



Universitetet
i Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study programme/specialisation: <i>Mathematics and physics</i>	Spring / Autumn semester, 20. ¹⁹ ... Open / Confidential
Author: <i>Levi Andrei Tallaksen</i>	<i>Levi Andrei Tallaksen</i> (signature of author)
Programme coordinator: <i>Anders Tranberg</i> Supervisor(s): <i>Alexander Ulanovskii</i>	
Title of master's thesis: <i>Uniform Sampling of Band-limited Signals</i>	
Credits: <i>60</i>	
Keywords: <i>Sampling, frames, Paley-Wiener spaces, Bernstein spaces</i>	Number of pages: <i>86</i> + supplemental material/other: Stavanger, <i>15.06/2019</i> date/year

Contents

1	Introduction	4
2	Some conventions and notations	5
2.1	Functions	5
2.2	Sequences and series	6
2.3	Vector spaces	7
3	Functions and operators	8
3.1	Operators	8
3.2	L^p -spaces	10
3.3	The Fourier transform and its inverse	11
4	Banach theory	15
4.1	The operator norm	15
4.2	Fundamental results of Banach theory	16
4.3	Dual spaces	18
4.4	The adjoint operator	18
4.5	Absolutely convergent series in Banach spaces	20
5	Hilbert spaces	22
5.1	The adjoint of operators between Hilbert spaces	22
5.2	Complete sequences	24
5.3	Bessel's inequality and its converse	25
5.4	Orthonormal sequences	26
5.5	Riesz bases	30
6	Frames	34
6.1	Introduction to frames	34
6.2	Alternate dual frames	36
6.3	The frame operator	37
6.4	The dual frame	40
6.5	Riesz bases and frames	43
7	Some function spaces	46
7.1	Some complex analysis	46
7.2	Paley-Wiener spaces	48
7.3	Extending Paley-Wiener functions to the complex plane	50
7.4	Bernstein spaces	53
7.5	Properties of Bernstein spaces	55
7.6	Comparing PW_σ and B_σ	60

8	Sampling	62
8.1	A discussion about sampling	62
8.2	The sampling problem	63
8.3	Bessel's inequality in Paley-Wiener spaces	65
8.4	Sampling in Paley-Wiener space	66
9	The strong sampling problem in PW_σ and B_σ	69
9.1	Weak convergence of sequences of u.d. sets	69
9.2	Weak limits of translates	73
9.3	Lower uniform densities	75
9.4	Sampling in B_σ	77
9.5	Perturbations of sets of SS	79
9.6	Beurling's sampling theorem	80
10	Uniform sampling in PW_σ	82
10.1	Uniform sets of SS for PW_σ	82
10.2	Reconstruction	82
10.3	Perturbations of $\frac{\mathbb{Z}}{2\sigma}$	84
11	Conclusion	86

1 Introduction

This master thesis presents a short introduction into the field of sampling of functions of one variable. It is not meant to come up with any new theory, just to give an overview over known material, and to prove most of the background theory. The main focus is on band-limited functions, i.e. L^2 -functions f whose Fourier transform vanishes outside a bounded set, called the *spectrum* of f . We will also consider the larger class of Bernstein functions, which are not necessarily square-integrable.

Sampling means reducing a continuous function (signal) f to a discrete set Λ . The set $\{f(\lambda) \text{ s.t. } \lambda \in \Lambda\}$ is called the samples of f on Λ , while Λ is called the sampling set.

The classical *weak sampling problem* is to determine when a given function, or class of functions, can be determined from knowledge of Λ and their samples. The *strong sampling problem* asks when the sampling is *stable*, which is about how small perturbations changes the results. For the class of band-limited L^2 -functions, whose spectrum lies inside a bounded set S , this is equivalent to asking for which sets Λ the exponential system

$$\{e^{2\pi i\lambda(\cdot)} \text{ s.t. } \lambda \in \Lambda\}$$

is a *frame* in the Hilbert space $L^2(S)$.

In the case of uniform sampling, i.e. when the sampling set Λ is an arithmetic progression, we have the classical Shannon's sampling theorem whenever S is an interval. It gives a precise answer to how close the sample points must be for the exponential system to be an orthonormal basis for $L^2(S)$. Also, it gives a reconstruction formula for the function we are sampling. We will focus on uniform sampling in section 10, which is when we will get to this theorem.

Until that point, we will be more general than uniform sampling. In that case, the sampling problem becomes more complicated. However, the Swedish mathematician Arne Beurling came up with a beautiful approach, which essentially gives the sampling properties in terms of a density of the sampling set. We will briefly present his theory, but not prove a lot of his theorems. It will just be introduced, and finally applied to the case of uniform sampling.

The sampling theory presented below is based on the Fourier transform, the theory of frames and a little bit of complex analysis. The theory of frames is itself based on the theory of Hilbert spaces, which are Banach spaces with an inner product. Especially operators between such spaces are important, which is why we will start our discussion there. We will go into some very basic definitions and theorems, and some less well-known ones that will be useful to us later. For that reason, we will not start looking at sampling before section 8.

2 Some conventions and notations

2.1 Functions

Functions are normally defined in a pointwise sense. To say that a function $f : \mathbb{R} \rightarrow \mathbb{R}$ squares the argument, we could write

$$f(x) := x^2, \forall x \in \mathbb{R}.$$

However, we will often use a dot in parenthesis when we talk about the function itself, rather than the value of the function at a specific point. That is, we might define the same function as $f := (\cdot)^2$. Note that if f is a function, then an expression like $g := fe^{2\pi i(\cdot)}$ means the function defined pointwise by

$$g(x) := f(x)e^{2\pi ix}, \forall x \in \mathbb{R}.$$

Sometimes, we will encounter functions that are not defined pointwise. These functions, we will view as equivalence classes of functions. If f is a function defined pointwise, then the equivalence class containing $f : \mathbb{R} \rightarrow \mathbb{C}$ will be denoted by $[f]$. It is defined to consist of all functions $g : \mathbb{R} \rightarrow \mathbb{C}$, defined pointwise, s.t. (such that) the set

$$\{x \in \mathbb{R} \text{ s.t. } g(x) \neq f(x)\} \subset \mathbb{R}$$

has measure zero. Functions that coincide, except on a set of measure zero, are said to be equal a.e. (almost everywhere).

We will not go into the details of what a measure is, but it is a function whose inputs are (measurable) sets, and whose output is, in some sense, the size of that set. In particular, the measure of any interval equals the length of that interval, and the measure of a disjoint union of sets equals the sum of their individual measures. A set of measure zero is a set that is, for any $\epsilon > 0$, contained in some union of finitely or countably many open intervals of length less than ϵ . Discrete sets, i.e. sets containing finitely or countably many elements, have measure zero. Also, the (Lebesgue) integral of the unit function over a set S is equal to the measure of S , denoted by $\mu(S)$. This is the only thing we really need to know about measures in this thesis, other than the specific case that the measure is zero. Note that an unbounded set can have infinite measure, so in general, $\mu(S) \in \overline{\mathbb{R}_0^+} = [0, \infty]$.

A real function f might only be defined on a subset $S \subset \mathbb{R}$. In that case, we will use the convention that f is defined on \mathbb{R} , but vanishes outside S . That is, we identify f by the function $f\chi_S$, where χ_S is the characteristic function of S , defined by

$$\chi_S(x) := \begin{cases} 1, & x \in S \\ 0, & x \in \mathbb{R} \setminus S \end{cases}$$

Normally, we will just denote this extension by f as well. That is also the case when we extend or restrict a function in a given different way, as long as it does not create any confusion. Sometimes, we will start by calling it something else, until we have justified extending the original function, but in general, we will denote them by the same letter. Unsurprisingly, we will do the same when we extend or restrict functions between other vector spaces, which

we will call *operators*. There is really no difference between the terms function, signal, map and operator, but we will use them somewhat differently.

Note that there is an important exception to using the same notation when we extend or restrict a function. If $f: \mathbb{R} \rightarrow \mathbb{C}$ is a function and $\Lambda \subset \mathbb{R}$ is discrete, we do NOT denote $f|_{\Lambda}$ by f . Instead, as we will see in section 8, we will denote it by $f|_{\Lambda}$. $f|_{\Lambda}$ will then be viewed as a sequence, indexed by Λ , rather than a function. This requires an ordering, as we will talk about in the following section.

2.2 Sequences and series

When we have a sequence, it might be important to know the order of the terms. For that reason, we will always view the index set as a sequence indexed by \mathbb{N} . Of course, \mathbb{N} is indexed in increasing order. That is, if $\Lambda \subset \mathbb{R}$ is our index set, we think of it as the sequence

$$\Lambda = \{\lambda_1, \lambda_2, \dots\} = \{\lambda_k\}_{k=1}^{\infty} = \{\lambda_k\}_{k \in \mathbb{N}}.$$

Lower case letters with subscripts in \mathbb{N} will always denote the elements of the set with the corresponding capital letter. However, we will nearly always use the notation $\{a_{\lambda}\}_{\lambda \in \Lambda}$, rather than $\{a_{\lambda_k}\}_{k \in \mathbb{N}}$, to denote a sequence indexed by Λ . Of course, the same is true for series, which are just limits of sequences of partial sums:

$$\sum_{\lambda \in \Lambda} c_{\lambda} := \sum_{k \in \mathbb{N}} c_{\lambda_k} := \lim_{n \rightarrow \infty} \sum_{k=1}^n c_{\lambda_k}.$$

In both cases, we are assuming to have decided on an ordering of Λ , even though it is not stated explicitly. Subscripts of subscripts are mainly used to denote subsequences, in which case not all the elements of Λ would be covered.

In the examples, our index set is often the integers, \mathbb{Z} . Then, we are summing over the symmetric partial sums, i.e.

$$\sum_{k \in \mathbb{Z}} c_k := \lim_{n \rightarrow \infty} \sum_{k=-n}^n c_k = c_0 + \lim_{n \rightarrow \infty} \sum_{k=1}^n (c_k + c_{-k}).$$

Strictly speaking, this does not correspond to an ordering, since we are grouping terms pairwise together, but the general theorems remain true for this convention.

When writing general results, until section 8, we will always index the sequences and series by \mathbb{N} . However, we may replace it by any other discrete set, since they have the same cardinality. The index set might be finite, but that does not create any problems, since finite sums can be viewed as infinite sums whose terms vanish from some point onwards. Note that for the same reason, we might still call it a sequence or series even if there are finitely many terms.

Finally, we will mention a sequence of functions that is going to be used so much that have a notation for it:

$$E(\Lambda) := \{e^{2\pi i \lambda(\cdot)}\}_{\lambda \in \Lambda}.$$

It is also a normal convention to drop the factor 2π , but we will include it here. The main reason is that we will do the same when we define the Fourier transform in section 3.6, in order not to need any normalization. Most references, however, are from books that drop that factor.

2.3 Vector spaces

Whenever we say space, we mean vector space. Through the whole thesis, in general, we will assume that a given vector space is complex. That is, we can take linear combinations with complex coefficients. However, in the finite-dimensional cases, we will often refer to both \mathbb{R}^n and \mathbb{C}^n , even though \mathbb{R}^n is a real vector space. In particular, some of our examples are using \mathbb{R}^2 . Since \mathbb{C} is an extension of \mathbb{R} , and is in fact a real vector space itself of dimension two, it should be no surprise that most theorems also work for real vector spaces. The reason for using \mathbb{R}^2 in many examples, is because it is easier to picture what is going on in that space, as it can be drawn on a two-dimensional paper. The examples work equally well for \mathbb{C}^2 , though, which is a complex vector space. It should be noted that in these examples, we will as a convention use subscripts to denote the two coordinates. That is, if $x \in \mathbb{R}^2$, then

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Also, quite sloppily, we will refer to both \mathbb{R}^n and \mathbb{C}^n as Euclidean space, since it will never be important for us to distinguish them.

In the case that we have a *normed* vector space X , we will denote the norm by $\|\cdot\|$ or $\|\cdot\|_X$. We will normally use the subscript if there are multiple normed spaces involved, except when we are talking about the *operator norm*, to be defined in section 4.1. However, in the case that X is $L^p(S)$ or $l^p(\Lambda)$, we will just write $\|\cdot\|_p$. Specifying whether the elements are functions or sequences, and what their domain or index set is, will normally be clear from the context. The sup-norm, which is typically used for spaces of bounded functions, is denoted by $\|\cdot\|_\infty$. Sometimes, we will denote inner products with subscripts as well, if we need to distinguish between two Hilbert spaces. However, we normally have just one, in which case we will always denote it by $\langle \cdot, \cdot \rangle$. It is well-known that norms and inner products are continuous.

It should be noted that when consider stable sampling in section 8.2, we will encounter two different norms on the same space. In that case, we give them different subscripts to distinguish them. As we will see, we have a particular way of doing it in that case, so it is always clear which of the two we are talking about.

We will never prove that a given vector space is actually a vector space. It is normally very easy. However, when we encounter Paley-Wiener spaces and Bernstein spaces, we will prove completeness. Another thing to note is that whenever we talk about a basis, we mean a *Schauder* basis, i.e. a sequence of vectors s.t. every vector can be uniquely expanded as a linear combination of *all* the basis vectors. Hence, if the basis is an infinite sequence, we must allow infinite linear combinations.

3 Functions and operators

There are lots of different properties that an operator between vector spaces may have. Section 3.1 presents 11 of them, and proves a few relations between them that will be useful later. A particularly important operator is the *Fourier transform*, which is defined on $L^1(\mathbb{R})$, and can be extended to $L^p(\mathbb{R})$ for any $p > 1$. We will, however, only consider its extension to the Hilbert space $L^2(\mathbb{R})$, which will be the most relevant one.

3.1 Operators

Definition 3.1. Let X, Y be vector spaces and $T: X \rightarrow Y$ be an operator.

- (i) T is called **linear** if $T(\alpha x + \beta y) = \alpha Tx + \beta Ty$, $\forall \alpha, \beta \in \mathbb{C}$, $\forall x, y \in X$.
- (ii) T is called **anti-linear** if $T(\alpha x + \beta y) = \bar{\alpha}Tx + \bar{\beta}Ty$, $\forall \alpha, \beta \in \mathbb{C}$, $\forall x, y \in X$.
- (iii) T is called **surjective** if $\forall y \in Y$, $\exists x \in X$ s.t. $Tx = y$.
- (iv) T is called **injective** if whenever $x, y \in X$ and $Tx = Ty$, we must have $x = y$.
- (v) T is called **bijective**, or **invertible**, if it is both surjective and injective. In that case, the **inverse** of T is the operator $T^{-1}: Y \rightarrow X$ defined by $T^{-1}(Tx) = x$, $\forall x \in X$.
- (vi) T is called an **isomorphism** if it is linear and bijective.
- (vii) T is called an **anti-isomorphism** if it is anti-linear and bijective.

Now, let X and Y be normed spaces.

- (viii) T is called **bounded (above)** if $\exists K > 0$ s.t. $\|Tx\|_Y \leq K\|x\|_X$, $\forall x \in X$. K is called an (upper) **bound** for T . The space of bounded linear operators mapping X into Y is denoted by $B(X, Y)$.
- (ix) T is called **bounded below** if $\exists K > 0$ s.t. $\|Tx\|_Y \geq K\|x\|_X$, $\forall x \in X$. K is called a **lower bound** for T .
- (x) T is called an **isometry** if $\|Tx\|_Y = \|x\|_X$, $\forall x \in X$.
- (xi) T is called **continuous** at $x \in X$ if for any sequence $\{x_k\}_{k \in \mathbb{N}} \subset X$ converging to x , it is true that the sequence $\{Tx_k\}_{k \in \mathbb{N}} \subset Y$ converges to Tx . If T is continuous at every element of X , we simply say that T is continuous.

These are the main properties we will go into here. It can be shown that (xi) is equivalent to the well-known ϵ - δ -definition of continuity. The rest of this section presents and proves some other general facts involving the properties above.

Lemma 3.2. Let X and Y be vector spaces and $T: X \rightarrow Y$ be a linear operator. Then, we have:

(i) T is injective if and only if the only $x \in X$ satisfying $Tx = 0$ is $x = 0$.

Now, let X and Y be normed spaces.

(ii) If T is bounded below, it is injective.

(iii) T is bounded if and only if it is continuous.

Proof. (i) The fact that $T0 = 0$ follows from linearity of T . And by definition, if T is injective, there can be no more than one $x \in X$ satisfying $Tx = 0$. This proves necessity.

For sufficiency, assume that $Tx = 0$ if and only if $x = 0$. If $x_1, x_2 \in X$ satisfy $Tx_1 = Tx_2$, then $0 = Tx_1 - Tx_2 = T(x_1 - x_2)$. Hence, our assumption implies that $x_1 - x_2 = 0$, i.e. that $x_1 = x_2$, showing that T is injective.

(ii) Assume that $K > 0$ is a lower bound for T . If some $x \in X$ satisfies $Tx = 0$, then $0 = \|Tx\|_Y \geq K\|x\|_X$. This can only happen if $\|x\|_X = 0$, i.e. if $x = 0$. Thus, T is injective by (i).

(iii) **[Wik2]** Assume that $K > 0$ is a bound for T , and pick an $x \in X$. Let $\{x_k\}_{k \in \mathbb{N}} \subset X$ be a sequence converging to x . Pick an $\epsilon > 0$, and find $N \in \mathbb{N}$ satisfying

$$\|x - x_k\|_X \leq \frac{\epsilon}{K}, \quad \forall k \geq N.$$

Then,

$$\|Tx - Tx_k\|_Y = \|T(x - x_k)\|_Y \leq K\|x - x_k\|_X \leq \epsilon, \quad \forall k \geq N.$$

That is, $Tx_k \rightarrow Tx$ in Y as $k \rightarrow \infty$, so T is continuous at x . Since $x \in X$ was arbitrary, this shows that T is continuous.

Conversely, assume T is continuous at 0. Find a $\delta > 0$ s.t. $\|Ty\|_Y \leq 1$ whenever $\|y\|_X \leq \delta$. Setting $K := \frac{1}{\delta}$, we have:

$$\|Tx\|_Y = \left\| T \left(\delta \frac{x}{\|x\|_X} \right) \right\|_Y \leq \frac{1}{\delta} \|x\|_X \cdot 1 = K\|x\|_X, \quad \forall x \in X,$$

since $y := \delta \frac{x}{\|x\|_X}$ satisfies $\|y\|_X = \delta$. Thus, K is a bound for T . □

From the proof of (iii), it is obvious that continuity for linear operators, continuity is implied by continuity at just the origin! This is also a well-known fact.

Before we conclude the introduction to the basic properties of operators, we will briefly talk about compositions of them. That is, given operators $R: X \rightarrow Y$ and $S: Y \rightarrow Z$, we will look at the operator $T := SR: X \rightarrow Z$. If we know that both R and S possess a given property, will that property necessarily be transferred to T ? Let us first show that boundedness will, whether it is above or below.

Lemma 3.3. Let X, Y and Z be normed spaces. Given two operators $R: X \rightarrow Y$ and $S: Y \rightarrow Z$, define $T := SR: X \rightarrow Z$.

- (i) If $K_1, K_2 > 0$ are bounds for R and S , respectively, then K_1K_2 is a bound for T .
- (ii) If $K_1, K_2 > 0$ are lower bounds for K_1 and K_2 , respectively, then K_1K_2 is a lower bound for T .

Proof. (i) $\|Tx\|_Z = \|S(Rx)\|_Z \leq K_2 \cdot \|Rx\|_Y \leq K_2K_1 \cdot \|x\|_X, \forall x \in X$.

The proof of (ii) is similar. □

What about the other properties? If R and S are anti-linear, then it actually turns out that T is linear, since $T(\alpha x + \beta y) = S(\overline{\alpha}Rx + \overline{\beta}Ry) = \alpha(SRx) + \beta(SRy) = \alpha Tx + \beta Ty$. Hence, anti-linearity cannot be transferred to T . Of course, this can also be said about anti-isomorphism, which is a stronger property. However, these turn out to be the only exceptions. We will not provide the proofs, but they are very trivial.

Proposition 3.4. *Any composition of operators sharing a property in definition 3.1, other than anti-linearity and anti-isomorphism, must itself possess that property.*

3.2 L^p -spaces

As usual, we define

$$L^p(S) := \{f: \mathbb{R} \rightarrow \mathbb{C} \text{ s.t. } \int_S |f(x)|^p dx < \infty\},$$

where $p \in [1, \infty)$ and $S \subseteq \mathbb{R}$ are given. The corresponding space of equivalence classes, defined in section 2.1, is also denoted by $L^p(S)$. Whenever necessary, we will point out which of the two we are talking about. These equivalence classes are crucial, since the norm defined by

$$\|[f]\|_p := \left(\int_S |f(x)|^p dx \right)^{\frac{1}{p}}, \forall f \in L^p(S). \quad (1)$$

will not be a norm if the elements of $L^p(S)$ are functions. The reason is that there exist non-zero functions $f \in L^p(S)$ satisfying $\|f\|_p = 0$, which is forbidden for norms. However, a nice thing about the equivalence classes is that any integral can be computed using any of its representatives. That is, if $f, g \in L^1(S)$ and $f = g$ a.e., then

$$\int_S f(x)dx = \int_S g(x)dx.$$

After all, any integral over a set of measure zero will vanish, so if the functions coincide outside a set of measure zero, then the integrals must also coincide. In particular, that is what allows us to define the norm, given by (1), using any representative of the equivalence class.

As mentioned in section 2.1, if $S \subset \mathbb{R}$, then any $f \in L^p(S)$ will be interpreted as the function $f\chi_S$, which is defined on \mathbb{R} . With this convention, it is clear that $L^p(S) \subseteq L^p(R)$ whenever $S \subset R$, with equality if and only if the sets differ only on a set of measure zero. Another useful property can be seen from Hölder's inequality ([Ch10], p. 19). If S is bounded, if $p > q$, if $r > 1$ is given so that $\frac{q}{p} + \frac{1}{r} = 1$, and if $f \in L^p(S)$, we have:

$$\|f\|_q^q = \int_S |1 \cdot f(x)|^q dx \leq \left(\int_S 1^{qr} dx \right)^{\frac{1}{r}} \left(\int_S |f(x)|^{q\frac{r}{q}} dx \right)^{\frac{q}{p}} = \mu(S)^{\frac{1}{r}} \|f\|_p^q < \infty,$$

where $\mu(S)$ is the measure of S . Hence, any L^p -function is also an L^q -function. To see that the converse does not hold, notice that $f(x) := \frac{1}{(x-x_0)^p}$, where x_0 is an inner point of S , defines a function that is in $L^q(S)$, but not in $L^p(S)$. We state these two observations as a theorem.

Theorem 3.5. *Let $p, q \in [1, \infty)$ and $R, S \subseteq \mathbb{R}$ be given.*

- (i) *If $S \subset R$, then $L^p(S) \subseteq L^p(R)$. Also, $L^p(S) = L^p(R)$ if and only if $\mu(R \setminus S) = 0$.*
- (ii) *If S is bounded, and if $p > q$, then $L^p(S) \subset L^q(S)$.*

Specifically, theorem 3.5 tells us that $L^2(S) \subset L^1(S) \subset L^1(\mathbb{R})$ for any bounded set $S \subset \mathbb{R}$. This is a fact will be used a lot when we get to Paley-Winer spaces.

3.3 The Fourier transform and its inverse

Definition 3.6. *The operators \mathfrak{F} and \mathfrak{G} are defined pointwise on $L^1(\mathbb{R})$ by:*

(i)

$$\hat{f} := \mathfrak{F}f := \int_{\mathbb{R}} f(x)e^{-2\pi i(\cdot)x} dx, \quad \forall f \in L^1(\mathbb{R}).$$

(ii)

$$\check{F} := \mathfrak{G}F := \int_{\mathbb{R}} F(t)e^{2\pi i(\cdot)t} dt, \quad \forall F \in L^1(\mathbb{R}).$$

\mathfrak{F} is called the **Fourier transform**.

The operators above are well-defined, since the exponential factors have absolute value 1. Since $L^1(S) \subseteq L^1(\mathbb{R})$ for any $S \subset \mathbb{R}$, we can specifically apply them to $f \in L^1(S)$. In that case, we only need to integrate over S , since f vanishes elsewhere.

Clearly, both operators are linear by the properties of integrals. Some well-known, less obvious properties of the Fourier-transform are given in the following theorem. Since \hat{f} and \check{f} are symmetric about the y-axis, it is clear that \mathfrak{G} must also have these properties.

Theorem 3.7 ([Ch10], p. 138-141). *For any $f \in L^1(\mathbb{R})$, we have:*

- (i) *\hat{f} is continuous on \mathbb{R} , and $\hat{f}(t) \rightarrow 0$ as $|t| \rightarrow \infty$.*
- (ii) *$\|\hat{f}\|_{\infty} \leq \|f\|_1$.*
- (iii) *If $\hat{f} \in L^1(\mathbb{R})$, then $f = \mathfrak{G}\hat{f}$ a.e.*

Property (ii) makes sense because the sup-norm is a norm on the space of continuous functions tending to 0 at $\pm\infty$. Note that it tells us that the Fourier transform is a bounded operator, and that 1 is a bound. In fact, 1 is the smallest bound, but that is not an important point to us yet. It is a lot more relevant what happens to the norm when we extend \mathfrak{F} to an operator on $L^2(\mathbb{R})$, which can be done in a very convenient way!

As we have seen, $L^2[-R, R] \subset L^1(\mathbb{R})$ for any $R > 0$. Thus, it makes sense to define an operator \mathfrak{F}_R on $L^2(\mathbb{R})$ by

$$\mathfrak{F}_R f := \int_{-R}^R f(x) e^{-2\pi i(\cdot)x} dx, \quad \forall f \in L^2(\mathbb{R}).$$

It is well-known that if we let $R \rightarrow \infty$, the integral converges a.e. Hence, up to a set of measure zero, we can define an operator T on $L^2(\mathbb{R})$ by

$$Tf := \lim_{R \rightarrow \infty} (\mathfrak{F}_R f), \quad \forall f \in L^2(\mathbb{R}).$$

That is, if $R \subseteq \mathbb{R}$ is the set of points where the integral converges, then

$$(Tf)(t) := \begin{cases} \int_{-\infty}^{\infty} f(x) e^{-2\pi ixt} dx, & t \in R \\ \text{undefined}, & t \in \mathbb{R} \setminus R \end{cases}, \quad \forall f \in L^2(\mathbb{R}).$$

Since Tf is defined a.e., we will not view it as a function, but as an equivalence class of functions.

One property of our new operator is that if $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, then $Tf = \hat{f}$. That is, T is an extension of the Fourier transform from $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ to $L^2(\mathbb{R})$. For that reason, we will also denote T by \mathfrak{F} , and refer to it as the Fourier transform on $L^2(\mathbb{R})$. Just as the Fourier transform on $L^1(\mathbb{R})$, it has some really nice properties. The most relevant ones for us are stated in the following theorem. One of them only makes sense if we extend \mathfrak{G} to $L^2(\mathbb{R})$ as well, which is of course possible to do in the same way.

Theorem 3.8 ([Ch10], p. 143-144). *The Fourier transform on $L^2(\mathbb{R})$ satisfies:*

- (i) $\hat{\hat{f}} \in L^2(\mathbb{R})$, $\forall f \in L^2(\mathbb{R})$.
- (ii) $\langle \hat{f}, \hat{g} \rangle = \langle f, g \rangle$, $\forall f, g \in L^2(\mathbb{R})$.
- (iii) $\mathfrak{F}: L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$ is bijective, and $\mathfrak{F}^{-1} = \mathfrak{G}$.

This theorem gives us an idea of how nice the Fourier transform on $L^2(\mathbb{R})$ is! Properties (i) and (iii) tell us that \mathfrak{F} maps $L^2(\mathbb{R})$ bijectively into itself, and (iii) also tells us what the inverse is. For that reason, \mathfrak{G} is usually referred to as the *Fourier inverse transform*, and denoted by \mathfrak{F}^{-1} . However, since it is not in general true for L^1 -functions, we will use a non-standard notation here to point out that they are not always inverses of each other.

Theorem 3.8 (ii) tells us that the Fourier transform is *unitary*, i.e. that it preserves inner products. In particular, it preserves norms and orthogonality. Later, we will take advantage of this to transform an orthonormal system in $L^2(\mathbb{R})$ into another one. Comparing with the Fourier transform on $L^1(\mathbb{R})$, they are both bounded linear operators, with 1 being the smallest bound, but only the Fourier transform on $L^2(\mathbb{R})$ is an isometry.

It should be noted that since $\hat{f} \in L^2(\mathbb{R})$, it makes sense to talk about convergence of $\mathfrak{F}_R f$ to f in the L^2 -norm. For that reason, it is very usual to express the Fourier transform as an integral, even though it is not defined everywhere. After all, an integral with infinite limits of integration simply means the limit of a corresponding integral as the limits of integration tend to infinity. Taking the limit in the L^2 -norm, rather than in a pointwise sense, we do not need to worry about sets of measure zero to say that $\lim_{R \rightarrow \infty} \mathfrak{F}_R f = \hat{f}$.

Example 3.9. For constants $c, \sigma > 0$, let $f := e^{2\pi i c(\cdot)} \chi_{[-\sigma, \sigma]} \in L^1[-\sigma, \sigma]$. Then,

$$\begin{aligned} \hat{f}(t) &= \int_{-\sigma}^{\sigma} e^{2\pi i c x} e^{-2\pi i x t} dx = \int_{-\sigma}^{\sigma} e^{2\pi i (c-t)x} dx = \frac{1}{2\pi i (c-t)} \left[e^{2\pi i (c-t)x} \right]_{x=-\sigma}^{x=\sigma} \\ &= \frac{1}{\pi(c-t)} \cdot \frac{e^{2\pi \sigma i (c-t)} - e^{-2\pi \sigma i (c-t)}}{2i} = \frac{\sin(2\pi \sigma (c-t))}{\pi(c-t)} = 2\sigma \operatorname{sinc}(2\sigma(c-t)), \quad \forall t \in \mathbb{R}, \end{aligned}$$

where the sinc-function is defined by

$$\operatorname{sinc}(x) := \begin{cases} \frac{\sin(\pi x)}{\pi x}, & x \in \mathbb{R} \setminus \{0\} \\ 1, & x = 0 \end{cases}$$

Example 3.10. For a constant $c \in \mathbb{R}$ and a function $f \in L^1(\mathbb{R})$, let $g := f e^{2\pi i c(\cdot)} \in L^1(\mathbb{R})$ and $h := f(\cdot - c) \in L^1(\mathbb{R})$. With a change of variable $y := x - c$, we get:

$$\begin{aligned} \hat{g}(t) &= \int_{\mathbb{R}} f(x) e^{2\pi i c x} e^{-2\pi i x t} dx = \int_{\mathbb{R}} f(x) e^{-2\pi i (t-c)x} dx = \hat{f}(t-c), \quad \forall t \in \mathbb{R}. \\ \hat{h}(t) &= \int_{\mathbb{R}} f(x-c) e^{-2\pi i x t} dx = \int_{\mathbb{R}} f(y) e^{-2\pi i (y+c)t} dy = \int_{\mathbb{R}} f(y) e^{-2\pi i y t} dy \cdot e^{-2\pi i c t} \\ &= \hat{f}(t) e^{-2\pi i c t}, \quad \forall t \in \mathbb{R}. \end{aligned}$$

That is, the Fourier transform turns multiplication by a complex exponential into translation, and vice versa. Clearly, the same calculations hold if $f \in L^2(\mathbb{R})$, as long as we only consider the values of t s.t. the integral converges. Alternatively, we can note that integrating over a bounded interval leads to the same integral in y . The change of variable changes the limits of integration, but when we let them tend to $\pm\infty$, it makes no difference. Thus, the Fourier transform on $L^2(\mathbb{R})$ must possess these properties as well.

Example 3.11. Given a bounded set $S \subset \mathbb{R}$, let $F \in L^2(S)$, and define $f := \check{F} \in L^2(\mathbb{R})$. Also, let $G(t) := tF(t)$, $\forall t \in \mathbb{R}$ and $g := \check{G}$. Since

$$\|G\|_2^2 = \int_S |G(t)|^2 dt = \int_S |t|^2 \cdot |F(t)|^2 dt \leq (\sup |S|)^2 \|F\|_2^2 < \infty,$$

we have $G \in L^2(S)$, and thus $g \in L^2(\mathbb{R})$. We will not justify interchanging derivative and integral yet, but we get back to this again in section 7.3. Assuming we can, and repeatedly using the fact that \mathfrak{F} and \mathfrak{G} preserve norms, we have:

$$f'(x) = \frac{d}{dx} \int_S F(t) e^{2\pi i x t} dt = 2\pi i \int_S t F(t) e^{2\pi i x t} dt = 2\pi i \int_S G(t) e^{2\pi i x t} dt = 2\pi i g(x),$$

$$\forall x \in \mathbb{R}.$$

$$\|f'\|_2 = 2\pi \|g\|_2 = 2\pi \|G\|_2 \leq 2\pi (\sup |S|) \|F\|_2 = 2\pi (\sup |S|) \|f\|_2.$$

From this, we see that if f is the Fourier inverse transform of an L^2 -function vanishing outside a bounded set, then f is differentiable on \mathbb{R} , $f' \in L^2(\mathbb{R})$, and we have found an upper bound on $\|f'\|_2$. Our upper bound also shows that if we fix the set S , then the differentiation operator on the space of such functions is bounded! We will come back to that in section 7.6.

4 Banach theory

A normed vector space X is said to be *Banach*, or *complete*, if every Cauchy-sequence in X converges to some element of X . Given any $n \in \mathbb{N}$ and any $p \in [1, \infty)$, it is well known that both \mathbb{R}^n , \mathbb{C}^n , $l^p(\mathbb{N})$ and $L^p(\mathbb{R})$ are all complete w.r.t. their usual norms. The same is true for $l^p(\Lambda)$ and $L^p(S)$, where $\Lambda \subset \mathbb{R}$ is discrete and $S \subset \mathbb{R}$. However, unlike Hilbert spaces, there are a lot of other structures that a Banach space may have, and a big problem for mathematicians has been to classify all of them. We will not consider that problem here, but we will look at some properties that all Banach spaces must have. The most important concept we will consider is the *adjoint* of an operator between two Banach spaces.

Operators mapping a vector space into \mathbb{R} or \mathbb{C} will be very relevant in this section. They are typically called *functionals*. Especially in section 4.3, the elements of $B(X, \mathbb{C})$ are central, i.e. the bounded linear functionals on X . Note that it is also normal to require X to be Banach or Hilbert, or that the functional is linear.

4.1 The operator norm

Recall that there are many vector spaces whose elements are functions or operators. An important one is the space $B(X, Y)$, defined in section 3.1, where X and Y are normed spaces. It has a well-known norm, typically called the *operator norm*, which can be defined equivalently in any of the following three ways:

$$\|T\|_{X \rightarrow Y} := \sup_{x \neq 0} \frac{\|Tx\|_Y}{\|x\|_X} = \sup_{\|x\|_X=1} \|Tx\|_Y = \sup_{\|x\|_X \leq 1} \|Tx\|_Y, \quad \forall T \in B(X, Y).$$

Equivalence of the three definitions follows from the fact that $\|\alpha x\|_X = |\alpha| \cdot \|x\|_X$, where $\alpha \in \mathbb{C}$ and $x \in X$, which is one of the axioms for norms. $\|T\|_{X \rightarrow Y}$ is simply the smallest bound for T . In all three expressions, we have skipped mentioning that we take supremum over $x \in X$. We will continue skipping that throughout this thesis, since it saves space under the supremum sign. It will always be clear anyway what space x is taken from. Also, we will normally drop the subscripts for the operator norm. However, if the operator may have different domains or co-domains, we now have a way of specifying them. Note that with this norm, lemma 3.3 (i) tells us that $\|SR\|_{X \rightarrow Z} \leq \|R\|_{X \rightarrow Y} \cdot \|S\|_{Y \rightarrow Z}$ for compositions of bounded operators.

An important property of $B(X, Y)$ is that it is a Banach space, assuming that Y is. This is our first theorem in this section. Note that if we want to take the limit of a real sequence, but we do not know whether it exists, it is sometimes convenient to replace \lim by \limsup . That always exists, and it coincides with the limit if it exists. We will use that trick a few times in this thesis, and the proof of the following theorem is the first time.

Theorem 4.1. *If X is a normed space and Y is a Banach space, then $B(X, Y)$ is complete w.r.t. the operator norm.*

Proof. Let $\{A_k\}_{k \in \mathbb{N}} \subset B(X, Y)$ be a Cauchy-sequence. Fix an $x \in X$, pick an $\epsilon > 0$ and find $N \in \mathbb{N}$ s.t. $\|A_m - A_n\| \leq \frac{\epsilon}{\|x\|_X}$ whenever $m, n \geq N$. Then,

$$\|A_mx - A_nx\|_Y = \|(A_m - A_n)x\|_Y \leq \|A_m - A_n\| \cdot \|x\|_X \leq \epsilon, \forall m, n \geq N,$$

showing that $\{A_kx\}_{k \in \mathbb{N}} \subset Y$ is also Cauchy. Since Y is complete, it converges to some element of Y , which we will denote by Ax . Letting $x \in X$ vary, this defines an operator $A : X \rightarrow Y$. We need to show that (i) A is bounded, (ii) A is linear and (iii) $\{A_k\}_{k \in \mathbb{N}}$ converges to A in the operator norm.

(i) Knowing that Cauchy-sequences are bounded, let $K > 0$ be an upper bound for the set $\{\|A_k\|, k \in \mathbb{N}\}$. Since norms are continuous, we have:

$$\|Ax\|_Y = \left\| \lim_{k \rightarrow \infty} (A_kx) \right\|_Y = \lim_{k \rightarrow \infty} \|A_kx\|_Y \leq \limsup_{k \rightarrow \infty} \|A_k\| \cdot \|x\|_X \leq K\|x\|_X, \forall x \in X.$$

$$\begin{aligned} \text{(ii) } A(\alpha x + \beta y) &= \lim_{k \rightarrow \infty} (A_k(\alpha x + \beta y)) = \lim_{k \rightarrow \infty} (\alpha A_kx + \beta A_ky) \\ &= \alpha \lim_{k \rightarrow \infty} (A_kx) + \beta \lim_{k \rightarrow \infty} (A_ky) = \alpha Ax + \beta Ay, \forall \alpha, \beta \in \mathbb{C}, \forall x, y \in X. \end{aligned}$$

(iii) As before, pick an $\epsilon > 0$, but this time, find $N \in \mathbb{N}$ s.t. $\|A_m - A_n\| \leq \epsilon$ whenever $m, n \geq N$. For any $x \in X$ with $\|x\|_X = 1$, we have:

$$\begin{aligned} \|(A - A_n)x\|_Y &= \|Ax - A_nx\|_Y = \left\| \lim_{m \rightarrow \infty} (A_mx) - A_nx \right\|_Y = \lim_{m \rightarrow \infty} \|A_mx - A_nx\|_Y \\ &= \lim_{m \rightarrow \infty} \|(A_m - A_n)x\|_Y \leq \lim_{m \rightarrow \infty} \|A_m - A_n\| \cdot \|x\|_X \leq \epsilon, \forall n \geq N. \end{aligned}$$

Hence, $\|A - A_n\| = \sup_{\|x\|_X=1} \|(A - A_n)x\|_Y \leq \epsilon, \forall n \geq N$, i.e. $A_n \rightarrow A$ as $n \rightarrow \infty$.

□

4.2 Fundamental results of Banach theory

Now, we will get a brief overview of the most important theorems in Banach theory. They are not so crucial in this thesis, so we will call them lemmas. However, two of them lead to other theorems that we will use later.

Lemma 4.2 (Banach-Steinhaus' theorem; [Sc02], p. 71). *Let X and Y be normed spaces, where X is Banach, and let $F \subseteq B(X, Y)$ be a family of operators.*

$$\text{If } \sup_{T \in F} \|Tx\|_Y < \infty, \forall x \in X, \text{ then } \sup_{T \in F} \|T\|_{X \rightarrow Y} < \infty.$$

Banach-Steinhaus' theorem is also referred to as the *uniform boundedness principle*, since it states that any pointwise bounded subset of $B(X, Y)$ is also uniformly bounded. Specifically, if $\{T_k\}_{k \in \mathbb{N}} \subset B(X, Y)$ is a pointwise bounded sequence converging to some $T \in B(X, Y)$, then by Banach-Steinhaus' theorem,

$$\|Tx\|_Y = \lim_{k \rightarrow \infty} \|T_k x\|_Y \leq \limsup_{k \rightarrow \infty} \|T_k\|_{X \rightarrow Y} \|x\| \leq \sup_{k \in \mathbb{N}} \|T_k\|_{X \rightarrow Y} \|x\|, \quad \forall x \in X.$$

Thus, T is also bounded, and $\sup_{k \in \mathbb{N}} \|T_k\|_{X \rightarrow Y}$ is a bound for T . In particular, this is useful to study the partial sum operator, but we will not do that here.

Lemma 4.3 (Hahn-Banach theorem; [Sc02], p. 148). *Let X be a vector space, let $V \subset X$ be a subspace, and let $p: X \rightarrow \mathbb{R}$ be a functional satisfying:*

- (i) $p(x + y) \leq p(x) + p(y)$, $\forall x, y \in X$.
- (ii) $p(\alpha x) = |\alpha| p(x)$, $\forall \alpha \in \mathbb{C}$, $\forall x \in X$.

Also, let $f: V \rightarrow \mathbb{C}$ be a linear functional satisfying $\operatorname{Re}(f) \leq p$ on V . Then, f can be extended to a linear functional on X satisfying $|f| \leq p$ on X .

Functionals with properties (i) and (ii) are called *sublinear*. Note that norms are examples of a sublinear functionals. We will take advantage of this observation to prove the following result.

Corollary 4.4 ([Sc02], p. 36). *Given a normed space X and a non-zero $x_0 \in X$, there exists a functional $f \in B(X, \mathbb{R})$ s.t. $\|f\|_{X \rightarrow \mathbb{R}} = 1$ and $f(x_0) = \|x_0\|_X$.*

Proof. Let $V := \{\alpha x_0, \alpha \in \mathbb{C}\} \subseteq X$, and define $g \in B(V, \mathbb{R})$ by the expression

$$g(\alpha x_0) := |\alpha| \cdot \|x_0\|_X, \quad \forall \alpha \in \mathbb{C}.$$

Then, $|g(\alpha x_0)| = |\alpha| \cdot \|x_0\|_X = \|\alpha x_0\|_X$, showing that $\|g\|_{V \rightarrow \mathbb{R}} = 1$. Also, the functional $p := \|\cdot\|_X: X \rightarrow \mathbb{R}$ is sublinear and satisfies $\operatorname{Re}(g) \leq |g| \leq p$ on V . Hence, by the Hahn-Banach theorem, g has a linear extension $f: X \rightarrow \mathbb{R}$ satisfying $|f| \leq p$ on X . Since $p(x)$ is just the norm of x , the last inequality tells us that f is bounded, and that $\|f\|_{X \rightarrow \mathbb{R}} \leq 1$. Also, since f is an extension of g , we have $\|f\|_{X \rightarrow \mathbb{R}} \geq \|g\|_{V \rightarrow \mathbb{R}} = 1$. We conclude that $\|f\|_{X \rightarrow \mathbb{R}} = 1$, and of course, $f(x_0) = g(x_0) = \|x_0\|_X$. \square

Lemma 4.5 (Open mapping theorem; [Sc02], p. 71). *Let X and Y be Banach spaces. If $T \in B(X, Y)$ is surjective, and if $U \subseteq X$ is open, then $T[U] \subseteq Y$ is open.*

Mappings with this property are typically called *open*, which is the reason for the name of the lemma. It is well-known that continuity can be defined topologically by the condition that the preimage $T^{-1}[V] \subset X$ is open whenever $V \subset Y$ is open. In that sense, openness is simply the converse of continuity. From this remark, it is obvious that if an operator T is both open and continuous, so is T^{-1} if it exists. Specifically, this holds for any bijective $T \in B(X, Y)$, since T is continuous by lemma 3.2 (iii). This gives us the following result.

Theorem 4.6. *Let X and Y be Banach spaces. If $T \in B(X, Y)$ is bijective, then $T^{-1} \in B(Y, X)$.*

Proof.

$$T^{-1}(\alpha Tx + \beta Ty) = T^{-1}T(\alpha x + \beta y) = \alpha x + \beta y = \alpha T^{-1}(Tx) + \beta T^{-1}(Ty), \forall \alpha, \beta \in \mathbb{C},$$

$\forall x, y \in X$, showing linearity. As remarked, T^{-1} is continuous by the open mapping theorem, hence bounded by lemma 3.2 (iii). \square

4.3 Dual spaces

With theorem 4.1 established, we are ready to encounter a special class of bounded linear operators, namely those having \mathbb{C} as their co-domain. That is, we will introduce the spaces of bounded linear functionals, specifically on Banach spaces.

Definition 4.7. *If X is a normed space, then $X^* := B(X, \mathbb{C})$ is called the **dual space** of X .*

If $X^{**} = X$, up to isomorphism, we say that X is *reflexive*. Since \mathbb{C} is complete, theorem 4.1 ensures us that X^* is complete as well. This already tells us that might not be reflexive, which we would often expect when using the name *dual*. We will see later that Hilbert spaces are reflexive, and our next example tells us that it is also the case for L^p -spaces when $p > 1$. On the other hand, $L^1(\mathbb{R})$ is not even the dual of a normed space, even though it is complete! We will not prove that, but it is good to be aware of the fact that such cases exist.

Example 4.8. *Given $S \subseteq \mathbb{R}$ and $p \in (1, \infty)$, find $q \in (1, \infty)$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$. Let $g: \mathbb{R} \rightarrow \mathbb{C}$ be a function, and consider the functional $\Phi_g: L^p(S) \rightarrow \mathbb{C}$ given by*

$$\Phi_g f = \int_S f(x) \overline{g(x)} dx, \forall f \in L^p(S).$$

Let us check whether Φ_g is even well-defined. For any $f \in L^p(S)$, by Hölder's inequality,

$$|\Phi_g f| \leq \int_S |f(x)g(x)| dx \leq \left(\int_S |f(x)|^p \right)^{\frac{1}{p}} \left(\int_S |g(x)|^q \right)^{\frac{1}{q}} = \|f\|_p \cdot \|g\|_q.$$

Hence, Φ_g is indeed well-defined, as long as $g \in L^q(S)$. Our calculation also shows that Φ_g is bounded, with $\|\Phi_g\| \leq \|g\|_q$. It is easily verified that Φ_g is linear as well, so $\Phi_g \in B(L^p(S), \mathbb{C}) = (L^p(S))^$. In fact, it can be shown that every element of $(L^p(S))^*$ is equal to Φ_g for some unique $g \in L^q(S)$. Hence, up to isomorphism, we conclude that $(L^p(S))^* = L^q(S)$. In particular, this shows that $L^p(S)$ is reflexive.*

4.4 The adjoint operator

We start with two normed spaces, X and Y , and an operator $T \in B(X, Y)$. Given any $y^* \in Y^*$, consider the composition y^*T . Note that it maps X into \mathbb{C} . Also, since it is composed of two bounded linear operators, it must itself be a bounded linear operator by proposition 3.4. Hence, we conclude that $y^*T \in B(X, \mathbb{C}) = X^*$. Letting $y^* \in Y^*$ vary, this observation allows us to define a very useful operator mapping Y^* into X^* .

Definition 4.9. *Let X and Y be normed spaces. If $T \in B(X, Y)$, then the **adjoint** of T is the operator $T^*: Y^* \rightarrow X^*$ given by $T^*y^* = y^*T, \forall y^* \in Y^*$.*

Theorem 4.10. *Let X and Y be normed spaces. If $T \in B(X, Y)$, then $T^* \in B(Y^*, X^*)$, and $\|T^*\| = \|T\|$.*

Proof. $T^*(\alpha x^* + \beta y^*) = (\alpha x^* + \beta y^*)T = \alpha x^*T + \beta y^*T = \alpha T^*x^* + \beta T^*y^*$, $\forall \alpha, \beta \in \mathbb{C}, \forall x^*, y^* \in Y^*$. Thus, T^* is linear. Also,

$$\|T^*y^*\|_{X^*} = \|y^*T\|_{X^*} \leq \|y^*\|_{Y^*} \cdot \|T\|, \forall y^* \in Y^*$$

by lemma 3.3 (i), showing that T^* is bounded and that $\|T^*\| \leq \|T\|$.

For the converse inequality, pick an $x \in X$ with $\|x\|_X = 1$. By corollary 4.4, there exists a $y^* \in B(Y, \mathbb{R}) \subset Y^*$ satisfying $\|y^*\|_{Y^*} = 1$ and $y^*(Tx) = \|Tx\|_Y$. Hence, we get:

$$\begin{aligned} \|Tx\|_Y = |y^*Tx| &\leq \|y^*T\|_{X^*} \cdot \|x\|_X = \|T^*y^*\|_{X^*} \cdot \|x\|_X \leq \|T^*\| \cdot \|y^*\|_{Y^*} \cdot \|x\|_X \\ &= \|T^*\| \cdot \|x\|_X. \end{aligned}$$

This shows that $\|T\| \leq \|T^*\|$, so we conclude that $\|T^*\| = \|T\|$. \square

We finish this subsection by giving two results that we will use later.

Theorem 4.11. *Given Banach spaces X and Y , let $T \in B(X, Y)$.*

(i) *T is injective with closed range if and only if it is bounded below ([Sc02], p. 67).*

(ii) *T is surjective if and only if T^* is bounded below ([Ru91], p. 100).*

Proof of (i). Assume that $K > 0$ is a lower bound for T . We have already seen in lemma 3.2 that T is injective. Pick a sequence $\{y_k\}_{k \in \mathbb{N}} \subset T[X]$ converging to some $y \in Y$, and find a sequence $\{x_k\}_{k \in \mathbb{N}}$ s.t. $y_k = Tx_k$ for $k \in \mathbb{N}$. Pick an $\epsilon > 0$, and find $N \in \mathbb{N}$ s.t. $\|y_k - y_l\|_Y \leq \epsilon K$ whenever $k, l \geq N$. This exists since convergent sequences are Cauchy. Now,

$$\|x_k - x_l\|_X \leq \frac{1}{K} \|T(x_k - x_l)\|_Y = \frac{1}{K} \|Tx_k - Tx_l\|_Y \leq \epsilon, \forall k, l \geq N.$$

Hence, $\{x_k\}_{k \in \mathbb{N}}$ is Cauchy, so it converges to some $x \in X$. Thus,

$$y = \lim_{k \rightarrow \infty} (Tx_k) = T(\lim_{k \rightarrow \infty} x_k) = Tx \in T[X].$$

This shows that $T[X]$ is closed.

Conversely, assume that T is injective with closed range. By injectivity, $T: X \rightarrow T[X]$ is bijective. Any Cauchy-sequence in $T[X]$ converges to some $y \in Y$, and since T has closed range, $y \in T[X]$. Hence, $T[X]$ is complete, so lemma 4.6 tells us that $T^{-1}: T[X] \rightarrow X$ is continuous. Hence,

$$\|x\|_X = \|T^{-1}Tx\|_X \leq \|T^{-1}\|_{T[X] \rightarrow X} \cdot \|Tx\|,$$

showing that $\|T^{-1}\|_{T[X] \rightarrow X}$ is a lower bound for T . \square

Using theorem 4.11, we might now be able to use boundedness below to show that an operator between Banach spaces X and Y is surjective and/or injective. Given $T \in B(X, Y)$, if we can find constants $K_1, K_2 > 0$ satisfying

$$\|Tx\|_Y \geq K_1\|x\|_X, \forall x \in X$$

and

$$\|T^*y^*\|_{X^*} \geq K_2\|y^*\|_{Y^*}, \forall y^* \in Y^*,$$

it will prove that T is bijective. This is something we will take advantage of when we study the frame operator in section 6.3.

4.5 Absolutely convergent series in Banach spaces

A series in a normed space X is said to converge *absolutely* if the corresponding series of norms converges in \mathbb{R} . That is, absolute convergence of $\sum_{k \in \mathbb{N}} x_k$ means that $\sum_{k \in \mathbb{N}} \|x_k\|_X$ converges. If a series converges to the same limit regardless of the ordering, it is said to converge *unconditionally*. These concepts will not be that important to us in general, but we will sometimes take advantage of an important theorem for the case that $X = \mathbb{C}$. We will, however, prove it for a general Banach space.

Lemma 4.12 ([Wik1]). *Let $\{x_k\}_{k \in \mathbb{N}}$ be a sequence in a Banach space X . If the series $\sum_{k \in \mathbb{N}} x_k$ converges absolutely, it converges unconditionally.*

Proof. Pick an $\epsilon > 0$. Since $\{\sum_{k=1}^n \|x_k\|_X\}_{n \in \mathbb{N}} \subset \mathbb{C}$ is Cauchy,

$$\left\| \sum_{k=1}^n x_k - \sum_{k=1}^m x_k \right\|_X = \left\| \sum_{k=m+1}^n x_k \right\|_X \leq \sum_{k=m+1}^n \|x_k\|_X = \left| \sum_{k=1}^n \|x_k\|_X - \sum_{k=1}^m \|x_k\|_X \right| \leq \epsilon$$

for sufficiently large $n, m \in \mathbb{N}$. Thus, $\{\sum_{k=1}^n x_k\}_{n \in \mathbb{N}} \subset X$ is also Cauchy, so it converges to some $x \in X$.

Now, find $N_1, N_2 \in \mathbb{N}$ satisfying:

$$\left\| x - \sum_{k=1}^n x_k \right\|_X \leq \frac{\epsilon}{2}, \forall n \geq N_1.$$

$$\sum_{k=n}^{\infty} \|x_k\|_X \leq \frac{\epsilon}{2}, \forall n \geq N_2.$$

N_2 exists because the tail of any convergent series converges to 0. Let $M := \max\{N_1, N_2\}$. Pick a reordering of \mathbb{N} , i.e. a bijective function $\sigma: \mathbb{N} \rightarrow \mathbb{N}$, and define:

$$J := \{\sigma^{-1}(1), \sigma^{-1}(2), \dots, \sigma^{-1}(M)\}.$$

$$N := \max(J).$$

$$I_n := \{1, 2, \dots, n\}, \forall n \in \mathbb{N}.$$

Then, for any $n \geq N$, we have:

$$\begin{aligned}
& \left\| x - \sum_{k=1}^n x_{\sigma(k)} \right\|_X = \left\| x - \sum_{k \in J} x_{\sigma(k)} - \sum_{k \in I_n \setminus J} x_{\sigma(k)} \right\|_X \\
& \leq \left\| x - \sum_{k \in J} x_{\sigma(k)} \right\|_X + \left\| \sum_{k \in I_n \setminus J} x_{\sigma(k)} \right\|_X = \left\| x - \sum_{j=1}^M x_j \right\|_X + \left\| \sum_{j=M+1}^n x_j \right\|_X \\
& \leq \left\| x - \sum_{j=1}^M x_j \right\|_X + \sum_{j=M+1}^n \|x_j\|_X \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

□

Lemma 4.12 is particularly useful in \mathbb{R} and \mathbb{C} , since it implies that if a series of non-negative real numbers converges, then the order of the terms does not matter. In particular, whether a given sequence is in $l^p(\mathbb{N})$ is independent of the order. This will be useful in section 6, since where adding elements to an l^2 -sequence will be relevant.

In \mathbb{R} , it is well-known that the converse is also the case, i.e. that unconditionally convergent series are absolutely convergent. That is not true for general Banach spaces. For example, if $\{e_k\}_{k \in \mathbb{N}}$ is an orthonormal basis for a Hilbert space H , we will see later that $\sum_{k \in \mathbb{N}} c_k e_k$ converges if and only if $\{c_k\}_{k \in \mathbb{N}} \in l^2(\mathbb{N})$. The convergence turns out to be unconditional ([Ch10], p. 81). However, absolute convergence would mean that $\sum_{k \in \mathbb{N}} \|c_k e_k\|_H = \sum_{k \in \mathbb{N}} |c_k| < \infty$, i.e. that $\{c_k\}_{k \in \mathbb{N}} \in l^1(\mathbb{N})$. Since not all l^2 -sequences are l^1 -sequences, this disproves the converse of lemma 4.12. Note, however, that this discussion tells us something else: If $\{c_k\}_{k \in \mathbb{N}} \in l^1(\mathbb{N})$, then $\sum_{k \in \mathbb{N}} c_k e_k$ is absolutely convergent, hence convergent, so $\{c_k\}_{k \in \mathbb{N}} \in l^2(\mathbb{N})$. That is, $l^2(\mathbb{N}) \subset l^1(\mathbb{N})$. This is no surprise, since the terms of a convergent series must tend to zero, so their squares tend to zero even more rapidly.

Example 4.13. We want to compute the sum $\sum_{k \in \mathbb{Z}} \frac{1}{(k - \frac{1}{2})^2}$ by taking advantage of the following well-known fact:

$$\sum_{n \in \mathbb{N}} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Since all the terms are positive, we do not need to worry about the order. Hence, we can split the integers into the positive and the non-negative ones, and split them again into even and odd numbers, to get:

$$\begin{aligned}
\sum_{n \in \mathbb{N}} \frac{1}{(2n-1)^2} &= \sum_{n \in \mathbb{N}} \frac{1}{n^2} - \sum_{n \in \mathbb{N}} \frac{1}{(2n)^2} = \sum_{n \in \mathbb{N}} \frac{1}{n^2} - \frac{1}{4} \sum_{n \in \mathbb{N}} \frac{1}{n^2} = \frac{\pi^2}{6} - \frac{1}{4} \cdot \frac{\pi^2}{6} = \frac{\pi^2}{8}. \\
\sum_{n \in \mathbb{N}} \frac{1}{(n - \frac{1}{2})^2} &= \sum_{n \in \mathbb{N}} \frac{1}{(\frac{1}{2}(2n-1))^2} = 4 \sum_{n \in \mathbb{N}} \frac{1}{(2n-1)^2} = 4 \cdot \frac{\pi^2}{8} = \frac{\pi^2}{2}. \\
\sum_{k \in \mathbb{Z}} \frac{1}{(k - \frac{1}{2})^2} &= 2 \sum_{n \in \mathbb{N}} \frac{1}{(n - \frac{1}{2})^2} = 2 \cdot \frac{\pi^2}{2} = \pi^2.
\end{aligned}$$

5 Hilbert spaces

A Hilbert space, which we will usually denote by H , is a Banach space whose norm is induced by an inner product. As is well-known, the inner product should be linear in the first slot and anti-linear in the second slot:

(i)

$$\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle, \quad \forall \alpha, \beta \in \mathbb{C}, \quad \forall x, y, z \in H.$$

(ii)

$$\langle z, \alpha x + \beta y \rangle = \bar{\alpha} \langle z, x \rangle + \bar{\beta} \langle z, y \rangle, \quad \forall \alpha, \beta \in \mathbb{C}, \quad \forall x, y, z \in H.$$

Property (i) is one of the axioms for inner products, and combining it with the axiom that $\langle x, y \rangle = \overline{\langle y, x \rangle}$, $\forall x, y \in H$ yields property (ii). By continuity of the inner product, which follows from Cauchy-Schwarz' inequality ([Ch10], p. 62), the two properties above are also satisfied for infinite sums. That is, if $\sum_{k \in \mathbb{N}} c_k x_k$ converges in H , then we have, for any $y \in H$:

(i)

$$\sum_{k \in \mathbb{N}} c_k \langle x_k, y \rangle \text{ converges in } \mathbb{C} \text{ to } \left\langle \sum_{k \in \mathbb{N}} c_k x_k, y \right\rangle.$$

(ii)

$$\sum_{k \in \mathbb{N}} \bar{c}_k \langle y, x_k \rangle \text{ converges in } \mathbb{C} \text{ to } \left\langle y, \sum_{k \in \mathbb{N}} c_k x_k \right\rangle.$$

Given any $n \in \mathbb{N}$, it is well known that \mathbb{R}^n , \mathbb{C}^n , $l^2(\mathbb{N})$ and $L^2(\mathbb{R})$ are Hilbert spaces w.r.t. their usual inner products. The same is true for $l^2(\Lambda)$ and $L^2(S)$ for any discrete $\Lambda \subset \mathbb{R}$ and any $S \subset \mathbb{R}$. We will later encounter Paley-Wiener spaces, which also turn out to be Hilbert spaces. It turns out that every finite-dimensional Hilbert space is isomorphic to \mathbb{C}^n for some $n \in \mathbb{N}$, while every infinite-dimensional, *separable* Hilbert space, to be defined in section 5.2, is isomorphic to $l^2(\mathbb{N})$ ([Ch10], p. 82). However, firstly, there are Hilbert spaces whose dimension is uncountable, in which case it is not separable. Secondly, isomorphic Hilbert spaces might be convenient to treat in different ways, depending on their applications. And in fact, for some of our purposes, it is more convenient to use anti-isomorphisms than isomorphisms, since we can use inner products for that!

5.1 The adjoint of operators between Hilbert spaces

Let H and K be Hilbert spaces. If $T: H \rightarrow K$ is a bounded linear operator, we know that we can define the adjoint operator $T^*: K^* \rightarrow H^*$. What do H^* and K^* look like? This question turns out to have a very simple answer, as we will now show. We will take advantage of the following property of norms and inner products:

$$\|y\|_H = \sup_{\|x\|_H=1} |\langle x, y \rangle|, \quad \forall y \in H. \quad (2)$$

It follows immediately from Cauchy-Schwarz' inequality and the fact that setting $x = \frac{y}{\|y\|_H}$ gives equality.

Lemma 5.1. Any Hilbert space H is anti-isomorphic to its dual space H^* , and the function $U_H: H \rightarrow H^*$ defined by $U_H y = \langle \cdot, y \rangle$, $\forall y \in H$ is an isometric anti-isomorphism.

Proof. It is easily verified that $U_H y$ is a bounded linear functional for any $y \in H$, i.e. that $U_H y \in H^*$. By Riesz' representation theorem ([Ch10], p. 70), any bounded linear functional on H can be expressed as a left inner product with a fixed $y \in H$. Thus, U_H is surjective. Since Riesz' representation theorem also states that the $y \in H$ is unique, we conclude that U_H is injective as well, hence bijective. It remains to be shown that U_H (i) is anti-linear and (ii) preserves norms.

$$(i) U_H(\alpha y + \beta z) = \langle \cdot, \alpha y + \beta z \rangle = \bar{\alpha} \langle \cdot, y \rangle + \bar{\beta} \langle \cdot, z \rangle = \bar{\alpha} U_H y + \bar{\beta} U_H z, \forall \alpha, \beta \in \mathbb{C}, \forall y, z \in H.$$

$$(ii) \|U_H y\|_{H^*} = \sup_{\|x\|_H=1} |(U_H y)x| = \sup_{\|x\|_H=1} |\langle x, y \rangle| = \|y\|_H, \forall y \in H$$

by (2). □

Lemma 5.1 allows us to interpret the adjoint operator in a very convenient way for Hilbert spaces. If $T: H \rightarrow K$ is a bounded linear operator between two Hilbert spaces, then T^* maps K into H , up to (anti)-isomorphism. The following theorem tells us what T^* looks like with that interpretation.

Theorem 5.2. Given two Hilbert spaces H and K , and a $T \in B(H, K)$, define the operator $S := U_H^{-1} T^* U_K: K \rightarrow H$. Then, S is the unique operator satisfying

$$\langle Tx, y \rangle_K = \langle x, Sy \rangle_H, \forall x \in H, \forall y \in K. \quad (3)$$

Also, $S \in B(K, H)$ and $\|S\| = \|T\|$.

Proof. For any $y \in K$, we have:

$$(T^* U_K)y = T^*(U_K y) = T^* \langle \cdot, y \rangle_K = \langle T(\cdot), y \rangle_K.$$

$$(U_H S)y = U_H(Sy) = \langle \cdot, Sy \rangle_H.$$

Since $U_H S = T^* U_K$ by definition of S , this proves (3). Uniqueness follows from the well-known fact that if $\langle x, y \rangle_H = \langle x, z \rangle_H$ for all $x \in H$, then $y = z$ ([Ch10], p. 70). Since S is a composition of bounded linear operators, we must have $S \in B(K, H)$ by proposition 3.4. Now, taking advantage of lemma 3.3 (i) and the fact that U_H, U_K and their inverses are isometries, we have:

$$\|S\| = \|U_H^{-1} T^* U_K\| \leq \|U_H^{-1}\| \cdot \|T^*\| \cdot \|U_K\| = \|T^*\|.$$

$$\|T^*\| = \|U_H S U_K^{-1}\| \leq \|U_H\| \cdot \|S\| \cdot \|U_K^{-1}\| = \|S\|.$$

The two inequalities above, together with theorem 4.10, tell us that $\|S\| = \|T^*\| = \|T\|$. □

For Hilbert spaces, it is normal to define T^* to be the operator satisfying $\langle Tx, y \rangle_K = \langle x, T^*y \rangle_H$. That is, to define T^* to be the operator we called S in theorem 5.2. Accordingly, we will simply denote S by T^* from now on. It will always be clear which operator we are talking about. Note that since both U_H and U_K are bijective, anyone of the two operators called T^* will be surjective/injective if and only if the other one is. Note that to identify T^* , we just need to find an operator satisfying (3), since T^* is UNIQUELY determined by that equation.

Some properties of the adjoint operator are given in the following proposition. The identity operator is denoted by I .

Proposition 5.3. *Given Hilbert spaces H, J and K , let $T \in B(H, J)$ and $S \in B(J, K)$ be operators. Then, we have:*

(i) $(ST)^* = T^*S^*$.

(ii) $(T^*)^* = T$.

(iii) If T is invertible, so is T^* , and $(T^*)^{-1} = (T^{-1})^*$.

Proof.

$$(i) \langle T^*S^*x, y \rangle_H = \langle S^*x, Ty \rangle_J = \langle x, STy \rangle_K, \forall x \in K, \forall y \in H.$$

$$(ii) \langle Tx, y \rangle_J = \langle x, T^*y \rangle_H = \overline{\langle T^*y, x \rangle_H} = \overline{\langle y, (T^*)^*x \rangle_J} = \langle (T^*)^*x, y \rangle_J, \forall x \in H, \forall y \in J.$$

(iii) By theorem 4.6, $T^{-1} \in B(Y, X)$, so its adjoint is defined. Also, by (i),

$$T^*(T^{-1})^* = (T^{-1}T)^* = I^* = I.$$

$$(T^{-1})^*T^* = (TT^{-1})^* = I^* = I.$$

□

5.2 Complete sequences

Definition 5.4. *Let X be a normed space. A sequence in X is said to be **complete** if finite linear combinations of its elements can approximate any element of X . If a complete sequence in X exists, then X is said to be **separable**.*

A basis is an example of a complete sequence. In that case, we can improve our approximations by just adding more coefficients, without changing the ones we have already used. We cannot do that in every separable space, though, since they might not possess any basis. Specifically, a complete sequence is not necessarily a basis, even if it is linearly independent. For example, it is well-known that if X is the space of continuous functions on a fixed closed interval, equipped with the sup-norm, then $\{1, (\cdot), (\cdot)^2, (\cdot)^3, \dots\}$ is a complete sequence in X that does not form a basis for X . But if a basis exists, it is certainly complete, so existence of a basis for a normed space definitely guarantees separability.

In Hilbert spaces, we say that two vectors are *orthogonal* if their inner product vanishes. This concept is very useful for a lot of reasons, and we will look at some of them in the next section. For now, we will see how it can be used to give an alternative definition of complete sequences.

Theorem 5.5 ([Ch10], p. 71). *Let H be a Hilbert space. A sequence $\{v_k\}_{k \in \mathbb{N}} \subset H$ is complete if and only if the only element of H orthogonal to all the terms is the zero-vector.*

We have pointed out that every basis in a normed space is complete. But if we have a basis and add extra elements, we obviously still have a complete sequence. Contrary to a basis, we do not require complete sequences to be linearly independent, so adding extra elements does no harm. However, if we assume linear independence, things are sometimes different. It is well-known that in n -dimensional Euclidean space, any linearly independent set of n vectors is a basis. A bigger set will not be linearly independent, while a smaller set will not be complete, so any complete, linearly independent set is a basis. What about separable, infinite-dimensional spaces?

5.3 Bessel's inequality and its converse

Definition 5.6. *Let $\{u_k\}_{k \in \mathbb{N}}$ be a sequence in a Hilbert space H . If there exists a $B > 0$ s.t.*

$$\sum_{k \in \mathbb{N}} |\langle x, u_k \rangle|^2 \leq B \|x\|^2, \quad \forall x \in H, \quad (4)$$

*then $\{u_k\}_{k \in \mathbb{N}}$ is called a **Bessel sequence** in H . B is called a **Bessel bound** for the sequence, and (4) is called **Bessel's inequality**.*

An important feature of Bessel sequences is that the inner products $\langle x, u_k \rangle$ always form an l^2 -sequence. After all, Bessel's inequality gives a finite upper bound for the l^2 -norm of that sequence. Less obvious is the fact that for Bessel sequences, any l^2 -sequence can be used as coefficients to get a convergent series.

Theorem 5.7 ([Ch10], p. 77). *If $\{u_k\}_{k \in \mathbb{N}}$ be a Bessel sequence in a Hilbert space H . If $\{c_k\}_{k \in \mathbb{N}} \in l^2(\mathbb{N})$, then $\sum_{k \in \mathbb{N}} c_k u_k$ converges in H .*

Proof. Pick a sequence $\{c_k\}_{k \in \mathbb{N}} \in l^2(\mathbb{N})$, and define

$$x_n := \sum_{k=1}^n c_k u_k, \quad \forall n \in \mathbb{N}.$$

Then, whenever $n > m$, using (2) and Hölder's inequality, we have:

$$\begin{aligned} \|x_n - x_m\|^2 &= \left\| \sum_{k=m+1}^n c_k u_k \right\|^2 = \sup_{\|x\|=1} \left| \langle x, \sum_{k=m+1}^n c_k u_k \rangle \right|^2 = \sup_{\|x\|=1} \left| \sum_{k=m+1}^n \overline{c_k} \langle x, u_k \rangle \right|^2 \\ &\leq \sup_{\|x\|=1} \sum_{k=m+1}^n |c_k|^2 \sum_{k=m+1}^n |\langle x, u_k \rangle|^2 \leq \sup_{\|x\|=1} \sum_{k=m+1}^{\infty} |c_k|^2 \sum_{k \in \mathbb{N}} |\langle x, u_k \rangle|^2 \\ &\leq \sup_{\|x\|=1} \sum_{k=m+1}^{\infty} |c_k|^2 \cdot B \|x\|^2 = B \sum_{k=m+1}^{\infty} |c_k|^2. \end{aligned}$$

Since the last sum is the tail of a convergent sequence, it becomes arbitrarily small for large $m \in \mathbb{N}$. Hence, $\{x_n\}_{k \in \mathbb{N}}$ is Cauchy, hence convergent. \square

Note that there might be convergent sequences whose coefficients do not form an l^2 -sequence. E.g. if $\{e_k\}_{k \in \mathbb{N}}$ is an orthonormal basis for H , then we will see in the next section that it is a Bessel sequence in H . Thus, $\{u_k\}_{k \in \mathbb{N}} := \{\frac{1}{k^2}e_k\}_{k \in \mathbb{N}}$ also satisfies Bessel's inequality. Clearly, $\{k\}_{k \in \mathbb{N}} \notin l^2(\mathbb{N})$, but still,

$$\sum_{k \in \mathbb{N}} k u_k = \sum_{k \in \mathbb{N}} k \frac{1}{k^2} e_k = \sum_{k \in \mathbb{N}} \frac{1}{k} e_k$$

converges, since $\{\frac{1}{k}\}_{k \in \mathbb{N}} \in l^2(\mathbb{N})$.

The converse of (4),

$$A\|x\|^2 \leq \sum_{k \in \mathbb{N}} |\langle x, u_k \rangle|^2, \quad \forall x \in H, \quad (5)$$

has no commonly used name, other than the converse of Bessel's inequality. Both of them, however, will be important for us when we get to *frames* in section 6. We notice that if an $x \in H$ is orthogonal to all the u_k , then the right side of (5) vanishes, showing that $x = 0$. That is, due to theorem 5.5, every system satisfying the converse of Bessel's inequality is complete. The converse is not true, though. E.g. if $\{e_k\}_{k \in \mathbb{N}}$ is an orthonormal basis for H , then $u_k := k e_k$ for $k \in \mathbb{N}$ defines another complete system. Still,

$$\sum_{k \in \mathbb{N}} |\langle e_n, u_k \rangle|^2 = \sum_{k \in \mathbb{N}} \frac{1}{k} |\langle e_n, e_k \rangle|^2 = \sum_{k \in \mathbb{N}} \frac{1}{k} \delta_{nk} = \frac{1}{n}$$

becomes arbitrarily small for large enough $n \in \mathbb{N}$. Since all the e_n have the same norm, this contradicts (5).

5.4 Orthonormal sequences

Orthonormal vectors in a Hilbert space are orthogonal vectors of norm 1. A system of orthonormal vectors will be abbreviated by ONS, and if it is also a basis, we will write ONB. A nice property of separable Hilbert spaces is that they always possess an ONB ([Ch10], p. 82). Not only do they always exist, but they also have some really nice properties. Before going specifically into ONBs, we will consider some more general properties of orthonormal systems.

Let $\{u_k\}_{k \in \mathbb{N}}$ be an ONS in a Hilbert space H . Then, for any $n \in \mathbb{N}$,

$$\left\| \sum_{k=1}^n c_k u_k \right\|^2 = \left\langle \sum_{k=1}^n c_k u_k, \sum_{l=1}^n c_l u_l \right\rangle = \sum_{k=1}^n \sum_{l=1}^n c_k \bar{c}_l \langle u_k, u_l \rangle = \sum_{k=1}^n \sum_{l=1}^n c_k \bar{c}_l \delta_{kl} = \sum_{k=1}^n |c_k|^2$$

for any sequence of coefficients. Since all the operators we are dealing with are continuous, the same must hold as $n \rightarrow \infty$, assuming that the linear combination actually converges. That is,

$$\left\| \sum_{k \in \mathbb{N}} c_k u_k \right\|^2 = \sum_{k \in \mathbb{N}} |c_k|^2 \quad (6)$$

whenever the linear combination converges in H . Hence, the sequence of coefficients must always be an l^2 -sequence.

Do ALL l^2 -sequences give us convergent linear combinations? The first part of the following theorem will give us an answer to that.

Theorem 5.8. *Let $\{u_k\}_{k \in \mathbb{N}}$ be an ONS in a Hilbert space H . Then,*

(i)

$$\sum_{k \in \mathbb{N}} |\langle x, u_k \rangle|^2 \leq \|x\|^2, \quad \forall x \in H.$$

(ii) $\sum_{k \in \mathbb{N}} c_k u_k$ converges in H if and only if $\{c_k\}_{k \in \mathbb{N}} \in l^2(\mathbb{N})$. In that case,

$$\left\| \sum_{k \in \mathbb{N}} c_k u_k \right\|^2 = \sum_{k \in \mathbb{N}} |c_k|^2.$$

Proof. (i) ([Mc07], p. 12) Given any $x \in H$, define $c_k := \langle x, u_k \rangle$, $\forall k \in \mathbb{N}$. For any $n \in \mathbb{N}$, we have:

$$\begin{aligned} \left\| x - \sum_{k=1}^n c_k u_k \right\|^2 &= \left\langle x - \sum_{k=1}^n c_k u_k, x - \sum_{k=1}^n c_k u_k \right\rangle \\ &= \|x\|^2 - \sum_{k=1}^n \overline{c_k} \langle x, u_k \rangle - \sum_{k=1}^n c_k \langle u_k, x \rangle + \left\| \sum_{k=1}^n c_k u_k \right\|^2 \\ &= \|x\|^2 - \sum_{k=1}^n \overline{c_k} c_k - \sum_{k=1}^n c_k \overline{c_k} + \sum_{k=1}^n |c_k|^2 = \|x\|^2 - \sum_{k=1}^n |c_k|^2, \end{aligned}$$

where we have taken advantage of the finite version of (6). Hence,

$$\sum_{k=1}^n |\langle x, u_k \rangle|^2 = \sum_{k=1}^n |c_k|^2 = \|x\|^2 - \left\| x - \sum_{k=1}^n c_k u_k \right\|^2 \leq \|x\|^2,$$

so letting $n \rightarrow \infty$ gives the desired result.

(ii) Assuming convergence, the equality is just (6), and it shows necessity. Since (i) is a special case of Bessel's inequality, theorem 5.7 shows sufficiency. \square

Property (i) shows, as pointed out in the proof, that any ONS is a Bessel sequence with Bessel bound 1. If the system is also a basis, we even have equality, as the next theorem shows.

Theorem 5.9. Let $\{e_k\}_{k \in \mathbb{N}}$ be an ONB for a Hilbert space H . Then, the following hold:

(i)

$$x = \sum_{k \in \mathbb{N}} \langle x, e_k \rangle e_k, \quad \forall x \in H.$$

(ii)

$$\|x\|^2 = \sum_{k \in \mathbb{N}} |\langle x, e_k \rangle|^2, \quad \forall x \in H.$$

Proof. (i) Pick an $x \in H$. Since $\{e_k\}_{k \in \mathbb{N}}$ is a basis for H , there exists a unique sequence $\{c_k\}_{k \in \mathbb{N}} \subset \mathbb{C}$ s.t. $x = \sum_{k \in \mathbb{N}} c_k e_k$. Taking inner products with the e_n , this yields:

$$\langle x, e_n \rangle = \left\langle \sum_{k \in \mathbb{N}} c_k e_k, e_n \right\rangle = \sum_{k \in \mathbb{N}} c_k \langle e_k, e_n \rangle = \sum_{k \in \mathbb{N}} c_k \delta_{kn} = c_n.$$

Hence,

$$x = \sum_{k \in \mathbb{N}} c_k e_k = \sum_{k \in \mathbb{N}} \langle x, e_k \rangle e_k.$$

$$(ii) \sum_{k \in \mathbb{N}} |\langle x, e_k \rangle|^2 = \sum_{k \in \mathbb{N}} \langle x, e_k \rangle \langle e_k, x \rangle = \left\langle \sum_{k \in \mathbb{N}} \langle x, e_k \rangle e_k, x \right\rangle = \langle x, x \rangle = \|x\|^2, \quad \forall x \in H.$$

□

Example 5.10. It is well-known that $E(\mathbb{Z})$ is an ONB for $L^2[-\frac{1}{2}, \frac{1}{2}]$ ([Ch10], p. 129). What about $L^2[-\sigma, \sigma]$ for a general $\sigma > 0$? We will show that $\frac{1}{\sqrt{2\sigma}} E(\frac{\mathbb{Z}+c}{2\sigma})$ is an ONB for any $c > 0$.

Firstly, pick an $f \in L^2[-\sigma, \sigma]$. Then,

$$g(x) := f(2\sigma x) e^{2\pi i c x}, \quad \forall x \in [-\frac{1}{2}, \frac{1}{2}]$$

defines a function $g \in L^2[-\frac{1}{2}, \frac{1}{2}]$. Find the unique sequence $\{c_k\}_{k \in \mathbb{Z}} \subset \mathbb{C}$ s.t.

$$g = \sum_{k \in \mathbb{Z}} c_k e^{2\pi i k(\cdot)}$$

in the L^2 -norm. Then, again in the L^2 -norm, we have:

$$f = g \left(\frac{1}{2\sigma}(\cdot) \right) e^{2\pi i c \frac{1}{2\sigma}(\cdot)} = \sum_{k \in \mathbb{Z}} c_k e^{2\pi i k \frac{1}{2\sigma}(\cdot)} e^{2\pi i c \frac{1}{2\sigma}(\cdot)} = \sum_{k \in \mathbb{Z}} c_k e^{2\pi i \frac{k+c}{2\sigma}(\cdot)}.$$

This shows that the arbitrary $f \in L^2[-\sigma, \sigma]$ can be uniquely expanded in $\{e^{2\pi i \frac{k+c}{2\sigma}(\cdot)}\}_{k \in \mathbb{Z}} = E(\frac{\mathbb{Z}+c}{2\sigma})$, so it is a basis for $L^2[-\sigma, \sigma]$.

Now, we need to calculate the inner products. If $k, l \in \mathbb{Z}$ and $k \neq l$, then

$$\begin{aligned} \langle e^{\frac{\pi}{\sigma}i(k+c)(\cdot)}, e^{\frac{\pi}{\sigma}i(l+c)(\cdot)} \rangle &= \int_{-\sigma}^{\sigma} e^{\frac{\pi}{\sigma}i(k+c)x} e^{-\frac{\pi}{\sigma}i(l+c)x} dx = \int_{-\sigma}^{\sigma} e^{\frac{\pi}{\sigma}i(k-l)x} dx \\ &= \frac{\sigma}{\pi i(k-l)} \left[e^{\frac{\pi}{\sigma}i(k-l)x} \right]_{x=-\sigma}^{x=\sigma} = \frac{\sigma(e^{\pi i(k-l)} - e^{-\pi i(k-l)})}{\pi i(k-l)} = \frac{\sigma((-1)^{k-l} - (-1)^{-(k-l)})}{\pi i(k-l)} = 0, \end{aligned}$$

since $k-l$ and $-(k-l)$ are either both even or both odd. Also,

$$\langle e^{\frac{\pi}{\sigma}i(k+c)(\cdot)}, e^{\frac{\pi}{\sigma}i(k+c)(\cdot)} \rangle = \int_{-\sigma}^{\sigma} e^{\frac{\pi}{\sigma}i(k+c)x} e^{\frac{\pi}{\sigma}i(k+c)x} dx = \int_{-\sigma}^{\sigma} dx = 2\sigma, \quad \forall k \in \mathbb{Z}.$$

Hence, if we normalize by dividing by $\frac{1}{\sqrt{2\sigma}}$, we get an ONB.

In fact, since all these exponential functions are periodic with period 2σ , both the expansion and the integrals must give the same result if we replace $[-\sigma, \sigma]$ by any interval S of length 2σ . Hence, we conclude that $\frac{1}{\sqrt{2\sigma}}E(\frac{\mathbb{Z}+c}{2\sigma})$ is an ONB for $L^2(S)$.

Example 5.11. Let $f(x) := 2\pi i x e^{\frac{\pi}{\sigma}ix}$, $\forall x \in [-\sigma, \sigma]$. Then, $f \in L^2[-\sigma, \sigma]$, which has the ONB $\frac{1}{\sqrt{2\sigma}}E(\frac{\mathbb{Z}}{2\sigma})$. We have, for $k \in \mathbb{Z}$, using integration by parts:

$$\begin{aligned} \langle f, e^{\frac{\pi}{\sigma}ik(\cdot)} \rangle &= \int_{-\sigma}^{\sigma} 2\pi i x e^{\frac{\pi}{\sigma}ix} e^{-\frac{\pi}{\sigma}ikx} dx = \int_{-\sigma}^{\sigma} 2\pi i x e^{-\frac{\pi}{\sigma}i(k-\frac{1}{2})x} dx \\ &= -\frac{2\sigma}{k-\frac{1}{2}} \int_{-\sigma}^{\sigma} x \left(\frac{d}{dx} e^{-\frac{\pi}{\sigma}i(k-\frac{1}{2})x} \right) dx \\ &= -\frac{2\sigma}{k-\frac{1}{2}} \left(\left[x e^{-\frac{\pi}{\sigma}i(k-\frac{1}{2})x} \right]_{x=-\sigma}^{x=\sigma} - \int_{-\sigma}^{\sigma} e^{-\frac{\pi}{\sigma}i(k-\frac{1}{2})x} dx \right) \\ &= -\frac{4\sigma^2}{k-\frac{1}{2}} \cdot \frac{e^{-\pi i(k-\frac{1}{2})} + e^{\pi i(k-\frac{1}{2})}}{2} - \frac{2\sigma^2}{\pi i(k-\frac{1}{2})^2} \left[e^{-\frac{\pi}{\sigma}i(k-\frac{1}{2})x} \right]_{x=-\sigma}^{x=\sigma} \\ &= -\frac{4\sigma^2 \cos(\pi(k-\frac{1}{2}))}{k-\frac{1}{2}} - \frac{4\sigma^2}{\pi(k-\frac{1}{2})^2} \cdot \frac{e^{-\pi i(k-\frac{1}{2})} - e^{\pi i(k-\frac{1}{2})}}{2i} \\ &= -\frac{4\sigma^2 \cos(\frac{\pi}{2}(2k-1))}{k-\frac{1}{2}} + \frac{4\sigma^2 \sin(\frac{\pi}{2}(2k-1))}{\pi(k-\frac{1}{2})^2}, \quad \forall k \in \mathbb{Z}. \end{aligned}$$

The cosine of any odd multiple of $\frac{\pi}{2}$ is zero, so the first term vanishes. The sine, on the other hand, oscillates between 1 and -1 , so we have $\sin(\frac{\pi}{2}(2k-1)) = (-1)^{k-1}$, $\forall k \in \mathbb{Z}$. Hence,

$$\langle f, e^{\frac{\pi}{\sigma}ik(\cdot)} \rangle = \frac{4\sigma^2(-1)^{k-1}}{\pi(k-\frac{1}{2})^2}, \quad \forall k \in \mathbb{Z},$$

so by theorem 5.9 (i),

$$f = \sum_{k \in \mathbb{Z}} \left\langle f, \frac{1}{\sqrt{2\sigma}} e^{\frac{\pi}{\sigma}ik(\cdot)} \right\rangle \frac{1}{\sqrt{2\sigma}} e^{\frac{\pi}{\sigma}ik(\cdot)} = \sum_{k \in \mathbb{Z}} \frac{2\sigma(-1)^{k-1}}{\pi(k-\frac{1}{2})^2} e^{\frac{\pi}{\sigma}ik(\cdot)}$$

in the L^2 -norm on $[-\sigma, \sigma]$. Since f is differentiable on $[-\sigma, \sigma]$ and has continuous derivative, this Fourier series converges to f also in a pointwise sense.

5.5 Riesz bases

We have seen that if we have an ONS, specifically an ONB, then the linear combinations that converge are exactly those with coefficients taken from l^2 . We will now look at another type of basis where that is the case. If H and K are Hilbert spaces, if $\{u_k\}_{k \in \mathbb{N}}$ is a Bessel sequence in H , and if $T \in B(H, K)$, then

$$\sum_{k \in \mathbb{N}} |\langle x, Tu_k \rangle_K|^2 = \sum_{k \in \mathbb{N}} |\langle T^*x, u_k \rangle_H|^2 \leq B \|T^*x\|_H^2 \leq B \|T^*\|^2 \|x\|_K^2, \quad \forall x \in K. \quad (7)$$

That is, bounded linear operators between Hilbert spaces map Bessel sequences into Bessel sequences. The basis we will consider is the image of an ONB under such an operator, so this observation guarantees that l^2 -coefficients give convergence. Also, the operator will be an isomorphism, which allows us to say more than that. The following lemma tells us exactly what we need in this respect.

Lemma 5.12 ([Yo01], p. 25). *Let $\{u_k\}_{k \in \mathbb{N}} \subset X$ and $\{v_k\}_{k \in \mathbb{N}} \subset Y$ be sequences in two Banach spaces, X and Y . If there exists a bounded isomorphism $T : X \rightarrow Y$ satisfying $Tu_k = v_k$ for any $k \in \mathbb{N}$, then*

$$\sum_{k \in \mathbb{N}} c_k u_k \text{ converges in } X \text{ if and only if } \sum_{k \in \mathbb{N}} c_k v_k \text{ converges in } Y.$$

Proof. Assume $\sum_{k \in \mathbb{N}} c_k u_k$ converges to $x \in X$, and define

$$x_n := \sum_{k=1}^n c_k u_k \in X, \quad \forall n \in \mathbb{N}.$$

Now, continuity of T tells us that $Tx_n \rightarrow Tx \in Y$ as $n \rightarrow \infty$. Since

$$Tx_n = T \left(\sum_{k=1}^n c_k u_k \right) = \sum_{k=1}^n c_k Tu_k = \sum_{k=1}^n c_k v_k$$

by linearity of T , this shows necessity.

Since T^{-1} is also a bounded isomorphism by theorem 4.6, we can use the same argument to prove sufficiency. \square

Definition 5.13. *Let H be a Hilbert space. $\{u_k\}_{k \in \mathbb{N}}$ is called a **Riesz basis (RB)** for H if there exists a bounded isomorphism $T : H \rightarrow H$ s.t. $\{Tu_k\}_{k \in \mathbb{N}}$ is an ONB for H .*

By theorem 5.8 (ii) and lemma 5.12, the convergent linear combinations of an RB are exactly those whose coefficients are taken from $l^2(\mathbb{N})$. A convenient fact to be aware of is that if $\{e_k\}_{k \in \mathbb{N}}$ is an ONB for H , and if $T : H \rightarrow H$ is a bounded isomorphism, then $\{T^{-1}e_k\}_{k \in \mathbb{N}} := \{u_k\}_{k \in \mathbb{N}}$ is automatically a basis for H . Hence, to verify that the u_k form an RB for H , we just need to find T . To see why, given any $x \in H$, define $y := Tx \in H$, and note:

$$x = T^{-1}y = T^{-1} \left(\sum_{k \in \mathbb{N}} \langle y, e_k \rangle e_k \right) = \sum_{k \in \mathbb{N}} \langle y, e_k \rangle T^{-1}e_k = \sum_{k \in \mathbb{N}} \langle y, e_k \rangle u_k.$$

This shows that x can be expanded in the u_k . Similarly, if $\{c_k\}_{k \in \mathbb{N}}$ are the coefficients in such an expansion for x , then

$$y = Tx = T \left(\sum_{k \in \mathbb{N}} c_k u_k \right) = \sum_{k \in \mathbb{N}} c_k T u_k = \sum_{k \in \mathbb{N}} c_k e_k.$$

If the coefficients were not unique, this would give us different expansions for y in the ONB, which is a contradiction. Hence, they must be unique. This proves that $\{u_k\}_{k \in \mathbb{N}}$ is indeed a basis, specifically an RB, as long as the desired operator exists. Note that all the sums above converge, due to theorem 5.12, while continuity of T and T^{-1} justifies moving them into the sums.

It is possible to define different inner products on the same Hilbert space. We will not do that, but it turns out that there exists an inner product for which the RB becomes an ONB ([Yo01], p. 27). This gives another view of how closely related RBs are to ONBs. The elements of an RB might not have the same norm, like an ONB has, but they must be bounded both above and below. That is, the norm of the elements of an RB can neither get arbitrarily large nor arbitrarily small. That is one consequence of the following theorem, which is a generalization of theorem 5.8 (ii) for ONBs. To see why boundedness of the RB follows, set $c_k = 0$ for all but one $k \in \mathbb{N}$.

Theorem 5.14 ([Yo01], p. 27). *A basis $\{u_k\}_{k \in \mathbb{N}}$ for a Hilbert space H is an RB if and only if there exist constants $A, B > 0$ s.t.*

$$\sum_{k \in \mathbb{N}} c_k u_k \text{ converges, and } A \sum_{k \in \mathbb{N}} |c_k|^2 \leq \left\| \sum_{k \in \mathbb{N}} c_k u_k \right\|^2 \leq B \sum_{k \in \mathbb{N}} |c_k|^2, \forall \{c_k\}_{k \in \mathbb{N}} \in l^2(\mathbb{N}). \quad (8)$$

Proof. Assume the u_k form an RB for H , and find a bounded isomorphism $T: H \rightarrow H$ s.t. the $e_k := T u_k$ form an ONB for H . Then, for any sequence $\{c_k\}_{k \in \mathbb{N}} \in l^2(\mathbb{N})$, we have:

$$\begin{aligned} \sum_{k \in \mathbb{N}} |c_k|^2 &= \left\| \sum_{k \in \mathbb{N}} c_k e_k \right\|^2 = \left\| \sum_{k \in \mathbb{N}} c_k T u_k \right\|^2 = \left\| T \left(\sum_{k \in \mathbb{N}} c_k u_k \right) \right\|^2 \leq \|T\|^2 \left\| \sum_{k \in \mathbb{N}} c_k u_k \right\|^2. \\ \left\| \sum_{k \in \mathbb{N}} c_k u_k \right\|^2 &= \left\| \sum_{k \in \mathbb{N}} c_k T^{-1} e_k \right\|^2 = \left\| T^{-1} \left(\sum_{k \in \mathbb{N}} c_k e_k \right) \right\|^2 \leq \|T^{-1}\|^2 \left\| \sum_{k \in \mathbb{N}} c_k e_k \right\|^2 \\ &= \|T^{-1}\|^2 \sum_{k \in \mathbb{N}} |c_k|^2. \end{aligned}$$

This proves necessity, where we may pick $A = \frac{1}{\|T\|^2}$ and $B = \|T^{-1}\|^2$.

For sufficiency, assume that $\{u_k\}_{k \in \mathbb{N}}$ is a basis for H satisfying (8). Existence of that basis implies that H is separable, so it has an ONB as well. Pick an ONB $\{e_k\}_{k \in \mathbb{N}}$ for H , and define an operator $S: H \rightarrow H$ by

$$Sx := \sum_{k \in \mathbb{N}} \langle x, e_k \rangle u_k, \forall x \in H.$$

Setting $c_k := \langle x, e_k \rangle$ for $k \in \mathbb{N}$, by theorem 5.9 (ii), our assumption says that

$$A\|x\|^2 \leq \|Sx\|^2 \leq B\|x\|^2, \forall x \in H.$$

Hence, S is bounded both above and below. Lemma 3.2 (ii) now shows injectivity. Clearly, S is linear as well, and it satisfies $Se_n = u_n$ for $n \in \mathbb{N}$. Now, pick a $y \in H$, and find the coefficients s.t. $y = \sum_{k \in \mathbb{N}} c_k u_k$. Define

$$y_n := \sum_{k=1}^n c_k u_k, \quad x_n := \sum_{k=1}^n c_k e_k \quad \forall n \in \mathbb{N}.$$

Applying (8) to a sequence of finitely many non-vanishing terms, and using theorem 5.8 (ii), we get:

$$\|y_n - y_m\|^2 = \left\| \sum_{k=m+1}^n c_k u_k \right\|^2 \geq A \sum_{k=m+1}^n |c_k|^2 = A \left\| \sum_{k=m+1}^n c_k e_k \right\|^2 = A \|x_n - x_m\|^2$$

whenever $n > m$. Since $\{y_k\}_{k \in \mathbb{N}}$ is Cauchy, this shows that $\{x_k\}_{k \in \mathbb{N}}$ is also Cauchy, so it converges to an $x \in H$. Also,

$$Sx_n = S \left(\sum_{k=1}^n c_k e_k \right) = \sum_{k=1}^n c_k S e_k = \sum_{k=1}^n c_k u_k \rightarrow y$$

as $n \rightarrow \infty$. Continuity of S tells us that $Sx = y$, so S is surjective. Due to theorem 4.6, we have now proven that S^{-1} satisfies the premises of the operator T in definition 5.13. \square

Something more general than RBs, called *Riesz sequences*, are in fact defined by (8). The difference is that we require the u_k to form a basis for their closed linear span, rather than for the whole of H . If a Riesz sequence is complete in H as well, it is obviously an RB for H .

It is also convenient to note how T^{-1} was constructed in the proof of theorem 5.14. Firstly, we picked the ONB, $\{e_k\}_{k \in \mathbb{N}}$, that we want to equal $\{Tu_k\}_{k \in \mathbb{N}}$. Then, we defined

$$T^{-1}y := \sum_{k \in \mathbb{N}} \langle y, u_k \rangle e_k, \quad \forall y \in H.$$

Constructing T might not be as easy, though. After all, there is no general way of constructing the inverse of a known bijective operator, even if it is bounded and linear. We could start by finding the inner product that makes $\{u_k\}_{k \in \mathbb{N}}$ an ONB, which would of course not change the fact that $\{e_k\}_{k \in \mathbb{N}}$ is an RB. That allows us to construct T in the same way, but finding that inner product is probably not any easier than finding T to begin with. However, if we pick a particularly convenient ONB, it might be easier to understand what T should look like. We will see an example in section 10.3, even though we will not attempt to find T in that case.

Example 5.15. Define $u_1, u_2, e_1, e_2 \in \mathbb{R}^2$ by:

$$u_1 := \begin{pmatrix} 2 \\ 0 \end{pmatrix}, u_2 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

$$e_1 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, e_2 := \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Then, $\{e_1, e_2\}$ is an ONB for \mathbb{R}^2 . We want to define an isomorphism $T \in B(\mathbb{R}^2, \mathbb{R}^2)$ s.t. $Tu_1 = e_1$ and $Tu_2 = e_2$. The inverse of T is given by:

$$T^{-1}y = \sum_{k=1}^2 \langle y, u_k \rangle e_k = 2y_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (y_1 + y_2) \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2y_1 \\ y_1 + y_2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \forall y \in \mathbb{R}^2.$$

Thus, T is given by the inverse of the 2×2 -matrix above, i.e.

$$Tx = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} x_1 \\ -x_1 + 2x_2 \end{pmatrix}, \forall x \in \mathbb{R}^2.$$

Clearly, T is linear. Since the matrix is invertible, T is bijective as well. For boundedness,

$$\|Tx\|^2 = \frac{1}{4}(x_1^2 + (-x_1 + 2x_2)^2) = \frac{1}{4}(x_1^2 + x_1^2 - 4x_1x_2 + 4x_2^2) \leq x_1^2 + x_2^2 - x_1x_2$$

$$\leq x_1^2 + x_2^2 + |x_1x_2| \leq x_1^2 + x_2^2 + (x_1^2 + x_2^2) = 2\|x\|^2.$$

Here, we have used the fact that either $|x_1x_2| \leq x_1^2$ or $|x_1x_2| \leq x_2^2$, showing that $|x_1x_2| \leq x_1^2 + x_2^2$. Thus, $\{u_1, u_2\}$ is an RB for \mathbb{R}^2 .

6 Frames

Frames are sequences in a Hilbert space H with some really nice properties. One of them is that every element of H can be expanded in the frame, just like a basis can. Also, like for the ONB, we have a general way of computing the coefficients in such an expansion. In fact, frame is an even more general concept than RB, which as we know is more general than ONB. The main difference is that a frame might be overcomplete, i.e. there might exist proper subsequence that is still a frame in H . This allows us to lose some frame vectors, and still be able to expand any element of H . Also, it means that the expansion is not unique. However, the frame operator, to be defined in section 6.3, gives a particularly convenient expansion, and that is the one that we have a general formula for.

As is well-known, $L^2(S)$ is a Hilbert space. If S is bounded, then there exist exponential systems that form frames in $L^2(S)$. These frames are particularly important to us, since the strong sampling problem in Paley-Wiener spaces is equivalent to finding all exponential frames in $L^2(S)$ (see section 8.4). That is our main reason for considering frames. This section is all about defining the most important concepts related to frames, proving some of their general properties and doing some examples.

6.1 Introduction to frames

Definition 6.1. A sequence $\{u_k\}_{k \in \mathbb{N}}$ in a Hilbert space H is called a **frame** in H if it satisfies both Bessel's inequality and its converse. That is, if there exist constants $A, B > 0$ satisfying

$$A\|x\|^2 \leq \sum_{k \in \mathbb{N}} |\langle x, u_k \rangle|^2 \leq B\|x\|^2, \quad \forall x \in H. \quad (9)$$

The largest lower bound and the smallest upper bound for the frame are called the **frame bounds**.

Note that whether $\{u_k\}_{k \in \mathbb{N}} \subset H$ is a frame or not, is independent of the ordering. The reason is that the sum in (9) contains only non-negative terms, and absolute convergence implies unconditional convergence. In particular, that means that if we add extra elements to the frame, we do not need to worry about where to put them in the sequence.

As we have seen in section 5.3, the converse of Bessel's inequality implies that every frame is complete. The converse is not true, though. E.g. if $\{e_k\}_{k \in \mathbb{N}}$ is an ONB for H , then $\{\frac{1}{k}e_k\}_{k \in \mathbb{N}}$ is also complete in H . However, given any $A > 0$, we have:

$$\sum_{k \in \mathbb{N}} \left| \langle e_n, \frac{1}{k}e_k \rangle \right|^2 = \frac{1}{n^2} < A = A\|e_n\|^2, \quad \forall n > \frac{1}{\sqrt{A}}.$$

Thus, the converse of Bessel's inequality is not satisfied for $\{\frac{1}{k}e_k\}_{k \in \mathbb{N}}$.

Another feature of the converse of Bessel's inequality is that it remains true if we add extra elements to our sequence, since it only adds extra non-negative terms to the series in the middle of (9). The same is not true for Bessel's inequality. For example, if we start with an ONB, and then add elements that do not form Bessel sequence, we clearly do not get a Bessel sequence. On the other hand, among the two inequalities, Bessel's inequality is

normally the easier one to check. Also, our first two examples give us a couple of ways to add elements to a Bessel sequence without disturbing Bessel's inequality.

Bessel's inequality tells us, by theorem 5.7, that every linear combination of the frame vectors with l^2 -coefficients converges. In particular, for any $x \in H$, Bessel's inequality ensures us that the sequence $\{\langle x, u_k \rangle\}_{k \in \mathbb{N}}$ is an element of $l^2(\mathbb{N})$. Hence, it can be used as coefficients in a linear combination of the u_k . In fact, this is exactly what we will do when defining alternate dual frames and the frame operator in the next two sections.

Before looking at examples, we will define two properties that a frame may have, so that we can check for those properties in our examples. However, we will not discuss them for general cases yet.

Definition 6.2. *A frame in a Hilbert space H is called*

- (i) **tight** if the two frame bounds are equal.
- (ii) **exact** if no proper subsequence of the frame is complete in H .

By theorem 5.9 (ii), every ONB is a tight frame with frame bound 1, and since it is a basis, it is also exact. Every ONS is a Bessel sequence with Bessel bound 1, but unless it is an ONB, it will not be complete, hence not a frame. We will see later that exact frames are always bases, in fact bases of a particular kind.

Example 6.3. *Let $J_1, J_2 \subset \mathbb{R}$ be disjoint, discrete sets. If $\{u_k\}_{k \in J_1}$ and $\{u_k\}_{k \in J_2}$ are Bessel sequences in a Hilbert space H , and their respective Bessel bounds are B_1 and B_2 , then*

$$\sum_{k \in J_1 \cup J_2} |\langle x, u_k \rangle|^2 = \sum_{k \in J_1} |\langle x, u_k \rangle|^2 + \sum_{k \in J_2} |\langle x, u_k \rangle|^2 \leq B_1 \|x\|^2 + B_2 \|x\|^2 = (B_1 + B_2) \|x\|^2,$$

for any $x \in H$. Hence, $\{u_k\}_{k \in J_1 \cup J_2}$ is a Bessel sequence in H , and $B_1 + B_2$ is a Bessel bound.

Example 6.4. *Let $\{u_k\}_{k \in \mathbb{N}}$ be a sequence in a Hilbert space H , and assume that $\{\|u_k\|\}_{k \in \mathbb{N}} \in l^2(\mathbb{N})$. Then,*

$$\sum_{k \in \mathbb{N}} |\langle x, u_k \rangle|^2 \leq \sum_{k \in \mathbb{N}} \|x\|^2 \cdot \|u_k\|^2$$

by Cauchy-Schwarz' inequality. Hence, $\{u_k\}_{k \in \mathbb{N}}$ is a Bessel sequence in H , and $\sum_{k \in \mathbb{N}} \|u_k\|^2$ is a Bessel bound.

These two examples, together with the fact that adding extra elements never destroys the converse of Bessel's inequality, give us some ways of adding extra elements. Example 6.3 tells us that we may add elements that form a Bessel sequence, and example 6.4 implies that we may pick that extra Bessel sequence to be any finite set. In particular, this shows that if we start with a finite number of ONBs and add finitely many other elements, we still get a frame.

Now, what about the case that H is a finite-dimensional Euclidean space? In that case, neither of the two inequalities is very difficult. We will give a theorem for that, before considering an example.

Theorem 6.5. Let $n \in \mathbb{N}$ and a finite set $J \subset \mathbb{R}$ be given. A sequence $\{u_j\}_{j \in J}$ in \mathbb{R}^n or \mathbb{C}^n is a frame if and only if it is complete.

Proof. Example 6.4 shows that the u_j satisfy Bessel's inequality, so we only need to check its converse. Assume that no lower bound for the u_j exists. Then, there exists a sequence $\{x_m\}_{m \in \mathbb{N}} \subset \mathbb{C}^n$ with $\|x_m\| = 1$ and

$$\frac{1}{m} = \frac{1}{m} \|x_m\|^2 > \sum_{j \in J} |\langle x_m, u_j \rangle|^2, \quad \forall m \in \mathbb{N}.$$

Since $\{x_m\}_{m \in \mathbb{N}}$ is a bounded sequence in Euclidean space, it has a subsequence $\{x_{m_k}\}_{k \in \mathbb{N}}$ converging to some $x \in \mathbb{C}^n$. Then, we have:

$$\sum_{j \in J} |\langle x, u_j \rangle|^2 = \sum_{j \in J} \left| \langle \lim_{k \rightarrow \infty} x_{m_k}, u_j \rangle \right|^2 = \lim_{k \rightarrow \infty} \sum_{j \in J} |\langle x_{m_k}, u_j \rangle|^2 \leq \lim_{k \rightarrow \infty} \frac{1}{m_k} = 0.$$

This shows that x is orthogonal to all the u_j , so $\{u_j\}_{j \in J}$ is not complete. \square

Example 6.6. In \mathbb{R}^2 , define the sequence $\{u_k\}_{k=1}^3 = \{u_1, u_2, u_3\}$ by:

$$u_1 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad u_2 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad u_3 := \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Since $\{u_1, u_2, u_3\}$ is a complete system in Euclidean space, it is a frame. However, let us verify that it is indeed the case. We have, for all $x \in \mathbb{R}^2$:

$$\sum_{k=1}^3 |\langle x, u_k \rangle|^2 = x_1^2 + (x_1 + x_2)^2 + x_2^2 \geq x_1^2 + x_2^2 = \|x\|^2.$$

$$\sum_{k=1}^3 |\langle x, u_k \rangle|^2 = x_1^2 + (x_1 + x_2)^2 + x_2^2 = 2(x_1^2 + x_2^2) + 2x_1x_2 \leq 2(x_1^2 + x_2^2) + 2(x_1^2 + x_2^2) = 4\|x\|^2.$$

Hence, $\{u_1, u_2, u_3\}$ is a frame in \mathbb{R}^2 , where we may pick $A = 1$ and $B = 4$. We will not worry about optimizing A and B .

6.2 Alternate dual frames

Definition 6.7. Let $\{u_k\}_{k \in \mathbb{N}}$ be a frame in a Hilbert space H . A frame $\{v_k\}_{k \in \mathbb{N}}$ in H is called an **alternate dual frame** for $\{u_k\}_{k \in \mathbb{N}}$ if it satisfies

$$x = \sum_{k \in \mathbb{N}} \langle x, v_k \rangle u_k, \quad \forall x \in H. \quad (10)$$

We will see later that any frame has at least one alternate dual frame. So if we can find one, it will give us a formula for expanding vectors in the frame! Note that the expansion in an ONB, given by theorem 5.9 (i), is in that form, showing that it is its own alternate dual frame. We will now show that, contrary to dual spaces, being an alternate dual frame is in general a reflexive property. Thus, we can talk about two frames as being each other's alternate dual

frames. It should be noted, though, that the alternate dual frame might not be unique. In fact, it can be shown that it is unique if and only if the frame is exact ([HL00], p. 35).

Theorem 6.8 ([HL00], p. 17). *Let $\{u_k\}_{k \in \mathbb{N}}$ be a frame in a Hilbert space H , and let $\{v_k\}_{k \in \mathbb{N}}$ be an alternate dual frame. Then, $x = \sum_{k \in \mathbb{N}} \langle x, u_k \rangle v_k$, $\forall x \in H$. That is, $\{u_k\}_{k \in \mathbb{N}}$ is an alternate dual frame for $\{v_k\}_{k \in \mathbb{N}}$.*

Partial proof. For any $x \in H$, we have:

$$\sum_{k \in \mathbb{N}} \langle x, u_k \rangle v_k = \sum_{k \in \mathbb{N}} \langle x, u_k \rangle \sum_{l \in \mathbb{N}} \langle v_k, v_l \rangle u_l = \sum_{l \in \mathbb{N}} \langle x, \sum_{k \in \mathbb{N}} \langle v_l, v_k \rangle u_k \rangle u_l = \sum_{l \in \mathbb{N}} \langle x, v_l \rangle u_l = x.$$

What remains, which we will not do here, is to justify interchanging the order of the two sums. \square

Note that the proof of theorem 6.8 never took advantage of the fact that $\{u_k\}_{k \in \mathbb{N}}$ and $\{v_k\}_{k \in \mathbb{N}}$ are frames, only that they satisfy the reconstruction formula (10). Hence, assuming that we can interchange the two sums in the proof, the reconstruction formula is symmetric for any two sequences satisfying it. Recall that as long as $\{u_k\}_{k \in \mathbb{N}}$ and $\{v_k\}_{k \in \mathbb{N}}$ are Bessel sequences, we know that (10) converges. Specifically, it does in the case of frames, which is what the theorem is actually about.

Example 6.9. *Let an ONB $\{e_k\}_{k \in \mathbb{N}}$ for H and a $c > 0$ be given. Clearly, $\{ce_k\}_{k \in \mathbb{N}}$ is a tight, exact frame in H with frame bounds c^2 . Its alternate dual frame is $\{\frac{1}{c}e_k\}_{k \in \mathbb{N}}$, as is easily seen by the expansion*

$$x = \sum_{k \in \mathbb{N}} \langle x, e_k \rangle e_k = \sum_{k \in \mathbb{N}} \langle x, \frac{1}{c}e_k \rangle ce_k, \forall x \in H.$$

This turns out to have a simple generalization. Given a tight frame $\{u_k\}_{k \in \mathbb{N}}$ with frame bounds $A > 0$, define, $v_k := \frac{1}{A}u_k$ for $k \in \mathbb{N}$. Clearly, $\{v_k\}_{k \in \mathbb{N}}$ is a tight frame with frame bounds $\frac{1}{A}$. It can be shown that the following reconstruction formula holds:

$$x = \frac{1}{A} \sum_{k \in \mathbb{N}} \langle x, u_k \rangle u_k = \sum_{k \in \mathbb{N}} \langle x, v_k \rangle u_k, \forall x \in H \quad (11)$$

([HL00], p. 14). *This shows that $\{v_k\}_{k \in \mathbb{N}}$ is an alternate dual frame for $\{u_k\}_{k \in \mathbb{N}}$. Setting $c := \sqrt{A}$, if the $\frac{1}{c}u_k$ form an ONB for H , we get the special case that we considered first.*

6.3 The frame operator

There is a particular alternate dual frame that is very convenient. It is defined in terms of a very important operator, which we will take a look at in this section. First, we need to define a related operator.

Definition 6.10. *Let $\{u_k\}_{k \in \mathbb{N}}$ be a Bessel sequence in a Hilbert space H . The operator $T: H \rightarrow l^2(\mathbb{N})$, defined by $Tx := \{\langle x, u_k \rangle\}_{k \in \mathbb{N}}$, $\forall x \in H$, is called the **analysis operator**.*

For the rest of this section, the only operator we will denote by T is the analysis operator. Note that a reformulation of (4) is that

$$\|Tx\|_2^2 \leq B\|x\|_H^2, \forall x \in H.$$

This is the same as saying that T is bounded, and that $\|T\| \leq \sqrt{B}$. Hence, Bessel sequences are exactly the sequences for which the analysis operator, as an operator mapping H into $l^2(\mathbb{N})$, is well-defined and bounded. Similarly, in the case of a frame, (9) can be written as

$$A\|x\|_H^2 \leq \|Tx\|_2^2 \leq B\|x\|_H^2, \forall x \in H.$$

From this, we see that a sequence in H is a frame if and only if T is well-defined and bounded both above and below. Obviously, T is linear as well, since inner products are linear in the first slot. Hence, $T \in B(H, l^2(\mathbb{N}))$, which allows us to talk about the adjoint of T . The two operators in the following definitions are well-defined for all Bessel sequences, but we will only consider them for frames. Otherwise, theorem 6.13 (iii) and (iv) will not be true, and those properties will be important!

Note that if H is \mathbb{R}^n or \mathbb{C}^n , and if the Bessel sequence contains m vectors, then T maps \mathbb{R}^n into \mathbb{R}^m or \mathbb{C}^n into \mathbb{C}^m .

Definition 6.11. Let $\{u_k\}_{k \in \mathbb{N}}$ be a frame in a Hilbert space H .

(i) The operator $T^* : l^2(\mathbb{N}) \rightarrow H$ is called the **synthesis operator**.

(ii) The operator $S := T^*T : H \rightarrow H$ is called the **frame operator**.

Just as for the analysis operator, for the rest of this section, we are always talking about the frame operator when we denote an operator by S . Now, what do T^* and S actually look like? We will see in the following theorem.

Theorem 6.12. Let $\{u_k\}_{k \in \mathbb{N}}$ be a frame in a Hilbert space H .

(i) The synthesis operator is given by

$$T^*\{c_k\}_{k \in \mathbb{N}} = \sum_{k \in \mathbb{N}} c_k u_k, \forall \{c_k\}_{k \in \mathbb{N}} \in l^2(\mathbb{N}).$$

(ii) The frame operator is given by

$$Sx = \sum_{k \in \mathbb{N}} \langle x, u_k \rangle u_k, \forall x \in H.$$

Proof. (i) For any $x \in H$ and any $\{c_k\}_{k \in \mathbb{N}} \in l^2(\mathbb{N})$, we have:

$$\langle Tx, \{c_k\}_{k \in \mathbb{N}} \rangle_2 = \sum_{k \in \mathbb{N}} (Tx)_k \overline{c_k} = \sum_{k \in \mathbb{N}} \langle x, u_k \rangle_H \overline{c_k} = \left\langle x, \sum_{k \in \mathbb{N}} c_k u_k \right\rangle_H.$$

Since this equals $\langle x, T^*\{c_k\}_{k \in \mathbb{N}} \rangle_H$ by theorem 5.2, this proves (i).

(ii) $Sx = T^*Tx = T^*\{\langle x, u_k \rangle\}_{k \in \mathbb{N}} = \sum_{k \in \mathbb{N}} \langle x, u_k \rangle u_k, \forall x \in H$ by (i). □

A very convenient observation is the following:

$$\|Tx\|_2^2 = \sum_{k \in \mathbb{N}} |\langle x, u_k \rangle|^2 = \sum_{k \in \mathbb{N}} \langle x, u_k \rangle \langle u_k, x \rangle = \left\langle \sum_{k \in \mathbb{N}} \langle x, u_k \rangle u_k, x \right\rangle = \langle Sx, x \rangle, \forall x \in H.$$

Hence, the sum in (9) may at any time be replaced by either $\|Tx\|_2^2$ or $\langle Sx, x \rangle$. This also shows that $\langle Sx, x \rangle$ is a non-negative real number for any $x \in H$, which is a general property of self-adjoint operators. This suggests that S might be self-adjoint, and it is not hard to show that it is. This is one of many really nice properties of the frame operator. Some of them are stated in the following theorem.

Theorem 6.13. *The frame operator is*

(i) *self-adjoint.*

(ii) *linear and bounded, with $\|S\| = \|T\|^2$.*

(iii) *bounded below. If $A > 0$ is a lower bound for the frame, then A is a lower bound for S .*

(iv) *bijective.*

Proof. (i) $S^* = (T^*T)^* = T^*T^{**} = T^*T = S$ by proposition 5.3 (i) and (ii).

(ii) Since S is composed of two bounded linear operators, S is itself bounded and linear by proposition 3.4. Also, we have:

$$\|S\| = \|T^*T\| \leq \|T\| \cdot \|T^*\| = \|T\|^2$$

by lemma 3.3 (i) and theorem 5.2.

$$\|Tx\|_2^2 = \langle Sx, x \rangle \leq \|Sx\|_H \cdot \|x\|_H \leq \|S\| \cdot \|x\|_H^2, \forall x \in H \Rightarrow \|T\|^2 \leq \|S\|$$

by Cauchy-Schwarz. The two inequalities together show that $\|S\| = \|T\|^2$.

(iii)

$$A\|x\|_H^2 \leq \langle Sx, x \rangle \leq \|Sx\|_H \cdot \|x\|_H \Rightarrow A\|x\|_H \leq \|Sx\|_H, \forall x \in H,$$

showing that S is bounded below, with A being a lower bound.

(iv) By (i) and (iii), both S and S^* are bounded below. Thus, S is bijective by theorem 4.11. \square

Note that if $A, B > 0$ are the frame bounds, recalling that $\|T\|^2 = B$, theorem 6.13 (ii) and (iii) state the following:

$$A\|x\| \leq \|Sx\| \leq B\|x\|, \forall x \in H. \quad (12)$$

We will take advantage of that to prove an important result in the next subsection.

6.4 The dual frame

We are now ready to introduce the particular alternate dual frame that the frame operator gives us. Recall from the previous subsection that S is bijective.

Theorem 6.14 ([OU16], p. 9). *Let $\{u_k\}_{k \in \mathbb{N}}$ be a frame in a Hilbert space H , and define $v_k := S^{-1}u_k$ for $k \in \mathbb{N}$. Then, $\{v_k\}_{k \in \mathbb{N}}$ is an alternate dual frame for $\{u_k\}_{k \in \mathbb{N}}$. If $A, B > 0$ are the frame bounds for $\{u_k\}_{k \in \mathbb{N}}$, then $\frac{1}{B}, \frac{1}{A}$ are the frame bounds for $\{v_k\}_{k \in \mathbb{N}}$.*

Partial proof. Whenever we pick a $y \in H$, we set $x := S^{-1}y$. We have:

$$\sum_{k \in \mathbb{N}} \langle y, v_k \rangle u_k = \sum_{k \in \mathbb{N}} \langle y, S^{-1}u_k \rangle u_k = \sum_{k \in \mathbb{N}} \langle S^{-1}y, u_k \rangle u_k = S(S^{-1}y) = y, \quad \forall y \in H,$$

where we have taken advantage of the fact that S^{-1} is self-adjoint by proposition 5.3 (iii). Thus, if $\{v_k\}_{k \in \mathbb{N}}$ is a frame in H , it is an alternate dual frame for $\{u_k\}_{k \in \mathbb{N}}$. Also,

$$\sum_{k \in \mathbb{N}} |\langle y, v_k \rangle|^2 = \sum_{k \in \mathbb{N}} |\langle y, S^{-1}u_k \rangle|^2 = \sum_{k \in \mathbb{N}} |\langle S^{-1}y, u_k \rangle|^2 = \sum_{k \in \mathbb{N}} |\langle x, u_k \rangle|^2 = \langle Sx, x \rangle, \quad \forall y \in H.$$

Hence, what remains is to find $c, C > 0$ s.t. $c\|y\|^2 \leq \langle Sx, x \rangle \leq C\|y\|^2$, $\forall y \in H$. What we will not do is to optimize them, but it is easily shown that $c := \frac{A}{B^2}$ and $C := \frac{B}{A^2}$ works. By (12),

$$\frac{A}{B^2}\|y\|^2 = \frac{A}{B^2}\|Sx\|^2 \leq A\|x\|^2 \leq \langle Sx, x \rangle \leq B\|x\|^2 \leq \frac{B}{A^2}\|Sx\|^2 = \frac{B}{A^2}\|y\|^2, \quad \forall y \in H.$$

□

Definition 6.15. *The alternate dual frame given in theorem 6.14 is called the (canonical) dual frame.*

Theorem 6.14 tells us that, as we have pointed out before, every frame does indeed have at least one alternate dual frame, namely the canonical dual frame. Note that the frame bounds are consistent with (11) for the case that $A = B$. With theorem 6.14, we have everything we need to expand any $x \in H$ in a given frame. The only remaining problem, in terms of finding the coefficients of this expansion, is that it requires finding S^{-1} , which we do not have a general expression for.

One nice property of the expansion given by the dual frame, is that it minimizes the l^2 -norm of the coefficients. That is, if $x = \sum_{k \in \mathbb{N}} c_k u_k = \sum_{k \in \mathbb{N}} \langle x, v_k \rangle u_k$ for some $x \in H$, then the l^2 -norm of the c_k cannot be smaller than the l^2 -norm of the $\langle x, v_k \rangle$. Of course, this does not require $\{c_k\}_{k \in \mathbb{N}}$ to be in $l^2(\mathbb{N})$, since it still holds if $\sum_{k \in \mathbb{N}} |c_k|^2 = \infty$. We state this property of the dual frame as a proposition, before turning to examples.

Proposition 6.16 ([OU16], p. 9). *Let $\{u_k\}_{k \in \mathbb{N}}$ be a frame in a Hilbert space H , and let $\{v_k\}_{k \in \mathbb{N}}$ be the dual frame. If $\{c_k\}_{k \in \mathbb{N}}$ is a sequence of complex numbers s.t. $\sum_{k \in \mathbb{N}} c_k u_k$ converges to an $x \in H$, then*

$$\sum_{k \in \mathbb{N}} |\langle x, v_k \rangle|^2 \leq \sum_{k \in \mathbb{N}} |c_k|^2.$$

Proof.

$$\langle x, S^{-1}x \rangle = \left\langle \sum_{k \in \mathbb{N}} \langle x, S^{-1}u_k \rangle u_k, S^{-1}x \right\rangle = \sum_{k \in \mathbb{N}} \langle S^{-1}x, u_k \rangle \langle u_k, S^{-1}x \rangle = \sum_{k \in \mathbb{N}} |\langle u_k, S^{-1}x \rangle|^2.$$

$$\langle x, S^{-1}x \rangle = \left\langle \sum_{k \in \mathbb{N}} c_k u_k, S^{-1}x \right\rangle = \sum_{k \in \mathbb{N}} c_k \langle u_k, S^{-1}x \rangle \leq \sqrt{\sum_{k \in \mathbb{N}} |c_k|^2} \sqrt{\sum_{k \in \mathbb{N}} |\langle u_k, S^{-1}x \rangle|^2}$$

by Hölder's inequality. Hence, combining the two lines above, we get:

$$\begin{aligned} \sum_{k \in \mathbb{N}} |c_k|^2 \cdot \sum_{k \in \mathbb{N}} |\langle u_k, S^{-1}x \rangle|^2 &\geq \left(\sum_{k \in \mathbb{N}} |\langle u_k, S^{-1}x \rangle|^2 \right)^2 \\ \Rightarrow \sum_{k \in \mathbb{N}} |c_k|^2 &\geq \sum_{k \in \mathbb{N}} |\langle S^{-1}x, u_k \rangle|^2 = \sum_{k \in \mathbb{N}} |\langle x, S^{-1}u_k \rangle|^2 = \sum_{k \in \mathbb{N}} |\langle x, v_k \rangle|^2. \end{aligned}$$

□

Example 6.17. Let $u_k := e^{\pi i k(\cdot)} \in L^2[-\frac{1}{2}, \frac{1}{2}]$ for $k \in \mathbb{Z}$. Since both $\{u_{2k}\}_{k \in \mathbb{Z}}$ and $\{u_{2k+1}\}_{k \in \mathbb{Z}}$ are ONBs for $L^2[-\frac{1}{2}, \frac{1}{2}]$, we have:

$$\sum_{k \in \mathbb{Z}} |\langle f, u_k \rangle|^2 = \sum_{k \in 2\mathbb{Z}} |\langle f, u_k \rangle|^2 + \sum_{k \in 2\mathbb{Z}+1} |\langle f, u_k \rangle|^2 = \|f\|_2^2 + \|f\|_2^2 = 2\|f\|_2^2, \forall f \in L^2[-\frac{1}{2}, \frac{1}{2}].$$

Thus, the u_k form a tight frame with frame bounds 2. Also,

$$Sf = \sum_{k \in \mathbb{Z}} \langle f, u_k \rangle u_k = \sum_{k \in 2\mathbb{Z}} \langle f, u_k \rangle u_k + \sum_{k \in 2\mathbb{Z}+1} \langle f, u_k \rangle u_k = f + f = 2f, \forall f \in L^2[-\frac{1}{2}, \frac{1}{2}].$$

That is, $S = 2I$, so $S^{-1} = \frac{1}{2}I$. Hence, the dual frame is $\{\frac{1}{2}u_k\}_{k \in \mathbb{Z}}$. Also,

$$\sum_{k \in \mathbb{Z}} |\langle f, v_k \rangle|^2 = \left(\frac{1}{2}\right)^2 \sum_{k \in \mathbb{Z}} |\langle f, u_k \rangle|^2 = \frac{1}{4} \cdot 2\|f\|_2^2 = \frac{1}{2}\|f\|_2^2, \forall f \in L^2[-\frac{1}{2}, \frac{1}{2}].$$

Hence, the minimum l^2 -norm of the coefficients in an expansion for f in the u_k is $\frac{1}{2}\|f\|_2^2$. Of course, every step in this example works equally well for any two ONBs for any separable Hilbert space.

Example 6.18. Define $u_1, u_2, u_3 \in \mathbb{R}^2$ as in example 6.6. The three operators of this section look like this:

$$\begin{aligned} Tx &= \{\langle x, u_k \rangle\}_{k=1}^3 = \{x_1, x_1 + x_2, x_2\}, \forall x \in \mathbb{R}^2. \\ T^*\{c_k\}_{k=1}^3 &= \sum_{k=1}^3 c_k u_k = c_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + c_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} c_1 + c_2 \\ c_2 + c_3 \end{pmatrix}, \forall c_1, c_2, c_3 \in \mathbb{R}. \end{aligned}$$

$$Sx = T^*\{x_1, x_1+x_2, x_2\} = \begin{pmatrix} x_1 + (x_1 + x_2) \\ (x_1 + x_2) + x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \forall x \in \mathbb{R}^2.$$

That is, S can be viewed as left-multiplication by the 2×2 -matrix above. Hence, S^{-1} is left-multiplication by the inverse of that matrix, i.e.

$$S^{-1}y = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \forall y \in \mathbb{R}^2.$$

The dual frame is:

$$v_1 = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

$$v_2 = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

$$v_3 = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$

Let us verify that (10) holds.

$$\begin{aligned} \sum_{k=1}^3 \langle x, v_k \rangle u_k &= \frac{1}{3} \left((2x_1 - x_2) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (x_1 + x_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + (-x_1 + 2x_2) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \\ &= \frac{1}{3} \left(\begin{pmatrix} 2x_1 - x_2 \\ 0 \end{pmatrix} + \begin{pmatrix} x_1 + x_2 \\ x_1 + x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ -x_1 + 2x_2 \end{pmatrix} \right) = \frac{1}{3} \begin{pmatrix} 3x_1 \\ 3x_2 \end{pmatrix} = x, \forall x \in \mathbb{R}^2. \end{aligned}$$

Now, for a special case, let $x_1 := 2$, $x_2 := 5$. Of course, $x \in \mathbb{R}^2$ can be expanded in the u_k as $x = \sum_{k=1}^3 c_k u_k$ by putting $c_1 = 2$, $c_2 = 0$, $c_3 = 5$. In that case,

$$\sum_{k=1}^3 |c_k|^2 = 2^2 + 5^2 = 29.$$

Another option is to put $c_1 = 0$, $c_2 = 2$, $c_3 = 3$. In that case,

$$\sum_{k=1}^3 |c_k|^2 = 2^2 + 3^2 = 13,$$

which is a definite improvement compared to just using the ONB. However, the optimal solution, in terms of minimizing the l^2 -norm of the coefficients, is given by proposition 6.16:

$$c_1 = \langle x, v_1 \rangle = \frac{1}{3}(2x_1 - x_2) = -\frac{1}{3}.$$

$$c_2 = \langle x, v_2 \rangle = \frac{1}{3}(x_1 + x_2) = \frac{7}{3}.$$

$$c_3 = \langle x, v_3 \rangle = \frac{1}{3}(-x_1 + 2x_2) = \frac{8}{3}.$$

$$\sum_{k=1}^3 |c_k|^2 = \frac{1}{9}(1^2 + 7^2 + 8^2) = \frac{114}{9} \approx 12.67.$$

One thing to note from example 6.18 is that the matrix for the frame operator is symmetric. That is consistent with the fact that S is in general a self-adjoint operator. After all, it is well-known that the complex, self-adjoint matrices are exactly the anti-symmetric ones, i.e. those that are equal to their conjugate transpose. In the real case, there is no difference between symmetry and anti-symmetry, so the result is as expected. We will see the same thing happening in example 6.20.

6.5 Riesz bases and frames

Now, we are ready to give an exact characterization of exact frames. We know that ONBs are exact frames, and we have mentioned in a sentence that all exact frames are a particular kind of basis. This is what we are going to show in our next theorem.

The theorem uses a term that we have not yet encountered. Two sequences, $\{u_k\}_{k \in \mathbb{N}}$ and $\{v_k\}_{k \in \mathbb{N}}$ in H , are said to be *biorthogonal* if $\langle u_k, v_l \rangle = \delta_{kl}$ for all $k, l \in \mathbb{N}$. That is, each element is orthogonal to all elements of the other sequence, except the one with the same index. As an example, any orthonormal system is biorthogonal to itself. More generally, if $\{u_k\}_{k \in \mathbb{N}}$ is an orthogonal system, it is biorthogonal to $\{\frac{u_k}{\|u_k\|^2}\}_{k \in \mathbb{N}}$. Note that a sequence might not have a biorthogonal sequence. E.g. if two of its elements are parallel, then there is no vector orthogonal to exactly one of them. However, in the case of exact frames, it always exists, and it is unique. In general, $\{v_k\}_{k \in \mathbb{N}}$ has a biorthogonal sequence if and only none of the v_k can be approximated by finite linear combinations of the others, and it is unique if and only if $\{v_k\}_{k \in \mathbb{N}}$ is complete in H as well ([Yo01], p. 24).

As pointed out in section 6.3, T is now referring to the analysis operator, rather than the operator in definition 5.13.

Theorem 6.19 ([Yo01], p. 157). *Let $\{u_k\}_{k \in \mathbb{N}}$ be a sequence in a Hilbert space H . Then, the following three are equivalent:*

- (i) $\{u_k\}_{k \in \mathbb{N}}$ is an RB for H .
- (ii) $\{u_k\}_{k \in \mathbb{N}}$ is an exact frame in H .
- (iii) $\{u_k\}_{k \in \mathbb{N}}$ is a frame in H , and it is biorthogonal to its dual frame.

Proof. (i) \Rightarrow (ii)

Let $L: l^2(\mathbb{N}) \rightarrow H$ be the synthesis operator,

$$L\{c_k\}_{k \in \mathbb{N}} := \sum_{k \in \mathbb{N}} c_k u_k, \quad \forall \{c_k\}_{k \in \mathbb{N}} \in l^2(\mathbb{N}).$$

Since the u_k form a basis for H , and since $l^2(\mathbb{N})$ is exactly the space of coefficients for the convergent linear combinations, L is well-defined and bijective. Theorem 5.14 tells us that L is bounded, and clearly, it is linear as well. Thus, L^* is defined. Every step in the proof of theorem 6.12 (i) still works here, so

$$L^*x = \{\langle x, u_k \rangle\}_{k \in \mathbb{N}}, \quad \forall x \in H.$$

By theorems 4.10 and 4.11 (ii), L^* is bounded both above and below. That is just a reformulation of (9), so $\{u_k\}_{k \in \mathbb{N}}$ is a frame in H . Since any proper subset of any basis is incomplete, it must be exact.

(ii) \Rightarrow (iii)

Pick an $n \in \mathbb{N}$. Since $\{u_k\}_{k \in \mathbb{N} \setminus \{n\}}$ is incomplete, there exists a $v_n \in H$ satisfying $\langle v_n, u_k \rangle = 0, \forall k \in \mathbb{N} \setminus \{n\}$. Since $\{u_k\}_{k \in \mathbb{N}}$ is complete, this implies $\langle v_n, u_n \rangle \neq 0$, so we may pick v_n to satisfy $\langle v_n, u_n \rangle = 1$. Repeating that for all $n \in \mathbb{N}$, we get a sequence $\{v_n\}_{n \in \mathbb{N}}$ biorthogonal to $\{u_k\}_{k \in \mathbb{N}}$. Also,

$$Sv_n = \sum_{k \in \mathbb{N}} \langle v_n, u_k \rangle u_k = \sum_{k \in \mathbb{N}} \delta_{nk} u_k = u_n, \quad \forall n \in \mathbb{N},$$

so $\{v_n\}_{n \in \mathbb{N}}$ is the dual frame.

(iii) \Rightarrow (i)

Pick a sequence $\{c_k\}_{k \in \mathbb{N}} \in l^2(\mathbb{N})$, and define

$$x := \sum_{k \in \mathbb{N}} c_k v_k,$$

where $\{v_k\}_{k \in \mathbb{N}}$ is the (biorthogonal) dual frame. Then,

$$(Tx)_n = \langle x, u_n \rangle = \left\langle \sum_{k \in \mathbb{N}} c_k v_k, u_n \right\rangle = \sum_{k \in \mathbb{N}} c_k \langle v_k, u_n \rangle = \sum_{k \in \mathbb{N}} c_k \delta_{kn} = c_n, \quad \forall n \in \mathbb{N}.$$

That is, $Tx = \{c_k\}_{k \in \mathbb{N}}$, so the analysis operator is surjective. Hence, theorem 4.11 (ii) shows that T^* is bounded below, and since it is also bounded above, this proves that $\{u_k\}_{k \in \mathbb{N}}$ satisfies (8). Thus, if it is a basis, it must be an RB by theorem 5.14.

Now, pick a $y \in H$, and find coefficients $\{d_k\}_{k \in \mathbb{N}} \subset \mathbb{C}$ s.t.

$$y = \sum_{k \in \mathbb{N}} d_k u_k.$$

Such an expansion exists by theorem 6.14. Now,

$$\langle S^{-1}y, u_n \rangle = \left\langle \sum_{k \in \mathbb{N}} d_k u_k, S^{-1}u_n \right\rangle = \sum_{k \in \mathbb{N}} d_k \langle u_k, v_n \rangle = \sum_{k \in \mathbb{N}} d_k \delta_{kn} = d_n, \quad \forall n \in \mathbb{N}.$$

Thus,

$$\sum_{n \in \mathbb{N}} |d_n|^2 = \sum_{n \in \mathbb{N}} |\langle S^{-1}y, u_n \rangle|^2 \leq B \|S^{-1}y\|^2 < \infty,$$

where B is a Bessel bound for $\{u_k\}_{k \in \mathbb{N}}$. This proves that such an expansion can only take l^2 -coefficients. Since T^* is bounded below, it is injective, showing that an expansion for y in the u_k with l^2 -coefficients must be unique. Hence, $\{u_k\}_{k \in \mathbb{N}}$ is indeed a basis for H . \square

Example 6.20. Define $u_1, u_2 \in \mathbb{R}^2$ as in example 5.15. We have seen that $\{u_1, u_2\}$ is an RB for \mathbb{R}^2 , so it should be a frame. Let us verify that it is indeed the case. For all $x \in \mathbb{R}^2$,

$$\sum_{k=1}^2 |\langle x, u_k \rangle|^2 = 2x_1^2 + (x_1 + x_2)^2 = 3x_1^2 + x_2^2 + 2x_1x_2 \leq 3x_1^2 + x_2^2 + 2(x_1^2 + x_2^2) \leq 5\|x\|^2$$

The three operators, plus the inverse of the frame operator, are:

$$Tx = \{\langle x, u_k \rangle\}_{k=1}^2 = \{2x_1, x_1 + x_2\}, \quad \forall x \in \mathbb{R}^2.$$

$$T^*\{c_1, c_2\} = \sum_{k=1}^2 c_k u_k = c_1 \begin{pmatrix} 2 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2c_1 + c_2 \\ c_2 \end{pmatrix}, \quad \forall c_1, c_2 \in \mathbb{R}.$$

$$Sx = \begin{pmatrix} 2(2x_1) + (x_1 + x_2) \\ x_1 + x_2 \end{pmatrix} = \begin{pmatrix} 5x_1 + x_2 \\ x_1 + x_2 \end{pmatrix} = \begin{pmatrix} 5 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \forall x \in \mathbb{R}^2.$$

$$S^{-1}y = \begin{pmatrix} 5 & 1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 1 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} y_1 - y_2 \\ -y_1 + 5y_2 \end{pmatrix}, \quad \forall y \in \mathbb{R}^2.$$

The dual frame is:

$$v_1 = \frac{1}{4} \begin{pmatrix} 1 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 2 \\ -2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

$$v_2 = \frac{1}{4} \begin{pmatrix} 1 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Now,

$$\sum_{k=1}^2 \langle x, v_k \rangle u_k = \frac{1}{2}(x_1 - x_2) \begin{pmatrix} 2 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} (x_1 - x_2) + x_2 \\ x_2 \end{pmatrix} = x, \quad \forall x \in \mathbb{R}^2,$$

verifying that $\{v_1, v_2\}$ does indeed satisfy (10). Since it is the only expansion in the u_k that exists, since they form a basis for \mathbb{R}^2 . It is easily seen that $\{v_1, v_2\}$ is biorthogonal to u_1, u_2 .

7 Some function spaces

Earlier, we have been looking at L^p -spaces as spaces of equivalence classes of functions. Soon, we want to consider separate functions. For that reason, we will in this section be very careful to distinguish f from $[f]$. Also, even though we will only do sampling on the real number line, it will sometimes be useful to think of it as sampling a function with complex domain. For that reason, we will start this section by introducing the complex analysis we need. Then, we will go into the actual spaces of functions we are interested in. The relevant function spaces are of two different kinds: Paley-Wiener spaces and Bernstein spaces.

7.1 Some complex analysis

Definition 7.1. A function mapping \mathbb{C} into \mathbb{C} is called

- (i) **analytic** at a point $z_0 \in \mathbb{C}$ if it is differentiable in some neighbourhood of z_0 .
- (ii) **entire** if it is analytic in the whole complex plane.

A well-known fact is that any power series, centered at some $z_0 \in \mathbb{C}$, converges uniformly to an analytic function inside some open ball B , centered at z_0 . Also, it is differentiable and integrable on B , and its derivative/anti-derivative can be obtained by differentiating/integrating term by term. The new power series has the same radius of convergence, although the behaviour at the boundary might be different. Limits can also be performed term by term, since the convergence is absolute.

Conversely, if f is analytic on B , it is well-known that its Taylor series about z_0 converges to f on B . Since the Taylor series is a power series, it has the mentioned properties. One implication of this is that f is infinitely many times differentiable on B , since its derivative is also a power series that converges on B . Another implication is that if f is entire, then its Taylor series about any point converges pointwise to f on \mathbb{C} . The reason is that any $z \in \mathbb{C}$ lies on some open ball B , centered at any point, and uniform convergence on B implies pointwise convergence at $z \in B$. Specifically, the Maclaurin series for f converges pointwise to f on \mathbb{C} , which we will take advantage of to prove our next lemma.

It should be noted that not all of the above is true for real functions. E.g. the function defined by $f(x) := \frac{1}{1+x^2}$, $\forall x \in \mathbb{R}$ is differentiable on \mathbb{R} , but its Maclaurin series only converges on $(-1, 1)$. If we extend f from \mathbb{R} to \mathbb{C} by defining $f(z) := \frac{1}{1+z^2}$, $\forall z \in \mathbb{C}$, then f has singularities at $z = \pm i$. Thus, the Maclaurin series is only guaranteed to converge (uniformly) to f for $|z| < 1$. In fact, it is also well-known that it diverges for $|z| > 1$.

Lemma 7.2. If $f: \mathbb{C} \rightarrow \mathbb{C}$ is entire, if $\{z_k\}_{k \in \mathbb{N}}$ is a sequence in $\mathbb{C} \setminus \{0\}$ converging to zero as $k \rightarrow \infty$, and if $f(z_k) = 0$ for any $k \in \mathbb{N}$, then $f = 0$ on \mathbb{C} .

Proof. Let $\sum_{k=0}^{\infty} c_k z^k$ be the Maclaurin series for f . We want to prove by induction that $c_k = 0$, $\forall k \in \mathbb{N}_0$.

For the basis case, we have:

$$0 = f(z_n) = \sum_{k=0}^{\infty} c_k z_n^k, \forall n \in \mathbb{N}.$$

Since $\lim_{n \rightarrow \infty} z_n = 0$, we get:

$$0 = \lim_{n \rightarrow \infty} f(z_n) = c_0 + \sum_{k=1}^{\infty} c_k \left(\lim_{n \rightarrow \infty} z_n \right)^k = c_0.$$

Now, for $m \in \mathbb{N}$, assume that $c_k = 0$ for all $k < m$. Then,

$$\frac{f(z)}{z^m} = \sum_{k=m}^{\infty} c_k z^{k-m} = \sum_{k=0}^{\infty} c_{k+m} z^k, \quad \forall z \in \mathbb{C} \setminus \{0\},$$

where we have translated the index of summation. Specifically,

$$\frac{f(z_n)}{z_n^m} = \sum_{k=0}^{\infty} c_{k+m} z_n^k, \quad \forall n \in \mathbb{N}.$$

Again, as $n \rightarrow \infty$, we get:

$$0 = \lim_{n \rightarrow \infty} \frac{f(z_n)}{z_n^m} = c_m + \sum_{k=1}^{\infty} c_{k+m} \left(\lim_{n \rightarrow \infty} z_n \right)^k = c_m.$$

This completes the proof by induction that all the coefficients of the Maclaurin series vanishes. Hence, the whole series vanishes for any $z \in \mathbb{C}$, showing that $f = 0$ on \mathbb{C} . \square

Lemma 7.3. (i) If $f: \mathbb{C} \rightarrow \mathbb{C}$ is an entire function that vanishes on \mathbb{R} , then $f = 0$.

(ii) If $f_1: \mathbb{C} \rightarrow \mathbb{C}$ and $f_2: \mathbb{C} \rightarrow \mathbb{C}$ are entire functions that coincide on \mathbb{R} , then $f_1 = f_2$.

Proof. (i) Let $z_k := \frac{1}{k+1}$, $\forall k \in \mathbb{N}$. Then, $z_k \rightarrow 0$ as $k \rightarrow \infty$, and $f(z_k) = 0$, $\forall k \in \mathbb{N}$. Hence, $f = 0$ by lemma 7.2.

(ii) Let $f := f_1 - f_2$. Then, f is an entire function that vanishes on \mathbb{R} . Hence, $f = 0$ by (i), i.e. $f_1 = f_2$. \square

Note that a consequence of lemma 7.3 is that any real, complex-valued function has at most one analytic extension to the complex plane. That is, if $f: \mathbb{R} \rightarrow \mathbb{C}$ has an extension to \mathbb{C} s.t. f is entire, then the extension is unique. This is the main knowledge we want to bring into the following sections. In particular, we will see that Paley-Wiener functions can be analytically extended to the complex plane.

Before finishing this section, we will give four more lemmas without proofs. All of them will be helpful when we are studying Bernstein functions, which is why we will mention the lemmas. However, none of them will be referred to more than once, so we will just continue to the next section without going more deeply into them.

Lemma 7.4 ([Ah66], p. 214). Let F be a family of continuous functions mapping \mathbb{C} into \mathbb{C} . Then, every sequence of functions in F has a subsequence that converges uniformly on every compact subset of \mathbb{C} , if and only if both of the following are satisfied.

(i) Given any compact set $E \subset \mathbb{C}$ and any $\epsilon > 0$, there exists a $\delta > 0$ satisfying $|f(z_2) - f(z_1)| \leq \epsilon$ for all $f \in F$ and all $z_1, z_2 \in E$ s.t. $|z_2 - z_1| \leq \delta$.

(ii) For any $z \in \mathbb{C}$, there exists a compact set $R \subset \mathbb{C}$ s.t. $f(z) \in R$, $\forall f \in F$.

Property (i) is often called *equicontinuity* on any compact set. It states that every element of F is uniformly continuous on E , where the same $\delta > 0$ works for all of them.

Lemma 7.5 (Maximum modulus principle; [Ah66], p. 134). *If $f : \mathbb{C} \rightarrow \mathbb{C}$ is analytic on a compact set $E \subset \mathbb{C}$, then $\sup_{z \in E} |f(z)|$ is attained on the boundary of E .*

Lemma 7.6 ([Yo01] p. 71 and 83). *Let $f : \mathbb{C} \rightarrow \mathbb{C}$ be entire, and assume that there exist constants $A, B > 0$ satisfying*

$$|f(z)| \leq Ae^{B|z|}, \quad \forall z \in \mathbb{C}. \quad (13)$$

(i) *Restricting f to the real number line, if $f \in L^p(\mathbb{R})$ for some $p \in [1, \infty)$, then $f(x) \rightarrow 0$ as $|x| \rightarrow \infty$.*

(ii) *If $f(x) \rightarrow 0$ as $|x| \rightarrow \infty$ along the real axis, then $f(x + iy) \rightarrow 0$ uniformly in y as $|x| \rightarrow \infty$ along the real axis. That is, for any bounded set $S \subset \mathbb{R}$, and for any $\epsilon > 0$, there exists an $R > 0$ s.t. $|f(x + iy)| \leq \epsilon$ whenever $|x| \geq R, y \in S$.*

Lemma 7.7 ([Ch10], p. 106). *If $\{f_k\}_{k \in \mathbb{N}}$ is a sequence of integrable functions mapping \mathbb{R} into \mathbb{C} , and if*

$$\sum_{k \in \mathbb{N}} \int_{\mathbb{R}} |f_k(x)| dx < \infty,$$

then

$$\int_{\mathbb{R}} \sum_{k \in \mathbb{N}} f_k(x) dx = \sum_{k \in \mathbb{N}} \int_{\mathbb{R}} f_k(x) dx.$$

Note that lemma 7.7 can also be applied to sequences of complex functions that are square-integrable on the real number line. After all, what matters here is their properties on \mathbb{R} , so we may consider their restrictions to the real number line and use the lemma.

7.2 Paley-Wiener spaces

In this subsection, we will sometimes encounter sets that are defined only up to a set of measure zero. To take care of that, we will view them as being equivalence classes of well-defined sets. The equivalence class containing a set $S \subseteq \mathbb{R}$ will be denoted by $[S]$, and it consists of the sets that equals S except possibly on a set of measure zero. Just like for functions, it is easily verified that this does indeed define an equivalence relation. We will say that $[S]$ is bounded if for any $S_c \in [S]$, there exists a set $\Lambda \subset \mathbb{R}$ of measure zero s.t. $S_c \setminus \Lambda$ is bounded. Similarly, given sets $U, V \subseteq \mathbb{R}$, we will say that $U \subseteq [S] \subseteq V$ if for any $S_c \in [S]$, there exist sets $\Lambda, \Gamma \subset \mathbb{R}$ of measure zero s.t. $U \subseteq S_c \setminus \Lambda$ and $S_c \setminus \Gamma \subseteq V$. We will never explicitly refer to equivalence classes of sets other than in this subsection. However, after the following definition, it will be clear that some of the sets to come are not really defined everywhere.

Definition 7.8. Let $f: \mathbb{R} \rightarrow \mathbb{C}$ be a function.

- (i) The set $\text{supp}(f) := \{x \in \mathbb{R} \text{ s.t. } f(x) \neq 0\}$ is called the **support** of f .
- (ii) The support of $[f]$ is defined, up to a set of measure zero, as $\text{supp}([f]) := \text{supp}(f)$.
- (iii) If $f \in L^1(\mathbb{R}) \cup L^2(\mathbb{R})$, the set $\text{spec}(f) := \text{supp}(\hat{f})$ is called the **spectrum** of f .
- (iv) If $\text{spec}(f)$ is bounded, then f is called **band-limited**.

Recall that the Fourier transform of an L^1 -function is a function, while an L^2 -function transforms to an equivalence class of functions. Thus, the spectrum is a well-defined set for L^1 -functions, but only defined up to a set of measure zero for L^2 -functions. However, for our purposes, that is sufficient, since integrals of functions that are equal a.e. will always coincide. That fact also means that we can integrate functions over equivalence classes of sets. We can even integrate an equivalence class $[f]$ of functions over an equivalence class $[S]$ of sets, since for any choice of $f_c \in [f]$ and $S_c \in [S]$, the integral $\int_{S_c} f_c(x) dx$ will take the same value. Still, to work with band-limited L^2 -functions, we need a way of saying that $\text{spec}(f)$ is bounded, even though it is only defined up to a set of measure zero. This is the main reason for our discussion about equivalence classes of sets.

Proposition 7.9 ([Ch10], p. 149). *If $S \subset \mathbb{R}$ is a bounded set, if $f \in L^2(\mathbb{R})$, and if $\text{spec}(f) \subseteq S$, then f is equal to a continuous function a.e.*

Proof. Since $f \in L^2(\mathbb{R}) \Rightarrow \hat{f} \in L^2(\mathbb{R})$ and $\text{supp}(\hat{f}) \subseteq S$, we have $\hat{f} \in L^2(S) \subset L^1(\mathbb{R})$ by theorem 3.5. Hence, $\mathfrak{G}\hat{f}$ is continuous and equals f a.e. by theorem 3.7 (i) and (iii), applied to the operator \mathfrak{G} (rather than to \mathfrak{F}). \square

Definition 7.10. *Given a bounded set $S \subset \mathbb{R}$, the **Paley-Wiener space** with spectrum S , denoted by PW_S , is defined by $PW_S := \{f \in L^2(\mathbb{R}) \text{ s.t. } f \text{ is continuous and } \text{spec}(f) \subseteq S\}$. If $\sigma > 0$ is given, then $PW_\sigma := PW_{[-\sigma, \sigma]}$.*

Proposition 7.9 ensures us that the definition of PW_S makes sense. Since different continuous functions cannot be equivalent, we could just as well have defined PW_S to be the corresponding space of equivalence classes of functions. Then it would by definition be a subspace of the Hilbert space $L^2(\mathbb{R})$. For our purposes, it is more convenient to let the elements be functions. Note that since the inner product on $L^2(\mathbb{R})$ is computed using representatives for the equivalence classes, the same inner product can be used for PW_S . Thus, it is an inner product space. It turns out that PW_S is complete w.r.t. the induced norm, i.e. the L^2 -norm. That is, PW_S is a Hilbert space. Before proving that, we will show how we might define PW_S in an equivalent way. In practice, that is the definition we normally use when we study Paley-Wiener functions.

Theorem 7.11 ([OU16], p. 13). *If $S \subset \mathbb{R}$ is a bounded set, then $PW_S = \{\mathfrak{G}F \text{ s.t. } F \in L^2(S)\}$.*

Proof. For $f \in PW_S$, let $F := \hat{f}$. Then, $F \in L^2(S)$, and the same argument as in the proof of proposition 7.9 shows that $\mathfrak{G}F$ is continuous and equals f a.e. Since f is also continuous, they must be equal everywhere, so $f \in \{\mathfrak{G}F \text{ s.t. } F \in L^2(S)\}$.

Conversely, assume $f = \mathfrak{G}F$ for some $F \in L^2(S) \subset L^1(\mathbb{R})$. Then, f is continuous by theorem 3.7 (i), and $\hat{f} = F$ a.e. by theorem 3.8 (iii). Hence, $\text{spec}(f) = \text{supp}(F) \subseteq S$, so $f \in PW_S$. \square

Theorem 7.12. *If $S \subset \mathbb{R}$ is a bounded set, then PW_S is complete w.r.t. the L^2 -norm.*

Proof. Let $\{f_k\}_{k \in \mathbb{N}} \subset PW_S$ be Cauchy. Then, $\{[f_k]\}_{k \in \mathbb{N}}$ is a Cauchy-sequence in the Hilbert space $L^2(\mathbb{R})$, so it converges to $[f] \in L^2(\mathbb{R})$ for some $f \in L^2(\mathbb{R})$. We need to show that f is equivalent to an element of PW_S . Since the Fourier transform on $L^2(\mathbb{R})$ is a linear isometry, we have:

$$\|\hat{f} - \hat{f}_k\|_2 = \|\mathfrak{F}f - \mathfrak{F}f_k\|_2 = \|\mathfrak{F}(f - f_k)\|_2 = \|f - f_k\|_2 \rightarrow 0$$

as $k \rightarrow \infty$. That is, $\hat{f}_k \rightarrow \hat{f}$ in the L^2 -norm as $k \rightarrow \infty$. Also, since all the \hat{f}_k vanish almost everywhere outside S , so does their limit class, \hat{f} . Thus, $\hat{f} \in L^2(S)$, so $g := \mathfrak{G}\hat{f} \in PW_S$ by theorem 7.12. Since $g = f$ a.e., this concludes the proof. \square

Obviously, if $S \subseteq R$, and if $\text{supp}(\hat{f}) \subseteq R$, then $\text{supp}(\hat{f}) \subseteq S$ as well. Thus, $PW_S \subseteq PW_R$. In particular, we may pick R to be an interval, centered at the origin, containing S . That is, $PW_S \subseteq PW_\sigma$ for any $\sigma \geq \sup|S|$. We will take advantage of that a lot of times, since Paley-Wiener spaces with such spectrums are often extra convenient. One reason becomes clear when we start comparing them to Bernstein spaces. However, for the moment, we will state most properties more generally, allowing S to be any bounded set.

7.3 Extending Paley-Wiener functions to the complex plane

As we have seen, the Paley-Wiener functions with spectrum $S \subset \mathbb{R}$ are the functions that can be expressed as $\int_S F(t)e^{2\pi ixt} dt$ for some $F \in L^2(S)$. What happens if we allow x to be any complex number? Is the integral still guaranteed to converge pointwise, and if it does, what properties will the defined function have? The following theorem gives part of the answer.

Theorem 7.13. *Given a bounded set $S \subset \mathbb{R}$, define $\sigma := \sup|S|$. If $F \in L^2(S)$, then $\int_S F(t)e^{2\pi i(\cdot)t} dt$ defines pointwise a function $f: \mathbb{C} \rightarrow \mathbb{C}$. f is entire, and it satisfies:*

(i) ([OU16], p. 14)

$$|f(z)| \leq \sqrt{\mu(S)} \|f\|_2 e^{2\pi\sigma|\text{Im}(z)|}, \quad \forall z \in \mathbb{C}.$$

(ii)

$$f' = 2\pi i \int_S tF(t)e^{2\pi i(\cdot)t} dt.$$

(iii)

$$f' \in PW_S, \text{ and } \|f'\|_2 \leq 2\pi\sigma \|f\|_2.$$

(iv)

$$|f'(z)| \leq 2\pi\sigma \sqrt{\mu(S)} \|f\|_2 e^{2\pi\sigma|\text{Im}(z)|}, \quad \forall z \in \mathbb{C}.$$

Proof. For any $c \in \mathbb{R}$, we have $0 < e^{ct} \leq e^{|c|\sigma}$ whenever $t \in S$. Hence, for any fixed $x, y \in \mathbb{R}$, setting $z = x + iy$, we have:

$$\begin{aligned} |f(z)| &= \left| \int_S F(t) e^{2\pi i(x+iy)t} dt \right| \leq \int_S |F(t) e^{2\pi ixt - 2\pi yt}| dt = \int_S |F(t)| \cdot |e^{2\pi ixt}| \cdot |e^{-2\pi yt}| dt \\ &\leq \int_S |F(t)| e^{2\pi\sigma|y|} dt \leq \left(\int_S 1^2 dt \right)^{\frac{1}{2}} \left(\int_S |F(t)|^2 dt \right)^{\frac{1}{2}} e^{2\pi\sigma|y|} = \sqrt{\mu(S)} \|F\|_2 e^{2\pi\sigma|y|} \\ &= \sqrt{\mu(S)} \|f\|_2 e^{2\pi\sigma|y|} < \infty, \end{aligned}$$

where we have used Hölder's inequality. This shows that the integral converges pointwise, so f is defined on \mathbb{C} . We have also proven (i).

Now, we need to show that f is differentiable at z . This involves a lot of interchanging integrals, sums and derivatives, so we will do the steps carefully. First, we note, for any $n \in \mathbb{N}$:

$$\begin{aligned} \sum_{n=0}^{\infty} \int_S \left| F(t) \frac{(2\pi izt)^n}{n!} \right| dt &\leq \sum_{n=0}^{\infty} \int_S |F(t)| \frac{(2\pi\sigma|z|)^n}{n!} dt = \|F\|_1 \sum_{n=0}^{\infty} \frac{(2\pi\sigma|z|)^n}{n!} < \infty \\ \sum_{n=0}^{\infty} \int_S \left| t F(t) \frac{(2\pi izt)^n}{n!} \right| dt &\leq \sum_{n=0}^{\infty} \int_S \sigma |F(t)| \frac{(2\pi\sigma|z|)^n}{n!} dt = \sigma \|F\|_1 \sum_{n=0}^{\infty} \frac{(2\pi\sigma|z|)^n}{n!} < \infty \end{aligned}$$

The fact that the last sum converges is easily verified by the ratio test, and in fact, it is well-known that its limit is $e^{2\pi\sigma|z|}$. We will soon take advantage of these observations to interchange integrals and sums. Assuming f is differentiable at z , we have:

$$\begin{aligned} f'(z) &= \frac{d}{dz} \int_S F(t) e^{2\pi izt} dt = \frac{d}{dz} \int_S F(t) \sum_{n=0}^{\infty} \frac{(2\pi izt)^n}{n!} dt = \frac{d}{dz} \int_S \sum_{n=0}^{\infty} F(t) \frac{(2\pi izt)^n}{n!} dt \\ &= \frac{d}{dz} \sum_{n=0}^{\infty} \int_S F(t) \frac{(2\pi izt)^n}{n!} dt = \frac{d}{dz} \left(\sum_{n=0}^{\infty} z^n \int_S F(t) \frac{(2\pi it)^n}{n!} dt \right) \\ &= \sum_{n=0}^{\infty} \left(\frac{d}{dz} z^n \int_S F(t) \frac{(2\pi it)^n}{n!} dt \right) = \sum_{n=1}^{\infty} \left(n z^{n-1} \int_S F(t) \frac{(2\pi it)^n}{n!} dt \right) \\ &= \sum_{n=1}^{\infty} \int_S 2\pi it F(t) \frac{(2\pi izt)^{n-1}}{(n-1)!} dt = 2\pi i \sum_{n=0}^{\infty} \int_S t F(t) \frac{(2\pi izt)^n}{n!} dt \\ &= 2\pi i \int_S \sum_{n=0}^{\infty} t F(t) \frac{(2\pi izt)^n}{n!} dt = 2\pi i \int_S t F(t) e^{2\pi izt} dt. \end{aligned}$$

Before continuing, we will justify some of the equalities above:

Fourth equality: Lemma 7.7.

Sixth equality: Any power series can be differentiated term by term.

Ninth equality: Translating the index of summation.

Tenth equality: Lemma 7.7 again.

Hence, we have found an expression for $f'(z)$ if it exists, i.e. if the last integral converges. To see that it does, note that

$$\int_S |tF(t)e^{2\pi izt}| dt \leq \int_S \sigma |F(t)| e^{-2\pi yt} dt \leq \sigma \sqrt{\mu(S)} \|f\|_2 e^{2\pi\sigma|y|}$$

by the same calculation as in the beginning of the proof. Thus, f is indeed differentiable on \mathbb{C} , i.e. entire. This also shows (ii) and (iv), since

$$|f'(z)| = \left| 2\pi i \int_S tF(t)e^{2\pi izt} dt \right| \leq 2\pi \int_S |tF(t)e^{2\pi izt}| dt \leq 2\pi\sigma \sqrt{\mu(S)} \|f\|_2 e^{2\pi\sigma|y|}.$$

Also, since $tF(t)$ defines a function in $L^2(S)$, (ii) implies that $f' \in PW_S$. Example 3.11 now shows that $\|f'\|_2 \leq 2\pi(\sup |S|)\|f\|_2 = 2\pi\sigma\|f\|_2$, which concludes (iii). \square

In accordance with theorem 7.13, we will view PW_S as a space of complex functions, rather than a space of real functions, whenever that is convenient. We are mainly interested in the case that $S = [-\sigma, \sigma]$, after which $\mu(S) = 2\sigma$. In any case, since $S \subseteq [-\sigma, \sigma]$, it is obvious that $PW_S \subseteq PW_\sigma$. Hence, if convenient, we can think of any Paley-Winer function as having spectrum inside some bounded interval.

Example 7.14. Given $\sigma > 0$, pick an $F \in L^2[-\sigma, \sigma]$, and define $f := \check{F} \in PW_\sigma$. Also, let $g: [-\sigma, \sigma] \rightarrow \mathbb{C}$ be the function given by

$$g(t) := 2\pi it e^{\frac{\pi}{2\sigma} it}, \quad \forall t \in [-\sigma, \sigma].$$

By example 5.11, recalling that the Fourier series converges pointwise to g ,

$$g(t) = \sum_{k \in \mathbb{Z}} \frac{2\sigma(-1)^{k-1}}{\pi(k - \frac{1}{2})^2} e^{\frac{\pi}{\sigma} ikt} \Rightarrow 2\pi it = \sum_{k \in \mathbb{Z}} \frac{(-1)^{k-1}}{\pi(k - \frac{1}{2})^2} e^{\frac{\pi}{\sigma} ikt} e^{-\frac{\pi}{2\sigma} it}, \quad \forall t \in [-\sigma, \sigma].$$

Combining this with theorem 7.13 (ii),

$$\begin{aligned} f'(x) &= \int_{-\sigma}^{\sigma} 2\pi it F(t) e^{2\pi ixt} dt = \int_{-\sigma}^{\sigma} \sum_{k \in \mathbb{Z}} \frac{2\sigma(-1)^{k-1}}{\pi(k - \frac{1}{2})^2} e^{\frac{\pi}{\sigma} ikt} e^{-\frac{\pi}{2\sigma} it} F(t) e^{2\pi ixt} dt \\ &= \frac{2\sigma}{\pi} \sum_{k \in \mathbb{Z}} \frac{(-1)^{k-1}}{(k - \frac{1}{2})^2} \int_{-\sigma}^{\sigma} F(t) e^{2\pi i(\frac{1}{2\sigma}(k - \frac{1}{2}) + x)t} dt = \frac{2\sigma}{\pi} \sum_{k \in \mathbb{Z}} \frac{(-1)^{k-1}}{(k - \frac{1}{2})^2} f\left(x + \frac{k - \frac{1}{2}}{2\sigma}\right), \end{aligned}$$

for all $x \in \mathbb{R}$. Exchanging integral and sum is justified by lemma 7.7, since the last sum converges absolutely:

$$\left| \frac{(-1)^{k-1}}{(k - \frac{1}{2})^2} f \left(x + \frac{k - \frac{1}{2}}{2\sigma} \right) \right| \leq \|f\|_\infty \frac{1}{(k - \frac{1}{2})^2}.$$

The integral test, applied to the function $\frac{\|f\|_\infty}{((\cdot) - \frac{1}{2})^2}$, now shows convergence. So we do indeed arrive at the pointwise formula

$$f'(x) = \frac{2\sigma}{\pi} \sum_{k \in \mathbb{Z}} \frac{(-1)^{k-1}}{(k - \frac{1}{2})^2} f \left(x + \frac{k - \frac{1}{2}}{2\sigma} \right), \quad \forall x \in \mathbb{R}. \quad (14)$$

7.4 Bernstein spaces

We are now ready to introduce a space of complex functions, defined by a that property that we have already encountered for Paley-Wiener functions.

Definition 7.15. Given a $\sigma > 0$, the **Bernstein space** B_σ is the space of all entire functions $f: \mathbb{C} \rightarrow \mathbb{C}$ with the property that there exists a constant $C > 0$ satisfying

$$|f(z)| \leq C e^{2\pi\sigma |\operatorname{Im}(z)|}, \quad \forall z \in \mathbb{C}. \quad (15)$$

Theorem 7.13 (i) shows that Paley-Wiener functions satisfy (15), so $PW_\sigma \subseteq B_\sigma$. The fact that the two spaces do not coincide is clear from the following example.

Example 7.16. Given a $\sigma > 0$, let $f(z) := \sin(2\pi\sigma z)$, $\forall z \in \mathbb{C}$. It is well-known that f is an entire function. Also, for any $x, y \in \mathbb{R}$,

$$\begin{aligned} |f(x + iy)| &= \left| \frac{e^{2\pi i \sigma (x+iy)} - e^{-2\pi i \sigma (x+iy)}}{2i} \right| = \frac{1}{2} |e^{2\pi i \sigma x - 2\pi \sigma y} - e^{-2\pi i \sigma x + 2\pi \sigma y}| \\ &\leq \frac{1}{2} (|e^{2\pi i \sigma x} e^{-2\pi \sigma y}| + |e^{-2\pi i \sigma x} e^{2\pi \sigma y}|) = \frac{1}{2} (e^{-2\pi \sigma y} + e^{2\pi \sigma y}) \leq e^{2\pi \sigma |y|}. \end{aligned}$$

Thus, $f \in B_\sigma$. However, f is not square-integrable over the real number line, since the integral of $|f|^2$ over one period is a positive number, $I > 0$. The integral over $n \in \mathbb{N}$ periods is then nI , which tends to infinity as $n \rightarrow \infty$. Hence, $f \notin PW_\sigma \subset L^2(\mathbb{R})$.

The fact that not all Bernstein functions are L^2 -functions means that we cannot use the L^2 -norm on B_σ . However, to find a norm that works, we can take advantage of a simple observation. If we restrict the domain of a Bernstein function to be the real number line, then (15) reduces to $|f(x)| \leq C$, $\forall x \in \mathbb{R}$. For spaces of bounded functions, it is typical to use the sup-norm, and that is exactly what we are going to do now. Just like the most well-known spaces of bounded functions, Bernstein spaces are complete w.r.t. the sup-norm, but does not have an inner product. We will prove that in the next section.

Theorem 7.17. If $\sigma > 0$, then, $\|\cdot\|_\infty: B_\sigma \rightarrow \mathbb{R}_0^+$, defined by

$$\|f\|_\infty = \sup_{x \in \mathbb{R}} |f(x)|, \quad \forall f \in B_\sigma,$$

is a norm on B_σ .

Proof. (i) Obviously, $\|f\|_\infty \geq 0$, $\forall f \in B_\sigma$, with equality if and only if f vanishes on the real number line. By lemma 7.3 (i), since Bernstein functions are entire, this happens if and only if $f = 0$.

$$(ii) \|\alpha f\|_\infty = \sup_{x \in \mathbb{R}} |\alpha f(x)| = |\alpha| \sup_{x \in \mathbb{R}} |f(x)| = |\alpha| \cdot \|f\|_\infty, \forall \alpha \in \mathbb{C}, \forall f \in B_\sigma.$$

$$(iii) \|f + g\|_\infty = \sup_{x \in \mathbb{R}} |f(x) + g(x)| \leq \sup_{x \in \mathbb{R}} (|f(x)| + |g(x)|) \\ \leq \sup_{x, y \in \mathbb{R}} (|f(x)| + |g(y)|) = \sup_{x \in \mathbb{R}} |f(x)| + \sup_{y \in \mathbb{R}} |g(y)| = \|f\|_\infty + \|g\|_\infty, \forall f, g \in B_\sigma.$$

□

Example 7.18. Let $m \in \mathbb{N}$, $\epsilon \in (0, \frac{1}{m})$, $\sigma > 0$ and $f \in B_\sigma$ be given. Define

$$g(z) := \text{sinc}(2\sigma\epsilon z), f_\epsilon(z) := f((1 - m\epsilon)z)(g(z))^m, \forall z \in \mathbb{C}.$$

For any $x, y \in \mathbb{R}$, setting $z = x + iy$, we have:

$$\int_{-\sigma\epsilon}^{\sigma\epsilon} e^{2\pi i z t} dt = \frac{1}{2\pi i z} [e^{2\pi i z t}]_{z=-\sigma\epsilon}^{z=\sigma\epsilon} = \frac{1}{\pi z} \frac{e^{2\pi i \sigma\epsilon z} - e^{-2\pi i \sigma\epsilon z}}{2i} = 2\sigma\epsilon \frac{\sin(2\pi\sigma\epsilon z)}{2\pi\sigma\epsilon z} \\ = 2\sigma\epsilon \text{sinc}(2\sigma\epsilon z) = 2\sigma\epsilon g(z).$$

$$2\sigma\epsilon |g(z)| \leq \int_{-\sigma\epsilon}^{\sigma\epsilon} |e^{2\pi i(x+iy)t}| dt = \int_{-\sigma\epsilon}^{\sigma\epsilon} |e^{2\pi i x t} e^{-2\pi y t}| dt \leq \int_{-\sigma\epsilon}^{\sigma\epsilon} e^{2\pi\sigma\epsilon|y|} dt = e^{2\pi\sigma\epsilon|y|}.$$

Hence, $g \in B_{\sigma\epsilon}$. Also,

$$|(g(z))^m| = |g(z)|^m \leq \left(\frac{1}{2\sigma\epsilon} e^{2\pi\sigma\epsilon|y|} \right)^m = \frac{1}{(2\sigma\epsilon)^m} e^{2\pi m\sigma\epsilon|y|},$$

showing that $g^m \in B_{m\sigma\epsilon}$. Now, since $f \in B_\sigma$, picking $C > 0$ satisfying (15), we have:

$$|f((1 - m\epsilon)z)| \leq C e^{2\pi\sigma|Im((1-m\epsilon)z)|} = C e^{2\pi\sigma(1-m\epsilon)|y|},$$

so $f((1 - m\epsilon)(\cdot)) \in B_{\sigma(1-m\epsilon)}$. Thus,

$$|f_\epsilon(z)| \leq \frac{C}{(2\sigma\epsilon)^m} e^{2\pi\sigma(1-m\epsilon)|y|} e^{2\pi m\sigma\epsilon|y|} = \frac{C}{(2\sigma\epsilon)^m} e^{2\pi\sigma|y|}.$$

Since $z \in \mathbb{C}$ was arbitrary, this shows that $f_\epsilon \in B_\sigma$.

7.5 Properties of Bernstein spaces

This subsection covers the most important properties of Bernstein spaces. Our ultimate goal is to show that the sup-norm makes B_σ a Banach space, which will be the last theorem in this section. However, there are lots of other properties, all requiring a little bit of complex analysis. We will carefully go through a proof for each of them, taking advantage of what we know from section 7.1.

As we have seen, $PW_\sigma \subset B_\sigma$. More specifically, $PW_\sigma \subseteq B_\sigma \cap L^2(\mathbb{R})$, where we interpret $L^2(\mathbb{R})$ as being the space of all complex functions that are square integrable on the real number line. In fact, the converse also turns out to be true. That is, ALL entire L^2 -functions satisfying (15) are elements of PW_σ , i.e. their spectrum lie inside $[-\sigma, \sigma]$. This is known as the *Paley-Wiener theorem*, which we will now show.

Theorem 7.19 (Paley-Wiener theorem; [Yo01], p. 85). *If $\sigma > 0$, then $PW_\sigma = B_\sigma \cap L^2(\mathbb{R})$.*

Proof. Pick an $f \in B_\sigma \cap L^2(\mathbb{R})$, and find a $C > 0$ satisfying (15). As pointed out, what remains to be shown is that $\text{spec}(f) \subseteq [-\sigma, \sigma]$, i.e. that \hat{f} vanishes almost everywhere on $\mathbb{R} \setminus [-\sigma, \sigma]$. Fix a $t < -\sigma$. For any $R > 0$ and any $r \in (0, R)$, consider the five line segments

$$\begin{aligned}\gamma_1 &:= [-R, -R + ir], \quad \gamma_2 := [-R + ir, -R + iR], \quad \gamma_3 := [-R + iR, R + iR], \\ \gamma_4 &:= [R + ir, R + iR], \quad \gamma_5 := [R, R + iR].\end{aligned}$$

Note that together, taking γ_4 and γ_5 in the negative direction, the five line segments make a contour clockwise around a rectangle, starting at $-R$ and ending at R . For $n \in \{1, 2, 3, 4, 5\}$, let I_n be the integral

$$I_n := \int_{\gamma_n} f(z) e^{-2\pi izt} dz = \int_{\gamma_n} f(x + iy) e^{-2\pi ixt} e^{2\pi yt} dz.$$

Our goal is to estimate the contour integral $I := I_1 + I_2 + I_3 - I_4 - I_5$ for large $R > 0$, which is independent of the chosen $r > 0$. Note that the function $g := f e^{-2\pi i(\cdot)t} \in L^2(\mathbb{R})$, defined on \mathbb{C} , satisfies, for all $x, y \in \mathbb{R}$:

$$|g(x + iy)| = |f(x + iy)| \cdot |e^{-2\pi ixt}| \cdot |e^{2\pi yt}| \leq C e^{2\pi\sigma|y|} e^{2\pi yt} \leq C e^{2\pi(\sigma-t)|y|} \leq C e^{2\pi(\sigma-t)|z|}.$$

Thus, g is of exponential type, meaning that it satisfies (13). Now, pick an $\epsilon > 0$, and fix $r > 0$ s.t. $e^{2\pi(\sigma+t)r} \leq -\frac{2\pi(\sigma+t)\epsilon}{5C}$. This is possible because $\sigma + t < 0$. Also, find $R_1, R_2, R_3 > 0$ satisfying:

$$\begin{aligned}|f(-R + iy)| e^{2\pi yt} &= |g(-R + iy)| \leq \frac{\epsilon}{5r}, \quad \forall R \geq R_1. \\ |f(R + iy)| e^{2\pi yt} &= |g(R + iy)| \leq \frac{\epsilon}{5r}, \quad \forall R \geq R_2. \\ R e^{2\pi(\sigma+t)R} &\leq \frac{\epsilon}{10C}, \quad \forall R \geq R_3.\end{aligned}$$

R_1 and R_2 exist due to lemma 7.6. Now, for any $R \geq \max\{r, R_1, R_2, R_3\}$, we have:

$$\begin{aligned} |I_1| &\leq \int_0^r |f(-R + iy)| e^{2\pi yt} dy \leq \int_0^r \frac{\epsilon}{5r} dy = \frac{\epsilon}{5}. \\ |I_2| &\leq \int_r^R |f(-R + iy)| e^{2\pi yt} dy \leq \int_r^R C e^{2\pi\sigma y} e^{2\pi yt} dy = \left[\frac{C}{2\pi(\sigma + t)} e^{2\pi(\sigma+t)y} \right]_{y=r}^{y=R} \\ &= -\frac{C}{2\pi(\sigma + t)} (e^{2\pi(\sigma+t)r} - e^{2\pi(\sigma+t)R}) < -\frac{C}{2\pi(\sigma + t)} e^{2\pi(\sigma+t)r} \leq \frac{\epsilon}{5}. \\ |I_3| &\leq \int_{-R}^R |f(x + iR)| e^{2\pi Rt} dx \leq \int_{-R}^R C e^{2\pi\sigma R} e^{2\pi Rt} dx = 2RC e^{2\pi(\sigma+t)R} \leq \frac{\epsilon}{5}. \end{aligned}$$

Similar calculations show that $|I_4|, |I_5| \leq \frac{\epsilon}{5}$ as well. Thus, $|I| \leq \sum_{n=1}^5 |I_n| \leq \epsilon$. By Cauchy's integral theorem, since f is entire,

$$I = \int_{-R}^R f(x) e^{-2\pi ixt} dx.$$

Thus, we have shown that this integral tends to zero as $R \rightarrow \infty$, assuming that $t < -\sigma$.

Now, $h(z) := f(-z)$ defines another function $h \in B_\sigma \cap L^2(\mathbb{R})$. The change of variable $X := -x$ gives, for $t > \sigma \Rightarrow -t < -\sigma$:

$$\int_{-R}^R f(x) e^{-2\pi ixt} dx = \int_{-R}^R h(-x) e^{-2\pi i(-x)(-t)} dx = - \int_{-R}^R h(X) e^{-2\pi iX(-t)} dX \rightarrow 0$$

as $R \rightarrow \infty$. Thus, for $|t| < \sigma$,

$$\hat{f}(t) = \lim_{R \rightarrow \infty} \int_{-R}^R f(x) e^{-2\pi ixt} dx = 0.$$

□

The next property optimizes the constant C in (15).

Proposition 7.20 ([OU16], p. 17). *If $\sigma > 0$ and $f \in B_\sigma$, then*

$$|f(z)| \leq \|f\|_\infty e^{2\pi\sigma |Im(z)|}, \quad \forall z \in \mathbb{C}.$$

Proof. The case that $f = 0$ is trivial. Otherwise, find $C > 0$ satisfying (15). For any $\epsilon > 0$, define

$$f_\epsilon(z) := \frac{f(z)}{1 - i\epsilon z} e^{2\pi i\sigma z}, \quad \forall z \in \overline{\mathbb{C}^+},$$

where $\overline{\mathbb{C}^+}$ is the set of all $z \in \mathbb{C}$ with $Im(z) \geq 0$. Note that $1 - i\epsilon z$ never vanishes on $\overline{\mathbb{C}^+}$, so f_ϵ is analytic. Now, for any $x \in \mathbb{R}$ and any $y \in \mathbb{R}_0^+$, we have:

$$f_\epsilon(x + iy) = \frac{f(x + iy)}{1 - i\epsilon(x + iy)} e^{2\pi i\sigma(x + iy)} = \frac{f(x + iy)}{1 + \epsilon y - i\epsilon x} e^{2\pi i\sigma x} e^{-2\pi\sigma y}.$$

$$|f_\epsilon(x + iy)| = \frac{|f(x + iy)|}{\sqrt{(1 + \epsilon y)^2 + (\epsilon x)^2}} e^{-2\pi\sigma y} \leq \frac{C}{\sqrt{1 + 2\epsilon y + \epsilon^2(x^2 + y^2)}}.$$

For $R > 0$, define $E_R := \{z \in \mathbb{C} \text{ s.t. } \text{Im}(z) \geq 0, |z| \leq R\}$. If $|z| = R$, then

$$|f_\epsilon(z)| \leq \frac{C}{\sqrt{1 + 2\epsilon \text{Im}(z) + \epsilon^2 R^2}} \leq \frac{C}{\sqrt{1 + \epsilon^2 R^2}} \rightarrow 0 \text{ as } R \rightarrow \infty.$$

Thus, for sufficiently large R , this cannot exceed the largest value that $|f_\epsilon|$ takes on the real interval $[-R, R]$. Hence, the maximum modulus principle (lemma 7.5) tells us that $\sup_{z \in E_R} |f_\epsilon(z)|$ must be attained on $[-R, R]$. Since

$$|f_\epsilon(x)| = \frac{|f(x)|}{\sqrt{1 + \epsilon^2 x^2}} \leq |f(x)|$$

for all $x \in \mathbb{R}$, this shows that

$$\sup_{z \in \overline{\mathbb{C}^+}} |f_\epsilon(z)| \leq \sup_{x \in \mathbb{R}} |f_\epsilon(x)| \leq \sup_{x \in \mathbb{R}} |f(x)| = \|f\|_\infty.$$

Thus,

$$|f(x + iy)| = |f_\epsilon(x + iy)| \sqrt{(1 + \epsilon y)^2 + (\epsilon x)^2} e^{2\pi\sigma y} \leq \sqrt{(1 + \epsilon y)^2 + (\epsilon x)^2} \|f\|_\infty e^{2\pi\sigma y}$$

for any $x, y \in \mathbb{R}$ with $y \geq 0$. Since this holds for arbitrarily small $\epsilon > 0$, we can get the square root arbitrarily close to 1, as long as x and y are fixed. Hence, we conclude that $|f(x + iy)| \leq \|f\|_\infty e^{2\pi\sigma y} = \|f\|_\infty e^{2\pi\sigma|y|}$.

For $y < 0$, define $h(z) := f(-z)$ for $z \in \mathbb{C}$. Clearly, $h \in B_\sigma$ and $\|h\|_\infty = \|f\|_\infty$, so we have shown:

$$|f(x + iy)| = |h(-x - iy)| \leq \|h\|_\infty e^{2\pi\sigma(-y)} = \|f\|_\infty e^{2\pi\sigma|y|}, \quad \forall x, y \in \mathbb{R} \text{ s.t. } y \leq 0.$$

□

It is well-known that on normed spaces of differentiable functions, the differentiation operator is usually not bounded. We need some sort of restriction if we want that to be the case. To do an example with the sup-norm, let

$$f_k(x) := \sin(kx), \quad \forall x \in \mathbb{R}, \quad \forall k \in \mathbb{N}.$$

Then, both f_k and f'_k are bounded on \mathbb{R} , where $\|f_k\|_\infty = 1$ and $\|f'_k\|_\infty = k$, $\forall k \in \mathbb{N}$. Hence, $\frac{\|f'_k\|_\infty}{\|f_k\|_\infty} \rightarrow \infty$ as $k \rightarrow \infty$, showing that the differentiation operator is not bounded. This must also be true for the corresponding space of functions that can be analytically extended to the complex plane, since all the f_k have that property. However, for sufficiently large k , we have $f_k \notin B_\sigma$, so this does not prove that the differentiation operator on B_σ is not bounded. In fact, it turns out to be bounded! This is known as *Bernstein's inequality*, which is our next theorem. It is very similar to the result of example 3.11, except that in Bernstein spaces, we are using the sup-norm rather than the L^2 -norm.

Theorem 7.21 (Bernstein's inequality; [OU16], p. 21). *If $\sigma > 0$ and $f \in B_\sigma$, then $f' \in B_\sigma$ and $\|f'\|_\infty \leq 2\pi\sigma\|f\|_\infty$.*

Proof. Since the derivative of any entire function is entire, it suffices to show that the inequality holds. Pick an $f \in B_\sigma$. First, assume that $f \in L^2(\mathbb{R})$. Then, $f \in PW_\sigma$ by the Paley-Wiener theorem. Now, by example 4.13,

$$\begin{aligned} \left| \frac{2\sigma}{\pi} \sum_{k \in \mathbb{Z}} \frac{(-1)^{k-1}}{(k - \frac{1}{2})^2} f\left(x + \frac{k - \frac{1}{2}}{2\sigma}\right) \right| &\leq \frac{2\sigma}{\pi} \sum_{k \in \mathbb{Z}} \frac{1}{(k - \frac{1}{2})^2} \left| f\left(x + \frac{k - \frac{1}{2}}{2\sigma}\right) \right| \\ &\leq \frac{2\sigma}{\pi} \cdot \pi^2 \sup_{y \in \mathbb{R}} |f(y)| \leq 2\pi\sigma\|f\|_\infty \end{aligned}$$

for all $x \in \mathbb{R}$. Applying (14) to f , we get $\|f'\|_\infty \leq 2\pi\sigma\|f\|_\infty$, which proves Bernstein's inequality for the space $B_\sigma \cap L^2(\mathbb{R})$.

On the other hand, if $f \notin L^2(\mathbb{R})$, we need to approximate f pointwise by a sequence in $B_\sigma \cap L^2(\mathbb{R})$. Define

$$f_k(z) := f\left(\left(1 - \frac{1}{k}\right)z\right) \operatorname{sinc}\left(\frac{2\sigma z}{k}\right), \quad \forall z \in \mathbb{C}, \forall k \in \mathbb{N}.$$

Example 7.18 shows that $f_k \in B_\sigma$, where we put $m := 1$ and $\epsilon := \frac{1}{k}$. Also, since f is bounded on \mathbb{R} and the sinc-function is square-integrable on the real number line, we have $f_k \in L^2(\mathbb{R})$. Thus, $\{f_k\}_{k \in \mathbb{N}}$ is a sequence in $B_\sigma \cap L^2(\mathbb{R})$, so we have shown that

$$\|f'_k\|_\infty \leq 2\pi\sigma\|f_k\|_\infty, \quad \forall k \in \mathbb{N}.$$

Now, for any $x \in \mathbb{R} \setminus \{0\}$, we have:

$$\begin{aligned} \frac{d}{dx} \operatorname{sinc}(2\sigma\epsilon x) &= \frac{d}{dx} \left(\frac{\sin(2\pi\sigma\epsilon x)}{2\pi\sigma\epsilon x} \right) = \frac{2\pi\sigma\epsilon \cos(2\pi\sigma\epsilon x) \cdot 2\pi\sigma\epsilon x - \sin(2\pi\sigma\epsilon x) \cdot 2\pi\sigma\epsilon}{(2\pi\sigma\epsilon x)^2} \\ &= \frac{\cos(2\pi\sigma\epsilon x)}{x} - \frac{\sin(2\pi\sigma\epsilon x)}{2\pi\sigma\epsilon x^2} = \frac{\cos(2\pi\sigma\epsilon x) - \operatorname{sinc}(2\sigma\epsilon x)}{x} \rightarrow \frac{1 - 1}{x} = 0 \end{aligned}$$

as $\epsilon \rightarrow 0$. Thus,

$$\begin{aligned} f'_k(x) &= \left(1 - \frac{1}{k}\right) f'\left(\left(1 - \frac{1}{k}\right)x\right) \operatorname{sinc}\left(\frac{2\sigma x}{k}\right) + f\left(\left(1 - \frac{1}{k}\right)x\right) \frac{d}{dx} \left(\operatorname{sinc}\left(\frac{2\sigma x}{k}\right) \right) \\ &\rightarrow 1 \cdot f'(1 \cdot x) \cdot 1 + f(1 \cdot x) \cdot 0 = f'(x) \text{ as } k \rightarrow \infty. \end{aligned}$$

For $x = 0$, the derivative of the sinc-function vanishes, so

$$f'_k(0) = \left(1 - \frac{1}{k}\right) f'(0) \operatorname{sinc}(0) \rightarrow f'(0) \text{ as } k \rightarrow \infty.$$

Hence, $f'_k \rightarrow f'$ pointwise on \mathbb{R} as $k \rightarrow \infty$. Also,

$$|f'(x)| = \left| \lim_{k \rightarrow \infty} f'_k(x) \right| = \lim_{k \rightarrow \infty} |f'_k(x)| = \limsup_{k \rightarrow \infty} |f'_k(x)| \leq \limsup_{k \rightarrow \infty} \|f'_k\|_\infty, \quad \forall x \in \mathbb{R}$$

$$\Rightarrow \|f'\|_\infty \leq \limsup_{n \rightarrow \infty} \|f'_k\|_\infty.$$

Clearly, $\|f_k\|_\infty \leq \|f\|_\infty$ for all $k \in \mathbb{N}$, since the sinc-function never takes a greater value than 1, while $f((1 - \frac{1}{k})(\cdot))$ is just a translation of f . Hence, we finally get:

$$\|f'\|_\infty \leq \limsup_{k \rightarrow \infty} \|f'_k\|_\infty \leq \limsup_{k \rightarrow \infty} (2\pi\sigma \|f_k\|_\infty) \leq 2\pi\sigma \|f\|_\infty.$$

□

Note that Bernstein's inequality also gives an upper bound for $|f'|$, in terms of the imaginary part of the argument. Since $f' \in B_\sigma$, proposition 7.20 tells us that

$$|f'(z)| \leq \|f'\|_\infty e^{2\pi\sigma |Im(z)|}, \quad \forall z \in \mathbb{C}.$$

Bernstein's inequality now allows us to replace $\|f'\|_\infty$ by $2\pi\sigma \|f\|_\infty$ in the inequality above.

Now, we get to the one theorem that uses equicontonuity on compact sets. This is exactly the property we need before we are ready to prove completeness. Note how crucial Bernstein's inequality is in the proof.

Theorem 7.22 (Compactness property of Bernstein spaces; [OU16], p. 19). *Given $\sigma > 0$, every bounded sequence in B_σ has a subsequence converging uniformly on every compact subset of \mathbb{C} to some $f \in B_\sigma$.*

Proof. Pick a $C > 0$, and define $F := \{f \in B_\sigma \text{ s.t. } \|f\|_\infty \leq C\}$. We need to show that F satisfies the premises of lemma 7.4.

(i) Pick a compact $E \subset \mathbb{C}$ and an $\epsilon > 0$. Let $r := \max_{u \in E} |Im(u)|$ and $\delta := \frac{\epsilon}{2\pi\sigma C} e^{-2\pi\sigma r}$. If $f \in F$, and if $z_1, z_2 \in E$ satisfy $|z_2 - z_1| \leq \delta$, then

$$|f(z_2) - f(z_1)| = \left| \int_{z_1}^{z_2} f'(u) du \right| \leq \sup_{u \in E} |f'(u)| \cdot |z_2 - z_1| \leq \sup_{u \in E} (\|f'\|_\infty e^{2\pi\sigma |Im(u)|}) \cdot |z_2 - z_1|$$

$$\leq 2\pi\sigma \|f\|_\infty e^{2\pi\sigma r} \delta \leq 2\pi\sigma C e^{2\pi\sigma r} \delta = \epsilon.$$

(ii) Pick a $z_0 \in \mathbb{C}$, and define

$$r := e^{2\pi\sigma |Im(z_0)|}, \quad R := \{z \in \mathbb{C} \text{ s.t. } |z| \leq Cr\}.$$

Then, $R \subset \mathbb{C}$ is compact, and

$$|f(z_0)| \leq \|f\|_\infty e^{2\pi\sigma |Im(z_0)|} \leq Cr \Rightarrow f(z_0) \in R, \quad \forall f \in F.$$

Hence, by lemma 7.4, any sequence $\{f_k\}_{k \in \mathbb{N}} \subset F$ converges uniformly on every compact subset of \mathbb{C} to a function $f : \mathbb{C} \rightarrow \mathbb{C}$. For any $z \in \mathbb{C}$, since uniform convergence implies pointwise convergence,

$$|f(z)| = \left| \lim_{k \rightarrow \infty} f_k(z) \right| = \lim_{k \rightarrow \infty} |f_k(z)| \leq \limsup_{k \rightarrow \infty} \|f_k\|_\infty e^{2\pi\sigma|Im(z)|} \leq C e^{2\pi\sigma|Im(z)|}.$$

Thus, $f \in F \subset B_\sigma$, which concludes the proof. \square

As noted in the proof, the compactness property specifically says that any bounded sequence in B_σ has a subsequence converging pointwise on \mathbb{C} to a function $f \in B_\sigma$. This is what we will now use to finish this section.

Theorem 7.23. *If $\sigma > 0$, then B_σ is complete w.r.t. the sup-norm.*

Proof. Let $\{f_j\}_{j \in \mathbb{N}} \subset B_\sigma$ be Cauchy. Since Cauchy-sequences are bounded, it has a subsequence $\{f_{j_k}\}_{k \in \mathbb{N}}$ converging pointwise to some $f \in B_\sigma$ by the compactness property. Pick an $\epsilon > 0$, and find $N \in \mathbb{N}$ s.t. $\|f_{j_k} - f_j\|_\infty \leq \epsilon$ whenever $j, k \geq N$. Then,

$$\begin{aligned} |f(x) - f_j(x)| &= \left| \lim_{k \rightarrow \infty} f_{j_k}(x) - f_j(x) \right| = \lim_{k \rightarrow \infty} |f_{j_k}(x) - f_j(x)| \\ &\leq \limsup_{k \rightarrow \infty} \|f_{j_k} - f_j\|_\infty \leq \epsilon, \end{aligned}$$

for all $x \in \mathbb{R}$ and all $j \geq N$. Thus,

$$\|f - f_j\|_\infty = \sup_{x \in \mathbb{R}} |f(x) - f_j(x)| \leq \epsilon, \quad \forall j \geq N,$$

showing that $f_j \rightarrow f$ in the sup-norm as $j \rightarrow \infty$. Hence, B_σ is complete w.r.t. that norm. \square

7.6 Comparing PW_σ and B_σ

Since $PW_\sigma \subset B_\sigma$, the two spaces are guaranteed to have some similar properties. For example, if functions with a given property exists in PW_σ , they must also exist in B_σ . Also, if there exists at most one function in B_σ with a given property, i.e. if we have uniqueness in B_σ , then there is at most one such function in PW_σ . Another implication is that any general property of all functions in B_σ must also be a general property in PW_σ . In particular, the properties of the sup-norm are satisfied in PW_σ . Thus, $\|\cdot\|_\infty$ is a norm on PW_σ , but it does not have as nice properties as the L^2 -norm. In particular, PW_σ is not complete w.r.t. the sup-norm. However, it turns out that there is an inequality relating the two norms on PW_σ . We will now show that for the more general space PW_S , where the bounded set S might not be an interval centered at the origin.

Lemma 7.24. *If $S \subset \mathbb{R}$ is bounded, and if $f \in PW_S$, then f is bounded on \mathbb{R} , and $\|f\|_\infty \leq \sqrt{\mu(S)}\|f\|_2$.*

Proof. Setting $F := \hat{f}$, we have:

$$\begin{aligned} |f(x)| &= \left| \int_S F(t) e^{2\pi i x t} dt \right| \leq \int_S |F(t) e^{2\pi i x t}| dt = \int_S |F(t)| dt \\ &\leq \left(\int_S 1^2 dt \right)^{\frac{1}{2}} \left(\int_S |F(t)|^2 dt \right)^{\frac{1}{2}} = \sqrt{\mu(S)} \|F\|_2 = \sqrt{\mu(S)} \|f\|_2, \quad \forall x \in \mathbb{R}. \end{aligned}$$

Taking supremum over $x \in \mathbb{R}$, we get the desired result. \square

In particular, if $S = [-\sigma, \sigma]$, lemma 7.24 states that $\|f\|_\infty \leq \sqrt{2\sigma}\|f\|_2$. Comparing theorem 7.13 (i) and (iv) with proposition 7.20 and Bernstein's inequality, this should be no surprise, even though they do not prove it. With the properties we have found, we get the following list of properties that hold for any $\sigma > 0$:

(i) $PW_\sigma = B_\sigma \cap L^2(\mathbb{R})$.

(ii) B_σ , equipped with the sup-norm, is a Banach space.

(iii) PW_σ , equipped with the L^2 -inner product, is a Hilbert space.

For $f \in B_\sigma$:

(iv) $|f(z)| \leq \|f\|_\infty e^{2\pi\sigma|Im(z)|}, \forall z \in \mathbb{C}$.

(v) $f' \in B_\sigma$, and $\|f'\|_\infty \leq 2\pi\sigma\|f\|_\infty$.

(vi) $|f'(z)| \leq 2\pi\sigma\|f\|_\infty e^{2\pi\sigma|Im(z)|}, \forall z \in \mathbb{C}$.

For $f \in PW_\sigma$:

(vii) $|f(z)| \leq \sqrt{2\sigma}\|f\|_2 e^{2\pi\sigma|Im(z)|}, \forall z \in \mathbb{C}$.

(viii) $f' \in PW_\sigma$, and $\|f'\|_2 \leq 2\pi\sigma\|f\|_2$.

(ix) $|f'(z)| \leq 2\pi\sigma\sqrt{2\sigma}\|f\|_2 e^{2\pi\sigma|Im(z)|}, \forall z \in \mathbb{C}$.

(x) $\|f'\|_\infty \leq 2\pi\sigma\sqrt{2\sigma}\|f\|_2$.

Note that since all the properties of f also apply to f' , we can iterate the properties of f' as many times as we like. For example, (viii) can be generalized to: $\|f^{(n)}\|_2 \leq (2\pi\sigma)^n\|f\|_2$ for all $f \in PW_\sigma$.

8 Sampling

Let X be a space of functions and $\Lambda \subset \mathbb{R}$ be a discrete set. Sampling theory addresses the question of whether any $f \in X$ can be uniquely determined if we know its values at Λ , and whether there are any significant changes under small perturbations. We will give a precise formulation of the two main problems in the next section. If f describes a process that we are measuring in real life, we will not be able to measure points arbitrarily close to each other, and our measurements might not be exact. After all, the instruments can only measure with a certain frequency and a certain precision. That is why we will always assume that Λ is *uniformly discrete*, and why we are interested in what happens under small perturbations. This section presents the so-called *weak* and the *strong* sampling problems in general, before concentrating on Paley-Wiener spaces.

8.1 A discussion about sampling

Definition 8.1. Let $\Lambda = \{\lambda_1, \lambda_2, \dots\} \subset \mathbb{R}$ be discrete. If $\delta(\Lambda) := \inf_{k \neq l} |\lambda_k - \lambda_l| > 0$, then Λ is called **uniformly discrete** (u.d.), and $\delta(\Lambda)$ is called the **separation constant** for Λ .

Uniformly discrete sets simply mean sets whose elements never get arbitrarily close to each other, and the separation constant is the the largest lower bound for the distance between them. For example, $\delta(\mathbb{Z}) = \delta(\mathbb{N}) = 1$, while sets like \mathbb{Q} and $\{\frac{1}{k} \text{ s.t. } k \in \mathbb{N}\}$ are not u.d. If we measure in a time-domain with an instrument that can measure 100 times every second, then $\delta(\Lambda)$ must be at least 0.01 seconds. That is, if we start at 0 and measure as often as we can, then

$$\Lambda = \{0, 0.01, 0.02, 0.03, \dots\} = \frac{1}{100}\mathbb{N}_0.$$

When Λ is an arithmetic progression like this, i.e. if the distance between consecutive points is constant, we are dealing with so-called *uniform* sampling. In practice, that is normally what we do, since we want as much data as possible. We will, however, study more general sampling than that, until we focus specifically on uniform sampling in Paley-Wiener space in section 10.

The set Λ is called the *sampling set*, and the values $\{f(\lambda) \text{ s.t. } \lambda \in \Lambda\}$ are called the *samples* of f at Λ . The restriction of f to the sampling set is denoted by $f|_{\Lambda}$. That is, $f|_{\Lambda}$ is defined on Λ by $f|_{\Lambda}(\lambda) := f(\lambda)$, $\forall \lambda \in \Lambda$. As before, we will assume that Λ is ordered, so that we can view Λ and $f|_{\Lambda}$ as sequences. This is important when we work with sampling in Paley-Wiener spaces, since we will encounter a lot of sums over Λ . If Λ is bounded below, we might pick ascending order, which simply means chronologically if the domain is time. Note that with our new notation, sampling is about whether knowledge of $f|_{\Lambda}$ is sufficient to determine f .

Note that a different question is whether f is guaranteed to exist. If Y is a space of sequences indexed by Λ , we could ask whether any sequence in Y is equal to $f|_{\Lambda}$ for some $f \in X$. That is called the *interpolation problem*, which is not addressed in this thesis. Our focus is on uniqueness, plus a stronger property that is related to small perturbations, assuming that the function we are searching for actually exists.

What do we need to know about the function f to have any chance of being able to determine it from knowledge of $f|_\Lambda$? With no assumptions on f at all, one solution would be:

$$f(x) = \begin{cases} f|_\Lambda(x), & x \in \Lambda \\ 0, & x \in \mathbb{R} \setminus \Lambda \end{cases}$$

This solution is probably not the one we are looking for, and it is definitely not unique. Usually, a convenient assumption is that f is continuous. However, that still allows us to draw straight lines between the samples, which is probably not the right solution either. That solution is still not differentiable, though. A good start is to assume that f is differentiable infinitely many times. In any case, it is clear that we need some sort of assumptions on f , which is why we always consider some given function space X .

As mentioned, we can never measure anything with arbitrary frequency. Also, we cannot measure anything forever. Hence, in reality, Λ is always a finite set. This is no problem if the dimension of X is also finite. For example, it is well-known that given any set of $n \in \mathbb{N}$ points in \mathbb{R}^2 , having different x -coordinates, there exists a unique polynomial of degree at most $n - 1$ going through all of them. Thus, as long as Λ contains at least n points, it can be used for sampling in the space of such polynomials. However, if we have no restriction on the degree of the polynomial, we get a problem. In fact, most convenient function spaces have infinite dimension, just like the space of all polynomials. If X has an infinite basis, $\{v_k\}_{k \in \mathbb{N}}$, then we need infinitely many coefficients $c_k \in \mathbb{C}$ to be able to write

$$f = \sum_{k \in \mathbb{N}} c_k v_k, \quad \forall f \in X,$$

and we need infinitely many data points to determine infinitely many coefficients. Since we only consider u.d. sampling sets, we must sample forever to get infinitely many data points. One attempt to get around that problem is to restrict ourselves to a convenient, finite-dimensional subspace of X that we expect to approximate f well. E.g. in Paley-Wiener spaces, there may be finitely many frequencies we expect to get, which allows us to use the discrete Fourier transform. Another idea is to measure finitely many points, and assume the function vanishes on the rest of our infinite sampling set. To be exact, that cannot work if the only element of X vanishing outside a bounded set is the zero-function. That is the case in Paley-Wiener spaces and Bernstein spaces. However, if we get a function that is very small outside that bounded set, it might be a good approximation of the function we are searching for. These are problems that will not be covered in this thesis. We will allow unbounded sampling sets, and see what it leads to mathematically.

8.2 The sampling problem

To give a precise formulation of the sampling problem, we will take advantage of a convenient operator.

Definition 8.2. Let $\Lambda \subset \mathbb{R}$ be a discrete set and X be a space of functions with real or complex domain containing Λ . The **restriction operator** R_Λ is defined by $R_\Lambda f := f|_\Lambda$, $\forall f \in X$.

That is, R_Λ maps f into its samples on Λ . From now on, whenever we denote an operator by R_Λ , we are always talking about the restriction operator. Linearity of R_Λ follows from how linear combinations of functions are defined. However, no other property in definition 3.1 makes sense to talk about before we specify a co-domain Y for R_Λ , and some of the properties also require both X and Y to be normed. The only thing we can say in general about Y is that it must be a space of sequences indexed by Λ . Assuming that Y is normed and R_Λ is injective, it turns out that we have a general way of inducing a norm on X .

Proposition 8.3. *Let $\Lambda \subset \mathbb{R}$ be a discrete set, let X be a space of functions, and let Y be a normed co-domain for R_Λ . If R_Λ is injective, then $\|f\| := \|R_\Lambda f\|_Y, \forall f \in X$ defines a norm on X .*

Proof. (i) Since $\|\cdot\|_Y$ is a norm on Y , we have $\|f\| \geq 0, \forall f \in X$, with equality if and only if $R_\Lambda f = 0$. By linearity and injectivity of R_Λ , this happens if and only if $f = 0$.

$$(ii) \|\alpha f\| = \|R_\Lambda(\alpha f)\|_Y = \|\alpha R_\Lambda f\|_Y = |\alpha| \cdot \|R_\Lambda f\|_Y = |\alpha| \cdot \|f\|, \forall f \in X, \forall \alpha \in \mathbb{C}.$$

$$(iii) \|f + g\| = \|R_\Lambda(f + g)\|_Y = \|R_\Lambda f + R_\Lambda g\|_Y \leq \|R_\Lambda f\|_Y + \|R_\Lambda g\|_Y = \|f\| + \|g\|,$$

$\forall f, g \in X.$ □

We will denote this specific norm by $\|\cdot\|_Y$, even though it is a norm on X rather than on Y . Whenever we use it, it will be clear whether we are taking the norm of a function in X or a sequence in Y . As we know, injectivity of R_Λ means that given a sequence $y \in Y$, there exists at most one function $f \in X$ satisfying $R_\Lambda f = y$. In other words, if it exists, it must be unique. This motivates part (i) of our next definition. Part (ii) is the main reason for introducing the norm given by proposition 8.3, as it defines a very convenient, even stronger, property.

Definition 8.4. *Let $\Lambda \subset \mathbb{R}$ be a u.d. set, let X be a space of functions, and let Y be a co-domain for R_Λ .*

(i) *If R_Λ is injective, then Λ is called a **uniqueness set (US)** for X .*

(ii) *If R_Λ is injective, if X and Y are normed, and if $\|\cdot\|_X$ and $\|\cdot\|_Y$ are equivalent norms on X , then Λ is called a set of **stable sampling (SS)** for X . By equivalent, we mean that there exist constants $c, C > 0$ satisfying*

$$c\|f\|_X \leq \|f\|_Y \leq C\|f\|_X, \forall f \in X. \tag{16}$$

These are exactly the sets that sampling theory tries to determine. The *weak sampling problem* is to find all uniqueness sets for X , and the *strong sampling problem* is to find all sets of stable sampling for X . Note that neither of the two problems involve actually finding a way to reconstruct f from $f|_\Lambda$, but if we can, then it is good to know that it is unique. Our main focus will be on the strong sampling problem. When we get to section 9.5, we will see one aspect of stable sampling that is very convenient. Note that we can check (16) even without knowing that R_Λ is injective. After all, the definition of $\|\cdot\|_Y : X \rightarrow \mathbb{R}_0^+$

makes sense, even if it does not define a norm on X . In fact, if the left inequality is satisfied, then R_Λ is necessarily injective. The reason is that $f|_\Lambda = 0$ still implies $\|f\|_Y = 0$, so the inequality states that $c\|f\|_X \leq 0$, i.e. $f = 0$. Thus, we do not need to check for injectivity, as long as we have verified (16).

Example 8.5. Given a u.d. set $\Lambda \subset \mathbb{R}$, let Y be the space of bounded sequences indexed by Λ . Define

$$\|y\|_Y := \sup_{\lambda \in \Lambda} |f(y)|, \quad \forall y \in Y.$$

Then, the same arguments as in the proof of theorem 7.17 show that $\|\cdot\|_Y$ is a norm on Y . Now, pick a $\sigma > 0$. Clearly, since Bernstein functions are bounded on \mathbb{R} , $R_\Lambda f \in Y$ for all $f \in B_\sigma$. That is, R_Λ maps B_σ into Y . Also, $\|f\|_\infty \geq \|f\|_Y = \|f|_\Lambda\|_Y$ for all $f \in B_\sigma$, since the supremum over a subset of \mathbb{R} cannot exceed the supremum over \mathbb{R} . Thus, the right inequality in (16) is always satisfied, where we may pick $C = 1$. Thus, to check whether Λ is a set of SS for B_σ , we just need to check whether there exists a $c > 0$ s.t.

$$c \sup_{x \in \mathbb{R}} |f(x)| \leq \sup_{\lambda \in \Lambda} |f(\lambda)|, \quad \forall f \in B_\sigma.$$

8.3 Bessel's inequality in Paley-Wiener spaces

We want to consider the strong sampling problem in Paley-Wiener spaces and Bernstein spaces. In order to do so, we need to find a normed co-domain for the restriction operator. Example 8.5 solves that problem for Bernstein spaces in such a way that the left inequality in (16) is all we need to check. It turns out that we can do the same in Paley-Wiener spaces. The following theorem shows us how.

Theorem 8.6 ([OU16], p. 15). *Let $S \subset \mathbb{R}$ be bounded, and let $\Lambda \subset \mathbb{R}$ be u.d. Then, there exists a $C > 0$, only dependent on S and $\delta(\Lambda)$, satisfying $\sum_{\lambda \in \Lambda} |f(\lambda)|^2 \leq C \|f\|_2^2$ for all $f \in PW_S$.*

Proof. Pick a $\sigma \in (0, \frac{1}{2}\delta(\Lambda)]$ satisfying $S \subseteq [-\frac{1}{4\sigma}, \frac{1}{4\sigma}]$. Such a σ always exists, since if we pick it to be sufficiently small, then the interval $[-\frac{1}{4\sigma}, \frac{1}{4\sigma}]$ will be large enough to contain S . We will prove that the theorem holds with $C = \frac{\pi^2}{8\sigma}$. Define $h \in PW_S$ and $\{h_\lambda\}_{\lambda \in \Lambda} \subset PW_S$ by $h := \frac{1}{\sqrt{2\sigma}} \chi_{[-\sigma, \sigma]}$ and $h_\lambda := h(\cdot + \lambda)$, $\forall \lambda \in \Lambda$. Then, we have, for any $t \in \mathbb{R}$:

$$\begin{aligned} \hat{h}(t) &= \int_{-\sigma}^{\sigma} \frac{\sqrt{2\sigma}}{2\sigma} e^{-2\pi i x t} dx = \frac{-\sqrt{2\sigma}}{2\sigma \cdot 2\pi i t} [e^{-2\pi i x t}]_{x=-\sigma}^{x=\sigma} = \sqrt{2\sigma} \cdot \frac{e^{2\pi i \sigma t} - e^{-2\pi i \sigma t}}{2i \cdot 2\pi \sigma t} \\ &= \sqrt{2\sigma} \cdot \frac{\sin(2\pi \sigma t)}{2\pi \sigma t} = \sqrt{2\sigma} \operatorname{sinc}(2\sigma t). \end{aligned}$$

If $t \in S$, then $2\sigma t \in [-\frac{1}{2}, \frac{1}{2}]$. On that interval, the sinc-function only takes positive values, the smallest one being $\frac{2}{\pi}$. Hence, the smallest value \hat{h} can take on S is $\sqrt{2\sigma} \cdot \frac{2}{\pi} = \frac{1}{\sqrt{C}}$.

Now, given any $\lambda \in \Lambda$, h_λ vanishes everywhere except on an interval of length $2\sigma \leq \delta(\Lambda)$. Thus, if $\kappa \in \Lambda$ and $\kappa \neq \lambda$, then h_λ and h_κ cannot both be non-zero at more than one single

point. Hence, $h_\lambda \overline{h_\kappa} = 0$ a.e., so $\langle h_\lambda, h_\kappa \rangle = 0$. Also, $\langle h_\lambda, h_\lambda \rangle = \frac{1}{2\sigma} \int_{-\lambda-\sigma}^{-\lambda+\sigma} dx = 1$. This shows that $\{h_\lambda\}_{\lambda \in \Lambda}$ is an orthonormal system in $L^2(\mathbb{R})$, since $\lambda, \kappa \in \Lambda$ were arbitrary.

Since the Fourier transform on $L^2(\mathbb{R})$ preserves inner-products, $\{\hat{h}_\lambda\}_{\lambda \in \Lambda}$ is an orthonormal system as well. Thus, it satisfies Bessel's inequality by theorem 5.8 (i). Also, by example 3.10, $\hat{h}_\lambda(t) = e^{2\pi i \lambda t} \hat{h}(t)$, $\forall t \in \mathbb{R}, \forall \lambda \in \Lambda$.

If $f \in PW_S$, then $g := \frac{\hat{f}}{\hat{h}} \in L^2(S)$, where we interpret $\frac{0}{0}$ as being zero. Thus,

$$|\langle \bar{g}, h_\lambda \rangle| = \left| \int_S \frac{\hat{f}(t)}{\hat{h}(t)} \hat{h}(t) e^{2\pi i \lambda t} dt \right| = \left| \int_{\mathbb{R}} \hat{f}(t) e^{2\pi i \lambda t} dt \right| = |\mathfrak{G} \hat{f}(\lambda)| = |f(\lambda)|, \forall \lambda \in \Lambda.$$

$$C \|f\|_2^2 = C \|\hat{f}\|_2^2 = C \|g \hat{h}\|_2^2 \geq C \left\| g \cdot \frac{1}{\sqrt{C}} \right\|_2^2 = \|\bar{g}\|_2^2 \geq \sum_{\lambda \in \Lambda} |\langle \bar{g}, h_\lambda \rangle|^2 = \sum_{\lambda \in \Lambda} |f(\lambda)|^2.$$

□

In the next section, we will see that theorem 8.6 actually states that Bessel's inequality is always satisfied in Paley-Wiener space. Note that $\sum_{\lambda \in \Lambda} |f(\lambda)|^2$ is simply the l^2 -norm of $f|_\Lambda$. Hence, one consequence is that the l^2 -norm of $f|_\Lambda$ is finite for any $f \in PW_S$. That is exactly what it means to be an element of $l^2(\Lambda)$. Thus, we can pick $l^2(\Lambda)$ to be the co-domain for the restriction operator, whenever the domain is a Paley-Wiener space. We finish this section by stating this as a theorem.

Corollary 8.7. *Let $S \subset \mathbb{R}$ be bounded, and let $\Lambda \subset \mathbb{R}$ be u.d. Then R_Λ maps PW_S into $l^2(\Lambda)$.*

8.4 Sampling in Paley-Wiener space

By definition, a set of SS for PW_S is a u.d. set $\Lambda \subset \mathbb{R}$ s.t. there exist constants $c, C > 0$ satisfying

$$c \|f\|_2^2 \leq \sum_{\lambda \in \Lambda} |f(\lambda)|^2 \leq C \|f\|_2^2, \forall f \in PW_S. \quad (17)$$

Theorem 8.6 ensures us that the right inequality is always satisfied. Hence, just as for Bernstein spaces, we only need to check the left one.

Since Paley-Wiener spaces are Hilbert spaces, we can talk about whether a sequence of Paley-Wiener functions satisfies Bessel's inequality or its converse. It turns out that the inner product in (9) has a very convenient form. Given a bounded set $S \subset \mathbb{R}$, if $F \in L^2(S)$, and if $f := \mathfrak{G} \bar{F} \in PW_S$, then

$$\langle F, e^{2\pi i \lambda (\cdot)} \rangle = \int_S F(t) e^{-2\pi i \lambda t} dt = \int_S \overline{F(t) e^{2\pi i \lambda t}} dt = \overline{f(\lambda)}, \forall \lambda \in \mathbb{R}. \quad (18)$$

Thus, since $\|f\|_2 = \|\bar{F}\|_2 = \|F\|_2$, a reformulation of (17) is:

$$c\|F\|_2^2 \leq \sum_{\lambda \in \Lambda} |\langle F, e^{2\pi i \lambda(\cdot)} \rangle|^2 \leq C\|F\|_2^2, \quad \forall F \in L^2(S). \quad (19)$$

Hence, theorem 8.6 actually states that $E(\Lambda)$ is a Bessel sequence in $L^2(S)$! Since the Fourier inverse transform preserves inner products, it also states that the Fourier inverse transform of that exponential system is a Bessel sequence in PW_S . Hence, we can think of it as being Bessel's inequality in any of the two spaces.

With this discussion in mind, it is now easy to give a precise characterization of both all US sets and all sets of SS for Paley-Wiener spaces. It turns out that the two sampling problems can be interpreted in the language of frames and complete sequences.

Theorem 8.8. *Let $S \subset \mathbb{R}$ be bounded and $\Lambda \subset \mathbb{R}$ be u.d.*

(i) Λ is a US for PW_S if and only if $E(\Lambda)$ is complete in $L^2(S)$.

(ii) Λ is a set of SS for PW_S if and only if $E(\Lambda)$ is a frame in $L^2(S)$.

Proof. (i) First, assume that Λ is not a US for PW_S . Find a non-zero $f \in PW_S$ that vanishes on Λ , and let $F := \hat{f}$. Then, by (18)

$$\langle \overline{F}, e^{2\pi i \lambda(\cdot)} \rangle = \overline{f(\lambda)} = 0, \quad \forall \lambda \in \Lambda.$$

That is, \overline{F} is orthogonal to $E(\Lambda)$. Since \overline{F} is a non-zero element of $L^2(S)$ by injectivity of \mathfrak{F} , this shows that $E(\Lambda)$ is not complete in $L^2(S)$.

Conversely, assume that $E(\Lambda)$ is not complete in $L^2(S)$, and find a non-zero $F \in L^2(S)$ orthogonal to $E(\Lambda)$. Then, by (18),

$$\overline{f(\lambda)} = \langle F, e^{2\pi i \lambda(\cdot)} \rangle = 0, \quad \forall \lambda \in \Lambda,$$

where $f := \mathfrak{G}\overline{F}$. Hence, by injectivity of \mathfrak{G} , f is a non-zero element of PW_S that vanishes on Λ , showing that Λ is not a US for PW_S .

(ii) As we have seen, Λ is a set of SS for PW_S if and only if (19) is satisfied. This is just (9) for the sequence $E(\Lambda)$, with $A = c$ and $B = C$. \square

This theorem is a good reason to study frames, in particular exponential frames in $L^2(S)$! In fact, not only can we find all sets of SS for PW_S by study frames. We can also reconstruct in PW_S , as long as we are able to find an alternate dual frame! The proof is a generalization of an argument from [Ch10], page 152.

Theorem 8.9. *Let $S \subset \mathbb{R}$ be bounded and $\Lambda \subset \mathbb{R}$ be a set of SS for PW_S . Let $\{v_\lambda\}_{\lambda \in \Lambda}$ be an alternate dual frame to $E(\Lambda)$ for $L^2(S)$, and define $v'_\lambda := \mathfrak{G}\overline{v_\lambda}$, $\forall \lambda \in \Lambda$. Then, the following reconstruction formula holds pointwise on \mathbb{R} :*

$$f = \sum_{\lambda \in \Lambda} f(\lambda)v'_\lambda, \quad \forall f \in PW_S.$$

Proof. Let $f \in PW_S$, and define $F := \hat{f}$. Since $\overline{F} \in L^2(S)$, we can expand it in the alternate dual frame as

$$\overline{F} = \langle \overline{F}, e^{2\pi i \lambda(\cdot)} \rangle v_\lambda = \sum_{\lambda \in \Lambda} \overline{f(\lambda)} v_\lambda = \sum_{k \in \mathbb{N}} \overline{f(\lambda_k)} v_{\lambda_k}.$$

This convergence is in the $L^2(S)$ -norm, so we get:

$$\begin{aligned} \left| f(x) - \sum_{k=1}^n f(\lambda_k) v'_{\lambda_k} \right|^2 &= \left| \int_S F(t) e^{2\pi i x t} dt - \sum_{k=1}^n f(\lambda_k) \int_S \overline{v_{\lambda_k}(t)} e^{2\pi i x t} dt \right|^2 \\ &= \left| \int_S \left(F(t) - \sum_{k=1}^n f(\lambda_k) \overline{v_{\lambda_k}(t)} \right) e^{2\pi i x t} dt \right|^2 \leq \int_S \left| F(t) - \sum_{k=1}^n f(\lambda_k) \overline{v_{\lambda_k}(t)} \right|^2 dt \\ &= \int_S \left| \overline{F(t)} - \sum_{k=1}^n \overline{f(\lambda_k)} v_{\lambda_k}(t) \right|^2 dt \rightarrow 0 \text{ as } n \rightarrow \infty, \forall x \in \mathbb{R}. \end{aligned}$$

Hence, $\sum_{k=1}^n f(\lambda_k) v'_{\lambda_k} \rightarrow f(x)$ as $n \rightarrow \infty$ for all $x \in \mathbb{R}$. □

We will get back to this formula in section 10.2, when we are specifically talking about uniform sampling.

9 The strong sampling problem in PW_σ and B_σ

Our main goal of this section is to understand Beurling's sampling theorem, which gives an almost complete answer to the strong sampling problem for PW_σ and B_σ . In the beginning, we will consider some of the tools that are needed to understand at to prove Beurling's sampling theorem. However, proving it requires some theory that is not presented here, in particular weak convergence of measures. For that reason, we will refer to [OU16], rather than actually proving the theorem. In that book, Paley-Wiener spaces and Bernstein spaces are defined slightly different than here. Therefore, we start section 9.6 by discussing how it changes the theorems before going into them.

9.1 Weak convergence of sequences of u.d. sets

The sampling properties of Paley-Wiener spaces, as we have seen, can be equivalently found by studying frames and complete sequences in $L^2(S)$. There is nothing similar for Bernstein spaces, so their sampling properties must be approached in a different way. We will do it via a concept called *weak limits of translates*. The goal of this section is to understand what that means.

Definition 9.1. (i) Given $A, B \subseteq \mathbb{R}$, pick an $\epsilon > 0$. B is called an ϵ -**perturbation** of A if there exists a bijective function $f: A \rightarrow B$ s.t. $|f(x) - x| \leq \epsilon$, $\forall x \in A$.

(ii) A sequence $\{\Lambda_k\}_{k \in \mathbb{N}}$ of u.d. sets is said to **converge weakly** to a set $\Lambda \subset \mathbb{R}$ if for any $\epsilon > 0$ and any $a, b \in \mathbb{R} \setminus \Lambda$, there exists an $N \in \mathbb{N}$ s.t. $\Lambda \cap (a, b)$ is an ϵ -perturbation of $\Lambda_k \cap (a, b)$ whenever $k \geq N$. Λ is called the **weak limit** of $\{\Lambda_k\}_{k \in \mathbb{N}}$.

Note that if $\Lambda \subset \mathbb{R}$ is discrete, and if $\Gamma \subset \mathbb{R}$ is an ϵ -perturbation of Λ , then part of definition 9.1 (i) says that

$$|\lambda_k - \gamma_k| \leq \epsilon, \forall k \in \mathbb{N} \quad (20)$$

for a convenient choice of ordering. If Λ is u.d. and $\epsilon < \frac{\delta(\Lambda)}{2}$, then the elements of Λ cannot lie within distance ϵ of more than one element of Γ . Surjectivity of f makes sure that an element of Γ always exists on the interval $(\lambda_k - \epsilon, \lambda_k + \epsilon)$, so by our previous comment, there is *exactly* one. This gives us an alternative way of viewing ϵ -perturbations of u.d. sets, as long as ϵ is small enough, and specifically when we consider weak limits.

Another view is obtained by considering distances between real numbers and sets of real numbers:

$$\text{dist}(x, B) := \inf\{|x - y| \text{ s.t. } y \in B\},$$

where $x \in \mathbb{R}$ and $B \subseteq \mathbb{R}$. If $\Lambda, \Gamma \subset \mathbb{R}$ are u.d. sets, then Γ being an ϵ -perturbation of Λ is almost the same as both of the following being true:

$$\text{dist}(\lambda, \Gamma) \leq \epsilon, \forall \lambda \in \Lambda.$$

$$\text{dist}(\gamma, \Lambda) \leq \epsilon, \forall \gamma \in \Gamma.$$

Together, they make sure that every element of any of the sets is close enough to some element of the other set. The missing part is that two elements of Λ can be close enough to the same element of Γ , or the other way round. However, as before, that problem is automatically fixed whenever $\epsilon < \frac{\delta(\Lambda)}{2}$.

Now, intuitively, what is a weak limit? If $J \subset \mathbb{R}$ is a finite set, we denote the number of elements in J by $\#J$. Also, we denote the j 'th element of Λ_k by $(\Lambda_k)_j$. Assume that $\Lambda_k \rightarrow \Lambda$ weakly as $k \rightarrow \infty$, and pick $a, b \in \mathbb{R} \setminus \Lambda$. By (20), for some ordering, $(\Lambda_k)_j \rightarrow \lambda_j$ as $k \rightarrow \infty$ for any $j \in \mathbb{N}$ s.t. $\lambda_j \in (a, b)$. Since we could have picked (a, b) to contain any element of Λ , this must be true for all $j \in \mathbb{N}$. From now on, whenever we know that we have weak convergence, we will assume such an ordering to be used. Bijectivity of the function f in definition 9.1 (i) is equivalent to saying that $\#(\Lambda_k \cap (a, b)) = \#(\Lambda \cap (a, b))$ for sufficiently large $k \in \mathbb{N}$. Thus, a reformulation of definition 9.1 (ii) is that for any $\epsilon > 0$ and any $a, b \in \mathbb{R} \setminus \Lambda$, there exists $N \in \mathbb{N}$ satisfying:

$$\#(\Lambda_k \cap (a, b)) = \#(\Lambda \cap (a, b)), |(\Lambda_k)_j - \lambda_j| \leq \epsilon, \forall k \geq N, \forall j \in \mathbb{N} \text{ s.t. } \lambda_j \in (a, b).$$

The fact that (a, b) contains only finitely elements of Λ ensures us that we may pick the same $N \in \mathbb{N}$ for all the $\lambda_j \in (a, b)$.

It is convenient to note that we may restrict ourselves to considering intervals centered at the origin. That is, we may assume that $-a = b$. To see why, consider an interval $(-R, R)$ containing (a, b) , where $\pm R \notin \Lambda$. For sufficiently large $k \in \mathbb{N}$, we have

$$\#(\Lambda \cap (-R, R)) = \#(\Lambda_k \cap (-R, R)).$$

Assuming $\epsilon < \frac{\delta(\Lambda)}{2}$, the $(\Lambda_k)_j$ approaching $\lambda_j \in (-R, R) \setminus (a, b)$ must themselves be outside (a, b) , so we must have $\#(\Lambda \cap (a, b)) = \#(\Lambda_k \cap (a, b))$ as well. Note that this argument even holds if we restrict R to be a natural number. Hence, we only need to consider a discrete family of open intervals to check for weak convergence. We will take advantage of this observation in the proof of theorem 9.4.

We have seen that if $\Lambda_k \rightarrow \Lambda$ weakly as $k \rightarrow \infty$, then $(\Lambda_k)_j \rightarrow \lambda_j$ in \mathbb{R} for any $j \in \mathbb{N}$. Conversely, if we know that $\{(\Lambda_k)_j\}_{k \in \mathbb{N}}$ converges to λ_j for any $j \in \mathbb{N}$, does that imply weak convergence? In other words, are we guaranteed to have $\#(\Lambda_k \cap (a, b)) = \#(\Lambda \cap (a, b))$ for sufficiently large $k \in \mathbb{N}$, regardless of our choice of interval? Let us check with an example.

Example 9.2. Define the sequence $\{\Lambda_k\}_{k \in \mathbb{N}}$ by:

$$\Lambda_k := \{-\text{mod}(k, 2), 1, 2, 3, \dots, k-1, k+1, \dots\}, \forall k \in \mathbb{N},$$

where $\text{mod}(x, y)$ is the remainder when x is divided by y . The first few elements look like this:

$$\Lambda_1 = \{-1, 2, 3, 4, 5, 6, \dots\}$$

$$\Lambda_2 = \{0, 1, 3, 4, 5, 6, \dots\}$$

$$\Lambda_3 = \{-1, 1, 2, 4, 5, 6, \dots\}$$

$$\Lambda_4 = \{0, 1, 2, 3, 5, 6, \dots\}$$

We note that the lower element jumps between -1 and 0 , so we do not expect $\{\Lambda_k\}_{k \in \mathbb{N}}$ to converge weakly. To absolutely prove it, note that no matter how big $k \in \mathbb{N}$ becomes, the number of elements in the set $\Lambda_k \cap (-\frac{1}{2}, \frac{1}{2})$ does NOT stop jumping between 0 and 1 . However, it turns out to be possible to order the elements in such a way that $\{(\Lambda_k)_j\}_{k \in \mathbb{N}} \subset \mathbb{R}$ converges for any $j \in \mathbb{N}$. Specifically, we define, for all $j \in \mathbb{N}$:

$$(\Lambda_k)_j := \begin{cases} j, & j \in \mathbb{N} \setminus \{k\} \\ -\text{mod}(j, 2), & j = k \end{cases}, \quad \forall k \in \mathbb{N}.$$

Then, Λ_k is the same set as before, for any $k \in \mathbb{N}$. Also, for any fixed $j \in \mathbb{N}$, the sequence $\{(\Lambda_k)_j\}_{k \in \mathbb{N}} \subset \mathbb{R}$ converges to j . Thus, we conclude that convergence of each sequence of corresponding terms does not imply weak convergence of the sequence of u.d. sets.

It should be noted, though, that there exists a subsequence converging weakly. Specifically, it is not hard to see that the subsequence $\{\Lambda_{2k}\}_{k \in \mathbb{N}}$ converges weakly to \mathbb{N}_0 . Also, $\{\Lambda_{2k-1}\}_{k \in \mathbb{N}}$ converges weakly to $\mathbb{N} \cup \{-1\}$.

Now, what can we say in general about the weak limit? It is natural to think that it is u.d., since points that are always far from each other cannot converge to points that are close to each other. That turns out to be the case. What can we say about the separation constant? This is what our next lemma is all about.

Lemma 9.3 ([OU16], p. 27)). *Let $\{\Lambda_k\}_{k \in \mathbb{N}}$ be a sequence of u.d. sets converging weakly to some set $\Lambda \subset \mathbb{R}$. Then, Λ is u.d., and*

$$\delta(\Lambda) \geq \limsup_{k \rightarrow \infty} \delta(\Lambda_k).$$

Proof. Choose different numbers $n, m \in \mathbb{N}$. Pick an $\epsilon > 0$ and find an $N \in \mathbb{N}$ s.t.

$$|(\Lambda_k)_n - \lambda_n| \leq \epsilon, \quad |(\Lambda_k)_m - \lambda_m| \leq \epsilon, \quad \forall k \geq N.$$

Then, by the reverse triangle inequality, we have:

$$\begin{aligned} |\lambda_n - \lambda_m| &= |((\Lambda_k)_n - (\Lambda_k)_m) - (\lambda_m - (\Lambda_k)_m - \lambda_n + (\Lambda_k)_n)| \\ &\geq |(\Lambda_k)_n - (\Lambda_k)_m| - |((\Lambda_k)_n - \lambda_n) + (\lambda_m - (\Lambda_k)_m)| \\ &\geq |(\Lambda_k)_n - (\Lambda_k)_m| - (|(\Lambda_k)_n - \lambda_n| + |\lambda_m - (\Lambda_k)_m|) \geq \delta(\Lambda_k) - 2\epsilon \end{aligned}$$

whenever $k \geq N$. Hence,

$$|\lambda_n - \lambda_m| \geq \limsup_{k \rightarrow \infty} \delta(\Lambda_k) - 2\epsilon,$$

and since this holds for any $\epsilon > 0$, we conclude that

$$|\lambda_n - \lambda_m| \geq \limsup_{k \rightarrow \infty} \delta(\Lambda_k).$$

This holds for any $n, m \in \mathbb{N}$ with $n \neq m$, so Λ is u.d. with the desired lower bound for $\delta(\Lambda)$. \square

This shows, in particular, that if all the Λ_k have the same separation constant, then the limit cannot have a smaller one. We will see in example 9.5 that it might be strictly larger. Note that if $\delta(\Lambda_k) \rightarrow 0$ as $k \rightarrow \infty$, then it MUST be strictly larger, as we cannot have $\delta(\Lambda) = 0$. However, if the $\delta(\Lambda_k)$ do not become arbitrarily small, then it turns out that the sequence satisfies a property that is very similar to Bolzano-Weierstrass' lemma for Euclidean space, which states that any bounded sequence has a convergent subsequence. In fact, our next theorem is a consequence of that lemma.

Theorem 9.4 ([OU16], p. 28). *Let $\{\Lambda_k\}_{k \in \mathbb{N}}$ be a sequence of u.d. sets satisfying $\limsup_{k \rightarrow \infty} \delta(\Lambda_k) > 0$. Then, $\{\Lambda_k\}_{k \in \mathbb{N}}$ has a weakly convergent subsequence.*

Proof. By passing down to a subsequence if necessary, we may assume that there exists a $\delta > 0$ s.t. $\delta(\Lambda_k) \geq \delta$ for all $k \in \mathbb{N}$. For $N \in \mathbb{N}$, define $I_N := (-N, N)$, except that if a sequence of elements in the Λ_k converges to N or $-N$, we shift that endpoint by $\min\{\frac{1}{2}, \frac{\delta}{2}\}$ units towards the origin. Now, define

$$n_N := \liminf_{k \rightarrow \infty} \#(\Lambda_k \cap I_N), \quad \forall N \in \mathbb{N}.$$

Pick an ordering s.t. for any $N \in \mathbb{N}$, the first n_N elements of (Λ_k) all lie inside I_N for sufficiently large $k \in \mathbb{N}$. Obviously, the sequence $\{(\Lambda_k)_j\}_{k \in \mathbb{N}}$ is bounded for all $j \in \{1, 2, \dots, n_1\}$. Hence, by Bolzano-Weierstrass' lemma, $\{(\Lambda_k)_1\}_{k \in \mathbb{N}}$ has a subsequence converging to a $c_1 \in I_1$. Then, the corresponding subsequence of $\{(\Lambda_k)_2\}_{k \in \mathbb{N}}$ itself has a subsequence converging to a $c_2 \in I_1$. Continuing this process n_1 times, we get a subsequence $\{\Lambda_{k_m}\}_{m \in \mathbb{N}}$ of $\{\Lambda_k\}_{k \in \mathbb{N}}$ s.t. $\{(\Lambda_{k_m})_j\}_{m \in \mathbb{N}}$ converges to $c_j \in I_1$ for all $j \in \{1, 2, \dots, n_1\}$. We will denote the indexes of that subsequence by $k_{1,m}$, i.e. $(\Lambda_{k_{1,m}})_j \rightarrow c_j$, for all $j \leq n_1$, as $m \rightarrow \infty$.

Now, we repeat the same process for the set I_2 to get a sub-indexing $\{k_{2,m}\}_{m \in \mathbb{N}}$, s.t. for all $j \in \{n_1 + 1, n_1 + 2, \dots, n_1 + n_2\}$, the sequence $(\Lambda_{k_{2,m}})_j$ converges to a $c_j \in I_2$. By sub-indexing, we mean that the indexing sequence, $\{k_{2,m}\}_{m \in \mathbb{N}}$, is a subsequence of $\{k_{1,m}\}_{m \in \mathbb{N}}$. This ensures us that for $j \leq n_2$, $(\Lambda_{k_{2,m}})_j \rightarrow c_j$ as $m \rightarrow \infty$, even if $j \leq n_1$.

For $N > 2$, doing the process for I_N , we get an indexing sequence, $\{k_{N,m}\}_{m \in \mathbb{N}}$, satisfying the following two criteria:

- (i) $\{k_{N,m}\}_{m \in \mathbb{N}}$ is a subsequence of $\{k_{N-1,m}\}_{m \in \mathbb{N}}$.
- (ii) $(\Lambda_{k_{N,m}})_j$ converges to a $c_j \in I_N$, for all $j \leq n_N$, as $m \rightarrow \infty$.

By construction, the c_j do not change as N increases. This yields a u.d. set $\Lambda \subset \mathbb{R}$, defined by $\lambda_j := c_j$, for all $j \in \mathbb{N}$.

We need to construct an indexing $\{k_m\}_{m \in \mathbb{N}}$ s.t. $\Lambda_{k_m} \rightarrow \Lambda$ weakly as $k \rightarrow \infty$. Let us look at the indexing sequences we have constructed so far.

$$\begin{array}{cccc} \mathbf{k}_{1,1} & k_{1,2} & k_{1,3} & \dots \\ k_{2,1} & \mathbf{k}_{2,2} & k_{2,3} & \dots \\ k_{3,1} & k_{3,2} & \mathbf{k}_{3,3} & \dots \\ \dots & \dots & \dots & \dots \end{array}$$

Here, we have highlighted the diagonal indexes, $\{k_{m,m}\}_{m \in \mathbb{N}}$, as these are the ones we will be using. Let $k_m := k_{m,m}$ for all $m \in \mathbb{N}$. Then, for any $N \in \mathbb{N}$, $\{k_m\}_{m \in \mathbb{N}}$ is a subsequence

of $\{k_{N,m}\}$, except for the first $N - 1$ terms. Thus, $(\Lambda_{k_m})_j \rightarrow c_j$, for all $j \in \mathbb{N}$, as $m \rightarrow \infty$. Also, by construction, Λ and Λ_{k_m} contain the same number of elements on each of the I_N for sufficiently large $m \in \mathbb{N}$. Thus, $\{\Lambda_{k_m}\}_{m \in \mathbb{N}}$ does indeed converge weakly to Λ . \square

Example 9.5. Let $\Lambda_k := \mathbb{Z} \cup \{k - \frac{1}{2}\}$ for $k \in \mathbb{N}$. Then, $\delta(\Lambda_k) = \frac{1}{2}$ for any $k \in \mathbb{N}$, since they all contain a point in the middle of two consecutive integers. Now, for any $k \in \mathbb{N}$, if I is an interval whose right endpoint is less than $k - \frac{1}{2}$, then $\Lambda_k \cap I$ contains exactly the integers lying on I . That is, $\Lambda_k \cap I = \mathbb{Z} \cap I$. Hence, as $k \rightarrow \infty$, we see that $\Lambda_k \rightarrow \mathbb{Z}$ weakly. However, $\delta(\mathbb{Z}) = 1 > \frac{1}{2}$, showing that a sequence of u.d. sets with equal separation constant may converge weakly to a set of larger separation constant.

Example 9.6. Let $\Lambda \subset \mathbb{R}$ be u.d., and let $\{c_k\}_{k \in \mathbb{N}} \subset \mathbb{R}$ be a sequence of numbers converging to some $c \in \mathbb{R}$. Let $\{\Lambda_k\}_{k \in \mathbb{N}} := \{\Lambda - c_k\}_{k \in \mathbb{N}}$ be the corresponding sequence of translates of Λ . Order the elements of Λ_k s.t. $(\Lambda_k)_j = \lambda_j - c_k$ for all $k, j \in \mathbb{N}$. Pick an interval $(a, b) \subset \mathbb{R}$ with $a, b \notin \Lambda - c$, and an $\epsilon > 0$ small enough that (a, b) and $(a, b) \pm \epsilon$ contain exactly the same elements of Λ . For sufficiently large $k \in \mathbb{N}$, we have:

$$|(\Lambda_k)_j - (\Lambda - c)_j| = |(\lambda_j - c_k) - (\lambda_j - c)| = |c - c_k| \leq \epsilon, \forall j \in \mathbb{N}.$$

Hence, $(\Lambda_k)_j \rightarrow (\Lambda - c)_j$ as $k \rightarrow \infty$. Also, by construction,

$$\#\Lambda_k \cap (a, b) = \#(\Lambda - c) \cap (a, b)$$

for equally large $k \in \mathbb{N}$. Hence, Λ_k converges weakly to $\Lambda - c$.

9.2 Weak limits of translates

The goal of this section is to understand what weak limits we can get from sequences of translates of a given u.d. set. That is, if $\Lambda \subset \mathbb{R}$ is a given u.d. set, and if $\{c_k\}_{k \in \mathbb{N}} \subset \mathbb{R}$ is an arbitrary sequence of numbers, what can the weak limit of $\{\Lambda - c_k\}_{k \in \mathbb{N}}$ possibly be if it exists? To get closer to an answer, we start by looking at some special cases of translates.

We have seen in example 9.6 that if $c_k \rightarrow c \in \mathbb{R}$ as $k \rightarrow \infty$, then $\Lambda - c_k \rightarrow \Lambda - c$ weakly as $k \rightarrow \infty$. This shows that we can get any translate of Λ , and a very easy, general way of generating the sequence. The easiest is to pick $\{c_k\}_{k \in \mathbb{N}}$ to be a constant sequence, but it can be any sequence converging to the number we want to translate by.

If $\{c_k\}_{k \in \mathbb{N}}$ does not converge, we can still make $\{\Lambda - c_k\}_{k \in \mathbb{N}}$ converge weakly. For example, if $\Lambda = \mathbb{Z}$ and all the c_k are integers, then all the $\Lambda - c_k$ are just \mathbb{Z} . Can we also make it converge to something different than a translate of Λ ? Let us look at an example.

Example 9.7. Let $\Lambda_k := \mathbb{N} - k$ and $\Gamma_k := \mathbb{N} + k$, $\forall k \in \mathbb{N}$. The first few elements of $\{\Lambda_k\}_{k \in \mathbb{N}}$ are:

$$\Lambda_1 = \{0, 1, 2, 3, \dots\}$$

$$\Lambda_2 = \{-1, 0, 1, 2, 3, \dots\}$$

$$\Lambda_3 = \{-2, 1, 0, 1, 2, 3, \dots\}$$

Any integer will be captured by Λ_k for sufficiently large $k \in \mathbb{N}$. Obviously, no non-integer is contained in any of the Λ_k , so we conclude that $\Lambda_k \rightarrow \mathbb{Z}$ weakly as $k \rightarrow \infty$.

Now, let us look at the first few elements of $\{\Gamma_k\}_{k \in \mathbb{N}}$.

$$\Gamma_1 = \{2, 3, 4, 5, 6, \dots\}$$

$$\Gamma_2 = \{3, 4, 5, 6, \dots\}$$

$$\Gamma_3 = \{4, 5, 6, \dots\}$$

This time, any integer disappears for sufficiently large $k \in \mathbb{N}$. Again, no non-integer is ever captured, so we conclude that $\Gamma_k \rightarrow \emptyset$ weakly, i.e. the empty set, as $k \rightarrow \infty$.

Example 9.7 shows that $\{\Lambda - c_k\}_{k \in \mathbb{N}}$ may converge weakly, even if $\{c_k\}_{k \in \mathbb{N}}$ is unbounded! If we replace \mathbb{N} by any other set that is bounded below, we will still get the empty set by putting $c_k := -k$ for $k \in \mathbb{N}$. Similarly, if Λ is bounded above, $c_k := k$, $\forall k \in \mathbb{N}$ would give the empty set. Had we picked $\Lambda := \mathbb{Z}$, though, we would just get \mathbb{Z} back again in both cases, since we never gain or lose any integer. The fact that \mathbb{Z} is neither bounded above nor below makes sure that even if $\{c_k\}_{k \in \mathbb{N}}$ is unbounded, we will never lose all the elements larger/smaller than a fixed number. However, it turns out that even sets that are unbounded both above and below may have the empty set as a weak limit of translates!

Example 9.8. Let $I := \{[2^{2n}, 2^{2n+1}] \cup [-2^{2n+1}, -2^{2n}] \text{ s.t. } n \in \mathbb{N}_0\}$, and let $\Lambda := \mathbb{Z} \cap I$. That is, Λ contains the integers on the intervals

$$[1, 2], [4, 8], [16, 32], [64, 128], [256, 512], \dots$$

and on the corresponding negative intervals. Define $a_k := \frac{2^{2k+1} + 2^{2k}}{2}$, $b_k := \frac{2^{2k} + 2^{2k-1}}{2}$ and $c_k := 2^{2k} - 1$ for $k \in \mathbb{N}$. Then, we have:

$$I - a_1 = I - \frac{8 + 4}{2} = I - 6 = \dots \cup [-5, -4] \cup [-2, 2] \cup [10, 26] \cup \dots$$

$$I - a_2 = I - \frac{32 + 16}{2} = I - 24 = \dots \cup [-20, -16] \cup [-8, 8] \cup [40, 104] \cup \dots$$

$$I - a_3 = I - \frac{128 + 64}{2} = I - 96 = \dots \cup [-80, -64] \cup [-32, 32] \cup [160, 416] \cup \dots$$

In general, $I - a_k$ contains the interval $[-2^{2k-1}, 2^{2k-1}]$. Hence, any real number is contained in $I - a_k$ for large enough $k \in \mathbb{N}$. Similarly, any integer is contained in $\Lambda - a_k$ for sufficiently large $k \in \mathbb{N}$. Since no number between them is ever captured, we conclude that $\Lambda - a_k \rightarrow \mathbb{Z}$ weakly as $k \rightarrow \infty$.

For the second sequence, we have:

$$I - b_1 = I - \frac{4 + 2}{2} = I - 3 = \dots \cup [-2, -1] \cup [1, 5] \cup \dots$$

$$I - b_2 = I - \frac{16 + 8}{2} = I - 12 = \dots \cup [-8, -4] \cup [4, 20] \cup \dots$$

$$I - b_3 = I - \frac{64 + 32}{2} = I - 48 = \dots \cup [-32, -16] \cup [16, 80] \cup \dots$$

In general, elements of the interval $(-2^{2k-2}, 2^{2k-2})$ are NOT contained in $I - b_k$. Similarly, $\Lambda - b_k$ does not capture any integer (or non-integer) on that interval, which becomes arbitrarily large as $k \rightarrow \infty$. Hence, $\Lambda - b_k \rightarrow \emptyset$ weakly as $k \rightarrow \infty$.

The third sequence yields:

$$I - c_1 = I - 3 = \dots \cup [-2, -1] \cup [1, 5] \cup \dots$$

$$I - c_2 = I - 15 = \dots \cup [-11, -7] \cup [1, 17] \cup \dots$$

$$I - c_3 = I - 63 = \dots \cup [-47, -31] \cup [1, 65] \cup \dots$$

In general, $I - c_k$ contains $[1, 2^{2k} + 1]$, but no elements of $(-2^{2k-1} + 1, 1)$. Restricting ourselves to integers, the same is true for $\Lambda - c_k$. Any number greater than or equal to 1 is contained in the former interval for sufficiently large $k \in \mathbb{N}$, while any number less than 1 is captured by the latter. Thus, we conclude that $\Lambda - c_k \rightarrow \mathbb{N}$ weakly as $k \rightarrow \infty$.

Note what allowed us to get both the integers and the empty set as weak limits in example 9.8. Λ contained the integers on infinitely many intervals whose length became arbitrarily large, allowing us to get \mathbb{Z} as a weak limit. The fact that the empty intervals also became arbitrarily large is what gave us the empty set as a weak limit. Also, in the example, translating Λ by the sequence $\{3, 6, 12, 24, 48, 96, \dots\} = \{3 \cdot 2^{k-1}\}_{k \in \mathbb{N}}$, we get both \mathbb{Z} and \emptyset as subsequential limits.

The set of ALL weak limits of translates of a u.d. set Λ is denoted by $W(\Lambda)$. As we have seen, $W(\Lambda)$ contains all translates of Λ . More generally, if $\Gamma \in W(\Lambda)$, then any translate of Γ is in $W(\Lambda)$ as well. After all, if $\Lambda - c_k \rightarrow \Gamma$ weakly as $k \rightarrow \infty$, then $\Lambda - (c_k + a)$ converges weakly to $\Gamma - a$. However, we have also demonstrated that even if Λ contains only integers, $W(\Lambda)$ may contain very different sets. In example 9.8, it contained both \mathbb{Z} , which has a separation of 1, and \emptyset , which has infinite separation. This is an example of how different separation in different regions may give weak limits of different separation. If the separation within the regions is not constant, it gets even trickier. However, in any case, a good indication is to look for different regions having the same separation constant, and see how the size of these regions changes.

9.3 Lower uniform densities

In this subsection, when considering a u.d. set $\Lambda \subset \mathbb{R}$, we define

$$n(R) := \inf_{a \in \mathbb{R}} \#(\Lambda \cap (a, a + R)), \quad \forall R > 0.$$

It satisfies:

- (i) $n(R_2) \geq n(R_1)$ whenever $R_2 \geq R_1 > 0$.
- (ii) $n(R_1 + R_2) \geq n(R_1) + n(R_2)$ for all $R_1, R_2 > 0$.
- (iii) $n(R) \leq \frac{R}{\delta(\Lambda)}$, rounded up, for all $R > 0$.

Property (iii) is a direct result of the fact that the elements cannot be arbitrarily close to each other. (ii) follows from the following observation: If I is an open interval of length

$R_1 + R_2$, containing n elements of Λ , then there exist two open intervals I_1, I_2 of length R_1 and R_2 , respectively, s.t. $\Lambda \cap (I_1 \cup I_2) = \Lambda \cap I$. The intervals may be picked so that they do not contain any common element of Λ . Hence, if I is the interval of length $R_1 + R_2$ that minimizes $\#(\Lambda \cap I)$, then I_1 and I_2 are two particular intervals of length R_1 and R_2 , respectively, satisfying $n(R_1 + R_2) \geq \#(\Lambda \cap I_1) + \#(\Lambda \cap I_2)$. Since $\#(\Lambda \cap I_1) \geq n(R_1)$ and $\#(\Lambda \cap I_2) \geq n(R_2)$, the property follows.

For any $R > 0$, applying property (i) multiple times, we have $\frac{n(2^k R)}{2^k R} \geq \frac{n(2^l R)}{2^l R}$ whenever $k \geq l$. That is, $\{\frac{n(2^k R)}{2^k R}\}_{k \in \mathbb{N}}$ is an increasing sequence. Since (iii) makes sure that it is also bounded above by $\frac{1}{\delta(\Lambda)}$, it must converge to some limit, $L \in [0, \frac{1}{\delta(\Lambda)}]$. It turns out that L is independent of the choice of R ([OU16], p. 29). This is exactly what we need to define the density used to formulate Beurling's sampling theorem in section 9.6.

Definition 9.9. If $\Lambda \subset \mathbb{R}$ is u.d., the **lower uniform density** of Λ is defined by

$$D^-(\Lambda) := \lim_{R \rightarrow \infty} \inf_{a \in \mathbb{R}} \frac{\#(\Lambda \cap (a, a + R))}{R} = \lim_{R \rightarrow \infty} \frac{n(R)}{R}.$$

If we replace inf by sup in definition 9.9, we get the *upper uniform density* $D^+(\Lambda)$, but we will not use that here. Step by step, what we do is:

- (i) Pick an arbitrary $R > 0$.
- (ii) Count the number of elements of Λ on every open interval of length R . The count is always at least 0 and at most $\frac{R}{\delta(\Lambda)}$ rounded up. Hence, the set of counts is bounded, so it has both supremum and infimum. We pick the infimum to get $n(R)$.
- (iii) Divide by R to get the smallest density that Λ can have on intervals of length R .
- (iv) Let $R \rightarrow \infty$.

Note that if Λ is an arithmetic progression, then $D^-(\Lambda) = \frac{1}{\delta(\Lambda)}$. Also, if we fix the separation constant, there is no way that we can make a u.d. set any sparser than the case of an arithmetic progression. This observation reflects property (iii) at the beginning of the subsection, and it helps us comparing the two quantities. The reciprocal of the separation constant is the largest number of terms we can have on an open interval I , relative to the length of I . In that sense, it can be thought of as an upper density, so the lower uniform density is kind of an opposite. With that interpretation, it is no surprise that the two coincide when the separation is constant. An important difference, though, is that in the *uniform* densities, we let the length of I go to infinity. For $\frac{1}{\delta(\Lambda)}$, on the other hand, we can just choose I so that it maximizes $\frac{\#(\Lambda \cap I)}{\mu(I)}$. E.g. if Λ is bounded above or below, then $D^-(\Lambda) = 0$. The same is true if there exists a sequence of empty regions growing arbitrarily large, as in example 9.8. However, that does not mean that $\delta(\Lambda)$ is infinite, since other regions have finite separation. As long as a set contains at least two points, there is no doubt that $\delta(\Lambda) < \infty$.

Example 9.10. Let $\Lambda := \mathbb{Z} \setminus 3\mathbb{Z} = \{\dots, -8, -7, -5, -4, -2, -1, 1, 2, 4, 5, 7, 8, \dots\}$. Then, we can verify that the counting function n satisfies, for all $k \in \mathbb{N}$:

$$n(N) = \begin{cases} 2k, & N = 3k \\ 2k & N = 3k + 1 \\ 2k + 1, & N = 3k + 2 \end{cases}$$

In particular, since $n(3) = n(4) = 2$, we obviously have $\frac{n(3)}{3} > \frac{n(4)}{4}$. Hence, $\frac{n(R)}{R}$ is not an increasing function of R . However, $\frac{n(3k)}{3k} = \frac{2k}{3k} = \frac{2}{3}$ for all $k \in \mathbb{N}$. Also, for $R \in (3k-3, 3k)$, we have $\frac{n(3k-3)}{3k} < \frac{n(R)}{R} < 2$. Since $\frac{n(3k-3)}{3k} = \frac{2k-2}{3k} \rightarrow \frac{2}{3}$ as $k \rightarrow \infty$, the squeeze lemma shows that $\frac{n(R)}{R} \rightarrow \frac{2}{3}$ as $R \rightarrow \infty$. Hence, the lower uniform density does exist, and $D^-(\Lambda) = \frac{2}{3}$.

Example 9.10 demonstrates the fact that $\frac{n(R)}{R}$ is not an increasing function of R , even though the sequence $\{\frac{n(2^k R)}{2^k R}\}_{k \in \mathbb{N}}$ is for fixed $R > 0$. The reason is that increasing R by a sufficiently small amount will normally not capture any new elements of Λ . Hence, the fraction will decrease in most small regions. However, we have also demonstrated (without proof) that as $R \rightarrow \infty$, it does not decrease enough for the limit to not exist.

Before turning back to the sampling problems, it is convenient to note that bounded perturbations do not change the lower uniform density. That is, if $\Lambda \subset \mathbb{R}$ is u.d., and if there exists a $\delta > 0$ s.t. $\delta_k \in [0, \delta]$ for all $k \in \mathbb{N}$, then $\Gamma := \{\lambda_k + \delta_k \text{ s.t. } k \in \mathbb{N}\}$ satisfies $D^-(\Gamma) = D^-(\Lambda)$. The reason is that for any interval $I \subset \mathbb{R}$, no matter how large it is, the number of elements inside I cannot change by more than a fixed amount.

Specifically, we can get at most $\delta\delta(\Lambda)$ new elements from the above, and similarly from below, so $|\#(\Lambda \cap I) - \#(\Gamma \cap I)| \leq 2\delta\delta(\Lambda)$. Hence, as $\mu(I) \rightarrow \infty$, the changes approach zero relative to $\mu(I)$. Of course, we are assuming that two points never overlap under the perturbations. The separation constant may change, since that does not involve a limit, but the lower uniform density stays fixed.

9.4 Sampling in B_σ

We are now ready to give a lemma about sampling sets for Bernstein spaces. It does not immediately help us identifying the uniqueness sets or the sets of SS, but it allows us move the problem into the space of weak limits of translates.

Lemma 9.11 ([OU16], p. 28). *Given a u.d. set $\Lambda \subset \mathbb{R}$ and a $\sigma > 0$, we have:*

- (i) *If Λ is a set of SS for B_σ , then every element of $W(\Lambda)$ is a set of SS for B_σ .*
- (ii) *If every element of $W(\Lambda)$ is a US for B_σ , then Λ is a set of SS for B_σ .*

proof of (ii). Assume that Λ is not a set of SS for B_σ . Then, there exists a sequence $\{f_k\}_{k \in \mathbb{N}} \subset B_\sigma$ s.t.

$$\|f_k\|_\infty = 1, \|f_k|_\Lambda\|_\infty \leq \frac{1}{k}, \forall k \in \mathbb{N}.$$

Since all the f_k take values arbitrarily close to 1 on \mathbb{R} , there exists a sequence $\{c_k\}_{k \in \mathbb{N}} \subset \mathbb{R}$ s.t. $\{f_k(c_k)\}_{k \in \mathbb{N}} \subset \{z \in \mathbb{C} \text{ s.t. } |z| \leq 1\}$ converges to 1 as $k \rightarrow \infty$. Now, define the sequence $\{g_k\}_{k \in \mathbb{N}} \subset B_\sigma$ by

$$g_k(z) := f_k(z + c_k), \forall z \in \mathbb{C}, \forall k \in \mathbb{N}.$$

By passing down to a subsequence if necessary, theorem 9.4 makes sure that $\{\Lambda - c_k\}_{k \in \mathbb{N}}$ converges weakly to some $\Gamma \in W(\Lambda)$. Also, again possibly passing down to a subsequence, $\{g_k\}_{k \in \mathbb{N}}$ converges pointwise on \mathbb{C} to some $g \in B_\sigma$ by the compactness property. Pick a $j \in \mathbb{N}$, and let $E := [\gamma_j - 1, \gamma_j + 1]$. We need to show that $g_k((\Lambda - c_k)_j) \rightarrow g(\gamma_j)$ as $k \rightarrow \infty$. Hence, pick an $\epsilon > 0$, and find an $N_1 \in \mathbb{N}$ and a $\delta \in (0, 1]$ satisfying:

$$|g(x) - g_k(x)| \leq \frac{\epsilon}{2}, \forall k \geq N_1, \forall x \in E.$$

$$|g(x) - g(\gamma_j)| \leq \frac{\epsilon}{2}, \forall x \in [\gamma_j - \delta, \gamma_j + \delta].$$

N_1 exists because $g_k \rightarrow g$ uniformly on E , while δ exists because g is continuous at γ_j . Also, find $N_2 \in \mathbb{N}$ s.t.

$$|\gamma_j - (\Lambda - c_k)_j| \leq \delta, \forall k \geq N_2.$$

Set $N := \max\{N_1, N_2\}$ and $x_k := (\Lambda - c_k)_j$ for $k \in \mathbb{N}$. Then, since $x_k \in E$ whenever $k \geq N$, we have:

$$\begin{aligned} |g(\gamma_j) - g_k(x_k)| &= |(g(\gamma_j) - g(x_k)) + (g(x_k) - g_k(x_k))| \\ &\leq |g(\gamma_j) - g(x_k)| + |g(x_k) - g_k(x_k)| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \forall k \geq N. \end{aligned}$$

This shows that indeed, $g_k(x_k) = g_k((\Lambda - c_k)_j) \rightarrow g(\gamma_j)$ as $k \rightarrow \infty$.

Now, define $j : \mathbb{N} \rightarrow \mathbb{N}$ so that $(\Lambda - c_k)_j = \lambda_{j(k)} - c_k$, $\forall k \in \mathbb{N}$. We get:

$$\begin{aligned} \lim_{k \rightarrow \infty} |f_k(\lambda_{j(k)})| &\leq \lim_{k \rightarrow \infty} \|f_k|_\Lambda\|_\infty \leq \lim_{k \rightarrow \infty} \frac{1}{k} = 0. \\ g(\gamma_j) &= \lim_{k \rightarrow \infty} g_k(\lambda_{j(k)} - c_k) = \lim_{k \rightarrow \infty} f_k(\lambda_{j(k)} - c_k + c_k) = \lim_{k \rightarrow \infty} f_k(\lambda_{j(k)}) = 0. \end{aligned}$$

Since $j \in \mathbb{N}$ was arbitrary, this shows that $g|_\Gamma = 0$. However, $g \neq 0$, since

$g_k(0) = f_k(c_k) \rightarrow 1 \neq 0$ as $k \rightarrow \infty$. Thus, we conclude that Γ is not a US for B_σ . \square

Note that one statement of lemma 9.11 is that if all elements of $W(\Lambda)$ are uniqueness sets, then they are also sets of SS. The reason is that by (ii), it implies that Λ is a set of SS, which by (i) implies that the elements of $W(\Lambda)$ are sets of SS. This is NOT the same as saying that in B_σ , being a US and being a set of SS are equivalent properties, as the following example shows. It is only guaranteed to be the case if ALL elements of $W(\Lambda)$ are uniqueness sets, including Λ itself. The remaining statement of the lemma is, of course, that Λ is that a set of SS for B_σ if and only if every element of $W(\Lambda)$ is.

Example 9.12. Let $\sigma := \frac{1}{2}$, and let $\Lambda := \mathbb{Z} \cup \{\frac{1}{2}\}$. It can be shown that if $f \in B_\sigma$ vanishes on \mathbb{Z} , then there exists a $C > 0$ s.t. $f(z) = C \sin(\pi z)$, $\forall z \in \mathbb{C}$ ([OU16], p. 18). Thus, if $f(\frac{1}{2}) = 0$ as well, then $f = 0$. This shows that Λ is a US for B_σ , while \mathbb{Z} is not. Example 9.5 shows that $\mathbb{Z} \in W(\Lambda)$. Hence, if Λ were a set of SS for B_σ , then lemma 9.11 would imply that \mathbb{Z} is a set of SS for B_σ , which is a contradiction. We conclude that Λ is a US set for B_σ , but not a set of SS.

9.5 Perturbations of sets of SS

In this subsection, we will consider differences between sequences of different indexings. The elements of that difference are defined in terms of the ordering when \mathbb{N} is used as the indexing set. That is, by $f|_{\Gamma} - f|_{\Lambda}$, we mean the sequence $\{f(\gamma_k) - f(\lambda_k)\}_{k \in \mathbb{N}}$. Note that the restriction operator can be defined the same way independently of which indexing set we choose.

Recall that if two u.d. sets $\Lambda, \Gamma \subset \mathbb{R}$ are ϵ -perturbations of each other if any element of any of the sets is within distance ϵ of an element of the other set. If $\epsilon < \min\{\frac{\delta(\Lambda)}{2}, \frac{\delta(\Gamma)}{2}\}$, then that condition is sufficient. Intuitively, when we say something is *stable*, we would think that it does not change under sufficiently small perturbations. In that sense, what do we mean by stable sampling, which we defined in a different way?

Let X be a normed function space, and let Y be a normed co-domain for the restriction operator. If $\Lambda, \Gamma \subset \mathbb{R}$ are u.d. sets, with Λ being a set of SS for X , and if we find $C > 0$ in (16), then

$$\|f|_{\Gamma}\|_Y = \|f|_{\Gamma} - f|_{\Lambda} + f|_{\Lambda}\|_Y \geq \|f|_{\Lambda}\|_Y - \|f|_{\Gamma} - f|_{\Lambda}\|_Y \geq C\|f\|_X - \|f|_{\Gamma} - f|_{\Lambda}\|_Y, \quad \forall f \in X.$$

Thus, if we can find a constant $K \in (0, C)$ satisfying $\|f|_{\Gamma} - f|_{\Lambda}\|_Y \leq K\|f\|_X$, it will show that the left inequality in (16) is satisfied for Γ . Of course, we do not expect every u.d. set to satisfy that, but there might very well be a $\delta > 0$ s.t. every δ -perturbation of Γ has that property.

Recall that in Paley-Wiener spaces and Bernstein spaces, we only need to check the right inequalities in (16) to figure out whether a given u.d. set is a set of SS. Thus, we can take advantage of our our previous discussion. This is exactly what we will do to prove the following two theorems.

Theorem 9.13. *Let $\sigma > 0$ and a u.d. set $\Lambda \subset \mathbb{R}$ be given. If Λ is a set of SS for B_{σ} , then there exists a $\delta > 0$ s.t. every δ -perturbation of Λ is also a set of SS for B_{σ} .*

Proof. Find $c > 0$ s.t. $\|f|_{\Lambda}\|_{\infty} \geq c\|f\|_{\infty}$, $\forall f \in B_{\sigma}$, and let $\delta := \min\{\frac{\delta(\Lambda)}{4}, \frac{c}{4\pi\sigma}\}$. If $\Gamma \subset \mathbb{R}$ is a δ -perturbation of Λ , then

$$\begin{aligned} |f(\gamma_k) - f(\lambda_k)| &= \left| \int_{\lambda_k}^{\gamma_k} f'(u) du \right| \leq \sup_{u \in [\lambda_k, \gamma_k]} |f'(u)| \cdot |\gamma_k - \lambda_k| \\ &\leq \|f'\|_{\infty} \cdot \delta \leq 2\pi\sigma \|f\|_{\infty} \cdot \frac{c}{4\pi\sigma} = \frac{c}{2} \|f\|_{\infty}, \quad \forall k \in \mathbb{N}. \end{aligned}$$

$$\text{Hence, } \|f|_{\Gamma}\|_{\infty} \geq \|f|_{\Lambda}\|_{\infty} - \sup_{k \in \mathbb{N}} |f(\gamma_k) - f(\lambda_k)| \geq c\|f\|_{\infty} - \frac{c}{2}\|f\|_{\infty} = \frac{c}{2}\|f\|_{\infty}.$$

□

Theorem 9.14. *Let a bounded set $S \subset \mathbb{R}$ and a u.d. set $\Lambda \subset \mathbb{R}$ be given. If Λ is a set of SS for PW_S , then there exists a $\delta > 0$ s.t. every δ -perturbation of Λ is also a set of SS for PW_S .*

Proof. Find $c > 0$ s.t. $\|f|_{\Lambda}\|_2 \geq c\|f\|_2$, $\forall f \in PW_S$. In accordance with theorem 8.6, find $C > 0$ s.t. $\|f|_X\|_2 \leq C\|f\|_2$ for any $f \in PW_S$ and any u.d. set $X \subset \mathbb{R}$ with $\delta(X) \geq \frac{\delta(\Lambda)}{2}$.

Now, define $\delta := \min\{\frac{\delta(\Lambda)}{4}, \frac{c}{4\pi(\sup|S|)C}\}$. For any δ -perturbation $\Gamma \subset \mathbb{R}$ of Λ , and any $f \in PW_S$, the mean value theorem ensures us that there exists a sequence $\{x_k\}_{k \in \mathbb{N}} \subset \mathbb{R}$ s.t.

$$x_k \in [\gamma_k, \lambda_k], f(\gamma_k) - f(\lambda_k) = f'(x_k)(\gamma_k - \lambda_k), \forall k \in \mathbb{N}.$$

Also, $X := \{x_k \text{ s.t. } k \in \mathbb{N}\}$ is u.d. with $\delta(X) \geq \frac{\delta(\Lambda)}{2}$. Thus, by theorem 7.13 (iii),

$$\begin{aligned} \|f|_{\Gamma} - f|_{\Lambda}\|_2 &= \sqrt{\sum_{k \in \mathbb{N}} |f(\gamma_k) - f(\lambda_k)|^2} = \sqrt{\sum_{k \in \mathbb{N}} |f'(x_k)|^2 \cdot |\gamma_k - \lambda_k|^2} \leq \delta \sqrt{\sum_{k \in \mathbb{N}} |f'(x_k)|^2} \\ &= \delta \|f'|_X\|_2 \leq \delta C \|f'\|_2 \leq 2\pi(\sup|S|)\delta C \|f\|_2 \leq \frac{c}{2} \|f\|_2. \\ \|f|_{\Gamma}\|_2 &\geq \|f|_{\Lambda}\|_2 - \|f|_{\Gamma} - f|_{\Lambda}\|_2 \geq c \|f\|_2 - \frac{c}{2} \|f\|_2 = \frac{c}{2} \|f\|_2. \end{aligned}$$

Since Γ and f were arbitrary, this proves the theorem. \square

We now have a couple of tools to determine some sets of SS from others. In both Bernstein spaces and Paley-Wiener spaces, sufficiently small perturbations still keep the property of being a set of SS. By lemma 9.11 (i), the same can be said for weak limits of translates, at least for Bernstein spaces. We have not seen whether the same can be said for Paley-Wiener spaces, but it turns out that the sets of stable sampling are almost the same for PW_{σ} as for B_{σ} . We will not prove it, but this is one of the crucial steps to prove our theorem for the next section.

Another very relevant question about small perturbations is the following: under small perturbations of the samples, are we still guaranteed to get a function, and will its changes be small? That is, if $\{\delta_{\lambda}\}_{\lambda \in \Lambda}$ is a sequence of complex numbers of small absolute value, and if we replace $f(\lambda)$ by $f(\lambda) + \delta_{\lambda}$, does there exist a function g in the desired space s.t. $g(\lambda) = f(\lambda) + \delta_{\lambda}$ for $\lambda \in \Lambda$? And if g exists, is $|f(x) - g(x)|$ small for $x \in \mathbb{C}$? In a practical situation, this is a question of whether high precision in our measurements gives us a function close to the one we want. However, this is not a question this thesis attempts to answer. Identifying the sets of SS is our main focus here.

9.6 Beurling's sampling theorem

We are almost ready to give the main theorem of this section. But as mentioned, [OU16] uses another definition of PW_{σ} and B_{σ} than this thesis. Thus, since the theorem is taken from that book, we need to know how to interpret it with our definitions. On page 13, [OU16] defines the Fourier transform by:

$$\hat{f} := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ix(\cdot)} dx.$$

One difference is the factor $\frac{1}{\sqrt{2\pi}}$, which is needed to make the Fourier transform on $L^2(\mathbb{R})$ unitary. However, the relevant detail here is that the argument is scaled, since that is what changes the support of \hat{f} . If $f \in PW_{\sigma}$ with our definition, then

$$\int_{\mathbb{R}} f(x)e^{-ixt} dx = \int_{\mathbb{R}} f(x)e^{-2\pi ix(\frac{t}{2\pi})} dx = 0$$

a.e. for $|\frac{t}{2\pi}| > \sigma$, i.e. for $|t| > 2\pi\sigma$. This shows that what we call PW_{σ} is the same space that [OU16] calls $PW_{2\pi\sigma}$.

Similarly, on page 17, the book defines B_{σ} to be the space of all entire functions s.t. there exists a constant $C > 0$ satisfying

$$|f(z)| \leq Ce^{\sigma|Im(z)|}, \forall z \in \mathbb{C}.$$

Thus, what we call B_{σ} is the same space that they call $B_{2\pi\sigma}$.

From these observations, we get what we need to interpret the theorems. E.g. assume that [OU16] states a u.d. set $\Lambda \subset \mathbb{R}$ to be a set of SS for PW_{σ} if and only if a given statement containing σ is satisfied. Replacing σ by $2\pi\sigma$ everywhere then gives the corresponding statement for $PW_{2\pi\sigma}$, which is what we call PW_{σ} . Of course, the same can be said about B_{σ} . With this in mind, we can now state the theorem.

Theorem 9.15. *Let a u.d. set $\Lambda \subset \mathbb{R}$ and $\sigma > 0$ be given.*

- (i) *If $D^{-}(\Lambda) > 2\sigma$, then Λ is a set of SS for both PW_{σ} and B_{σ} .*
- (ii) *If $D^{-}(\Lambda) < 2\sigma$, then Λ is neither a set of SS for PW_{σ} nor B_{σ} .*
- (iii) *If $D^{-}(\Lambda) = 2\sigma$, then Λ is not a set of SS for B_{σ} .*

This is found in [OU16], pages 30 and 33, where we have replaced $\frac{\sigma}{\pi}$ by $\frac{2\pi\sigma}{\pi} = 2\sigma$. Note that in the limiting case that $D^{-}(\Lambda) = 2\sigma$, Λ may or may not be a set of SS for PW_{σ} . For example, we know that in $L^2[-\frac{1}{2}, \frac{1}{2}]$, $E(\mathbb{Z})$ is an ONB, while $E(\mathbb{Z} \setminus \{0\})$ is incomplete. Thus, by theorem 8.8 (ii), \mathbb{Z} is a set of SS for $PW_{\frac{1}{2}}$, while $E(\mathbb{Z} \setminus \{0\})$ is not. However, they both have lower uniform density 1, which is the limiting case for $\sigma = \frac{1}{2}$.

In the beginning of this section, we mentioned that Beurling's sampling theorem almost solves the strong sampling problem completely for PW_{σ} and B_{σ} . Now, we see that it does solve the problem completely for B_{σ} . The only remaining part of the question is what u.d. sets $\Lambda \subset \mathbb{R}$ with $D^{-}(\Lambda) = 2\sigma$ are sets of SS for PW_{σ} . We will not address that problem in general, but as we will argue in the next section, the arithmetic progressions with $D^{-}(\Lambda) = 2\sigma$ are sets of SS for PW_{σ} .

Note that the version of Beurling's sampling theorem presented here does not say anything about stable sampling in PW_S , unless S is an interval centered at the origin. However, since $PW_S \subseteq PW_{\sigma}$ for any $\sigma > 0$ satisfying $S \subseteq [-\sigma, \sigma]$, sampling in PW_{σ} implies sampling in PW_S . Also, the two Paley-Wiener spaces have the same norm, so SS in PW_{σ} implies SS in PW_S . What remains unanswered here is whether there are any other sets of SS for PW_S .

10 Uniform sampling in PW_σ

We will now limit ourselves to considering uniform sampling, i.e. cases where the distance between consecutive elements of the sampling set is fixed. Also, we will only consider PW_σ , i.e. the Paley-Wiener spaces whose spectrum is an interval centered at the origin. What does the theory we have look like in that case?

10.1 Uniform sets of SS for PW_σ

As we know, for any $c \in \mathbb{R}$, $E(\frac{\mathbb{Z}+c}{2\sigma})$ is a multiple of an ONB for $L^2[-\sigma, \sigma]$, hence a (tight, exact) frame. By theorem 8.8 (ii), this is equivalent to saying that $\frac{\mathbb{Z}+c}{2\sigma}$ is a set of SS for PW_σ . As noted before, $D^-(\Lambda) = \frac{1}{\delta(\Lambda)}$ whenever Λ is an arithmetic progression, so $D^-(\frac{\mathbb{Z}+c}{2\sigma}) = 2\sigma$. This is exactly the limiting case in Beurling's sampling theorem. Clearly, every arithmetic progression with separation $\frac{1}{2\sigma}$ can be written as $\frac{\mathbb{Z}+c}{2\sigma}$ for some $c \in \mathbb{R}$, where we may pick $c \in [0, 2\sigma)$. Thus, in the case of uniform sampling, Beurling's sampling theorem leads to the following result.

Corollary 10.1. *Given a $\sigma > 0$, let $\Lambda \subset \mathbb{R}$ be an arithmetic progression. Then, Λ is a set of SS for PW_σ if and only if $D^-(\Lambda) \geq 2\sigma$, i.e. if and only if $\delta(\Lambda) \leq \frac{1}{2\sigma}$.*

Hence, depending on how much data we want and how much we are able to get, we may pick any set of no more separation than $\frac{1}{2\sigma}$.

10.2 Reconstruction

We have mainly looked at what the sets of stable sampling are, without considering ways to actually find the function. However, from theorem 8.9, any set of SS for a Paley-Wiener spaces automatically gives us a reconstruction formula. In the case that $S = [-\sigma, \sigma]$, and the sampling set is $\frac{\mathbb{Z}}{2\sigma}$, we get a classical result, known as *Shannon's sampling theorem*.

Theorem 10.2 (Shannon's sampling theorem). *Given a $\sigma > 0$, if $f \in PW_\sigma$, then*

$$f(x) = \sqrt{2\sigma} \sum_{k \in \mathbb{Z}} f\left(\frac{k}{2\sigma}\right) \text{sinc}(2\sigma x - k), \quad \forall x \in \mathbb{R}. \quad (21)$$

Proof. For any $k \in \mathbb{Z}$, we have:

$$\begin{aligned} \mathfrak{G}\left(\overline{e^{2\pi i \frac{k}{2\sigma}(\cdot)}} \chi_{[-\sigma, \sigma]}\right)(x) &= \int_{-\sigma}^{\sigma} e^{-2\pi i \frac{k}{2\sigma} t} e^{2\pi i x t} dt = \int_{-\sigma}^{\sigma} e^{2\pi i (x - \frac{k}{2\sigma}) t} dt \\ &= \frac{1}{2\pi i (x - \frac{k}{2\sigma})} \left[e^{2\pi i (x - \frac{k}{2\sigma}) t} \right]_{t=-\sigma}^{t=\sigma} = \frac{1}{\pi (x - \frac{k}{2\sigma})} \cdot \frac{e^{2\pi \sigma i (x - \frac{k}{2\sigma})} - e^{-2\pi \sigma i (x - \frac{k}{2\sigma})}}{2i} \\ &= 2\sigma \frac{\sin(2\pi \sigma (x - \frac{k}{2\sigma}))}{2\pi \sigma (x - \frac{k}{2\sigma})} = 2\sigma \text{sinc}\left(2\sigma \left(x - \frac{k}{2\sigma}\right)\right) = 2\sigma \text{sinc}(2\sigma x - k), \quad \forall x \in \mathbb{R}. \end{aligned}$$

Since $\frac{1}{\sqrt{2\sigma}} E(\frac{\mathbb{Z}}{2\sigma})$ is an ONB for $L^2[-\sigma, \sigma]$, and thus equals its own dual frame, Shannon's sampling theorem follows from theorem 8.9. \square

Shannon's sampling theorem is one of the most well-known theorems in sampling theory. It is often stated for the case that $\sigma = \frac{1}{2}$, after which (21) becomes

$$f(x) = \sum_{k \in \mathbb{Z}} f(k) \operatorname{sinc}(x - k), \quad \forall x \in \mathbb{R}, \quad \forall f \in PW_{\frac{1}{2}}.$$

If we translate $\frac{\mathbb{Z}}{2\sigma}$ to get $\frac{\mathbb{Z}+c}{2\sigma}$, then the same calculation as before gives the exact same combination of sinc-functions, except that k is replaced by $k+c$. In the rest of this discussion, we will just assume that $c = 0$, remembering that cases where $c \neq 0$ are easy to get around.

How about if Λ is a uniform sampling set of smaller separation than $\frac{1}{2\sigma}$? In that case, $E(\Lambda)$ is not an ONB, so it might not be its own dual frame. In the case that we halve the separation, as demonstrated in example 6.17, the dual frame of $E(\Lambda)$ is simply $\frac{1}{2}E(\Lambda)$. Hence, if we replace k by $\frac{k}{2}$ to get a separation of $\frac{1}{4\sigma}$, we just need to halve every term in (21):

$$f(x) = \frac{\sqrt{2\sigma}}{2} \sum_{k \in \mathbb{Z}} f\left(\frac{k}{4\sigma}\right) \operatorname{sinc}\left(2\sigma x - \frac{k}{2}\right), \quad \forall x \in \mathbb{R}, \quad \forall f \in PW_{\sigma}.$$

The same argument works equally well if we scale down the separation by another natural number $m > 2$. Thus, the reconstruction formula becomes:

$$f(x) = \frac{\sqrt{2\sigma}}{m} \sum_{k \in \mathbb{Z}} f\left(\frac{k}{2\sigma m}\right) \operatorname{sinc}\left(2\sigma x - \frac{k}{m}\right), \quad \forall x \in \mathbb{R}, \quad \forall f \in PW_{\sigma}. \quad (22)$$

So far, the density $\frac{1}{\delta(\Lambda)}$ has been a multiple of 2σ . This has allowed us to view $E(\Lambda)$ as a disjoint union of ONBs, which is what we took advantage of to compute the frame operator in example 6.17. The situation is a bit trickier otherwise. In that case, we need another way to find the dual frame, or another alternate dual frame, for $E(\Lambda)$. One possibility to get around that problem is to take advantage of the fact that $PW_{\sigma} \subseteq PW_{\gamma}$ whenever $\gamma \geq \sigma$. As long as $\gamma \leq \frac{1}{2\delta(\Lambda)}$, $E(\Lambda)$ is a set of SS for PW_{γ} as well. Picking $\gamma \in [\sigma, \frac{1}{2\delta(\Lambda)}]$ s.t. $\frac{1}{\delta(\Lambda)}$ is a multiple of 2γ , we can use (21), with γ in place of σ . Note that such a γ always exists. Firstly, for Λ to be a set of SS for PW_{σ} , we must have $\sigma \leq \frac{1}{2\delta(\Lambda)}$ by Beurling's sampling theorem. Secondly, picking $\gamma := \frac{1}{2\delta(\Lambda)}$ yields one solution. If a $\gamma < \frac{1}{2\delta(\Lambda)}$ satisfies our conditions, we may also use (22) for the space $PW_{\frac{1}{2\delta(\Lambda)}}$.

Example 10.3. Let $\sigma := \frac{2}{\pi}$, let $\Lambda := \frac{\mathbb{Z}}{2}$, and assume that $f \in PW_{\sigma}$ satisfies:

$$f(\lambda) = \begin{cases} \pi, & \lambda = 0 \\ 0, & \lambda \in \mathbb{Z} \setminus \{0\} \\ \frac{1}{\lambda}, & \lambda \in 2\mathbb{Z} + \frac{1}{2} \\ -\frac{1}{\lambda}, & \lambda \in 2\mathbb{Z} - \frac{1}{2} \end{cases}$$

Since $\frac{1}{2\delta(\Lambda)} = 1 > \sigma$, we know that Λ is a set of SS for PW_{σ} . Thus, f is uniquely determined from $f|_{\Lambda}$, which is known, and we have a few ways of reconstructing f . If we pick a $\gamma \in [\frac{2}{\pi}, 1]$, then Λ is a set of SS for $PW_{\gamma} \supseteq PW_{\sigma}$ as well. Picking $\gamma = 1$ allows us to use (21) with either $m = 1$ or $m = 2$. to get:

$$\begin{aligned}
f(x) &= \sum_{k \in \mathbb{Z}} f\left(\frac{k}{2}\right) \operatorname{sinc}(2x - k) = \pi \operatorname{sinc}(2x) + \sum_{n \in \mathbb{N}} (f(n) \operatorname{sinc}(2x - n) + f(-n) \operatorname{sinc}(2x + n)) \\
&= \pi \operatorname{sinc}(2x) + \sum_{n \in \mathbb{N}} \frac{1}{n} (\operatorname{sinc}(2x - n) - \operatorname{sinc}(2x + n)), \quad \forall x \in \mathbb{R}.
\end{aligned}$$

Had we sampled on $\frac{\mathbb{Z}}{4}$, i.e. twice as frequently, we could still get f in the same way. However, we would also have the option of using (22), with $n = 2$. In addition, we could use $\gamma = 2$, since the density would be doubled. In any case, the solution is:

$$f(x) = \pi \operatorname{sinc}(x), \quad \forall x \in \mathbb{R}.$$

It can be checked that f coincides with $f|_{\Lambda}$ on $\frac{\mathbb{Z}}{2}$, and it is well-known that $f \in PW_{\frac{1}{2}} \subset PW_{\sigma}$. Note that finding f would be even easier had we known that $f \in PW_{\frac{1}{2}}$, since we could then use the values on \mathbb{Z} , where all but one of them vanish. However, since $PW_{\frac{1}{2}} \not\subseteq PW_{\sigma}$, we did not know that to begin with.

10.3 Perturbations of $\frac{\mathbb{Z}}{2\sigma}$

By theorem 9.14, if $\Lambda \subset \mathbb{R}$ is a set of SS for PW_{σ} , then every sufficiently small perturbation of Λ is also a set of SS for PW_{σ} . That is, if $\{\delta_k\}_{k \in \mathbb{N}}$ is a sequence of non-negative real numbers not exceeding δ , then $\Gamma := \{\lambda_k - \delta_k \text{ s.t. } k \in \mathbb{N}\}$ is a set of SS for PW_{σ} . By theorem 8.8 (ii), this is equivalent to saying that if $E(\Lambda)$ is a frame in $L^2(S)$, so is $E(\Gamma) = \{e^{2\pi i(\lambda_k - \delta_k)(\cdot)}\}_{k \in \mathbb{N}}$. But what is sufficiently small, i.e. how small does δ need to be? The lower uniform density does not change, so if $D^-(\Lambda) > 2\sigma$, Beurling's sampling theorem makes sure that we still have a set of SS. Thus, the only interesting cases are the ones where $D^-(\Lambda) = 2\sigma$.

Answering that question in general is hard, and we will not consider that problem. However, in the case that $\Lambda = \frac{\mathbb{Z}}{2\sigma}$, we actually have an exact answer! This is known as Kadec's $\frac{1}{4}$ -theorem, and it is known that the number $\frac{1}{4}$ cannot be improved.

Theorem 10.4 (Kadec's $\frac{1}{4}$ -theorem). *Let $\sigma > 0$ and a u.d. set $\Lambda \subset \mathbb{R}$ be given. If there exists a $\delta \in [0, \frac{1}{4})$ satisfying*

$$|\lambda_k - k| \leq \delta, \quad \forall k \in \mathbb{Z},$$

then $E(\frac{\Lambda}{2\sigma})$ is an RB for $L^2[-\sigma, \sigma]$.

Proof. See [Yo01] page 36 for the case that $\sigma = \pi$. Otherwise, if $f \in L^2[-\sigma, \sigma]$, then

$$g(x) := f\left(\frac{\sigma}{\pi}x\right), \quad \forall x \in [-\pi, \pi]$$

defines a function $g \in L^2[-\pi, \pi]$. Hence, there exists a unique sequence $\{c_k\}_{k \in \mathbb{N}}$ s.t.

$$g = \sum_{k \in \mathbb{Z}} c_k e^{i\lambda_k(\cdot)}$$

in the L^2 -norm. Hence, again in the L^2 -norm, setting $\gamma_k := \frac{\lambda_k}{2\sigma}$, we have:

$$f(x) = g\left(\frac{\pi}{\sigma}x\right) = \sum_{k \in \mathbb{Z}} c_k e^{i\lambda_k \frac{\pi}{\sigma}x} = \sum_{k \in \mathbb{Z}} c_k e^{2\pi i \gamma_k x}.$$

Existence and uniqueness of this expansion shows that $E(\frac{\Lambda}{2\sigma})$ is a basis for $L^2[-\sigma, \sigma]$. Also,

$$\begin{aligned} \left\| \sum_{k \in \mathbb{Z}} c_k e^{2\pi i \gamma_k (\cdot)} \right\|_{L^2[-\sigma, \sigma]}^2 &= \int_{-\sigma}^{\sigma} \left| \sum_{k \in \mathbb{Z}} c_k e^{2\pi i \gamma_k x} \right|^2 dx = \frac{\sigma}{\pi} \int_{-\pi}^{\pi} \left| \sum_{k \in \mathbb{Z}} c_k e^{i\lambda_k t} \right|^2 dt \\ &= \frac{\sigma}{\pi} \left\| \sum_{k \in \mathbb{Z}} c_k e^{i\lambda_k (\cdot)} \right\|_{L^2[-\pi, \pi]}^2, \end{aligned}$$

where we have been using the change of variable $t := \frac{\pi}{\sigma}x$. Since $E(\frac{\Lambda}{2\pi})$ satisfies (8) in $L^2[-\pi, \pi]$, we conclude that $E(\frac{\Lambda}{2\sigma})$ also satisfies (8) in $L^2[-\sigma, \sigma]$, where we scale A and B by the factor $\frac{\sigma}{\pi}$. \square

We now know how small the perturbations need to be for every UNIFORM sampling set. Kadec's $\frac{1}{4}$ -theorem solves it for the case that $\frac{1}{\delta(\Lambda)} = 2\sigma$, while Beurling's sampling theorem is sufficient to answer the cases of larger density. Note that since the frame is $E(\frac{\Lambda}{2\sigma})$, rather than $E(\Lambda)$, the upper bound for the perturbations is $2\sigma \cdot \frac{1}{4} = \frac{\sigma}{2}$, rather than $\frac{1}{4}$. Kadec's $\frac{1}{4}$ -theorem also tells us that the new frame is still an RB, i.e. it is exact. We will not find the operator mapping $E(\frac{\Lambda}{2\sigma})$ into the ONB $E(\frac{\mathbb{Z}}{2\sigma})$, but Kadec showed that it does exist, leading to the theorem.

11 Conclusion

This concludes the theory of this thesis. We have seen that the strong sampling problem in PW_S are equivalent to identifying the exponential frames in $L^2(S)$. Beurling's sampling theorem solves the strong sampling theorem for B_σ , and it almost solves it for PW_σ as well. For the case of uniform sampling in PW_σ , we have an exact answer. Small perturbations do not change the sampling properties in any of the two spaces, and for uniform sampling in PW_σ , we also know how small the perturbations need to be.

References

- [Ah66] Ahlfors, L. (1966). *Complex analysis*, 2nd edition. USA: McGraw-Hill Book Company.
- [Ch10] Christensen, O. (2010). *Functions, Spaces, and Expansions*. Birkhäuser.
- [HL00] Han, D. & Larson, D. R. (2000). *Frames, Bases and Group Representations*. Providence, Rhode Island: American Mathematical Society.
- [Mc07] McCormick, D. (12.02.2007). *Orthonormal Bases in Hilbert Spaces*. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.586.2734&rep=rep1&type=pdf>
- [OU16] Olevskii, A. & Ulanovskii, A. (2016). *Functions with Disconnected Spectrum*. Providence, Rhode Island: American Mathematical Society.
- [Ru91] Rudin, W. (1991). *Functional analysis*, 2nd edition. Singapore: McGraw-Hill, Inc.
- [Sc02] Schechter, M. (2002). *Principles of Functional analysis*, 2nd edition. Providence, Rhode Island: American Mathematical Society.
- [Yo01] Young, R. M. (2001). *An introduction to nonharmonic fourier series*, revised 1st edition. USA: Academic press.
- [Wik1] *Absolute convergence*. https://en.wikipedia.org/wiki/Absolute_convergence (edited 27.05.2019)
- [Wik2] *Bounded operator*. https://en.wikipedia.org/wiki/Bounded_operator (edited 25.01.2019)