence of a L such that $A_t = \varnothing$ for $t \geq L$.
Then we have with layer cake representation:

$$\begin{aligned}
\sum_{u \in U} u^r &= \sum_{i=1}^{|U|} f(i)^r \\
&= \int_{\{0,\ldots,|U|\}} f^r \, d\# \\
&= r \int_{(0,\infty)} s^{r-1} \#\{i \in \{0,\ldots,|U|\} \mid f(i) > s\} \, ds \\
&= r \int_{(0,\infty)} s^{r-1} |A_s| \, ds \\
&= r \left(\int_{(0,L]} s^{r-1} |A_s| \, ds + \int_{(L,\infty)} s^{r-1} |A_s| \, ds \right) \\
&= r \left(\int_{(0,L]} s^{r-1} |A_s| \, ds + \int_{(L,\infty)} s^{r-1} \cdot 0 \, ds \right) \\
&= r \int_{(0,L]} s^{r-1} |A_s| \, ds \\
&\leq r \int_{(0,L]} s^{r-1} \cdot C^{\frac{1}{1-b}} d(F)^{\frac{b}{1-b}} \cdot s^{-\frac{1}{1-b}} \, ds \\
&= (r \cdot C^{\frac{1}{1-b}} d(F)^{\frac{b}{1-b}}) \cdot \int_{(0,L]} s^{r-1-\frac{1}{1-b}} \, ds
\end{aligned}$$

Where the integral converges as from our assumption $r > \frac{1}{1-b}$.
Note that nothing in the last expression depends on U anymore, so we can take

$$C_r := (r \cdot C^{\frac{1}{1-b}} d(F)^{\frac{b}{1-b}}) \cdot \int_{(0,L]} s^{r-1-\frac{1}{1-b}} \, ds.$$

□

Remark 3.3. *The lemma does in general not hold for* $r = \frac{1}{1-b}$. *As a counterexample take* $X = \{\sqrt{i} - \sqrt{i-1} \mid i \in \mathbb{N}\}$, $F = Q(X)$ *and* $b = \frac{1}{2}$.

Now connect this lemma to norm variation.
Consider some totally ordered set J. Let $A \subseteq Q(J \times J)$ be such that $\{(a_{1,1}, a_{1,2}), \ldots, (a_{n,1}, a_{n,2})\} \in A$ iff, $a_{i,1} < a_{i,2}$ and there being a $\pi \in S^n$, such that $a_{\pi(i),2} = a_{\pi(i+1),1}$. One can think of A as finite sequences of tuples, such that the respective endpoints of adjacent tuples are the same.
Then $(J \times J, A)$ forms a finclosed space, as if we take some sequence $s \in A$ and remove some elements of s to get some subsequence $s' \subseteq s$, we can "fill up" each (inclusion maximal) hole

with one pair, giving us an $s' \subseteq s'' \in A$ with $|s''| \leq 2|s|$.

Recall from the introduction that Theorem 1.4 is proved as a deduction from (1.2). Now with $g(m) := m^{\frac{1}{2}}$, we can apply the transference principle 2.5 to (1.2) to get (1.3). Now via above construction with $J = \mathbb{N}_{>0}$, we can apply Lemma 3.2 with $b = \frac{1}{2}$ and

$$e((a,b)) := \left\| M_{n_j}(f_1, f_2, f_3) - M_{n_{j-1}}(f_1, f_2, f_3) \right\|^2_{\mathrm{L}^2(X)}$$

to (1.3), giving us exactly Theorem 1.4.
Theorem 1.6 can similarly be proven by applying Lemma 3.2 to (1.2) directly.

3.1 Formalisation

We formalise finclosed spaces in Lean as a structure

```
def IsFinclosed {X : Type u} (F : Set (Finset X)): Prop :=
  ∅ ∈ F ∧ ∃ k : ℝ, ∀(u : Finset X), (∃v ∈ F, u ⊆ v) →
  (∃u' ∈ F, u ⊆ u' ∧ (↑u' : Finset X).card ≤ k * u.card)

structure FinclosedSpace where
  X : Type u
  F : Set (Finset X)
  SpaceFinclosed : IsFinclosed F
```

An example of a finclosed space is to take some type and F as the set of all finite subsets.

```
instance univ (U : Type u): FinclosedSpace where
  X := U
  F := @Set.univ (Finset U)
  SpaceFinclosed := ...
```

Looking back, it potentially would have been more natural to avoid F in the definition of a finclosed space and instead use a predicate on the type Finset X, similar to the definition of a matroid in mathlib.
Now importantly, we formalise Lemma 3.2 as

```
def fun_closed {S : FinclosedSpace} (t : ℝ) (f : S.X → NNReal): Prop :=
  ∃ C : ℝ, ∀(u : Finset S.X), IsClosed u → ∑ i ∈ u, f i ≤ C * (u.card) ^ t

theorem layercake_overkill {S : FinclosedSpace} {b r : ℝ}
    (hb : b < 1) (hb2 : 0 ≤ b)
    {f : S.X → NNReal} (hf : fun_closed b f) (hr : 1 / (1 - b) < r):
    fun_closed 0 (f ^ r) := ..
```

Here, `IsClosed u`, means `u` closed with respect to the finclosed space `S`.
Once again, we omit the formalisation itself and refer to `https://github.com/felixpernegger/ErgodicAverages`.

Due to time constraints, we did not yet fully formalise the connection to norm-variation, which is not a trivial task in Lean. We have however mostly formalised the "filling up" gaps arguments, see for example the `FillUp` definition in the github file.

4 r-variation

After application of the Calderón transference principle with one function S and function f where g is constant we essentially end up with the following statement as seen in [12]

$$\sup_{0<n_0<n_1<\cdots<n_m} \sum_{j=1}^{m} \left\| M_{n_j} f - M_{n_{j-1}} f \right\|_2^2 \leq C \left\| f \right\|_2^2$$

For some C independent of f.

The main result of [9], where $r > 4$, can be similarly stated as

$$\sup_{0<n_0<n_1<\cdots<n_m} \sum_{j=1}^{m} \left\| M_{n_j}(F_1, F_2, F_3) - M_{n_{j-1}}(F_1, F_2, F_3) \right\|_2^r \leq C_r \left\| F_1 \right\|_8^r \left\| F_2 \right\|_8^r \left\| F_3 \right\|_4^r .$$

Both of these, and many other similar results, can be written more concisely by using the r-variation of a function.

Definition 4.1. *Let J be an ordered set, $I \subseteq J$, B a normed space*[1]*, $a : I \to B$, $r \in (0, \infty)$, then the* r-variation *of a is*

$$\|a\|_{V^r(I,B)} := \|a(t)\|_{V_t^r(I,B)} := \sup_{\substack{m \in \mathbb{N} \\ t_0, t_1, \ldots, t_m \in I \\ t_0 < t_1 < \cdots < t_m}} \left(\sum_{j=1}^{m} \|a(t_j) - a(t_{j-1})\|_B^r \right)^{1/r} .$$

In particular, for $r = 1$ this is the well-known variation of a function, which for $J = I = \mathbb{R}$ and $B = \mathbb{R}^n$ can geometrically be interpreted as the length of the curve f.
Total variation and the related notion of bounded variation have been implemented in mathlib, while r-variation as of writing of this paper has not.
Note that in the definition, instead of $t_0 < t_1 < \cdots < t_m$ we can also require $t_0 \leq t_1 < \cdots \leq t_m$.

A version of r-variation, often also called p-variation, with $r = 2$ has first been introduced in [16] in 1924 by Wiener and been expanded in [13] by Lépingle in his work about martingales. Due to its statistical properties, r-variation has been used in harmonic analysis, probability theory and related areas. More historical remarks can be found in [14].

[1] We will later see a slightly more general version, but for convenience focus on this case for now

For $1 \leq r$, we also refer to the r-variation as the *variation norm*, although it is technically only an extended seminorm, in particular for all $r \in (0, \infty)$ we have

$$\|a\|_{V^r(I,B)} = 0 \Leftrightarrow a \text{ is constant on } I$$

as can be easily seen from the definition. Adding a norm to the r-variation, most notably a supremum norm, alternatively also turns this into an extended norm.
Notably and maybe counter-intuitively, for $r > 1$, finer subdivisions do *not* give us better approximations of the r-variation in general.
Furthermore, note that for $I = \mathbb{N}$ finite r-variation and B Banach space implies convergence.

Using the variation norm, the examples above can now be written briefly as

$$\|M_n f\|_{V_n^2(\mathbb{N}-\{0\},X)} \leq C \|f\|_2$$

and

$$\|M_n(F_1, F_2, F_3)\|_{V_n^r(\mathbb{N}-\{0\},X)} \leq C_r \|F_1\|_8 \|F_2\|_8 \|F_3\|_4$$

Thus, in this section we first establish elementary results about r-variation and subsequently present a formalisation of some of those in Lean.

4.1 Properties of the r-variation

First, we want to think about r-variation in a slightly different manner.
Recall that for a function $g : \{1, \dots, n\} \to [0, \infty]$ and $p \in [0, \infty]$ the ℓ^p-norm (i.e. L^p norm with respect to counting measure) of g is defined as:

$$\|g\|_p := \begin{cases} |\operatorname{supp}(g)|, & p = 0 \\ \left(\sum_{k=1}^n g(k)^p\right)^{\frac{1}{p}}, & 0 < p < \infty, \\ \max\limits_{k \in \{1,\dots,m\}} g(k), & p = \infty. \end{cases}$$

With I, J and B as above, now for a monotone sequence $t : \{0, \dots, m\} \to I$ and $a : J \to B$, let

$$a_t : \{0, \dots, m-1\} \to B$$
$$k \mapsto \|a(t_{k+1}) - a(t_k)\|_B$$

Comparing definitions then shows that for $0 < r < \infty$

$$\|a\|_{V^r(I,B)} = \sup_{\substack{m \in \mathbb{N} \\ t_0, t_1, \dots, t_m \in I \\ t_0 < t_1 < \dots < t_m}} \|a_t\|_r$$

Therefore we can naturally extend the definition of r-variation to $r = \infty$, in which case we obtain

$$\|a\|_{V^\infty(I,B)} = \sup_{x,y \in I} \|a(x) - a(y)\|_B$$

Similarly, we can extend the definition to $r = 0$ in which case the variation norm is the number of fluctuations of a on I.[2]
Note that we still have, $\|a\|_{V^r(I,B)} = 0 \Leftrightarrow a$ is constant on I.

4.1.1 Change in r

Next we fix a, I, J and B and want to describe

$$S_a := \{r \in [0, \infty] : \|a\|_{V^r(I,B)} < \infty\}$$

First we show this is a right-sided interval:

Lemma 4.2. *Let $u \in S_a$ and $v > u$, then $v \in S_a$*

Proof. The cases $u = 0$ and $u = \infty$ are trivial. So let $u > 0$.
We distinguish by cases on v,
For $v < \infty$: Minkowski's inequality for sums implies

$$\left(\sum_{k=1}^{n} b_k^v\right)^{\frac{1}{v}} \leq \left(\sum_{k=1}^{n} b_k^u\right)^{\frac{1}{u}},$$

which implies the result.

For $v = \infty$: Assume $\infty \notin S_a$. Therefore for arbitrarily large $L > 0$ we find $x_L, y_L \in I$ with w.l.o.g $x_L < y_L$ and $L < \|a(x_L) - a(y_L)\|$. But then

$$L < \|a(x_L) - a(y_L)\| = (\|a(x_L) - a(y_L)\|^u)^{\frac{1}{u}} \leq \|a\|_{V^u(I,B)}$$

for arbitrarily large L, contradicting $u \in S_a$ Therefore S_a is a left-bounded interval. □

Remark 4.3. *Note that more generally, the proof also implies $\|a\|_{V^r(I,B)} \leq \|a\|_{V^{r'}(I,B)}$ for $0 < r' \leq r$.*

The interval can however both be open and closed, for the first case consider $a : \mathbb{N}_{>0} \to \mathbb{R}_{>0}$, $a(2n) = \frac{1}{n}, a(2n+1) = 0$.
Evidently for $0 < r < \infty$, $\|a\|^r_{V^r(\mathbb{N}_{>0},\mathbb{R}_{>0})} = \sum_{n=1}^{\infty} \frac{1}{n^r}$, implying $S_a = (1, \infty]$.

For the second case consider $a(2n) = \frac{1}{n \ln(n)^2}$ for $n \geq 4$, $a(0) = a(2) = 0$ and $a(2n+1) = 0$.
Again we have $\|a\|^r_{V^r(\mathbb{N},\mathbb{R}_{>0})} = \sum_{n=2}^{\infty} \frac{1}{(n \ln(n)^2)^r}$.
By Cauchy condensation test this converges for $r = 1$, but, by comparing it to $\frac{1}{n}$, diverges for $r < 1$, thus $S_a = [1, \infty]$

[2] Compared to the $r = \infty$ case the notion of a L^0-space/norm is a far less common and natural one. However this is the convention in mathlib, so we choose this to make our formalisation more natural. As we will later see, some properties hold everywhere but in $r = 0$.

By adding appropriate exponents to these examples, we find a such that $S_a = [r, \infty]$ and $S_a = (r, \infty]$ for all $0 < r < \infty$.
$S_a = \{\infty\} = [\infty, \infty]$ is achieved by $a(n) = (-1)^n$, $S_a = \varnothing$ by some unbounded sequence, i.e. $a(n) = n$.
$a \equiv 0$ satisfies $S_a = [0, \infty]$. Finally for $S_a = (0, \infty]$ consider $a(2n) = 2^{-n}, a(2n+1) = 0$. This has infinite 0-variation, as the sequence jumps infinitely often, but $\|a\|^r_{V^r(\mathbb{N},\mathbb{R}_{>0})} = \sum_{n=0}^{\infty} 2^{-nr}$ converges for all $0 < r < \infty$.

Putting these results together we get

Theorem 4.4. *Let $S \subseteq [0, \infty]$. There exist I, J, a as above with $S_a = S$ if and only if $S = (r, \infty]$ or $S = [r, \infty]$ for some $r \in [0, \infty]$.*

Next, we investigate the values the r-variation takes on S_a. Namely, for $a : J \to B$, let

$$\begin{aligned} e_a : [0, \infty] &\to [0, \infty] \\ r &\mapsto \|a\|_{V^r(I,B)} \end{aligned}$$

By Remark 4.3, e_a is monotonically decreasing on $(0, \infty]$, so one-sided limits exist.
From examples above, we know that e_a is not continuous in general (i.e. $S_a = [1, \infty]$).
It is also not continuous on S_a, consider $I = \{0, 1, 2\}$, $a(n) = \frac{(-1)^n}{2}$. Then $\|a\|_{V^r(I,B)} = 2^{\frac{1}{r}}$ for $0 < r \leq 1$ but $\|a\|_{V^0(I,B)} = 2$. Therefore the main theorem about continuity becomes

Theorem 4.5. *e_a is continuous on $S_a \cap (0, \infty]$.*

Proof. First we consider continuity at ∞. The case $e_a(\infty) = 0$ is trivial, same with $S_a = \{\infty\}$.
Assume $\lim_{r\to\infty} e_a(r) > e_a(\infty)$, so $\lim_{r\to\infty} e_a(r) = c' \cdot e_a(\infty)$ with $c' > 1$.
Let $1 < c < c'$. By assumption, for all $r \in (0, \infty)$, we find $t_0 < \cdots < t_n$ with

$$\sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|_B^r > (c \cdot e_a(\infty))^r$$

meaning

$$\sum_{j=1}^{n} (\|a(t_j) - a(t_{j-1})\|_B / (e_a(\infty)))^r > c^r.$$

Now let $r_0 \in (0, \infty) \cap S_a$. Note that for $r > r_0$, as $\|a(t_j) - a(t_{j-1})\|_B \leq e_a(\infty)$,

$$\sum_{j=1}^{n} (\|a(t_j) - a(t_{j-1})\|_B / (e_a(\infty)))^r \leq \sum_{j=1}^{n} (\|a(t_j) - a(t_{j-1})\|_B / (e_a(\infty)))^{r_0} \leq e_a(\infty)^{-r_0} e_a(r_0)^{r_0}$$

So

$$c^r e_a(\infty)^{r_0} \leq e_a(r_0)^{r_0}$$

for all $r > r_0$, so $e_a(r_0) = \infty$, contradicting $r_0 \in S_a$.

Now for $s \in S_a \cap (0, \infty)$, we first show upper semi-continuity.
Let $\varepsilon > 0$ be arbitrary and $t_0 < \cdots < t_n \in I$ such that

$$e_a(s) - \left(\sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|_B^s \right)^{\frac{1}{s}} < \frac{\varepsilon}{2}.$$

Now as $r \mapsto \left(\sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|_B^r \right)^{\frac{1}{r}}$ is continuous, we find a $s' > s$ with

$$\left(\sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|_B^{s'} \right)^{\frac{1}{s'}} > -\frac{\varepsilon}{2} + \left(\sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|_B^s \right)^{\frac{1}{s}}$$

implying, using monotonicity of e_a,

$$e_a(s') + \varepsilon > e_a(s) \geq e_a(s')$$

which again using monotonicity implies upper semi-continuity of e_a.

For lower semi-continuity, we can assume $S_a \neq [s, \infty]$. By homogeneity, we can also assume $e_a(\infty) < 1$.[3] Therefore $\|a(x) - a(y)\|_B < 1$ for all $x, y \in I$, in particular $r \mapsto \|a(x) - a(y)\|_B^r$ is monotonically decreasing and convex.

It suffices to show $f_a(r) := e_a(r)^r$ is lower semi-continuous, note f_a is still monotone by our assumption.
Assume $\lim_{r \nearrow s} f_a(r) = c' \cdot f_a(s)$ with $c' > c > 1$.
Fix $r_0 \in S_a \cap (0, s)$.
Again, for all $r_0 < r < s$ with $r \in S_a$, we find $t_0 < \cdots < t_n$ such that

$$\sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|_B^r \geq c \cdot f_a(s) \geq c \cdot \sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|^s$$

Now let $g(x) := \sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|_B^x$. We know g is monotone and continuous, satisfies $f_a(x) \geq g(x)$ and $g(r) \geq c \cdot g(s)$. By our assumption, g is convex and monotone.
Therefore

$$g(r_0) \geq \frac{c-1}{r-s} \cdot g(s) \cdot r_0 + g(s) \cdot \left(1 - \frac{s}{r-s}(c-1) \right) = g(s) \cdot \left(\frac{c-1}{r-s} \cdot (r_0 - s) + 1 \right)$$

For $r \nearrow s$ this implies $g(s) = 0$, in which case a is constant and the statement is trivial, or $g(r_0) = \infty$, contradiction to our construction of g.[4]

□

In a way, e_a is also continuous to the endpoint of S_a:

[3] We can avoid this assumption by dividing through $e_a(\infty)$ like we did in the $s = \infty$ case. This generalises the proof slightly. For simplicity however, we stick to the assumption.

[4] This also works if B only contains a seminorm, in which case we have a contradiction to $r_0 \in S_a$.

Lemma 4.6. *If $S_a = (s, \infty]$, then $\lim_{r \to s} e_a(r) = \infty$.*

Proof. If $s = 0$, we find $x < y < z$, with $\|e(y) - e(x)\|_B > 0$ and $\|e(z) - e(y)\|_B > 0$, so the statement follows from the fact that

$$\lim_{r \to 0} (\|e(y) - e(x)\|_B^r + \|e(z) - e(y)\|^r)^{\frac{1}{r}} = \infty.$$

So assume $\infty > s > 0$.
As $s \notin S_a$, for every $L \in \mathbb{R}_{>0}$, we find $t_1 < \cdots < t_n$ with

$$\left(\sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|_B^s \right)^{\frac{1}{s}} > L,$$

by continuity of $r \mapsto \left(\sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|_B^r \right)^{\frac{1}{r}}$ we find an $s' > s$ with

$$\left(\sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|_B^{s'} \right)^{\frac{1}{s'}} > L$$

so $e_a(s') > L$. By monotonicity of e_a, this gives $\lim_{r \to s} e_a(r) > L$. As L is arbitrary, we conclude $\lim_{r \to s} e_a(r) = \infty$.

□

Corollary 4.7. *If $S_a = (s, \infty]$ for some $s \in [0, \infty]$, then e_a is continuous on $[0, \infty]$.*

Proof. Follows directly from Lemma 4.6 and Theorem 4.5. □

Notably, recall that $S_a = [s, \infty]$ is also possible, in which case e_a is never continuous on $[0, \infty]$ as long as $s > 0$.

Essentially, all proofs we used mostly relied on properties of the function $r \mapsto a^r$ for some $a > 0$. As this map is also smooth, this gives us reason to believe the following conjecture:

Conjecture 4.8. *e_a is smooth on $S_a \cap (0, \infty)$.*

Due to the variation norm being quite technical and general, a classification of all functions $[0, \infty] \to [0, \infty]$ which are of the form e_a for some a, I, J seems hard.

4.1.2 Change in I

First, note that directly from the definition we get for $I \subseteq I' \subseteq J$ and arbitrary $r \in [0, \infty]$

$$\|a\|_{V^r(I,B)} \leq \|a\|_{V^r(I',B)}$$

and[5]

$$\|a\|_{V^r(\varnothing,B)} = 0$$

Nevertheless, $I \mapsto \|a\|_{V^r(I,B)}$ does not define an (outer) measure in any meaningful way.
Another interesting inequality is the following:

[5] For $r = \infty$, we use the convention that for any $\mathbb{R}$-valued function f, $\sup_{x \in \varnothing} f(x) := 0$.

Theorem 4.9 (Super-additivity)**.** *Let J be an ordered set, B normed space, $a : J \to B$ and $I, I' \subseteq J$ with $i \leq i'$ for all $i \in I$ and $i' \in I'$. Then we have for any $r \in (0, \infty)$*

$$\|a\|^r_{V^r(I,B)} + \|a\|^r_{V^r(I',B)} \leq \|a\|^r_{V^r(I \cup I',B)}$$

Proof. If one of the expressions is infinite the inequality becomes obvious.
Let $\varepsilon > 0$ be arbitrary.
Find sequences $t_0 < \cdots < t_n$ and $t'_0 < \cdots < t'_m$ such that $t_i \in I$, $t'_i \in I'$ and $\frac{\varepsilon}{2} + \sum_{j=1}^n \|a(t_j) - a(t_{j-1})\|^r_B > \|a\|^r_{V^r(I,B)}$ and $\frac{\varepsilon}{2} + \sum_{j=1}^m \|a(t'_j) - a(t'_{j-1})\|^r_B > \|a\|^r_{V^r(I',B)}$.
Note we have $t_0 < t_1 < \cdots < t_n \leq t'_0 < \dots t'_m$ therefore

$$\begin{aligned}
\|a\|^r_{V^r(I,B)} + \|a\|^r_{V^r(I',B)} - \varepsilon &\leq \left(\sum_{j=1}^n \|a(t_j) - a(t_{j-1})\|^r_B\right) + \left(\sum_{j=1}^m \|a(t'_j) - a(t'_{j-1})\|^r_B\right) \\
&\leq \left(\sum_{j=1}^n \|a(t_j) - a(t_{j-1})\|^r_B\right) + \|a(t'_0) - a(t_n)\|^r_B \\
&\quad + \left(\sum_{j=1}^m \|a(t'_j) - a(t'_{j-1})\|^r_B\right) \\
&\leq \|a\|^r_{V^r(I \cup I',B)}.
\end{aligned}$$

□

This inequality is useful for intervals, i.e. $I = [a, b]$ and $I' = [b, c]$.
In a special case, equality holds.

Corollary 4.10. *In the setting of Theorem 4.9, if $I \cap I' \neq \varnothing$ and $r \in (0, 1]$*

$$\|a\|^r_{V^r(I,B)} + \|a\|^r_{V^r(I',B)} = \|a\|^r_{V^r(I \cup I',B)}.$$

Proof. If one of the variations is infinite, this is not hard.
"$\leq$" follows from Theorem 4.9.
Let $\varepsilon > 0$ be arbitrary. So we find $t_0 < \cdots < t_n$ with $t_i \in I \cup I'$. It suffices to show

$$\sum_{j=1}^n \|a(t_j) - a(t_{j-1})\|^r_B \leq \|a\|_{V^r(I,B)} + \|a\|_{V^r(I',B)}.$$

If $t_0 \in I'$ or $t_n \in I$, we easily see that therefore $t_0 < \dots t_n$ is a valid sequence in I' or I which gives the result.
Otherwise, note that $I \cap I' = \{i_0\}$ for some i_0.
Because for $r \leq 1$, finer subdivisions give us better approximations of the variation norm (as for $r \leq 1$, $\|a + b\|^r \leq (\|a\| + \|b\|)^r \leq \|a\|^r + \|b\|^r$), we can without loss of generality assume that for some $0 < k < n$, $t_k = i_0$.
Then $t_0 < \cdots < t_k$ and $t_k < \cdots < t_n$ are valid subdivisions in I and I' respectively so:

$$\begin{aligned}
\sum_{j=1}^n \|a(t_j) - a(t_{j-1})\|^r_B &= \sum_{j=1}^k \|a(t_j) - a(t_{j-1})\|^r_B + \sum_{j=k+1}^n \|a(t_j) - a(t_{j-1})\|^r_B \\
&\leq \|a\|^r_{V^r(I,B)} + \|a\|^r_{V^r(I',B)}
\end{aligned}$$

□

If we remove the exponents r from Theorem 4.9, the inequalities change direction.

Theorem 4.11. *Let J be an ordered set, B normed space, $a : J \to B$ and $I, I' \subseteq J$ with $I \cap I' \neq \varnothing$ and $i \leq i'$ for all $i \in I$ and $i' \in I'$. Then we have for any $r \in [0, \infty]$*

$$\|a\|_{V^r(I\cup I',B)} \leq \|a\|_{V^r(I,B)} + \|a\|_{V^r(I',B)}$$

Proof. If one of the expressions is infinite, the inequality becomes easy (if $\|a\|_{V^r(I\cup I',B)} = \infty$, one can use a slightly modified version of the proof below).
Similarly, the cases $r = 0$ and $r = \infty$ are easy to check.

Again, we can assume $I \cap I' = \{i_0\}$.
Let $t_0 < \cdots < t_n$ be arbitrary with $t_i \in I \cup I'$.
Once again, it suffices to show

$$\left(\sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|_B^r\right)^{\frac{1}{r}} \overset{!}{\leq} \|a\|_{V^r(I,B)} + \|a\|_{V^r(I',B)}.$$

If $t_0 \in I'$ or $t_n \in I$ this is again trivial.

Else we can find a $0 \leq k < n$ such that $t_k \in I$ and $t_{k+1} \in I'$. Therefore $t_0 < \cdots < t_k \leq i_0 \leq t_{k+1} < \cdots < t_n$. But then $t_0 < \cdots < t_k \leq i_0$ and $i_0 \leq t_{k+1} < \cdots < t_n$ are subdivisions in I and I', so

$$\left(\|i_0 - t_k\|_B^r + \sum_{j=1}^{k} \|a(t_j) - a(t_{j-1})\|_B^r\right)^{\frac{1}{r}} \leq \|a\|_{V^r(I,B)}$$

and

$$\left(\|t_{k+1} - i_0\|_B^r + \sum_{j=k+1}^{n} \|a(t_j) - a(t_{j-1})\|_B^r\right)^{\frac{1}{r}} \leq \|a\|_{V^r(I',B)}.$$

So we need to show if $u := \sum_{j=1}^{k} \|a(t_j) - a(t_{j-1})\|_B^r$, $v := \sum_{j=k+1}^{n} \|a(t_j) - a(t_{j-1})\|_B^r$, $x := \|i_0 - t_k\|_B$ and $y := \|t_{k+1} - i_0\|_B$ that

$$(u + v + \|t_{k+1} - t_k\|_B^r)^{\frac{1}{r}} \overset{!}{\leq} (u + x^r)^{\frac{1}{r}} + (v + y^r)^{\frac{1}{r}}.$$

Note that by triangle inequality $\|t_{k+1} - t_k\|_B \leq x + y$, so it suffices to show (exponentiate with r):

$$u + v + (x + y)^r \overset{!}{\leq} \left((u + x^r)^{\frac{1}{r}} + (v + y^r)^{\frac{1}{r}}\right)^r$$

We show this inequality for all $u, v, x, y \geq 0$ and $r > 0$.
For $r \leq 1$ we have by Minkowski,

$$\begin{aligned} u + v + (x + y)^r &\leq u + v + x^r + y^r \\ &= \left((u + x^r)^{\frac{1}{r}}\right)^r + \left((v + y^r)^{\frac{1}{r}}\right)^r \\ &\leq \left((u + x^r)^{\frac{1}{r}} + (v + y^r)^{\frac{1}{r}}\right)^r. \end{aligned}$$

For $1 \leq r$ we use a more complicated argument.
We first prove the inequality for $u = 0$.
In this case the inequality follows, as the derivative with respect to v of the LHS is smaller than the derivative of the RHS with respect to v and equality at $v = 0$ holds:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}v}\left(v + (x+y)^r\right) = 1 \\ &= \left((v+y^r)^{\frac{1}{r}}\right)^{r-1} \cdot (v+y^r)^{\frac{1}{r}-1} \\ &\leq \left(x + (v+y^r)^{\frac{1}{r}}\right)^{r-1} \cdot (v+y^r)^{\frac{1}{r}-1} \\ &= \frac{\mathrm{d}}{\mathrm{d}v}\left(x + (v+y^r)^{\frac{1}{r}}\right)^r\end{aligned}$$

Having proved the result for $u = 0$, repeating the above argument with v proves the result. □

We have seen earlier that the variation norm is more-or-less continuous with r as the function argument.
For the case of I being a real interval, we get a similar statement:

Theorem 4.12. *Let $u, v \in \mathbb{R}$ and $a : [u, v] \to B$ continuous such that $\|a\|_{V^r([a,b],B)} < \infty$. Then*

$$g(x) := \|a\|_{V^r([u,x),B)}$$

is monotone and continuous on $[u, v]$.

Proof. Monotonicity follows from the observation at the beginning of this section, that the r-variation is continuous on increasing sets.
For continuity, first check the easy cases $r = 0$ and $r = \infty$ (notably, if $a : [u, v] \to B$ is continuous, $\|a\|_{V^0([a,b],B)} < \infty \Leftrightarrow a$ constant $\Leftrightarrow \|a\|_{V^0([a,b],B)} = 0$).

So for $r \in (0, \infty)$ we show lower and upper semi-continuity. Furthermore, to simplify our calculations, we show continuity of $h(x) = g(x)^r$.
First we show lower semi-continuity at $e \in (u, v]$. For all $\varepsilon > 0$, we find a sequence $u \leq t_0 < \dots t_n \leq e$ with $\varepsilon + \sum_{j=1}^n \|a(t_j) - a(t_{j-1})\|_B^r > h(e)$. We can assume that $t_n = e$.
Now by continuity of a, we find a $\delta > 0$, such that for all $y \in [u, v]$ with $|e - y| < 2\delta, \|a(e) - a(y)\|_B < \varepsilon$. W.l.o.g. $\delta + t_{n-1} < e$.
Consider $t_0 < \cdots < t_{n-1} < t_n - \delta < t_n$.

Then we have

$$
\begin{aligned}
&h(t) - \left(\|a(t_n) - a(t_n - \delta)\|_B^r + \|a(t_n - \delta) - a(t_{n-1})\|_B^r + \sum_{j=1}^{n-1} \|a(t_j) - a(t_{j-1})\|_B^r \right) \\
&\quad < \left(\varepsilon + \sum_{j=1}^{n-1} \|a(t_j) - a(t_{j-1})\|_B^r \right) \\
&\qquad - \left(\|a(t_n) - a(t_n - \delta)\|_B^r + \|a(t_n - \delta) - a(t_{n-1})\|_B^r + \sum_{j=1}^{n-1} \|a(t_j) - a(t_{j-1})\|_B^r \right) \\
&\quad = \varepsilon + \|a(t_n) - a(t_{n-1})\|^r - \left(\|a(t_n) - a(t_n - \delta)\|_B^r + \|a(t_n - \delta) - a(t_{n-1})\|_B^r\right) \\
&\quad \leq \varepsilon + \left(\|a(t_n) - a(t_n - \varepsilon)\| + \|a(t_n - \varepsilon) - a(t_{n-1})\|_B\right)^r \\
&\qquad - \left(\|a(t_n) - a(t_n - \delta)\|_B^r + \|a(t_n - \delta) - a(t_{n-1})\|_B^r\right) \\
&\quad \leq \varepsilon + \left(\|a\|_{V^\infty([a,b],B)} + \varepsilon\right)^r - \left(\|a\|_{V^\infty([a,b],B)}^r + \varepsilon^r\right).
\end{aligned}
$$

As ε goes to zero, the last expression goes to zero (note $\|a\|_{V^\infty([a,b],B)} < \infty$ as a has finite r-variation). With monotonicity of g, this proves $\lim_{x\nearrow e} g(x) = g(e)$.

The argument for upper semi-continuity at $e \in [u, v)$ uses a similar idea but is more complicated.
Assume $\lim_{x\searrow e} h(x) = h(e) + 2\varepsilon$ with $\varepsilon > 0$. By continuity of e, we again find a $\delta > 0$, such that for all $y \in [u, v]$ with $|e - y| < 2\delta$, $\|a(e) - a(y)\|_B < \varepsilon$. Without loss of generality $\delta + t_k < e$ and $\delta + e < t_{k+1}$. So by monotonicity of h, for every $x_0 := x \in (e, e + \delta]$, we find $u \leq t_0 < \cdots < t_n \leq x$ with

$$\sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|_B^r > h(e) + \varepsilon.$$

Note that $e < t_n$ and w.l.o.g. $t_0 \leq e$.
Thus, let $0 \leq k < n$ be such that $t_k \leq e$ but $t_{k+1} > e$.
Note that as before

$$p(x) := \|a(t_{k+1} - t_k)\|_B^r - \left(\|a(t_{k+1}) - a(e))\|_B^r + \|a(e) - a(t_k)\|_B^r\right)$$

is bounded above by some expression only depending on $\|a\|_{V^\infty([a,b],B)}$ and $\|a(t_{k+1} - a(e)\|$. As $t_{k+1} \leq x$ and a continuous, the expression gets arbitrarily small as x goes to e. In particular, by choosing δ small enough, we can guarantee that $p(y) \leq \frac{\varepsilon}{2}$ for all $y \in [e, e + \delta)$ (note that the value of $p(x)$ still depends on the sequence $t_0 < \cdots < t_n$ chosen).
Furthermore we know that

$$\sum_{j=1}^{k} \|a(t_j) - a(t_{j-1})\|_B^r \leq h(e),$$

implying

$$\sum_{j=k+2}^{n} \|a(t_j) - a(t_{j-1})\|_B^r \geq \varepsilon - p(x) \geq \frac{\varepsilon}{2}.$$

By choosing δ small enough, we can guarantee $\varepsilon - p(x) \geq \frac{\varepsilon}{2}$. Now if $e < x_1 := t_{k+1} \leq x_0 < e + \delta$ this implies

$$\varepsilon \leq \|a\|^r_{V^r([x_1,x_0],B)}.$$

But now we can apply the same argument with x_1 instead of $x = x_0$ (and recursively define $x_{n+1} := t_{k+1}$ to get

$$\varepsilon \leq \|a\|^r_{V^r([x_{n+1},x_n],B)}.$$

But then by Theorem 4.9

$$m \cdot \frac{\varepsilon}{2} \leq \sum_{n=0}^{m-1} \|a\|^r_{V^r([x_{n+1},x_n],B)} \leq \|a\|^r_{V^r([u,v],B)}.$$

Which for $m \to \infty$ gives a contradiction to $\|a\|_{V^r([u,v],B)}$ finite.

□

Remark 4.13. *This does not naturally generalise for $r < 1$ without the condition that $\|a\|_{V^r([a,b],B)} < \infty$. Consider $a : [-1,1] \to \mathbb{R}$ with $a(x) = \max(0,x)$. Then $g(x) = 0$ for all $x \leq 0$ and $r \in [0,\infty]$, but $g(x) = \infty$ for $x > 0$ and $r < 1$.*

It is easy to check, that if a is not semi-continuous, g is not as well, so as a corollary we get:

Corollary 4.14. *Let $u, v \in \mathbb{R}$ and $a : [u,v] \to B$ such that $\|a\|_{V^r([a,b],B)} < \infty$. Then*

$$g(x) := \|a\|_{V^r([u,x),B)}$$

is continuous if and only if a is continuous.

Recall that a function $f : X \to Y$ between metric spaces is called Hölder continuous to the exponent α if there is a constant $C > 0$ such that for all $x, x' \in X$

$$d_Y(f(x), f(x')) \leq C \cdot d_X(x,x')^\alpha.$$

There are several connections between r-variations, Hölder continuous functions and the related Hölder norm. A discussion of many results can be found in [4]. We conclude this section by discussing one interesting result thereof:

Theorem 4.15. *Let $r \in (0,\infty)$. Assume $a : [u,v] \to \mathbb{R}$ is continuous, has finite r-variation and is not constant on any interval $[x,y] \subseteq [u,v]$.*
Then the function g from Theorem 4.12 forms a bijection from $[u,v]$ to $[0, \|a\|_{V^r([u,v],B)}]$ and the function $h := a \circ (g^r)^{-1}$ is Hölder continuous with exponent $\frac{1}{r}$.

Proof. Trivially, $g(u) = 0$ and because a is continuous it is easy to see that $g(v) = \|a\|_{V^r([u,v],B)}$. Therefore, to show g is a bijection, by Theorem 4.12 and intermediate value theorem, it suffices to show it is monotone.
So let $x < y$. As a is nowhere constant, we find a $x < z < y$ with $a(z) \neq a(x)$. So consider a sequence $u \leq t_0 < \cdots < t_n < x$ with

$$\sum_{j=1}^{n} \|a(t_j) - a(t_{j-1})\|^r_B > g(x)^r - \|a(z) - a(x)\|^r_B\,. \qquad (*)$$

Then we can use the sequence $t_0 < \cdots < t_n \leq x < z$ to get a lower bound of $g(y)$, which together with $(*)$ implies $g(x)^r < g(y)^r$.

For Hölder, let $x, y \in [u, v]$ with $x \geq y$. As g^r is bijective and monotone, we can write $x = g(p)^r, y = g(q)^r$ with $p \geq q$, so $h(x) = p, h(y) = q$.
But therefore by Theorem 4.9

$$\begin{aligned}
|h(x) - h(y)| &= |f(p) - f(q)| \\
&\leq \left(\|a\|^r_{V^r([p,q),B)} \right)^{\frac{1}{r}} \\
&\leq \left(\|a\|^r_{V^r([u,p),B)} - \|a\|^r_{V^r([u,q),B)} \right)^{\frac{1}{r}} \\
&= (g(p)^r - g(q)^r)^{\frac{1}{r}} \\
&= |x - y|^{\frac{1}{r}} .
\end{aligned}$$

□

4.1.3 Change in a

First and importantly, we notice that r-variation is invariant under monotone increasing bijections.

Proposition 4.16. *Let $a : J \to B$, $I \subseteq J$ and $g : I' \to I$ be an increasing bijection of ordered sets. Then*

$$\|a \circ g\|_{V^r(I',B)} = \|a\|_{V^r(I,B)}.$$

Proof. Edge cases $r = 0$ and $r = \infty$ are easily checked.
Otherwise, this observation follows from the fact that g induces a bijection between finite monotone sequences in I and I' respectively.

□

As mentioned before, the r-variation forms an extended seminorm when $1 \leq r \leq \infty$. For $r = \infty$ this is easily checked, for $r \in (0, \infty)$, triangle inequality follows from Minkowski's inequality for sums, the other conditions are very easy to check as well.

Theorem 4.17. *For $r \in [1, \infty]$ and fixed I, $a \mapsto \|a\|_{V^r(I,B)}$ forms an extended seminorm on $J \to B$.*

In particular, the set of functions $J \to B$ with finite r-variation (on $I \subseteq J$) form a vector space, in case $r = 1$, these are called functions with bounded variation or BV functions.
Properties of functions with finite r-variation for some fixed r are extensively covered in [4].

4.2 Formalisation of the r-variation

We now present a formalisation of the r-variation in Lean.
As stated earlier, the usual variation of a function, meaning $r = 1$, has been implemented in mathlib, while the r-variation has not. We therefore will try to stay close to the already implemented version. Here, we only present the most interesting results and leave out the proofs.
To avoid making three different cases, $r = 0$, $r = \infty$ and $0 < r < \infty$, we implement the r-variation with the alternative ℓ^p characterisation we discussed at the beginning of section 4.1.
So for simplicity's sake, we redefine a suitable version of the ℓ^p (extended) norm as stated in section 4.1.

```
def lp_enorm {α : Type*} (p : ℝ≥0∞) (f : α → ℝ≥0∞): ℝ≥0∞ :=
  if p = 0 then
    (Function.support f).encard
  else if p = ∞ then
    ⨆ a : α, f a
  else
    (∑' x : α, (f x) ^ p.toReal) ^ (p.toReal)⁻¹
```

Using the counting measure, it should be easy to replace this redundant ℓ^p with an appropriate instance of the general, measure-theoretic version of the extended ℓ^p-norm.

Now the a_t as in section 4.1 is defined as

```
def diff_fun (n : ℕ) (t : ℕ → J) (a : J → B): Fin n → ℝ≥0∞ :=
  fun i ↦ edist (a (t (i.val + 1))) (a (t i.val))
```

There is a small formal difference to our mathematical definition. Namely, t formally takes values in the entirety of $\mathbb{N}$, instead of just $\{0, \dots, m\}$. We do this to stay consistent with the $r = 1$ case as implemented in mathlib.

Having defined a_t and the ℓ^p extended norm, we can state the r-variation simply as

```
def r_eVariation
    (r : ℝ≥0∞) {J B : Type*} [LinearOrder J][WeakPseudoEMetricSpace B]
    (I : Set J) (a : J → B): ℝ≥0∞ :=
  ⨆ p : ℕ  × { t : ℕ → J // Monotone t ∧ ∀ i, t i ∈ I},
    lp_enorm r (diff_fun p.1 p.2.1 a)
```

Once again, the details of the definition are heavily inspired by the already implemented $r = 1$ version. There are two main differences to our mathematical definition:

First, t again formally takes values of the entirety of $\mathbb{N}$ and is monotone on it. However the values of t bigger than "m" (meaning $p.1$) are entirely irrelevant for the definition, as we can extend every $t_0 \leq t_1 \leq \dots \leq t_m$ to a monotone t on $\mathbb{N}$, via $t_0 \leq \dots \leq t_m = t_{m+1} = t_{m+2} = \dots$.
Secondly and more importantly, the Lean definition is more general. While before, we exclusively considered B as a normed space, here we allow it to be a pseudo extended metric

space[6], note every normed space via $d(x, y) := \|x - y\|$ induces a pseudo extended metric space structure, so the Lean definition is in fact just a more general version.
Additionally, almost all of the propositions we proved (except Theorem 4.17), can be proved identically if we use a pseudo extended metric instead of a norm. As noted in a footnote, solely the proof of Theorem 4.5 has to be changed slightly, furthermore Lemma 4.6 and deductions thereof might fail in the case $r = 0$.

For the rest of this chapter, we will formalise the proof that this ℓ^p definition is equivalent to the original definition given in the beginning of the chapter, prove some technical details and use them to formalise the proof of Theorem 4.9 and some smaller results, like characterisation of the case $r = \infty$ and Remark 4.3. Our code for the r-variation spans about 900 lines.

First, note we defined the r-variation on the class `WeakPseudoEMetricSpace`.
We define it as:

```
class WeakPseudoEMetricSpace (α : Type u) : Type u extends EDist α where
  edist_self : ∀ x : α, edist x x = 0
  edist_comm : ∀ x y : α, edist x y = edist y x
  edist_triangle : ∀ x y z : α, edist x z ≤ edist x y + edist y z
```

The already formalised $r = 1$ instead uses the (stronger) class `PseudoEMetricSpace`, which adds topological constraints, making certain constructions simpler. However we then lose the important case of $\alpha =$ `ENNReal` (nonnegative real numbers with infinity), which we have in `WeakPseudoEMetricSpace` with $d(x, y) := |x - y|$[7] or simply, using that $x - y := 0$ for $x \leq y$ in `ENNReal`, $d(x, y) := (x - y) + (y - x)$.

```
instance : WeakPseudoEMetricSpace ℝ≥0∞ where
  edist := fun x y ↦ (x - y) + (y - x)
  edist_self := by simp only [tsub_self, add_zero, implies_true]
  edist_comm := by simp only [add_comm, implies_true]
  edist_triangle := ...

instance {α : Type*} [PseudoEMetricSpace α] : WeakPseudoEMetricSpace α where
  edist := edist
  edist_self := fun x ↦ edist_self x
  edist_comm := fun x ↦ edist_comm x
  edist_triangle := fun x ↦ edist_triangle x
```

Now going to the actual r-Variation, note that directly from the definition we can prove the observations stated at the beginning of section 4.1.2.

```
theorem r_eVariationOn_mono (h : I ⊆ I'):
    r_eVariationOn r I a ≤ r_eVariationOn r I' a := ...

@[simp] theorem r_eVariationOn_emptyset : r_eVariationOn r ∅ a = 0 := by
  simp [r_eVariationOn]
```

[6] Meaning, compared to a metric space $d(a, b) = \infty$ is possible and $d(a, b) = 0$ for $a \neq b$ as well.
[7] To use this in Lean, we first have to do a coercion to `EReal`

Unfolding definitions, we can show a usable version of the r-variation in the case $r \in (0, \infty)$.

```
theorem r_eVariationOn_ne_zero_ne_top (h : r ≠ 0) (h' : r ≠ ∞):
    r_eVariationOn r I a =
    ⨆ p : ℕ × { u : ℕ → J // Monotone u ∧ ∀ i, u i ∈ I },
    (∑ i ∈ Finset.range p.1,
      edist (a (p.2.1 (i + 1))) (a (p.2.1 i)) ^ r.toReal)^ r.toReal ⁻¹ := ...
```

In the case $r = 1$ and the `WeakPseudoEMetricSpace` being induced by a `PseudoEMetricSpace`, this coincides with the mathlib formalisation of the variation of a function.

```
theorem r_eVariationOn_one_eq_eVariationOn
    {B : Type*} [PseudoEMetricSpace B] (a : J → B) :
    r_eVariationOn 1 I a = eVariationOn a I := by simp
```

Importantly in the above characterisation of the $0 < r < \infty$ case, the outer exponent r.toReal^{-1} is *inside* the supremum. For many proofs however, like super-additivity, it is more convenient to get rid of the outer exponent thus instead consider $\|a\|^r_{V^r(I,B)}$.
So we want another characterisation of the case $0 < r < \infty$ with the exponent *outside* the supremum. For this, note that if $g : \mathbb{R}_{\geq 0} \cup \{\infty\} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ is a monotone, continuous function and $\varnothing \neq S \subseteq \mathbb{R}$, we have

$$\sup_{s \in S} g(s) = g\left(\sup_{s \in S} s\right)$$

We use a version of this to pull out the exponent of the supremum

```
theorem iSup_cont_ennreal
    {α : Type*} [Nonempty α] (f : α → ℝ≥0∞) (g : ℝ≥0∞ → ℝ≥0∞) (h :
    Continuous g) (h' : Monotone g):
    (⨆ a : α, g (f a)) = g (⨆ a : α, (f a)) := by sorry

theorem r_eVariationOn_pow_r_eq_sup (h : r ≠ 0) (h' : r ≠ ∞) :
  (r_eVariationOn r I a) ^ r.toReal =
  ⨆ p : ℕ × { u : ℕ → J // Monotone u ∧ ∀ i, u i ∈ I },
  ∑ i ∈ Finset.range p.1,
  edist (a (p.2.1 (i + 1))) (a (p.2.1 i))^r.toReal := ...
```

Critically, as the r-variation is the supremum of certain finite sums over monotone sequences, each of such a sum is smaller than the r-variation and, if the variation is finite, for every $\varepsilon > 0$ there is a sum at most ε smaller.

```
theorem sum_le_eVariationOn
    (h : r ≠ 0) (h' : r ≠ ∞) (m : ℕ) {u : ℕ → J} (mon : Monotone u) (hu : ∀ i,
     u i ∈ I) : (∑ i ∈ Finset.range m,
    edist (a (u (i + 1))) (a (u i)) ^ r.toReal)^  (r.toReal⁻¹)
    ≤ r_eVariationOn r I a := ...

lemma ENNReal_sup_lt_infty
    {α : Type*} [Inhabited α] {f : α → ℝ≥0∞}
    (h : ⨆ a : α, f a ≠ ∞) {ε : NNReal} (hε : 0 < ε) :
    ∃ b : α, ⨆ a : α, f a < f b + ε := ...
```

```
theorem r_eVariationOn_lt_infty (h : r ≠ 0) (h' : r ≠ ∞)
    (hv : r_eVariationOn r I a ≠ ∞) {ε : NNReal} (hε : 0 < ε) (hI: I.Nonempty):
    ∃ p : ℕ × { u : ℕ → J // Monotone u ∧ ∀ i, u i ∈ I }, r_eVariationOn r I a
    < (∑ i ∈ Finset.range p.1,
    edist (a (p.2.1 (i + 1))) (a (p.2.1 i))^r.toReal)^(r.toReal⁻¹) + ε := ...
```

The case $r = \infty$ allows for similar inequalities.
However, the fact that the r-variation for $r = \infty$ is defined in our Lean implementation via functions $u : \mathbb{N} \to J$ as well (instead of say $u : \{0, 1\} \to J$), first makes it necessary to extend pairs $(i, i') \in I \times I$ to a function $u : \mathbb{N} \to J$ in an appropriate manner.

```
def extend_fun (m : ℕ) (u : Fin (m+1) → J): ℕ → J :=
  fun n ↦ if hn : n < m + 1 then u ⟨n, hn⟩ else u m

def pair_fun (I : Set J) (j j' : I) : Fin 2 → J :=
  fun n ↦ if n = 0 then min j j' else max j j'

def pair_fun' (I : Set J) (j j' : I) :
    ℕ × { t // Monotone t ∧ ∀ (i : ℕ), t i ∈ I } :=
  ⟨2, extend_fun 1 (pair_fun I j j'),
  extend_fun_mon 1 (pair_fun I j j') (pair_fun_mon I j j'),
  fun i ↦ extend_fun_mem I 1 (pair_fun I j j') (pair_fun_mem I j j') i⟩
```

Notably, extend_fun takes monotone functions to monotone functions and functions with image contained in I to functions with image contained in I, which is what we need to use it for the r-variation and $\texttt{edist}(j, j') = \sum_{k=0}^{\infty} \texttt{edist}(a(\texttt{pair\_fun'}(j, j')(k+1), a(\texttt{pair\_fun'}(j, j')(k))$.
These are now the main ingredients for the characterisation of the $r = \infty$ case.

```
theorem infty_eVariationOn_eq_sup_edist :
    (r_eVariationOn ∞ I a) = ⨆ j : I × I, edist (a j.1) (a j.2) := by ...

theorem pair_le_r_eVariationOn_infty (j : I × I) :
    edist (a j.1) (a j.2) ≤ r_eVariationOn ∞ I a := by
  rw[infty_eVariationOn_eq_sup_edist]
  exact le_iSup_iff.mpr fun b a ↦ a j
```

```
theorem r_eVariationOn_infty_lt_infty
    (hv : r_eVariationOn ∞ I a ≠ ∞) {ε : NNReal} (hε : 0 < ε) (hI: I.Nonempty):
    ∃ j : I × I, r_eVariationOn ∞ I a < edist (a j.1) (a j.2) + ε := ...
```

To prove super-additivity, we need a way to "combine" finite sequences appropriately. Essentially, the above construction concatenates $t_0 \leq \cdots \leq t_n$ and $t'_1 \leq \cdots \leq t'_m$ with $t_n \leq t'_0$ to $t_0 \leq \cdots \leq t_n \leq t'_0 \leq \cdots \leq t'_m$.

```
def comb : ℕ × { u // Monotone u ∧ ∀ (i : ℕ), u i ∈ I } →
    ℕ × { u // Monotone u ∧ ∀ (i : ℕ), u i ∈ I' } → (ℕ → J) :=
  fun u u' k ↦ if k ≤ u.1 then u.2.1 k else u'.2.1 (k-u.1-1)

def comb' (p : ℕ × { u // Monotone u ∧ ∀ (i : ℕ), u i ∈ I })
    (p' : ℕ × { u // Monotone u ∧ ∀ (i : ℕ), u i ∈ I' })
    (h : p.2.1 p.1 ≤ p'.2.1 0):
    ℕ × { u // Monotone u ∧ ∀ (i : ℕ), u i ∈ I ∪ I' } :=
  ⟨p.1 + p'.1 + 1, comb I I' p p', comb_mono I I' p p' h, comb_mem I I' p p'⟩
```

Super-additivity / Theorem 4.9 then boils down to

```
theorem comb'_sum (p : ℕ × { u // Monotone u ∧ ∀ (i : ℕ), u i ∈ I })
    (p' : ℕ × { u // Monotone u ∧ ∀ (i : ℕ), u i ∈ I' })
    (h : p.2.1 p.1 ≤ p'.2.1 0):
    ∑ i ∈ Finset.range p.1, edist (a (p.2.1 (i + 1))) (a (p.2.1 i))^r.toReal +
    ∑ i ∈ Finset.range p'.1, edist (a (p'.2.1 (i + 1))) (a (p'.2.1 i))^r.toReal
    ≤ ∑ i ∈ Finset.range (comb' I I' p p' h).1,
    edist (a ((comb' I I' p p' h).2.1 (i + 1))) (a ((comb' I I' p p' h).2.1 i))^r.
    toReal := ...

theorem r_eVariationOn_superadditivity
    (h : r ≠ 0) (h' : r ≠ ∞) (II': ∀ i ∈ I, ∀ i' ∈ I', i ≤ i') :
    (r_eVariationOn r I a)^r.toReal + (r_eVariationOn r I' a)^r.toReal
    ≤ (r_eVariationOn r (I ∪ I') a)^r.toReal := ...
```

Finally, we want to prove Remark 4.3 (and by extension Theorem 4.2 barring the $r = 0$ case).

```
lemma summand_le_fin_sum_lp {a : α} (ha : a ∈ s) (f : α → ℝ≥0∞) {p : ℝ}
    (hp : 0 < p): f a ≤ (∑ x ∈ s, (f x)^p) ^ p⁻¹ := ...

theorem r_eVariation_on_mono (hr : r ≠ 0) (h : r ≤ r') :
    r_eVariationOn r' I a ≤ r_eVariationOn r I a := ...
```

Unfortunately, the proof is relatively tedious as we allow for $r' = \infty$ and never require the r-variation to be finite, the latter thus requiring us to characterise the non-finite case with appropriate sums as well, again for both $0 < r < \infty$ and $r = \infty$.

Future interesting work for formalisation of the r-variation and related topics might include a proof of Theorem 4.15, a proof of the extended seminorm properties in the case when `edist` is (induced by) a norm and $r \leq 1$ and finally, improving the class `WeakPseudoEMetricSpace`, such that it can be used for topology and in particular ε-δ proofs of continuity.

5 Iterate

5.1 Taking a finite amount of something

A difference between formalising mathematics and writing mathematics down on paper that was encountered in this formalisation, is taking an arbitrary finite amount of something. In particular for the transference principle, we have some arbitrary but fixed nonnegative integer m for which we take transformations $S_1, \ldots, S_m$, functions $f_1, \ldots, f_m$ and so on.

In Lean we can do this by specifying the indexing set in the following way:

Say we want to take objects $a_1, \ldots, a_m$ of some type X.
We can do this by introducing some indexing type ι and introduce some a with type $\iota \to X$. We can then "access" the elements $a_1, \ldots, a_m$, by taking i with type ι and look at (Lean notation) $a\ i$.

Crucially, this approach works for arbitrary types ι, so in order to ensure finiteness, we have to make the indexing type ι finite.
We can do this by giving ι the class predicate [`Fintype` ι].
Lean remembers class predicates. So by introducing ι globally via "variable" and giving it the `Fintype` predicate there, Lean will automatically detect the `Fintype` predicate when we work with ι and therefore in the vast majority of constructions which only work for finite types, we will not be forced to prove or even specify finiteness of ι again.

5.2 Iteration

However one construction that is very easy to think about on paper but hard to formalise is the following:

In the transference principle, we encounter (measure-preserving) transformations $S_1, S_2, \ldots, S_m : X \to X$ which commute, i.e. $S_i \circ S_j = S_j \circ S_i$.
In the transference principle, we then define some push-forward function $f_x : \mathbb{N}^m \to X$ which satisfies:

$$f_x(k_1, \ldots, k_m) = S_1^{k_1} \ldots S_m^{k_m} x$$

(this is a slightly simplified version)
Keeping the above section in mind, we want to construct some f of type $(\iota \to X \to X) \to (\iota \to \mathbb{N}) \to X \to X$. Then

$$f_x(k_1, \ldots, k_m) = f(S, k, x)$$

In Lean notation the latter becomes `f S k x`.
Notably we have two functional equations:

(1) For $i \in \iota$, $f(S, \delta_i, -) = S_i$

(2) $f(S, k + k', -) = f(S, k, -) \circ f(S, k', -)$

Where (1) follows immediately from the definition and (2) follows from the transformations being pairwise commuting, as we have:

$$\begin{aligned} f(S, k + k', -) &= S_1^{k_1 + k'_1} \dots S_m^{k_m + k'_m} \\ &= S_1^{k_1} S_2^{k_2 + k'_2} \dots S_m^{k_m + k'_m} S_1^{k'_1} \\ &= S_1^{k_1} S_2^{k_2} \dots S_m^{k_m + k'_m} S_1^{k'_1} S_2^{k'_2} \\ &\dots \\ &= S_1^{k_1} \dots S_m^{k_m} S_1^{k'_1} \dots S_m^{k'_m} \\ &= f(S, k, -) \circ f(S, k', -) \end{aligned}$$

Now these functional equations determine f uniquely:
Let f' be a function satisfying (1) and (2). To simplify notation, we write $f'(S, k, -) + f'(S, k', -) = f'(S, k, -) \circ f'(S, k', -)$.
In this setting $+$ is commutative from (2) and associative.
Now let $S : \iota \to X \to X$ and $k : \iota \to \mathbb{N}$ be arbitrary. We have:

$$\begin{aligned} f'(S, k, -) &= f'\left(S, \sum_{i \in \iota} k_i \cdot \delta_i, -\right) \\ &\overset{(2)}{=} \sum_{i \in \iota} f'(S, k_i \cdot \delta_i, -) \\ &\overset{(2)}{=} \sum_{i \in \iota} \sum_{j=1}^{k_i} f'(S, \delta_i, -) \\ &\overset{(2)}{=} f'\left(S, \sum_{i \in \iota} k_i \cdot \delta_i, -\right) \\ &= f(S, k, -) \end{aligned}$$

And therefore $f = f'$.

For our formalisation, we can therefore construct some f satisfying (1) and (2) via some complicated and indirect method, but then ignore the construction and deduce all information only using (1) and (2).

Defining such a function and proving (1) and (2) however is a non-obvious task.
A different approach to a very similar problem (iterated integration) than we are taking can be found in [8].

We define such a function in multiple steps.
First we need to put some order on ι, which, as an arbitrary finite type, is not properly ordered per-se, ideally we also want some order which we can access via natural numbers.

In Lean for some natural number `m`, `Fin m` is a type consisting of natural numbers less than `m`.
For some natural number `n<m`, where `n` has type $\mathbb{N}$, we can construct the corresponding member in `Fin m` via the constructor $\langle$`n`, `hn`$\rangle$, where `hn` is a proof of `n<m`.

Now for finite types α, mathlib has type equivalence (bijection) from α to `Fin (Nat.card` α`)`, where `Nat.card` is the cardinality.
So we can order ι, by saying for $i, j \in \iota$, that $i < j$ iff the corresponding images in the type equivalence satisfy this.

This is the main ingredient for the main help function, `iterate'_up_to`.

```
def iterate'_up_to:
    Fin (Nat.card ι + 1) → (ι → X → X)  → (Fin (Nat.card ι) → ℕ) → X → X :=
    fun m S k x ↦ by
  by_cases hm: m.val = 0
  · use x
  obtain ⟨m,hm'⟩ := m
  have t1: m - 1 < Nat.card ι := by
    simp only [Nat.card_eq_fintype_card]
    rw[Fintype.card_eq_nat_card]
    exact Nat.sub_one_lt_of_le (Nat.zero_lt_of_ne_zero hm) (Nat.le_of_lt_succ hm')
  have t2: m - 1 < Nat.card ι + 1 := by
    calc
      m - 1 ≤ m := Nat.sub_le m 1
        _   < Nat.card ι + 1 := hm'
    use (S ((Finite.equivFin ι).symm ⟨m-1,t1⟩))^[k ⟨m-1,t1⟩] (iterate'_up_to ⟨m-1,
    t2⟩ S k x)
  termination_by m => m.val
  decreasing_by
    exact Nat.sub_one_lt hm
```

Roughly speaking, this function iterates the transformations up to some index, i.e. if it takes input m, S, k, x it outputs $S_m^{k_m} S_{m-1}^{k_{m-1}} \dots S_1^{k_1} x$.
The function is defined recursively, if $m = 0$, it gives the identity otherwise it outputs $S_m^{k_m}$(`iterate'_up_to`$(m-1, S, k, x)$).

Here, Lean does not detect automatically that this recursion is well-defined, so we specify explicitly with `termination_by` and `decreasing_by` that the recursion ends, as we have the mono-variant `m.val`.

Using this, we define `iterate'` as `iterate'_up_to Nat.card` ι:

```
def iterate' : (ι → X → X)  → (Fin (Nat.card ι) → ℕ) → X → X :=
  iterate'_up_to ⟨Nat.card ι, lt_add_one (Nat.card ι)⟩
```

And finally, `iterate` by precomposition with the type equivalence between ι and `Fin (Nat.card ι)`:

```
def iterate: (ι → X → X)  → (ι → ℕ) → X → X :=
  fun S k x ↦ iterate' S (fun m ↦ k ((Finite.equivFin ι).symm m)) x
```

Notably, we still need to prove that `iterate` satisfies (1) and (2).
As this version is essentially a special case of `iterate'_up_to`, we will first prove analogue statements for the latter function.
`iterate'_up_to` is defined via a recursion, so the proof strategy for both (1) and (2) is an induction on m (how far we iterate up to).

For (1) the respective statement is:

```
theorem iterate'_up_to_unit
(m : Fin (Nat.card ι + 1)) (S : ι → X → X) (i : Fin (Nat.card ι)) (r : ℕ):
  iterate'_up_to m S (unit i r) =
  if (↑i : ℕ) < m then (S ((Finite.equivFin ι).symm i))^[r] else id := by
    obtain ⟨m,hm⟩ := m
    induction' m with m hm'
    · simp only [Fin.zero_eta, iterate'_up_to_zero, not_lt_zero', ↓reduceIte]
    ...
```

Note this is slightly stronger than (1), unit i r can be interpreted as δ_i^r.
For the induction, we first split up `m`, which has type `Fin (Nat.card` $\iota + 1$`)`, into ⟨`m`, `hm`⟩, where `m` has type ℕ and `hm` is a proof of $m < \text{Nat.card}\ \iota + 1$, and then apply induction to the (natural) `m`.
While the proof of the statement is almost trivial on paper, it is a little bit tedious in Lean, as one has to go back and forth between the types ℕ and types of the form `Fin k`, so we skip it here.

As a corollary, we get a version for iterate' and iterate:

```
theorem iterate'_unit (S : ι → X → X) (i : Fin (Nat.card ι)) (r : ℕ): iterate' S
     (unit i r) = (S ((Finite.equivFin ι).symm i))^[r] := by
  unfold iterate'
  rw[iterate'_up_to_unit]
  suffices : (↑i : ℕ) < (⟨Nat.card ι, lt_add_one (Nat.card ι)⟩ : Fin (Nat.card ι +
    1))
  · simp only [this, ↓reduceIte]
  obtain ⟨i,ih⟩ := i
  simp only
  contrapose ih
  simp_all only [not_lt]
```

```
theorem iterate_unit (S : ι → X → X) (i : ι) (r : ℕ):
    iterate S (unit i r) = (S i)^[r] := ...
```

Proving (2) is significantly more difficult, we again mostly omit the details of the proof steps.
First we need to introduce the notion of pairwise commuting transformations:

```
def pairwise_commuting {α : Type*} (S : α → X → X): Prop :=
  ∀ a b : α, (S a) ∘ (S b) = (S b) ∘ (S a)
```

For the eventual induction we need to note that if we iterate up to m, but $k_m = 0$ it is the same as iterating up to $m - 1$:

```
theorem iterate'_up_to_end_zero (m : Fin (Nat.card ι + 1)) (hm : m ≠ 0)
 (S: ι → X → X) (k: Fin (Nat.card ι) → ℕ) (hk: k ⟨m-1,fin_add_neg_one hm⟩ = 0):
  iterate'_up_to m S k = iterate'_up_to (m-1) S k := ...
```

Now by the observation

$$\text{iterate'\_up\_to}(m, S, k) = S_m^{k_m}(\text{iterate'\_up\_to}(m, S, k - k_m \cdot \delta_i))$$

```
theorem iterate'_up_to_add_end (m : Fin (Nat.card ι + 1)) (hm : m ≠ 0)
  {S : ι → X → X} (k: Fin (Nat.card ι) → ℕ):
  iterate'_up_to m S k = (S ((Finite.equivFin ι).symm
  ⟨m-1, fin_add_neg_one hm⟩))^[k ⟨m-1,fin_add_neg_one hm⟩]
  ∘ iterate'_up_to
  m S (k- unit (a := ⟨m-1,fin_add_neg_one hm⟩) (k ⟨m-1, fin_add_neg_one hm⟩)) :=
    ...
```

To apply the induction hypothesis for the final statement, we also need the observation that if $k = (k_1, \dots, k_m, \dots, k_{|\iota|})$ and $k' = (k_1, \dots, k_m, k'_{m+1}, \dots, k'_{|\iota|})$ that

$$\text{iterate'\_up\_to } m \ S \ k = \text{iterate'\_up\_to } m \ S \ k'$$

```
theorem iterate'_up_to_same (m : Fin (Nat.card ι + 1)) (S : ι → X → X)
  {k k': Fin (Nat.card ι) → ℕ} (h: ∀ n : Fin (Nat.card ι),
  (↑n : ℕ) < m → k n = k' n):
  iterate'_up_to m S k = iterate'_up_to m S k' := ...
```

This is enough to now prove the analogue of (2) for `iterate'_up_to` via induction on m:

```
theorem iterate'_up_to_add (m : Fin (Nat.card ι + 1)) {S : ι → X → X}
    (hS: pairwise_commuting S) (k k': Fin (Nat.card ι) → ℕ):
    iterate'_up_to m S (k+k') = iterate'_up_to m S k ∘ iterate'_up_to m S k' := by
    obtain ⟨m,hm⟩ := m
    revert k k'
    induction' m with m hm'
    · ...
    ...
```

Here we use the `revert` tactic, which removes k, k' from context and changes a goal of the form $P(k, k')$ for some proposition P to $\forall k\ \forall k', P(k, k')$.
This is sometimes necessary to obtain the "correct" induction hypothesis when doing induction in Lean.

Finally, as a corollary to this, we get compact versions of (2) for `iterate'` and `iterate`:

```
theorem iterate'_add
 {S : ι → X → X} (hS: pairwise_commuting S) (k k': Fin (Nat.card ι) → ℕ):
  iterate' S (k+k') = iterate' S k ∘ iterate' S k' :=
    iterate'_up_to_add ⟨Nat.card ι, lt_add_one (Nat.card ι)⟩ hS k k'
```

```
theorem iterate_add'
    {S : ι → X → X} (hS: pairwise_commuting S) (k k': ι → ℕ):
    iterate S (k + k') = iterate S k ∘ iterate S k' := by
  unfold iterate
  rw[← iterate'_add]
  suffices : ((fun m ↦ k ((Finite.equivFin ι).symm m)) + fun m ↦ k' ((Finite.
    equivFin ι).symm m))
    = (fun m ↦ (k + k') ((Finite.equivFin ι).symm m))
  · rw[this]
  ext m
  simp only [Pi.add_apply]
  exact hS
```

Now `iterate_unit` and `iterate_add'` prove that iterate satisfies (1) and (2), thus, as explained earlier, making the explicit construction obsolete.
And while in fact in the actual proof of the Calderón transference principle, we are not using the explicit construction of iterate, for reasons of convenience, we still use it to show that if $S_1, \ldots, S_n$ are all measurable (measure-preserving), that `iterate`$(S, k, -)$ is measurable (measure-preserving) as well:

```
@[fun_prop] theorem iterate_measurable {S : ι → Z → Z}
    (hS: ∀ i : ι, Measurable (S i)) (k : ι → ℕ):
  Measurable (iterate S k) := by
  unfold iterate
  fun_prop

@[simp] theorem iterate_measurepreserving {S : ι → Z → Z}
    (hS: MMeasurePreserving μ S) (k : ι → ℕ):
  MeasurePreserving (iterate S k) μ μ := ...
```

6 Conclusion

The main aim of this thesis was to formalise the Calderón transference principle as stated in Theorem 2.5. This provides a first step of a formalisation towards proving norm-variational estimates about ergodic averages, as in Theorem 1.3 from [10] and Theorem 1.4 from [9]. Therefore a natural extension of our formalisation would be to formalise other parts of the respective proofs or a generalisation thereof.
On the other end, one could formalise an appropriate proof going from one tuple of exponents to others, i.e. $(6, 6, 6)$. For this one might use the real interpolation theorem, which has been formalised in Lean as part of a formalisation of the Carleson theorem, see [7].

Additionally, in chapter 3 we introduced finclosed spaces, which form a generalisation of independence systems. As of now, the lemma we proved about them, lemma 3.2, is the only nontrivial result we found. It is therefore natural to ask, whether there are more interesting properties about them. As of writing, we also do not know about many examples of interesting finclosed spaces, which are not just independence systems.

In chapter 4 we also covered the r-variation of a function and in particular looked at what happens if we fix the function and change the exponent r. While most of our statements from the chapter are found in the literature and in particular in [4], we have most likely given the most general version of the r-variation so far. Theorem 4.5 appears to be new as well. It would be interesting to see if conjecture 4.8 and similar statements hold.
We have not formalised most statements from the chapter so far, giving us another source of potential future work.

On a more personal note, the Lean code for the Calderón transference principle and finclosed spaces is still not of very good quality and should be improved if it ever becomes part of a bigger formalisation.

The contents of chapter 2 and 3 are probably too technical for mathlib and would not find many applications elsewhere. Meanwhile, a formalisation of the r-variation, chapter 4, in mathlib seems potentially desirable, especially since the special case $r = 1$ has been implemented already and many statements from this special case generalise naturally.

Overall, the proof of the transference principle encompassed about 7000 lines of code in Lean, the definition of finclosed spaces and the proof of lemma 3.2 about 2000 and the formalisation of the r-variation and properties about it took about 900 lines of code.

Bibliography

[1] Jeremy Avigad, Leonardo De Moura, and Soonho Kong. "Theorem proving in Lean". In: (2021).

[2] Jeremy Avigad and Patrick Massot. *Mathematics in lean.* 2020.

[3] Alberto P Calderón. "Ergodic theory and translation-invariant operators". In: *Proceedings of the National Academy of Sciences* 59.2 (1968), pp. 349–353.

[4] VV Chistyakov and OE Galkin. "On maps of bounded p-variation with p> 1". In: *Positivity* 2.1 (1998), pp. 19–45.

[5] Jean-Pierre Conze and Emmanuel Lesigne. "Théoremes ergodiques pour des mesures diagonales". In: *Bulletin de la Société Mathématique de France* 112 (1984).

[6] Haskell B Curry. "Functionality in combinatory logic". In: *Proceedings of the National Academy of Sciences* 20.11 (1934), pp. 584–590.

[7] Floris van Doorn. *Carleson operators on doubling metric measure spaces.* URL: `https://florisvandoorn.com/carleson/` (visited on 08/31/2025).

[8] Floris van Doorn and Heather Macbeth. "Integrals Within Integrals: A Formalization of the Gagliardo-Nirenberg-Sobolev Inequality". In: *15th International Conference on Interactive Theorem Proving (ITP 2024)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. 2024, pp. 37–1.

[9] Polona Durcik, Lenka Slavićková, and Christoph Thiele. *Norm-variation of triple ergodic averages for commuting transformations.* arXiv preprint arXiv:2307.07372.

[10] Polona Durcik et al. "Norm variation of ergodic averages with respect to two commuting transformations". In: *Ergodic Theory and Dynamical Systems* 39.3 (2019), pp. 658–688.

[11] Nikos Frantzikinakis and Bryna Kra. "Convergence of multiple ergodic averages for some commuting transformations". In: *Ergodic Theory and Dynamical Systems* 25.3 (2005), pp. 799–809.

[12] Roger L Jones, Iosif V Ostrovskii, and Joseph M Rosenblatt. "Square functions in ergodic theory". In: *Ergodic theory and Dynamical systems* 16.2 (1996), pp. 267–305.

[13] Dominique Lépingle. "La variation d'ordre p des semi-martingales". In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 36.4 (1976), pp. 295–316.

[14] Richard Oberlin et al. "A variation norm Carleson theorem." In: *Journal of the European Mathematical Society (EMS Publishing)* 14.2 (2012).

[15] Terence Tao. "Norm convergence of multiple ergodic averages for commuting transformations". In: *Ergodic Theory and Dynamical Systems* 28.2 (2008), pp. 657–688.

[16] Norbert Wiener. "The quadratic variation of a function and its Fourier coefficients". In: *Journal of Mathematics and Physics* 3.2 (1924), pp. 72–94.