



BOUND DIRICHLET BELIEF NETWORK FOR TOPIC RECONSTRUCTION

Huiqiang Zhong

Supervisor: Dr. Xuhui Fan

School of Mathematics and Statistics
UNSW Sydney

July 2020

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF THE DEGREE OF
MASTER OF STATISTICS

Plagiarism statement

I declare that this thesis is my own work, except where acknowledged, and has not been submitted for academic credit elsewhere.

I acknowledge that the assessor of this thesis may, for the purpose of assessing it:

- Reproduce it and provide a copy to another member of the University; and/or,
- Communicate a copy of it to a plagiarism checking service (which may then retain a copy of it on its database for the purpose of future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct, and am aware of any potential plagiarism penalties which may apply.

By signing this declaration I am agreeing to the statements and conditions above.

Signed: _____

Date: _____

Abstract

Topic model, especially latent Dirichlet allocation(LDA), has aroused great research interest in the past two decades which is a useful tool for document classification. The LDA model assumes that words of a document are determined by the topic of the document and the topic-word matrix. Some multi-layer generative processes on word distribution have been proposed to improve LDA. However, they suffer from information decay and complicated sampling methods. Here we present a Dirichlet-Belief Network to address the problems above. By inserting auxiliary Poisson random variables into the layerwise connections and appropriate design, we can infer the latent parameters in an efficient way. As a Bayesian generative model, it is more interpretable and Gibbs sampling can be used in the training model.

Contents

Chapter 1	Introduction	1
1.1	Background and Importance	1
1.2	topic of research paper	2
1.3	Highlight the approach and principal finding	2
1.4	Organization of this paper	2
Chapter 2	Basic language	3
2.1	Bayesian Theorem	3
2.1.1	Prior Probability	4
2.1.2	Likelihood Probability	4
2.1.3	Posterior probability	4
2.2	Topologies in Banach Spaces	4
2.3	Bases in Banach spaces	4
Chapter 3	Classical Harmonic Analysis	8
3.1	Harmonic Analysis on Locally Compact Abelian Groups . . .	8
Chapter 4	Conclusion	13
References		14

CHAPTER 1

Introduction

1.1 Background and Importance

With the development of technology and almost all information is digitized, we are more and more likely to be overwhelmed by a large amount of information, and it is difficult to find an effective way to find, organize and understand a large amount of information[1]. Then, how to effectively obtain better information and how to automatically classify vast amounts of text data, organization and management become more challenging. Therefore, in the face of these problems and needs, the use of computers for language information processing has been extensively studied. As a research hotspot in the field of natural language processing, automatic text classification technology has been rapidly developed and widely used.

Such data is sparse and discrete data which is hard for computer to process. The bag of words models are proposed for simplifying representation in natural language processing. This model can convert a sentence into a vector representation, which is a relatively straightforward method. It does not consider the order of words in the sentence, only considers the number of occurrences of words in the vocabulary in this sentence so that it will miss the relationship between each individual vocabulary and induces huge problem of dimensional disaster. However it still a popular way in representing document.

After representing document by matrix, scientists hope to construct a generative model to imitate document writing. It reduces a complex procedure into some probabilistic steps and specify a sample distribution for the topic of document[?]. We will introduce a very well-known method in natural language processing - Latent Dirichlet Allocation(LDA) which can extract abstract "themes" from a series of documents through the mechanism of generating models.

This model assumes the implicit mechanism of published documents written by humans: each document is composed of a few topics, and each topic can be composed of a few important words description. In other words, When writing an article, we will first decide the topics related to our document, then each word we write in text is closely related to these topics. Let's follow the generative process of LDA step by step:

- For each document, we extract a theme from the theme distribution.
- Extracting a word from the word distribution corresponding to the above-mentioned theme.
- Repeat the above process until each word in the document is traversed.

This generative process is mainly composed of two distributions-the topic distribution of each document and the word distribution of each topic. In LDA, the topic distribution and word distribution are uncertain. The authors of LDA adopt the Bayesian idea that they should obey a distribution. The topic distribution and word distribution are subject to polynomial distributions, so the topic distribution and word distribution use Dirichlet distribution as their conjugate prior distribution because the polynomial distribution and Dirichlet distribution is a conjugate structure.

As mentioned above, one commonly-used prior for topic distribution and word distribution is Dirichlet distribution and recently lots of distributions have been imposed on topic distribution. For example, The correlation topic model (CTM) [?], which exhibits the correlation of topic by introducing logistic normal distribution. It allows each document to show different topics with different proportions and capture the relationship in ground data. Besides, there are hierarchical document representation based on Dirichlet process and Boltzmann machines and neural network. [?].

?

1.2 topic of research paper

Announce the research topic/question being addressed in research paper

1.3 Highlight the approach and principal finding

Description of your paper approach and why you chose it
brief summary of your major findings

1.4 Organization of this paper

CHAPTER 2

Basic language

In this chapter we outline the general context in which our study of multiplier theory in Banach spaces takes place. We begin by revising the elementary concepts associated with functional analysis in Banach spaces. Then Section 2.3 introduces the notions of bases, multiplier transforms and unconditional convergence of sequences, setting the scene for the rest of the thesis. The chapter ends by introducing a technique which is frequently used to obtain results pertaining to the unconditionality of sequences.

2.1 Bayesian Theorem

We can never be completely sure of this world, because it is a constantly changing being, and change is the essence of reality. However, what we can do is, as expressed by this theorem, as we obtain more and more data or evidence, our knowledge of reality has been updated and improved

$$\begin{aligned} P(A|B) &= \frac{P(A, B)}{P(B)} \\ &= \frac{P(B|A) * P(A)}{P(B)} \end{aligned}$$

- $P(A|B)$ is the conditional probability of A after the occurrence of B is known, and is also excluded from the posterior probability of A due to the value obtained from B, indicating the confidence that event A will occur after event B occurs.
- $P(A)$ is the a priori probability or edge probability of A, and represents the confidence that event A occurs.
- $P(B|A)$ is the conditional probability of B after the occurrence of A. It is also called the posterior probability of B because of the value obtained from A, and is also considered as a likelihood function.
- $P(B)$ is the prior probability or edge probability of B, which is called a standardized constant.
- $P(B|A)P(B)$ is called the standard likelihood ratio (there are many names, and no unified standard name is found), which indicates the degree of support provided by event B for the occurrence of event A.

2.1.1 Prior Probability

2.1.2 Likelihood Probability

2.1.3 Posterior probability

2.2 Topologies in Banach Spaces

A Banach space X is easily made a topological space with the topology induced by its norm, called the *norm topology* of X . We say a sequence of vectors $\{x_n\}$ *converges in* X if it converges in the norm topology of X . Another topology is the *weak topology* of X . A sequence $\{x_n\}_{n=1}^{\infty}$ of vectors in X converges to $x \in X$ in the weak topology of X if $\lim_{n \rightarrow \infty} \langle x_n - x, x^* \rangle = 0$ for all $x^* \in X^*$. As their names suggest, convergence in the norm topology implies convergence in the weak topology, or in other words, if a sequence converges then it also converges weakly.

We shall consider three topologies in the Banach space $(X, \|\cdot\|)$. Suppose $\{T_n\}_{n=1}^{\infty}$ is a sequence of operators in $\mathcal{B}(X, Y)$. We say it converges to T *in norm* if $\|T - T_n\| \rightarrow 0$, *strongly* if $\|Tx - T_nx\| \rightarrow 0$ for all $x \in X$, and *weakly* if $\langle Tx - T_nx, x^* \rangle \rightarrow 0$ for all $x \in X$ and $x^* \in X^*$. The topologies induced are called the norm operator topology, the strong operator topology and the weak operator topology of $\mathcal{B}(X, Y)$ respectively. A sequence which converges strongly will converge weakly, and in turn, convergence in the norm topology implies strong convergence. All three topologies have unique limits and distinguish points.

The following proposition is easy to verify and is included for familiarisation with the concepts discussed thus far. It shall also be used to establish later results.

Let X be a Banach space and suppose $\{T_n\}_{n=1}^{\infty}$ is a sequence of operators uniformly bounded in norm by some $K > 0$ and converging to T in the strong operator topology of $\mathcal{B}(X, X)$. Then $\|T\| \leq K$.

Proof. We want to show $\|Tx\| \leq K\|x\|$ for all $x \in X$. Fix $x \in X$ and $\epsilon > 0$ and choose an $n \in \mathbb{N}$ such that $\|Tx - T_nx\| < \epsilon$. Then

$$\|Tx\| \leq \|Tx - T_nx\| + \|T_nx\| < \epsilon + K\|x\|,$$

and the proposition follows. \square

2.3 Bases in Banach spaces

The definition and usefulness of a basis in a finite dimensional vector space is well-known. It is natural then to want an analogous concept for Banach spaces. The most useful approach is found in the notion of a Schauder basis. A comprehensive introduction to such bases is [24, pp. 1–52], upon which most of the material in this section is based.

A sequence $\{x_n\}_{n=1}^{\infty}$ in a Banach space X is called a *Schauder basis of* X if, for every $x \in X$, there is a unique sequence of scalars $\{a_n\}_{n=1}^{\infty}$ such that $x = \sum_{n=1}^{\infty} a_n x_n$. A sequence $\{x_n\}_{n=1}^{\infty}$ which is a Schauder basis for its closed linear span is called a *basic sequence*.

A Banach space X is called *separable* if it has a countable dense subset. Thus if X has a Schauder basis then X is separable. In general the converse is not true.

Since the class of Schauder bases is the only type of bases considered in this thesis, it will henceforth be implicit that any discussion about bases refers only to Schauder bases.

The ordered set of unit vectors $e_n = (0, 0, 0, \dots, 1, 0, \dots)$, where the 1 occurs in the n th coordinate, forms a basis in each of the spaces c , c_0 and ℓ_∞ . Another example of a basis in c is given by $x_1 = (1, 1, 1, \dots)$ and $x_n = e_{n-1}$ for $n > 1$. If $x = (b_1, b_2, \dots) \in c$, then the expansion of x with respect to this basis is

$$x = \left(\lim_{k \rightarrow \infty} a_k\right)x_1 + (a_1 - \lim_{k \rightarrow \infty} a_k)x_2 + (a_2 - \lim_{k \rightarrow \infty} a_k)x_3 + \dots$$

If a Banach space has a basis, we may consider it as a sequence space via an identification of each element $x = \sum_{n=1}^{\infty} a_n x_n$ in with the unique sequence of coefficients (a_1, a_2, a_3, \dots) . To do this note that it is essential to describe the basis as an ordered sequence rather than merely as a set. The following example highlights a deeper reason for why we must specify the ordering of the vectors in a basis.

Consider the sequence of vectors $\{x_n\}_{n=1}^{\infty}$ in c given by

$$x_n = (\underbrace{0, 0, \dots, 0}_{n-1}, 1, 1, \dots),$$

for $n \in \mathbb{N}$. It is not hard to see that this is a basis for c . It is called the *summing basis* in c . Consider the sequence b defined by

$$b_n = \sum_{k=1}^n (-1)^{k+1} \frac{1}{k}$$

for $n \in \mathbb{N}$. By Leibniz' test for alternating series, b_n has a limit as $n \rightarrow \infty$, so $b \in c$. The expansion of b with respect to the summing basis is $b = \sum_{n=1}^{\infty} a_n x_n$ where $a_n = (-1)^{n+1} \frac{1}{n}$. Now consider the following rearrangement of the series. Define $\pi : \mathbb{N} \rightarrow \mathbb{N}$ by

$$\pi(n) = \begin{cases} 2k & \text{if } n = 3k \\ 4k + 1 & \text{if } n = 3k + 1 \\ 4k + 3 & \text{if } n = 3k + 2. \end{cases}$$

Then π is clearly a permutation of \mathbb{N} . However, the expansion

$$b' \equiv \sum_{n=1}^{\infty} a_{\pi(n)} x_{\pi(n)}$$

does not lie in c since

$$b'_{3n} = \sum_{k=1}^n \left(\frac{1}{2k-1} + \frac{1}{2k+1} - \frac{1}{2k} \right) \geq 2 \sum_{k=1}^n \frac{1}{k}$$

diverges as $n \rightarrow \infty$.

Not every basis has the defect that the convergence of an expansion with respect to the basis is dependent on the order of summation of the expansion. For example, expansions in finite dimensional spaces can be summed in any order without altering convergence. The following proposition gives various methods for checking whether a sum of vectors can be freely rearranged while still respecting convergence.

[24, Proposition 1.c.1] Let $\{x_n\}_{n=1}^\infty$ be a sequence of vectors in a Banach space . Then the following are equivalent.

- (i) The series $\sum_{n=1}^\infty x_{\pi(n)}$ converges for every permutation π of the integers.
- (ii) The series $\sum_{k=1}^\infty x_{n_k}$ converges for every choice of $n_1 < n_2 < n_3 \dots$
- (iii) The series $\sum_{n=1}^\infty \varepsilon_n x_n$ converges for every choice of signs $\varepsilon_n = \pm 1$.
- (iv) For every $\epsilon > 0$ there exists an integer n_0 such that $\sum_{j \in J} x_j < \epsilon$ for every finite set of integers J which satisfies $\min\{j \in J\} > n_0$.

With the aid of the proposition, there are now easier ways to verify that the series $\sum_{n=1}^\infty a_n x_n$ given in Example basessumming basis is dependent on the order of summation. Simply observe that $\sum_{n=1}^\infty a_{2n-1} x_{2n-1}$ does not converge in c and hence statement (ii) of the proposition fails to hold.

Let $\{x_n\}_{n=1}^\infty$ be a sequence of vectors in a Banach space . A series $\sum_{n=1}^\infty x_n$ which satisfies any of the conditions (i), (ii), (iii) or (iv) in Proposition basesunconditional conv is said to be *unconditionally convergent*. A basis $\{x_n\}_{n=1}^\infty$ of a Banach space is *unconditional* if for every $x \in$, its expansion $x = \sum_{n=1}^\infty a_n x_n$ converges unconditionally. Otherwise the basis is said to be *conditional*.

Thus the standard ordered bases $\{e_n\}_{n=1}^\infty$ for c , c_0 and , $1 \leq p < \infty$ are unconditional bases. This means that any reordering of $\{e_n\}_{n=1}^\infty$ also yields unconditional bases in these spaces. The summing basis for c is clearly a conditional basis.

Given a basis $\{x_n\}_{n=1}^\infty$, one might ask what complex sequences $\{\phi_n\}_{n=1}^\infty$ give rise to a bounded multiplier transform? That is, is there a constant $C > 0$ such that for all $x \in$,

$$\sum_{n=1}^\infty \phi_n a_{n,x} x_n \leq Cx,$$

where $\sum_{n=1}^\infty a_{n,x} x_n$ is the expansion of each $x \in$ with respect to the basis? If that basis is unconditional, then the space of such sequences is just ℓ^∞ (see [24, Proposition 1.c.7]).

We have seen that conditional bases do exist, and some of the resulting difficulties that arise when working with a conditional basis. Is there any way in which we can avoid such problems in separable Banach spaces? In other words, can we always find an unconditional basis for a separable Banach space? The answer is no. The space $L^1[0, 1]$ has no unconditional basis. In fact, $L^1[0, 1]$ cannot even be embedded in a space with an unconditional basis [24, Propostion 1.d.1].

There are many open problems regarding unconditional bases. For example, suppose is a Banach space with an unconditional basis, and let be a complemented subspace of . Does have an unconditional basis? One outstanding problem has recently been solved (see [23]). Does every infinite dimensional

Banach space contain an unconditional basic sequence? The answer is no. We shall not pursue such broad problems in this thesis. Instead, we shall use tools from harmonic analysis and spectral theory to examine whether or not particular sequences in certain Banach spaces are conditional, and whether or not they form bases for those spaces.

CHAPTER 3

Classical Harmonic Analysis

We now leave the broad discussion about bases and conditionality in Banach spaces and instead focus our attention to some particular classes of the L^p Banach spaces. The theory of classical harmonic analysis provides a good starting point from which to tackle problems relating to bases and conditionality in this setting.

In this chapter we give an overview of some basic results from classical harmonic analysis and Fourier theory. We begin by introducing the fundamental concepts of harmonic analysis on locally compact Abelian groups. The Fourier transform and Fourier multipliers will be defined. Connections between multiplier transforms and convolution operators will be drawn in Section 3.1, as well as important examples given. In Section 3.1 we discuss some of the classical results of M. Riesz, Littlewood–Paley and Marcinkiewicz.

The general theory given in this chapter can be used to consider the set of functions $\{\varphi_n\}_{n \in \mathbb{Z}} \subset L^p()$, where $\varphi_n : t \mapsto e^{int}$. Is this a basis for the Banach space $L^p()$? Is it an unconditional basic sequence? If not, how close is it to being an unconditional sequence? The results stated in Section 3.1 answer such questions.

3.1 Harmonic Analysis on Locally Compact Abelian Groups

In this section we give the necessary background from which we can begin to answer the questions from above. The theory mentioned here can be found in [21] and is quite standard. In what follows, if G is a group, the inverse of an element $x \in G$ will be denoted by $-x$, and the group operation from $G \times G$ into G will be denoted by $(x, y) \mapsto x + y$.

A *locally compact Abelian group* (or an LCA group) G is an Abelian group which is also a locally compact Hausdorff space such that the group operations $x \mapsto -x$ from G onto G and $(x, y) \mapsto x + y$ from $G \times G$ onto G are continuous. The most-studied examples of LCA groups are the integers, the circle group and real line with their usual topologies. It is easy to verify that any Abelian group can be made into an LCA group when endowed with the discrete topology. In this thesis the circle group will feature most often. It can be modelled by the unit circle $\{\omega \in \mathbb{C} : |\omega| = 1\}$ in the complex plane, or as the quotient group $\mathbb{R}/2\pi$. We shall usually adopt the latter model, thinking of $\mathbb{R}/2\pi$ as the interval $[0, 2\pi]$ with endpoints identified and addition as the group operation.

A good reason to study LCA groups is that we can use them to construct measure spaces that are translation invariant.

Let G be a locally compact Abelian group. A *Haar measure* on G is a positive regular Borel measure μ having the following two properties:

- (i) $\mu(E) < \infty$ if $E \subseteq G$ is compact; and
- (ii) $\mu(E + x) = \mu(E)$ for all measurable $E \subseteq G$ and all $x \in G$.

[21, Chapter VII, §2] Let G be an LCA group. Then a Haar measure on G exists and is unique up to multiplication of a positive constant.

Hence one often speaks of *the* Haar measure. For $G = \mathbb{R}$, Haar measure is usually taken to be normalised Lebesgue measure, $(2\pi)^{-1}dt$. If G is discrete and infinite, Haar measure is usually normalised to have mass one at each point. We denote the L^p space of functions on G with respect to Haar measure by $L^p(G)$.

As mentioned in Example Banach Spaces L^p example, we will not usually distinguish between equivalence classes.

Using Haar measure, we may turn $L^1(G)$ into a Banach algebra.

Let G be an LCA group with Haar measure dy and suppose $f, g \in L^1(G)$. Then for almost all $x \in G$, the function $y \mapsto f(x - y)g(y)$ for $y \in G$ is integrable on G . If we write

$$h(x) = \int_G f(x - y)g(y)dy,$$

then $h \in L^1(G)$ and $h_1 \leq f_1 g_1$.

Let G, f, g be as in Theorem 1 on convolution. Then the *convolution* of f and g , denoted $f * g$, is given by

$$(f * g)(x) = \int_G f(x - y)g(y)dy$$

for almost all $x \in G$

Let G be an LCA group. Then $L^1(G)$ is a Banach Algebra under convolution and pointwise addition.

Our next aim is to define the Fourier transform of a function in $L^1(G)$. We start by introducing the characters of G .

A *character* on an LCA group G is a continuous mapping ξ from G into \mathbb{C} such that $|\xi(x)| = 1$ and $\xi(x + y) = \xi(x)\xi(y)$ for all $x, y \in G$. The set of all characters of G is denoted \hat{G} .

Thus a character is a continuous homomorphism of G into \mathbb{C} . It can be shown that the set of characters on \mathbb{R} is the set $\{\varphi_n\}_{n \in \mathbb{Z}}$, where $\varphi_n(t) = e^{int}$ for all $n \in \mathbb{Z}$ and all $t \in [0, 2\pi]$. Every character of \mathbb{R} has the form $n \mapsto e^{int}$ for some $t \in [0, 2\pi]$. The characters of \mathbb{R} are all of the form $x \mapsto e^{ixy}$ for some $y \in \mathbb{R}$.

For any LCA group G , the set of characters \hat{G} can be made an Abelian group. We shall denote its group operation by $+$ (in what follows, this should cause no confusion with the group operation $+$ for G) and define it by $(\xi_1 + \xi_2)(x) = \xi_1(x)\xi_2(x)$ for all $x \in G$. We write $\langle x, \xi \rangle$ for $\xi(x)$.

A topology is defined on \hat{G} by specifying that $\{\xi_n\}_{n=1}^\infty \subset \hat{G}$ converges to $\xi \in \hat{G}$ if $\{\xi_n\}_{n=1}^\infty \subset \hat{G}$ converges uniformly to ξ on all compact subsets of G . It can be shown that this topology turns \hat{G} into a locally compact Abelian group (see [21, Chapter VII, §3]). We say that \hat{G} is the *dual group* of G .

The *Pontryagin Duality Theorem* (see [21, p.189]) asserts that, for $x \in G$ and $\xi \in G$, the mapping $\xi \mapsto \langle x, \xi \rangle$ is a character on G , and every character on G is of this form. Moreover, the topology induced by uniform convergence on compact subsets of G coincides with the original topology on G . In other words, if G is the dual group of G , then G is the dual of G .

We illustrate the Pontryagin Duality Theorem for the LCA group \mathbb{T} . We saw that there exists a bijection between \mathbb{T} and \mathbb{T} , since given any $\xi \in \mathbb{T}$, there is a unique $n \in \mathbb{Z}$ such that $\langle t, \xi \rangle = e^{int}$ for all $t \in \mathbb{T}$. Moreover, the topology T is the discrete topology. Thus $\mathbb{T} \simeq \mathbb{T}$. One may similarly show that $\mathbb{Z} \simeq \mathbb{Z}$. This duality may be described (by abuse of notation) as $\langle n, t \rangle = e^{int} = \langle t, n \rangle$ for all $t \in \mathbb{T}$ and all $n \in \mathbb{Z}$. (On the left hand side, we regard n as an element of \mathbb{Z} while on the right hand side we regard it as a function on \mathbb{T} .)

We now have all the tools needed to define the Fourier transform of an integrable function on an LCA group.

Let G be an LCA group. Then the *Fourier transform* of a function $f \in L^1(G)$ is defined by

$$f(\xi) = \int_G \overline{\langle y, \xi \rangle} f(y) dy$$

for all $\xi \in G$, where dy is Haar measure on G .

Thus the Fourier transform of a function $f \in L^1(\mathbb{T})$ is

$$f(t) = \sum_{n=-\infty}^{\infty} e^{-int} f(n)$$

for all $t \in \mathbb{T}$, since Haar measure on \mathbb{T} is unit point-mass measure. Similarly, the Fourier transform of a function $f \in L^1(\mathbb{Z})$ is

$$f(n) = \frac{1}{2\pi} \int_0^{2\pi} e^{-int} f(t) dt$$

for all $n \in \mathbb{Z}$, where dt is Lebesgue measure. The *Fourier series* of f is the corresponding formal expression

$$\sum_{n=-\infty}^{\infty} f(n) \varphi_n$$

where $\varphi_n(t) = e^{int}$ for all $t \in \mathbb{T}$ and all $n \in \mathbb{Z}$.

We aim to use the Fourier transform to construct a large class of operators which act on the L^p space of a given LCA group. We may do this via *Plancherel's Theorem*.

[21, Chapter VII, §4] *Plancherel's Theorem*. Let G be a locally compact Abelian group. Then the Fourier transform on $L^1(G)$ is an isometry of $L^1 \cap L^2(G)$ onto a dense subspace of $L^2(G)$. Hence it can be extended to an isometry of $L^2(G)$ onto $L^2(G)$.

For the circle group \mathbb{T} , Plancherel's theorem takes the following form.

[21, Theorem I.5.5]

(i) Let $f \in L^2()$. Then

$$\sum_{n=-\infty}^{\infty} |f(n)|^2 = \frac{1}{2\pi} \int_0^{2\pi} |f(t)|^2 dt,$$

that is, $f_{L^2()} = f_{L^2()}$.

(ii) For each $f \in L^2()$,

$$f = \lim_{N \rightarrow \infty} \sum_{n=-N}^N f_n \varphi_n$$

in $L^2()$ norm, where $\varphi_n : t \mapsto e^{int}$.

(iii) Given any sequence $\{a_n\}_{n=-\infty}^{\infty}$ of complex numbers in $L^2()$, then there exists a unique $f \in L^2()$ such that $a_n = f(n)$.

Thus the correspondence $f \leftrightarrow \{f(n)\}_{n=-\infty}^{\infty}$ is an isometry between $L^2()$ and $L^2()$.

A by-product of the above theorem is that the set $\{\varphi_n\}_{n \in \mathbb{Z}}$ of functions in the Hilbert space $L^2()$ forms a complete orthonormal system (and hence an unconditional basis) with respect to the inner product

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(t) \overline{g(t)} dt.$$

It is harder to establish whether or not $\{\varphi_n\}_{n \in \mathbb{Z}}$ is a basis for $L^p()$ when $1 \leq p < \infty$ and $p \neq 2$. Standard results about Fejér's kernel (see Section 3.1 and [21, I.2.6]) give the following facts. First recall that a *trigonometric polynomial* is a function on of the form $\sum_{n=-N}^N a_n \varphi_n$ where $N \in \mathbb{Z}$ and $a_n \in \mathbb{C}$.

Fix $1 \leq p < \infty$. Then

- (i) the trigonometric polynomials are dense in $L^p()$;
- (ii) if $f, g \in L^p()$ and $f(n) = g(n)$ for all $n \in \mathbb{Z}$, then $f = g$; and
- (iii) if the Fourier series of a function $f \in L^p()$ does converge in L^p -norm, then it converges to f .

So if the Fourier series does converge for each function in $L^p()$, $\{\varphi_n\}_{n \in \mathbb{Z}}$ is a basis for $L^p()$. Otherwise, one would like to know whether or not $\{\varphi_n\}_{n \in \mathbb{Z}}$ is a basic sequence in $L^p()$ for $p \neq 2$, and whether this sequence is unconditional. One way to study such problems is through multiplier transforms.

Let G be an LCA group and let $\phi : G \rightarrow \mathbb{C}$ be bounded and measurable on G . By Plancherel's Theorem, define S_ϕ to be the continuous linear mapping of $L^2(G)$ into itself for which $(S_\phi f)^\wedge = \phi f^\wedge$ for all $f \in L^2(G)$. Then for $1 \leq p \leq \infty$, ϕ is said to be an $L^p(G)$ *multiplier* if and only if there is a constant $C > 0$ such that

$$S_\phi f_p \leq C_p f_p \tag{3.1}$$

for all $f \in L^2 \cap L^p()$. In this case S_ϕ is called the *multiplier transform corresponding to ϕ on $L^p(G)$* and ϕ may also be referred to as a *Fourier multiplier* or a *multiplier function* for $L^p(G)$.

The space of Fourier multipliers for $L^p(G)$ will be denoted by $M_p(G)$. We turn $M_p(G)$ into a Banach space by equipping it with the following norm:

for $\phi \in M_p(G)$ define $\phi_{M_p(G)}$ to be the usual operator norm on $L^p(G)$ for the operator S_ϕ . Thus $\phi_{M_p(G)}$ is the smallest possible $C \geq 0$ for which (3.1) holds. It is trivial to check from the definitions that if $\phi_1, \phi_2 \in M_p(G)$, then the product $\phi_1\phi_2$, defined by pointwise operations, is in $M_p(G)$ and

$$\phi_1\phi_2_{M_p(G)} \leq \phi_{1_{M_p(G)}}\phi_{2_{M_p(G)}}.$$

In fact, we have the following result.

[5, §3] Let G be an LCA group. Then for $1 \leq p < \infty$, $M_p(G)$ is a Banach algebra under pointwise operations, and the mapping $\phi \mapsto S_\phi$ is an identity-preserving algebra homomorphism of $M_p(G)$ into $(L^p(G))$.

The next example illustrates some of the concepts that have been introduced so far.

Let $G = \mathbb{T}$ and let $\varphi_n : t \mapsto e^{int}$ for all $t \in \mathbb{T}$ and $n \in \mathbb{Z}$. Fix an $m \in \mathbb{Z}$ and consider the characteristic function $\chi_{\{m\}}$ on \mathbb{T} . Then for any $f \in L^2(\mathbb{T})$ we have, by Definition 1.1, $\chi_{\{m\}} f = f(m)\chi_{\{m\}}$.

$$(S_{\chi_{\{m\}}} f)(n) = \chi_{\{m\}}(n)f(n) = \chi_{\{m\}}(n)f(m) = \varphi_m(n)f = (f(m)\varphi_m)(n)$$

for all $n \in \mathbb{Z}$. Thus $S_{\chi_{\{m\}}} f = f(m)\varphi_m$, and $S_{\chi_{\{m\}}}$ projects each f onto the m th term of its Fourier series.

Now we consider the convolution product of f with φ_m . For almost all $t \in \mathbb{T}$ we have

$$\begin{aligned} (\varphi_m * f)(t) &= \frac{1}{2\pi} \int_0^{2\pi} \varphi_m(t-s)f(s)ds \\ &= e^{imt} \frac{1}{2\pi} \int_0^{2\pi} e^{-ims}f(s)ds \\ &= e^{imt}f(m). \end{aligned}$$

Thus $S_{\chi_{\{m\}}}$ and the operator $f \mapsto \varphi_m * f$ coincide on $L^2(\mathbb{T})$. It is easy to see that $S_{\chi_{\{m\}}} = 1$ and hence $\chi_{\{m\}}_{M_p(G)} = 1$.

We now give one reason why multiplier transforms are of interest to us. Let $\varphi_n(t) = e^{int}$ for each $n \in \mathbb{Z}$ and $t \in \mathbb{T}$. One might hope that $\{\varphi_n\}_{n \in \mathbb{Z}}$ is a basis for $L^p(\mathbb{T})$. Assume for the moment that this is the case. The expansion of a function $f \in L^p(\mathbb{T})$ with respect to this basis is then given by $\sum_{n \in \mathbb{Z}} f(n)\varphi_n$. If one wanted to prove that the basis was unconditional, it would suffice to find (by Proposition 1.1) a constant C_p depending only on p such that

$$\sum_{n \in \mathbb{Z}} \varepsilon_n f(n) \varphi_n \leq C_p f_p$$

for all $f \in L^p(\mathbb{T})$ and all choices $\varepsilon_n = \pm 1$. This would be equivalent to finding a constant C_p such that

$$S_\varepsilon f_p \leq C_p f_p$$

for all $f \in L^2 \cap L^p(\mathbb{T})$ and all functions $\varepsilon : \mathbb{Z} \rightarrow \{-1, 1\}$. Motivated by this, we develop the general theory of multipliers a little further.

CHAPTER 4

Conclusion

In mathematics, certain kinds of mistaken proof are often exhibited, and sometimes collected, as illustrations of a concept of mathematical fallacy. There is a distinction between a simple mistake and a mathematical fallacy in a proof: a mistake in a proof leads to an invalid proof just in the same way, but in the best-known examples of mathematical fallacies, there is some concealment in the presentation of the proof. For example, the reason validity fails may be a division by zero that is hidden by algebraic notation. There is a striking quality of the mathematical fallacy: as typically presented, it leads not only to an absurd result, but does so in a crafty or clever way. Therefore these fallacies, for pedagogic reasons, usually take the form of spurious proofs of obvious contradictions. Although the proofs are flawed, the errors, usually by design, are comparatively subtle, or designed to show that certain steps are conditional, and should not be applied in the cases that are the exceptions to the rules.

The traditional way of presenting a mathematical fallacy is to give an invalid step of deduction mixed in with valid steps, so that the meaning of fallacy is here slightly different from the logical fallacy. The latter applies normally to a form of argument that is not a genuine rule of logic, where the problematic mathematical step is typically a correct rule applied with a tacit wrong assumption. Beyond pedagogy, the resolution of a fallacy can lead to deeper insights into a subject (such as the introduction of Pasch's axiom of Euclidean geometry and the five color theorem of graph theory). *Pseudaria*, an ancient lost book of false proofs, is attributed to Euclid.

Mathematical fallacies exist in many branches of mathematics. In elementary algebra, typical examples may involve a step where division by zero is performed, where a root is incorrectly extracted or, more generally, where different values of a multiple valued function are equated. Well-known fallacies also exist in elementary Euclidean geometry and calculus.

References

- [1] Blei, David M. "Probabilistic topic models." *Communications of the ACM* 55.4 (2012): 77-84.
- [2] Barron, K., Kung, E. and Proserpio, D., 2018, June. The Sharing Economy and Housing Affordability: Evidence from Airbnb. In *EC* (p. 5).
- [3] Benzinger, H., Berkson, E. and Gillespie, T.A., Spectral families of projections, semigroups, and differential operators, *Tran. Amer. Math. Soc.* **275** (1983), 431–475.
- [4] Berkson, E. and Gillespie, T.A., Fourier series criteria for operator decomposability, *Integral Equations Operator Theory* **9** (1986), 767–789.
- [5] Berkson, E., and Gillespie, T.A., Spectral decompositions and harmonic analysis on UMD spaces, *Studia Mathematica* **112**(1) (1994), 12–49.
- [6] Berkson, E. Gillespie, T.A. and Muhly, P.S., Abstract spectral decompositions guaranteed by the Hilbert transform, *Proc. London Math. Soc.* (3) **53** (1986), 489–517.
- [7] Bourgain, J., *Martingale transforms and geometry of Banach spaces*, Israel seminar on geometrical aspects of functional analysis (1983/84), XIV, 16 pp., Tel Aviv Univ., Tel Aviv, 1984.
- [8] Bourgain, J., Some remarks on Banach spaces in which martingale difference sequences are unconditional, *Ark. Mat.* **21** (1983), 163–168.
- [9] Bourgain, J., Vector-valued singular integrals and the H^1 -BMO duality, *Probability Theory and Harmonic Analysis* (ed. Chao, J. A.) (1986), 1–19.
- [10] Burkholder, D.L., A geometrical characterisation of Banach spaces in which martingale difference sequences are unconditional, *Ann. Probability* **9** (1981), 997–1011.
- [11] Burkholder, D.L., A geometric condition that implies the existence of certain singular integrals of Banach-space valued functions, *Proceedings of Conference on Harmonic Analysis in Honor of Antoni Zygmund* (Chicago, Illinois, 1981), Wadsworth Publishers: Belmont, 1983, pp. 270–286.
- [12] Burkholder, D.L., Martingales and Fourier analysis in Banach spaces, *Lecture Notes in Mathematics* **1206** (1986), 61–108.
- [13] Burkholder, D.L., Martingale transforms, *Ann. Math. Statist.* **37** (1966), 1494–1504.
- [14] Coifman R.R., and Weiss, G., *Transference methods in analysis*, Regional Conference Series in Mathematics 31, Amer. Math. Soc., Providence, R.I., 1977.
- [15] Conway, J., *A Course in Functional Analysis*, Springer-Verlag: New York, 1985.

- [16] Dowson, H., *Spectral theory of linear operators*, London Math. Soc. Monographs, No. 12. Academic Press: New York, 1978.
- [17] Doust, I., Norms of 0-1 matrices in $_p$, pp 50-55, Proc. Centre Math. Appl. Austral. Nat. Univ., 39, Austral. Nat. Univ., Canberra, 2000.
- [18] Doust, I. and Gillespie, T. A., Schur multipliers on $_p$ spaces, in preparation.
- [19] Dunford, N., and Schwartz, J.T. *Linear operators I: General theory*, Pure and Applied Mathematics 7. Interscience: New York, 1958.
- [20] Edwards, R.E., and Gaudry, G.I., *Littlewood–Paley and Multiplier Theory*, Springer-Verlag: Berlin, 1977.
- [21] Katznelson, Y., *An introduction to harmonic analysis*, Second corrected edition, Dover: New York, 1976.
- [22] Gohberg, I.C., and Kreĭn, M. G., *Introduction to the theory of linear nonselfadjoint operators*, Translations of Mathematical Monographs, Vol. 18, Amer. Math. Soc., Providence, R.I., 1969.
- [23] Gowers, W. T., and Maurey, B., The unconditional basic sequence problem, *J. Amer. Math. Soc.* **6** (1993), 851–874.
- [24] Lindenstrauss, J., and Tzafriri, L., *Classical Banach Spaces I and II*, Springer: Berlin, 1996.
- [25] Pedersen, G. K., *Analysis Now*, Springer: New York, 1989.
- [26] Stein, E. M., The development of square functions in the work of A. Zygmund, *Bulletin (New Series) of the Amer. Math. Soc.* **6** (1982), 5–30.
- [27] Stein, E. M., *Topics in Harmonic Analysis Related to the Littlewood–Paley Theory*, Princeton UP: Princeton, 1970.