TRACES OF OPERATORS WITH INTEGRABLE KERNELS

by

Engin Başakoğlu

B.S., in Mathematics, Middle East Technical University, 2013

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Mathematics
Boğaziçi University
2016

ACKNOWLEDGEMENTS

I would like to express my gratitude to my thesis supervisor Prof. Selçuk Demir for his guidance and support during the preparation of this study.

Also, I would like to thank Assoc. Prof. Burak Gürel, Prof. O. Teoman Turgut, Assist. Prof. Kemal I. Eroğlu for participating my thesis committee.

I am grateful to my parents for all their love and endless support.

My special thanks is to my friend Mustafa Türköz who always become a sincere friend, supporter and colleague.

I am grateful to Gamze Baykaldı for her priceless love and endless encouragement.

ABSTRACT

TRACES OF OPERATORS WITH INTEGRABLE KERNELS

Trace class operators are compact operators whose singular values are absolutely summable. C. Brislawn proved in [1] and [2] that a trace formula for trace class operator on $L^2(X)$ is determined by it's corresponding kernel where $X \subseteq \mathbb{R}^n$ $\big($in [1]$\big)$ or X is countably generated measure space $\big($in [2]$\big)$. We will benefit from [1] and [2] in this thesis. Yet the main aim of this thesis is to establish tools in order to find a trace formula for trace class operators on $L^2(X)$ with a kernel and to be given the proofs in detail by utilizing these tools. Once $X \subseteq \mathbb{R}^n$, proofs make use of Hardy-Littlewood Maximal Theorem; otherwise, when X is countably generated measure space, Hardy-Littlewood Maximal Theorem should be replaced by Doob's Maximal Theorem for martingales which is used to generalize the former results.

ÖZET

İNTEGRALLENEBİLİR ÇEKİRDEKLİ OPERATÖRLERİN İZLERİ

İz sınıfı operatörler tekil değerlerinin mutlak değerleri toplanılabilir tıkız operatorlerdir. C. Brislawn $L^2(X)$ üzerinde tanımlı bir iz sınıfı operatörün iz formülünün, bu operatorün alakalı çekirdeği tarafından belirlendiğini [1] ve [2]'de, X $\mathbb{R}^n$'in bir altkümesi olduğunda $\big([1]'de\big)$ ya da X, topolojisinin bazı sayılabilir, bir ölçü uzayı olduğunda $\big([2]'de\big)$ ispatlamıştır. Bu tezde [1] ve [2]'den yararlanacağız. Fakat bu tezin asıl amacı $L^2(X)$ üzerinde tanımlı çekirdekle verilen iz sınıfı operatörlerin izini bulmak için bir yol oluşturmak ve ispatların ayrıntılı bir şekilde bu yollardan faydalanılarak verilebilmesidir. $X \subseteq \mathbb{R}^n$ olduğunda ispatlar Hardy-Littlewood Maximal Teoremi yardımıyla verilecektir; ancak X, topolojisinin bazı sayılabilir, bir ölçü uzayı olduğunda ise Hardy-Littlewood Maximal Teoremini önceki sonuçları genelleştirmede kullanılan, martingale'ler için Doob'un Maximal Teoremiyle değiştirmek gerekecektir.

TABLE OF CONTENTS

LIST OF SYMBOLS

A^*	Adjoint of an operator A
$A_r f$	Linear operator that averages a function $f \in L^1_{loc}$
$A_{r(n)}L$	The n-dimensional averaging operator acting on the first variable of L
$A^{r(n)}L$	The n-dimensional averaging operator acting on the second variable of L
$\mathcal{B}_X$	Borel σ-algebra on X
$B(r, x)$	The ball of radius r centered at x
$\mathbb{C}$	Set of complex numbers
$\mathrm{cl}S$	Closure of the set S
$E(f\|\mathcal{M})$	Conditional expectation of f relative to σ-algebra $\mathcal{M}$
$E[X]$	Expectation of a random variable X
$\overline{\phi(x)}$	Conjugate of the function $\phi(x)$
$\mathrm{Im}K$	Image of an operator K
$\ker K$	Kernel of an operator K
L^p	Space of measurable complex valued functions whose L^p-norms are finite
L^1_{loc}	Space of locally integrable functions
$\mathscr{L}(\mathscr{H})$	Space of linear operators on the Hilbert space $\mathscr{H}$
$L * J(x, y)$	Convolution of kernels L and J
λ_i	The ith eigenvalue of some operator
$\lambda(\alpha)$	Distribution function of a measurable function on some measurable space
m	Lebesgue measure
$\max_{x\in\mathcal{I}} f(x)$	Maximum of a function f over $\mathcal{I}$
Mf	Maximal function of a function f
$M_{r(n)}L$	The n-dimensional maximal operator acting on the first variable of L
$M^{r(n)}L$	The n-dimensional maximal operator acting on the second variable of L

$\mathbb{N}$	Set of natural numbers
P	Probability measure
$\mathbb{R}^n$	The n-dimensional Euclidean space
$S^{\perp}$	Orthogonal complement of the set S
$\sup_{x\in\mathcal{I}} f(x)$	Supremum of a function $f(x)$ over $\mathcal{I}$
$\sigma(P)$	σ-algebra generated by the set P
T_K	An operator that indicates it's corresponding kernel $K(x, y)$
$\mathrm{tr}T$	Trace of an operator T
$Q_r(x)$	The n-dimensional cube of radius r centered at x
$[dx]$	Lebesgue measure
$\otimes$	Tensor product
$\langle ., . \rangle$	Usual inner product
$\|.\|$	Usual Euclidean norm induced by inner product
$\|.\|_{HS}$	Hilbert-Schmidt norm
$\|.\|_p$	L^p-norm
$\|.\|_\infty$	Essential supremum
$:=$	Equality that includes definition

1. INTRODUCTION

Let T_K be a compact operator on a Hilbert space $\mathscr{H}$. Then singular value decomposition of T_K shows us that

$$T_K = \sum_{i=1}^{\infty} \lambda_i \phi_i \otimes \overline{\psi_i} \tag{1.1}$$

is norm-convergent in $\mathscr{L}(\mathscr{H})$, where $\{\phi_i : i \in \mathbb{N}\}$ and $\{\psi_i : i \in \mathbb{N}\}$ are orthonormal sequences.

Therefore if T_P is a positive-definite Hilbert-Schmidt operator on $L^2(X)$ which is given by the kernel $P(x, y)$, then T_P has an eigenfunction expansion

$$T_P = \sum_{i=1}^{\infty} \lambda_i \phi_i \otimes \overline{\phi_i} \tag{1.2}$$

with corresponding expansion for the kernel

$$P(x, y) = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \overline{\phi_i(y)} \tag{1.3}$$

which is convergent in $L^2(X \times X)$, where $\{\phi_i : i \in \mathbb{N}\}$ is orthonormal sequence in $L^2(X)$.

In this thesis our main concern is to find a trace formula for an arbitrary trace class operator in terms of trace class kernel. Since any trace class operator is Hilbert-Schmidt as well and can be written as a linear combination of positive-definite trace class operators, expansions (1.2) and (1.3) make sense and they are going to play fundamental role once proving Theorem 3.2. For continuous trace class kernel $K(x, y)$ on a σ-compact, locally compact space X with Radon measure μ, Duflo [3, Theorem V.3.1.1] proves that

$$\text{tr}T_K = \int_X K(x,x)d\mu(x) \tag{1.4}$$

whereas Theorem 3.2 and Theorem 4.8 generalize (1.4) by carrying hypothesis of Duflo further and proving it for any trace class operator on $L^2(\mathbb{R}^n)$ and $L^2(\mu)$ respectively. Also an important idea which Theorem 3.2 and Theorem 4.8 put forward is defining the diagonal values of the trace class kernel by an averaging process where Weidmann introduces it in his paper [4].

In chapter 2, Hardy-Littlewood maximal and averaging operators are defined and their properties, which we need in chapter 3, are given in detail.

In chapter 3 it will become clear that trace formula for arbitrary trace class operator, in Theorem 3.2, could equally well be written in terms of the convolution of Hilbert-Schmidt kernels that stems from the fact that any trace class operator K factors as $K = LJ$, where L and J are Hilbert-Schmidt operators.

All the results in chapter 3 are presented for the trace class operators on $L^2(X)$, $X \subseteq \mathbb{R}^n$ whereas in chapter 4 the validity of these results are shown to hold for countably generated measure space X with σ-finite measure. In other words, chapter 4 reformulate the whole theory from the beginning in a purely measure theoretic way. Moreover in chapter 4 we will figure out that Doob's Maximal and Doob's Martingale Convergence Theorems are analogous versions of Hardy-Littlewood Maximal and Lebesgue Differentiation Theorems respectively.

2. AVERAGING ON HYPERCUBES

2.1. Positiveness of a Kernel

Let $I = [0,1]$. A Hilbert-Schmidt kernel is a summable function $P : I \times I \to \mathbb{C}$ satisfying

$$\int_0^1 \int_0^1 |P(x,y)|^2 dxdy < \infty.$$

In this case we write $P(x,y) \in L^2(I \times I)$, and note that

$$Q(x) = \left(\int_0^1 P^2(x,y)dy \right)^{\frac{1}{2}} \text{ and } R(y) = \left(\int_0^1 P^2(x,y)dx \right)^{\frac{1}{2}}$$

exist in the interval I almost everywhere and they belong to the Hilbert space $L^2(I)$. In the remaining text sometimes we will encounter with the notion of positiveness or positive definiteness of a kernel $P(x,y)$, therefore meanings of these must be clarified quite well for our purpose. Assume that

$$P(x,y) = \sum_{j=1}^{\infty} \lambda_j \phi_j(x) \overline{\phi_j(y)} \tag{2.1}$$

is the orthonormal expansion of the kernel $P(x,y)$ and let us introduce the double integral :

$$J(f,f) = \iint P(x,y) f(x) \overline{f(y)} dxdy \tag{2.2}$$

hence putting (2.1) into the above double integral (2.2) gives rise to the formula

$$J(f,f) = \sum_{j=1}^{\infty} \lambda_j |a_j|^2 \tag{2.3}$$

where

$$a_j = \int \phi_j(x) f(x) dx \tag{2.4}$$

is such that the orthonormal sequence of eigenfunctions $\{\phi_j\}_j$ corresponds to sequence of eigenvalues $\{\lambda_j\}_j$. Moreover,

$$J(\phi_i, \phi_i) = \iint \sum_j \lambda_j \phi_j(x) \overline{\phi_j(y)} \phi_i(x) \overline{\phi_i(y)} dx dy \tag{2.5}$$

$$= \sum_j \lambda_j \int \phi_j(x) \phi_i(x) dx \int \overline{\phi_j(y) \phi_i(y)} dy \tag{2.6}$$

$$= \lambda_i. \tag{2.7}$$

Next we suppose that

$$|\lambda_1| \geq |\lambda_2| \geq ... \geq |\lambda_n| \geq ... \tag{2.8}$$

thus,

$$J(f, f) \leq \sum_{j=1}^{\infty} |\lambda_j| |a_j|^2 \tag{2.9}$$

$$\leq |\lambda_1| \sum_{j=1}^{\infty} |a_j|^2 \tag{2.10}$$

$$\leq |\lambda_1| \|f\|_2^2 \tag{2.11}$$

where (2.11) is due to the Bessel's inequality with $a_j = \langle \phi_j, f \rangle$. Letting $\|f\|_2 = 1$ yields $J(f, f) \leq |\lambda_1|$. Note that for all normalized functions $f \in L^2$, $\max_f |J(f, f)| = |\lambda_1|$ (for this, take $f = \phi_1$). Since $J(\phi_j, \phi_j) = \lambda_j$ by (2.7), to have $J(f, f) = \sum_{j=1}^{\infty} \lambda_j |a_j|^2 \geq 0$ for all $f \in L^2$, we must take $\lambda_j \geq 0$ for all $j \in \mathbb{N}$ where (2.8) becomes $\lambda_1 \geq \lambda_2 \geq ... > 0$. In such a situation, that's to say if (2.2) is non-negative then $P(x, y)$ is called positive, and it is called positive-definite whenever $J(f, f) > 0$ for $\|f\|_2 > 0$. Let's start with Mercer's Theorem which shows the importance of uniform convergence of the positive-definite expansion (2.1) as calculating the trace of the corresponding positive operator of the expansion (2.1). In the statement of the Mercer's Theorem let $T_P = \sum_{i=1}^{\infty} \lambda_i \phi_i \otimes \overline{\phi_i}$ be the positive operator associated with kernel (2.1).

Theorem 2.1 (Mercer's Theorem)**.** *Suppose that $P(x, y) \in L^2(I \times I)$ is a continuous, positive-definite Hilbert-Schmidt kernel for $I = [0, 1]$; then the series $\sum_{i=1}^{\infty} \lambda_i \phi_i(x)\overline{\phi_i(y)}$ converges absolutely and uniformly, and $P(x, y) = \sum_{i=1}^{\infty} \lambda_i \phi_i(x)\overline{\phi_i(y)}$ holds. Consequently,* $\mathrm{tr} T_P = \int_0^1 P(x, x)dx$.

Proof. Suppose that $P : I \times I \to \mathbb{C}$ is a continuous positive-definite function and let $T_P : L^2(I) \to L^2(I)$ be the corresponding Hilbert-Schmidt operator. T_P, generated by the kernel $P(x, y)$, defined by

$$T_P f(x) = \int_0^1 P(x, y)f(y)dy$$

is continuous since $P(x, y)$ is continuous by assumption. Thus, all the eigenfunctions $\phi_i(x) = \frac{1}{\lambda_i} T_P \phi_i(x)$ are continuous for $i = 1, 2, ...$ This makes the following function in (2.12) continuous for each $n \in \mathbb{N}$,

$$P_n(x, y) = P(x, y) - \sum_{i=1}^{n} \lambda_i \phi_i(x)\overline{\phi_i(y)}. \tag{2.12}$$

Due to the theorem in [5] (page 243), the series $\sum_{i=1}^{\infty} \lambda_i \phi_i(x)\overline{\phi_i(y)}$ converges in the mean to $P(x, y)$. Hence, we may write $P_n(x, y) = \sum_{i=n+1}^{\infty} \lambda_i \phi_i(x)\overline{\phi_i(y)}$ in the sense of mean convergence[1]. Also we claim that $P_n(x, x) \geq 0$ for all $x \in I$ and all $n \in \mathbb{N}$; otherwise, if there was a point $x_0 \in I$ such that $P_n(x_0, x_0) < 0$, then by continuity of $P_n(x, y)$ we would have $P_n(x, y) < 0$ in a neighbourhood

$$N_\delta = \{(x, y) : x_0 - \delta < x < x_0 + \delta,\ x_0 - \delta < y < x_0 + \delta\}$$

for sufficiently small $\delta > 0$. To get a contradiction, consider the following computation

[1]Mean convergence is same as the norm convergence on the related normed space, in this text it is L^p convergence for p=2.

for $f \in L^2(I)$:

$$\int_0^1\int_0^1 P_n(x,y)f(y)\overline{f(x)}dxdy = \int_0^1\int_0^1 \sum_{i=n+1}^{\infty} \lambda_i\phi_i(x)\overline{\phi_i(y)}f(y)\overline{f(x)}dxdy \tag{2.13}$$

$$= \sum_{i=n+1}^{\infty} \lambda_i \int_0^1 \phi_i(x)\overline{f(x)}dx \int_0^1 \overline{\phi_i(y)}f(y)dy \tag{2.14}$$

$$= \sum_{i=n+1}^{\infty} \lambda_i \langle \phi_i, f\rangle\langle f, \phi_i\rangle \tag{2.15}$$

$$= \sum_{i=n+1}^{\infty} \lambda_i \langle \phi_i, f\rangle\overline{\langle\ \phi_i, f\rangle} \tag{2.16}$$

$$= \sum_{i=n+1}^{\infty} \lambda_i |\langle \phi_i, f\rangle|^2 \geq 0. \tag{2.17}$$

Since the above double integral (2.13) is always non-negative and the result (2.17) does not depend on the function $f \in L^2(I)$, it is enough to find an L^2 function f such that the double integral (2.13) is negative. We see letting

$$f(x) = \begin{cases} 1, & \text{if } x \in N_\delta^x = \{x : x_0 - \delta < x < x_0 + \delta\} \\ 0, & \text{if } x \in I \setminus N_\delta^x \end{cases}$$

works. Putting this f into the double integral (2.13) gives $\int_0^1\int_0^1 P_n(x,y)dxdy < 0$ in N_δ, a contradiction to (2.17). Therefore, $P_n(x,x) \geq 0$ for $n = 1, 2, ..$ which also yields

$$\sum_{i=1}^{m} \lambda_i|\phi_i(x)|^2 \leq P(x,x), \text{ for m=1,2,...}$$

This proves that $\sum_i \lambda_i|\phi_i(x)|^2$ converges. Let's denote $\max_{x\in I}(P(x,x)) = M$. For any $m, n \in \mathbb{N}$ with $m \geq n$ we have

$$\left(\sum_{i=n}^{m} |\lambda_i\phi_i(x)\overline{\phi_i(y)}|\right)^2 \leq \sum_{i=n}^{m} \lambda_i|\phi_i(x)|^2 \sum_{i=n}^{m} \lambda_i|\phi_i(y)|^2 \tag{2.18}$$

$$\leq M\sum_{i=n}^{m} \lambda_i|\phi_i(x)|^2 \tag{2.19}$$

by the Cauchy-Schwartz inequality. Therefore, for fixed y, $\sum_i \lambda_i \phi_i(x)\overline{\phi_i(y)}$ converges absolutely in x, or for fixed x vice versa. Since $\lambda_i|\phi_i(x)|^2$ is non-negative continuous function over I for all $i \in \mathbb{N}$ and the sum $\sum_i \lambda_i|\phi_i(x)|$ is continuous, by Dini's Theorem [2] $\sum_{i=1}^{\infty} \lambda_i|\phi_i(x)|^2$ is uniformly convergent there. Let $\tilde{P}(x,y) = \sum_{i=1}^{\infty} \lambda_i \phi_i(x)\overline{\phi_i(y)}$ be such that $\tilde{P}$ is the uniform limit of the series in y, then for any $f \in L^2$ we have

$$\begin{aligned}
\int_0^1 \tilde{P}(x,y)f(y)dy &= \int_0^1 \sum_{i=1}^{\infty} \lambda_i \phi_i(x)\overline{\phi_i(y)}f(y)dy \\
&= \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \int_0^1 f(y)\overline{\phi_i(y)}dy \\
&= \sum_{i=1}^{\infty} \lambda_i \phi_i(x)\langle f, \phi_i\rangle \\
&= \sum_{i=1}^{\infty} \phi_i(x)\langle f, \lambda_i\phi_i\rangle \\
&= \sum_{i=1}^{\infty} \phi_i(x)\langle f, T_P\phi_i\rangle \\
&= \sum_{i=1}^{\infty} \phi_i(x)\langle T_P f, \phi_i\rangle,
\end{aligned}$$

this series converges to $T_P f(x)$ thanks to the Hilbert-Schmidt Theorem which can be found in Tricomi [6] page 110, note that in the above equalities we have used positivity of the operator T_P. Consequently,

$$\begin{aligned}
&\int_0^1 \tilde{P}(x,y)f(y)dy = T_P f(x) \\
&\qquad = \int_0^1 P(x,y)f(y)dy \Rightarrow \int_0^1 [\tilde{P}(x,y) - P(x,y)]f(y)dy = 0
\end{aligned}$$

Since $\tilde{P} - P \in L^2$, define $f(y) := \overline{\tilde{P}(x,y) - P(x,y)}$ to obtain

$$\int_0^1 |\tilde{P}(x,y) - P(x,y)|^2 dy = 0 \Rightarrow \ \tilde{P}(x,y) - P(x,y) = 0$$

for fixed x and $0 \leq y \leq 1$. Thus, let $x = y$ to have $P(x,x) = \tilde{P}(x,x) = \sum_{i=1}^{\infty} \lambda_i|\phi_i(x)|^2$.

[2]Dini's Theorem: If the sum of an infinite series of non-negative continuous functions is a continuous function of x in a closed interval then the series converges uniformly there.

Thus by all the above arguments, convergence of the series $\sum_{i=1}^{\infty} \lambda_i \phi_i(x)\overline{\phi_i(y)}$ is uniform even once x and y both vary over I. As a result, we have the nice conclusion for the integral operator T_P generated by the kernel $P(x, y)$, that is

$$\begin{aligned} \text{tr}T_P &= \sum_{i=1}^{\infty} \lambda_i \\ &= \sum_{i=1}^{\infty} \lambda_i \int_0^1 |\phi_i(x)|^2 dx \\ &= \int_0^1 \sum_{i=1}^{\infty} \lambda_i |\phi_i(x)|^2 dx \\ &= \int_0^1 P(x, x) dx. \end{aligned}$$

□

Let's move on to the properties of the maximal and averaging operators which will have utmost significance once proving the particular theorems about kernels of trace class operator.

2.2. Averaging and Maximal Operators

We say that the family $\mathcal{F}$ of measurable subsets of $\mathbb{R}^n$ is regular whenever there exists a constant $c > 0$ such that $S \subset B$ with $m(S) \geq cm(B)$ where $S \in \mathcal{F}$ and B is the ball centered at the origin in $\mathbb{R}^n$. Let $\mathcal{F}_x$ be the family of all n-dimensional cubes $\{Q_r(x)\}_{r>0}$ centered at x. Since we can find a suitable open ball centered at origin in $\mathbb{R}^n$ and a constant $c > 0$ such that $Q_r(x) \subset B$ with $m(Q_r(x)) \geq cm(B)$ for each $r > 0$ and $x \in \mathbb{R}^n$ (e.g. let the radius of the ball be a little bit longer than sum of x's distance to the origin and the diagonal length of the cube), $\mathcal{F}_x$ becomes regular. Next, we shall define the linear operator A_r which averages $f \in L^1_{loc}(\mathbb{R}^n)$ [3]

[3] L^1_{loc} is the space of locally integrable functions $f : \mathbb{R}^n \to \mathbb{C}$ such that $\int_{B(r,x)} |f(x)|dx < \infty$ for every $r > 0$ and $x \in \mathbb{R}^n$, also instead of $B(r, x)$ choosing any bounded measurable set in $\mathbb{R}^n$ changes nothing.

over $Q_r = [-r, r]^n$ or equivalently over, the translation of it, $Q_r(x) = x + Q_r$:

$$A_r f(x) = \frac{1}{m(Q_r)} \int_{Q_r} f(x+s)ds$$
$$= \frac{1}{m(Q_r(x))} \int_{Q_r(x)} f(s)ds.$$

Note that according to [7, Lemma 3.16], $A_r f(x)$ is jointly continuous in x and r ($r > 0$, $x \in \mathbb{R}^n$, $f \in L^1_{loc}$). In what follows Mf is the Hardy-Littlewood maximal function of f:

$$Mf(x) = \sup_{r>0} \frac{1}{m(Q_r(x))} \int_{Q_r(x)} |f(s)|ds$$
$$= \sup_{r>0} \frac{1}{m(Q_r)} \int_{Q_r} |f(x+s)|ds.$$

It must be remarked that Mf is not necessarily finite, it can attain infinite values at some points $x \in \mathbb{R}^n$. Besides, maximal function of f over balls is defined similarly. By the regularity of the family $\mathcal{F}_x$, there exists r_0 such that $Q_r(x) \subset B(r_0, x)$ with $m(Q_r(x)) \geq cm(B(r_0, x))$ and hence

$$Mf(x) \leq c^{-1} \sup_{r_0>0} \frac{1}{m(B(r_0, x))} \int_{B(r_0,x)} |f(t)|dt =: M_{B_{r_0}} f(x).$$

This shows us that proving the following theorem for the maximal operator over balls instead of over cubes does not change anything in principle. I.e. the validity of Theorem 2.3 for the maximal operator over cubes is preserved.

Lemma 2.2 (Covering Lemma)**.** *Let $\mathcal{C}$ be the collection of open balls in $\mathbb{R}^n$, and for an open set $U \in \mathbb{R}^n$ if $U \subset \bigcup_{B_\alpha \in \mathcal{C}} B_\alpha$, then there exists a set of disjoint open balls $\{B_i\}_{i \in I} \subset \mathcal{C}$ (with cardinality finite or infinite) such that $\sum_{i \in I} m(B_i) > Cm(U)$.*

For the proof of Lemma 2.2 see [8, p. 9]. Furthermore C is a constant depending on the dimension n, for instance let $C = 3^{-n}$.

Theorem 2.3 (Maximal Theorem). *(i) If $f \in L^1(\mathbb{R}^n)$, then*

$$m\{x : Mf(x) > \alpha\} \leq \frac{A}{\alpha} \int_{\mathbb{R}^n} |f| dx$$

for every $\alpha > 0$, and $A = 3^n$.

(ii) If $f \in L^p(\mathbb{R}^n)$, then $Mf \in L^p(\mathbb{R}^n)$ and $\|Mf\|_p \leq C_p \|f\|_p$ where C_p depends on the dimension n and p for which $1 < p \leq \infty$.

(iii) If $f \in L^p(\mathbb{R}^n)$ then Mf is finite a.e., $1 \leq p \leq \infty$.

Proof. (i) Let $E_\alpha = \{x : Mf(x) > \alpha\}$. For each $x \in E_\alpha$, there exists $r > 0$ such that

$$A_r|f|(x) = \frac{1}{m(B(r,x))} \int_{B(r,x)} |f(y)| dy > \alpha \Rightarrow \int_{B(r,x)} |f(y)| dy > \alpha m(B(r,x))$$

which gives $m(B(r,x)) < \frac{1}{\alpha}\|f\|_1$. Let $B(r,x) = B_x$ for convenience, and for each such $x \in E_\alpha$, it is clear $\cup_{x \in E_\alpha} B_x \supset E_\alpha$. By the Covering Lemma there exists a mutually disjoint sequence of balls $\{B_k\}_k$ (finite or infinite) such that $\sum_k m(B_k) \geq Cm(E_\alpha)$ with $C = 3^{-n}$, this yields

$$\begin{aligned} \|f\|_1 &\geq \int_{\cup_k B_k} |f(y)| dy \\ &> \alpha \sum_k m(B_k) \\ &\geq \alpha C m(E_\alpha), \end{aligned}$$

it follows

$$\frac{1}{\alpha C} \int_{\mathbb{R}^n} |f(y)| dy \geq m(E_\alpha),$$

put $C = \frac{1}{A}$ to get (i).

(ii) For the case $p = 1$, consider the set E_α in which $Mf(x) > \alpha$ for every $\alpha > 0$ so that $Mf(x) \geq \alpha |x|^{-n}$ holds for $|x| \geq 1$, $n \in \mathbb{N}$. Once c in [7, Corollory 2.52] is equal to 1, we have $B = \{x \in \mathbb{R}^n : |x| < 1\}$ and $B^c = \{x \in \mathbb{R}^n : |x| \geq 1\}$. Since

$Mf(x) \geq \alpha|x|^{-n}$ for every $\alpha > 0$ and $|x| \geq 1$, by part (*b*) of [7, Corollory 2.52] we have $Mf \notin L^1(B^c)$. Hence, $Mf \notin L^1(\mathbb{R}^n)$. Thus, Mf is never integrable on $\mathbb{R}^n$ unless $f \equiv 0$. For the case $p = \infty$, the conclusion $\|Mf\|_\infty \leq A_\infty \|f\|_\infty$ is satisfied trivially by definition $\|f\|_\infty = \inf\{a \geq 0 : m(\{x : |f(x)| > a\}) = 0\}$ and the properties of Mf, so let $A_\infty = 1$. Now, suppose $1 < p < \infty$ and define

$$f_1(x) = \begin{cases} f(x), & \text{if } |f(x)| \geq \frac{\alpha}{2} \\ 0, & \text{otherwise} \end{cases}$$

Then

$$|f(x)| \leq |f_1(x)| + \frac{\alpha}{2} \Rightarrow Mf(x) \leq Mf_1(x) + \frac{\alpha}{2} \Rightarrow Mf_1(x) > \frac{\alpha}{2}. \tag{2.20}$$

Therefore, by (2.20) $E_\alpha = \{x : Mf(x) > \alpha\} \subset \{x : Mf_1(x) > \frac{\alpha}{2}\}$. Let

$$A_\alpha = \{x \in \mathbb{R}^n : |f(x)| > \frac{\alpha}{2},\ f \in L^p(\mathbb{R}^n)\}.$$

Then,

$$\begin{aligned} \infty &> \int_{\mathbb{R}^n} |f(x)|^p dx \\ &\geq \int_{A_\alpha} |f(x)|^p dx \\ &\geq \int_{A_\alpha} (\frac{\alpha}{2})^p dx \\ &= (\frac{\alpha}{2})^p m(A_\alpha) \Rightarrow m(A_\alpha) < \infty. \end{aligned}$$

Thus by [7, Proposition 6.12], $L^1 \supset L^p$ on A_α, so pick f_1 in L^1, once $f \in L^p$ to use the part (i) of the theorem:

$$m(E_\alpha) \leq \frac{2A}{\alpha} \int_{\mathbb{R}^n} |f_1| dx \tag{2.21}$$

$$= \frac{2A}{\alpha} \int_{|f| > \frac{\alpha}{2}} |f| dx. \tag{2.22}$$

Let $\lambda(\alpha) = m\{x : Mf(x) > \alpha\}$ be the distribution function of Mf, then

$$\int_{\mathbb{R}^n} (Mf)^p dx = -\int_0^\infty \alpha^p d\lambda(\alpha) \tag{2.23}$$
$$= p\int_0^\infty \alpha^{p-1}\lambda(\alpha)d\alpha. \tag{2.24}$$

Note that in the second equation we used integration by parts. By (2.22) and (2.24) we have the following inequality:

$$\|Mf\|_p^p = p\int_0^\infty \alpha^{p-1} m(E_\alpha)d\alpha \tag{2.25}$$
$$\leq p\int_0^\infty \alpha^{p-1}\left(\frac{2A}{\alpha}\int_{|f|>\frac{\alpha}{2}} |f(x)|dx\right)d\alpha \tag{2.26}$$
$$= 2Ap\int_0^\infty \alpha^{p-2}\left(\int_{|f|>\frac{\alpha}{2}} |f(x)|dx\right)d\alpha. \tag{2.27}$$

Due to $|2f(x)| > \alpha$ and by Fubini's Theorem (2.27) becomes

$$\begin{aligned} 2Ap\int_{\mathbb{R}^n}\int_0^{2|f(x)|} \alpha^{p-2}|f(x)|d\alpha dx &= \frac{2Ap}{p-1}\int_{\mathbb{R}^n} |2f|^{p-1}|f|dx \\ &= \frac{2^pAp}{p-1}\int_{\mathbb{R}^n} |f(x)|^p dx \\ &= (C_p)^p\|f\|_p^p \end{aligned}$$

where $C_p = 2\left(\frac{3^n p}{p-1}\right)^{\frac{1}{p}}$.

(iii) follows from (ii).

□

Corollary 2.4. *If $f \in L^p(\mathbb{R}^n)$, $1 \leq p \leq \infty$ or more generally, if $f \in L^1_{loc}(\mathbb{R}^n)$ then*

$$\lim_{r\to 0} \frac{1}{m(Q_r(x))}\int_{Q_r(x)} f(s)ds = f(x) \ \ a.e. \ [dx].$$

Proof. See [8, p. 8]. □

By definitions $|A_r f(x)| \leq Mf(x)$, so the Maximal Theorem implies A_r is bounded on $L^p(\mathbb{R}^n)$ whenever $1 < p \leq \infty$. Extend $A_r f(x)$ by defining $A_0 f(x) = f(x)$ on the set $\{x : \lim_{r\to 0} A_r f(x) = f(x) \text{ a.e. } [dx]\}$. At those x's, $|A_0(x)| \leq Mf(x)$ also holds, and an extension of $A_r f$ on this set is also a continuous function of $r \in [0, \infty)$. Thus, if we define $\tilde{f}(x) = \lim_{r\to 0} A_r f(x)$, then $\tilde{f}$ exists a.e. and $\tilde{f}(x) = f(x)$ a.e. $[dx]$.

Lemma 2.5. *Let $\{f_n\}_{n\in\mathbb{N}} \subset L^p(\mathbb{R}^n)$ be such that $f_n \to f$ in $L^p(\mathbb{R}^n)$ with $1 \leq p \leq \infty$ and $r > 0$, then $A_r f_n \to A_r f$ uniformly.*

Proof. Let $x + t = y$ and without loss of generality assume that $m(Q_r) \geq 1$, then

$$\begin{aligned}
|A_r f_n(x) - A_r f(x)| &= \frac{1}{m(Q_r)} \left| \int_{Q_r} (f_n - f)(y) dy \right| \\
&\leq \int_{Q_r} |(f_n - f)(y)| dy \\
&\leq \int_{\mathbb{R}^n} |(f_n - f)(y)| dy = \|f_n - f\|_1 \\
&\leq \|f_n - f\|_q^{\lambda} \|f_n - f\|_p^{1-\lambda}
\end{aligned}$$

where

$$\lambda = \frac{1 - p^{-1}}{q^{-1} - p^{-1}}, \ \lambda \in (0, 1) \text{ and } 0 < q < 1 < p \leq \infty.$$

We see in either case whether $f_n \to f$ in L^1 or $f_n \to f$ in L^p, $1 < p \leq \infty$, we are done with [7, Proposition 6.10]. Thus, $A_r f_n \to A_r f$ uniformly as $n \to \infty$. □

Corollary 2.6. *If $\sum f_n$ converges to f in L^p norm, then $\sum A_r f_n$ converges uniformly to $A_r f$.*

Proof. Use Lemma 2.5. □

Earlier on we defined linear operator A_r that averages $f \in L^1_{loc}(\mathbb{R}^n)$ over cubes, this time for $f \in L^1_{loc}(\mathbb{R}^{2n})$, to indicate the dimension we define it as follows:

$$A_r^{(2n)} f(x,y) = \frac{1}{(m(Q_r))^2} \int_{Q_r} \int_{Q_r} f(x+s, y+t) ds dt,$$

and the reason why it is defined on cubes is the following:

Lemma 2.7. *Let ϕ, ψ be in $L^2(\mathbb{R}^n)$; then*

$$A_r^{(2n)}(\phi \otimes \psi)(x,y) = A_r^{(n)}\phi(x) A_r^{(n)}\psi(y).$$

Proof.

$$\begin{aligned} A_r^{(2n)}(\phi \otimes \psi)(x,y) &= \frac{1}{(m(Q_r))^2} \int_{Q_r} \int_{Q_r} (\phi \otimes \psi)(x+s, y+t) ds dt \\ &= \frac{1}{(m(Q_r))^2} \int_{Q_r} \int_{Q_r} \phi(x+s)\psi(y+t) ds dt \\ &= \left(\frac{1}{m(Q_r)} \int_{Q_r} \phi(x+s) ds \right) \left(\frac{1}{m(Q_r)} \int_{Q_r} \psi(y+t) dt \right) \\ &= A_r^{(n)}\phi(x) A_r^{(n)}\psi(y) \end{aligned}$$

for ϕ and $\psi \in L^2(\mathbb{R}^n)$. □

Lemma 2.8. *The maximal function is submultiplicative and subadditive.*

Proof. It is submultiplicative:

$$\begin{aligned} A_r^{(2n)}(|\phi \otimes \psi|)(x,y) &= A_r^{(n)}|\phi(x)| A_r^{(n)}|\psi(y)| \\ &\leq M^{(n)}\phi(x) M^{(n)}\psi(y) \end{aligned}$$

and taking supremum of the two sides with respect to r gives

$$M^{(2n)}(\phi \otimes \psi)(x,y) \leq M^{(n)}\phi(x) M^{(n)}\psi(y).$$

It is subadditive:

$$\begin{aligned} M(\phi+\psi)(x) &= \sup_{r>0} \frac{1}{m(Q_r)} \int_{Q_r} |(\phi+\psi)(x+t)|dt \\ &\leq \sup_{r>0} \frac{1}{m(Q_r)} \int_{Q_r} |\phi(x+t)|dt + \sup_{r>0} \frac{1}{m(Q_r)} \int_{Q_r} |\psi(x+t)|dt \\ &= M\phi(x) + M\psi(x). \end{aligned}$$

□

Remark 1. Next chapter theorems are proven for trace class operators on $L^2(X)$ where $X = \mathbb{R}^n$, but proofs will be still valid if X is any measurable subset of $\mathbb{R}^n$. Therefore let $X = \mathbb{R}^n$ in chapter 3 in order to avoid any confusion.

3. TRACE CLASS OPERATORS ON $L^2(\mathbb{R}^n)$

3.1. Hilbert-Schmidt Operators

A Hilbert-Schmidt operator K on any Hilbert space to another is a bounded operator satisfying $\sum_i {\lambda_i}^2 < \infty$, where $\lambda_i' s$ are singular values of K, the eigenvalues of $|K| = (K^*K)^{\frac{1}{2}}$. An operator T from one Hilbert space $\mathscr{H}$ to itself which satisfies $\langle Th, h\rangle > 0$ for all $h \in \mathscr{H}$ with $\|h\| > 0$ is called positive-definite, if merely $\langle Th, h\rangle \geq 0$ for all $h \in H$, then it is called positive. Because it is a fact that a bounded operator T is self-adjoint if and only if $\langle Th, h\rangle \in \mathbb{R}$ for all $h \in H$, it is easily seen that K^*K is positive and K being Hilbert-Schmidt implies that K^*K is compact self adjoint operator. Therefore by the Spectral Theorem for compact self adjoint operators, K^*K becomes diagonalizable. Assume now that K is also positive-definite so that $K^*K = K^2$. Then K^2 is diagonalized by an orthonormal sequence $\{\phi_i\}_{i\in\mathbb{N}}$ of eigenvectors with corresponding eigenvalues $\{\mu_i\}_{i\in\mathbb{N}}$, $\mu_i > 0$. Since K^2 is positive it has a unique positive root $K = |K|$. Define $\lambda_i = \sqrt{\mu_i}$ in which λ_i's are eigenvalues of $|K|$. Since $\text{clIm}(K^2) = \ker(K^2)^\perp$ and $\ker(K^2) = \ker(K)$ we get $\text{clspan}\{\phi_i\}_{i\in\mathbb{N}} = \ker(K)^\perp$. Thus by the Spectral Theorem, for any $v \in \ker(K)^\perp$ we can write $Kv = \sum_{i=1}^{\infty} \lambda_i \langle v, \phi_i\rangle \phi_i$. As a result, by using rank 1 operator $\phi \otimes \overline{\phi} : v \mapsto \langle v, \phi\rangle\phi$ one can conclude that every positive-definite Hilbert-Schmidt operator K has the eigenfunction expansion $K = \sum_{i=1}^{\infty} \lambda_i \phi_i \otimes \overline{\phi_i}$ where $\{\phi_i\}_{i\in\mathbb{N}}$ is an orthonormal sequence and the eigenvalues, λ_i's, are positive and square summable.

Moreover Hilbert-Schmidt operators on $L^2(X)$ is considered as integral operators generated by the given integral kernels $K(x, y) \in L^2(X \times X)$. Let $\{\phi_i\}_{i\in\mathbb{N}}$ be the orthonormal sequence in $L^2(X)$ with the corresponding sequence of positive eigenvalues $\{\lambda_i\}_{i\in\mathbb{N}}$ such that $\sum_{i\in\mathbb{N}} \lambda_i^2 < \infty$. It is known from Hilbert space theory that $\sum_{i\in\mathbb{N}} \lambda_i^2$ converges iff $\sum_{i\in\mathbb{N}} \lambda_i\phi_i$ converges in the mean to an element of L^2 (here mean convergence means norm convergence for the space induced by that norm). Since $\{\phi_i(x)\overline{\phi_i(y)}\}_{i\in\mathbb{N}}$ form an orthonormal sequence of $L^2(X \times X)$, $\sum_i \lambda_i\phi_i(x)\overline{\phi_i(y)}$ converges in the mean to some function in $L^2(X \times X)$. Let this function be $\tilde{P}(x, y)$. Our aim for a while is to show that $\tilde{P}(x,y)$ equals pointwise almost everywhere to the kernel $P(x,y)$ that generates positive-definite Hilbert-Schmidt operator $T_P = \sum_{i=1}^{\infty} \lambda_i \phi_i \otimes \overline{\phi_i}$ where the orthonormal sequence $\{\phi_i\}_{i\in\mathbb{N}}$ corresponds to sequence of eigenvalues $\{\lambda_i\}_{i\in\mathbb{N}}$ of T_P. For any f and g in $L^2(X)$ let $F(x,y) = g(x)\overline{f(y)}$ be in $L^2(X \times X)$. Then

$$\begin{aligned}
\langle\, \tilde{P}, F\rangle &= \int_X \int_X \sum_i \lambda_i \phi_i(x)\overline{\phi_i(y)} f(y)\overline{g(x)}dxdy \\
&= \sum_i \lambda_i \int_X \phi_i(x)\overline{g(x)}dx \int_X f(y)\overline{\phi_i(y)} \\
&= \sum_i \lambda_i \langle\, \phi_i, g\rangle \langle\, f, \phi_i\rangle.
\end{aligned}$$

Since g is an element of separable Hilbert space $L^2(X)$ we can write $g(x) = h(x) + \sum_i \langle\, g, \phi_i\rangle \phi_i(x)$ where $h \in L^2(X)$ such that $T_P h(x) = 0$, for this positive T_P, we have the following calculation:

$$\begin{aligned}
\langle\, T_P f, g\rangle &= \langle\, f, T_P h\rangle + \langle\, T_P f, \sum_i \langle\, g, \phi_i\rangle \phi_i\rangle \\
&= \sum_i \overline{\langle\, g, \phi_i\rangle} \langle\, T_P f, \phi_i\rangle \\
&= \sum_i \langle\, \phi_i, g\rangle \langle\, T_P f, \phi_i\rangle \\
&= \sum_i \langle\, \phi_i, g\rangle \langle\, f, T_P \phi_i\rangle \\
&= \sum_i \langle\, \phi_i, g\rangle \langle\, f, \lambda_i \phi_i\rangle \\
&= \sum_i \lambda_i \langle\, \phi_i, g\rangle \langle\, f, \phi_i\rangle
\end{aligned}$$

also,

$$\begin{aligned}
\langle\, T_P f, g\rangle &= \int_X \int_X P(x,y) f(y)\overline{g(x)}dxdy \\
&= \langle\, P, F\rangle.
\end{aligned}$$

Hence, $\langle\, \tilde{P}, F\rangle = \langle\, P, F\rangle \Leftrightarrow \langle\, \tilde{P} - P, F\rangle = 0$. Since finite rank kernels are of the type $\sum_{i=1}^{n} \phi_i(x)\overline{\psi_i(y)}$ where ϕ_i's and ψ_i's are given elements of $L^2(X)$, $(\tilde{P} - P)(x,y)$

is orthogonal to all kernels of finite rank. Furthermore, set of finite rank kernels are everywhere dense in $L^2(X \times X)$ by a theorem in [5, p. 158]. Therefore, $(\tilde{P} - P)(x, y)$ is orthogonal to itself. I.e., $\tilde{P}(x, y) = P(x, y)$ a.e., so

$$\sum_i \lambda_i \phi_i(x) \overline{\phi_i(y)}$$

converges in the mean to $P(x, y)$ in $L^2(X \times X)$. Moreover, the series $T_P = \sum_i \lambda_i \phi_i \otimes \overline{\phi_i}$ converges in the Hilbert-Schmidt norm:

$$\begin{aligned}
\|T_P\|_{HS}^2 &= tr(T_P^2) \\
&= \sum_j \langle\, T_P^* T_P \phi_j, \phi_j \rangle \\
&= \sum_j \left\| T_P \phi_j \right\|^2 \\
&= \sum_j \sum_k |\langle\, T_P \phi_j, \phi_k \rangle|^2 \\
&= \int_X \int_X |P(x, y)|^2 d\mu(x) d\mu(y).
\end{aligned}$$

Let $\psi_{kj} = \phi_k \overline{\phi_j}$. Then, last equality above holds due to the followings :

$$\begin{aligned}
\langle\, T_P \phi_j, \phi_k \rangle &= \int_X \int_X P(x, y) \phi_j(y) \overline{\phi_k(x)} d\mu(x) d\mu(y) \\
&= \int_X \int_X P(x, y) \overline{\psi}_{kj}(x, y) d\mu(x) d\mu(y) \\
&= \langle\, P, \psi_{kj} \rangle.
\end{aligned}$$

Since $\{\psi_{kj}\}_{k,j=1}^{\infty}$ is an orthonormal basis of $L^2(X \times X)$ and $P(x, y) \in L^2(X \times X)$, by the Parseval's equality we obtain

$$\sum_{k,j} |\langle\, P, \psi_{kj} \rangle|^2 = \int_X \int_X |P(x, y)|^2 d\mu(x) d\mu(y)$$

that is convergent. The Hilbert-Schmidt norm dominates the operator norm, hence $\sum_i \lambda_i \phi_i \otimes \overline{\phi_i}$ converges in the operator norm as well. Let K be a trace-class operator.

I.e.,

$$\sum_j \lambda_j < \infty$$

where $\{\lambda_j\}_j$ is a sequence of non-zero eigenvalues of $|K|$ (multiplicities taken into account), then whenever

$$\text{tr}K = \int_X K(x,x)d\mu(x) \tag{3.1}$$

holds? Or is (3.1) even well defined for the given kernel $K(x,y)$ in $L^2(X \times X)$? The answer to the latter question is unfortunately negative in the sense that $K(x,y)$ being just the element of $L^2(X \times X)$ is not enough to have the formula (3.1) in general. On the other hand, when the kernel $K(x,y)$ is continuous square-integrable on a σ-compact, locally compact space X with Radon measure μ, Duflo [3, Theorem V.3.1.1] discovered that (3.1) holds. Bearing in mind Corollory 2.4 we shall define a pointwise representative for $K(x,y)$

$$\lim_{r\to 0} A_r^{(2n)} K(x,y) = \tilde{K}(x,y) \;\; \text{a.e. } [dxdy].$$

Definition 3.1. Let $P(x,y) = \sum_i \lambda_i \phi_i(x)\overline{\phi_i(y)}$ be an expansion for $T_P \geq 0$ where $\{\phi_i\}_i$ is a sequence of pointwise representatives for the eigenfunctions in $L^2(\mathbb{R}^n)$. If $\lim_{r\to 0} A_r\phi_i(x) = \phi_i(x)$ for all $i \in \mathbb{N}$, then $x \in \mathbb{R}^n$ is called a regular point of that expansion.

Since $\{\phi_i\}_i \subset L^2(\mathbb{R}^n)$, $\lim_{r\to 0} A_r\phi_i(x) = \phi_i(x)$ for almost every x by Corollory 2.4. Thus, almost every point in $\mathbb{R}^n$ is a regular point of the given kernel expansion. Earlier on we showed that $\sum_i \lambda_i\phi_i(x)\overline{\phi_i(y)}$ converges to $P(x,y) \in L^2(X \times X)$ in the mean, i.e., in L^2-norm. Hence for every $(x,y) \in \mathbb{R}^{2n}$ and $r > 0$, $\sum_i A_r^{(2n)}\lambda_i\phi_i(x)\overline{\phi_i(y)}$

converges uniformly to $A_r^{(2n)}P(x,y)$ owing to Corollory 2.6. By means of these facts,

$$\begin{aligned} A_r^{(2n)}P(x,y) &= \sum_i \lambda_i A_r^{(2n)}\phi_i(x)\overline{\phi_i(y)} \\ &= \sum_i \lambda_i A_r^{(2n)}(\phi_i \otimes \overline{\phi_i})(x,y) \\ &= \sum_i \lambda_i A_r^{(n)}\phi_i(x) A_r^{(n)}\overline{\phi_i(y)} \end{aligned}$$

at every $(x,y) \in \mathbb{R}^{2n}$ and $r > 0$.

Theorem 3.2. *Let K be a trace-class operator on $L^2(\mathbb{R}^n)$; then $M^{(2n)} \in L^1(\mathbb{R}^n)$, $\tilde{K}(x,x)$ exist a.e. $[dx]$, and* $\mathrm{tr}K = \int \tilde{K}(x,x)dx$.

Proof. It is certain that every bounded operator T can be expressed as $B + iC$ where B and C are self adjoint (to see put $B = \frac{1}{2}(T+T^*)$ and $C = \frac{1}{2i}(T-T^*)$). Also owing to [9 Theorem 2.7.15], for every self adjoint compact operator T, there are unique positive compact operators A and B such that $T = A - B$. As every trace class operator K is compact, and can be written as $K = B + iC$ via self adjoint operators B and C, it is immediate that B and C are compact. Hence there exist unique positive compact operators P_1, P_2, P_3, P_4 on $L^2(\mathbb{R}^n)$ such that $K = P_1 - P_2 + i(P_3 - P_4)$. Knowing the fact that every trace-class operator is Hilbert-Schmidt and maximal function M is subadditive, A_r and trace formula are linear, it is sufficient to prove the theorem for the positive-definite trace-class operator $P \geq 0$ that has an eigenfunction expansion $P = \sum_{i=1}^{\infty} \lambda_i \phi_i \otimes \overline{\phi_i}$ where ϕ_i denotes representatives for the eigenfunctions defined pointwise everywhere in $L^2(\mathbb{R}^n)$. As P is trace-class and the corresponding expansion $\sum_i \lambda_i |\phi_i(x)|^2$ of the kernel $P(x,x)$ on the diagonal of $\mathbb{R}^{2n}$ is convergent, we have the following result:

$$\int P(x,x)dx = \int \sum_i \lambda_i |\phi_i(x)|^2 dx \tag{3.2}$$

$$= \sum_i \lambda_i \int |\phi_i(x)|^2 dx \tag{3.3}$$

$$= \sum_i \lambda_i < \infty \tag{3.4}$$

which further shows that the series $\sum_i \lambda_i |\phi_i(x)|^2$ is finite a.e., since maximal function is subadditive and submultiplicative we have

$$\begin{aligned} M^{(2n)}P(x,x) &= M^{(2n)} \sum_i \lambda_i \phi_i(x) \overline{\phi_i(x)} \\ &\leq \sum_i M^{(2n)}(\phi_i \otimes \overline{\phi_i})(x,x) \\ &\leq \sum_i \lambda_i M^{(n)} \phi_i(x) M^{(n)} \overline{\phi_i(x)} \\ &= \sum_i \lambda_i (M^{(n)} \phi_i(x))^2 \end{aligned}$$

and

$$\begin{aligned} \int M^{(n)} \phi_i(x) M^{(n)} \overline{\phi_i(x)} dx &\leq \left(\int |M^{(n)} \phi_i(x)|^2 dx \right)^{\frac{1}{2}} \left(\int |M^{(n)} \overline{\phi_i(x)}|^2 dx \right)^{\frac{1}{2}} \\ &= \|M\phi_i\|_2 \left\| M\overline{\phi_i} \right\|_2 . \end{aligned}$$

So we get the significant result by the Maximal Theorem

$$\begin{aligned} \int M^{(2n)} P(x,x) dx &\leq \sum_i \lambda_i \|M\phi_i\|_2 \left\| M\overline{\phi_i} \right\|_2 \\ &\leq C_2^2 \sum_i \lambda_i \|\phi_i\|_2^2 \\ &< \infty \end{aligned}$$

which proves $M^{(2n)} K(x,x) \in L^1(\mathbb{R}^n)$, besides note that $\sum_i \lambda_i M^{(n)} \phi_i(x) M^{(n)} \overline{\phi_i(x)} < \infty$ a.e. $[dx]$. Next, we will prove that $\tilde{P}(x,x)$ exists a.e. $[dx]$ in order to conclude that

$$\text{tr} P = \int \tilde{P}(x,x) dx.$$

Because $\sum_i \lambda_i |\phi_i(x)|^2$ and $\sum_i \lambda_i (M^{(n)} \phi_i(x))^2$ are finite almost everywhere, let us pick regular point set $Y \subset \mathbb{R}^n$ so that $\sum_i \lambda_i |\phi_i(x)|^2$, $\sum_i \lambda_i (M^{(n)} \phi_i(x))^2$ are finite on it. It is known that for each $x \in \mathbb{R}^n$, $A_r \phi_i(x)$ is a continuous function of $r \in (0, \infty)$, $\phi_i \in L^2(\mathbb{R}^n)$ for each $i \in \mathbb{N}$ and moreover, $A_r \phi_i(x)$ can be extended on Y to a continuous function of

$r \in [0,\infty)$ by defining $A_0\phi_i(x) = \phi_i(x)$ for all i. We know that $|A_r\phi_i(x)|^2 \leq (M\phi_i(x))^2$ for each $x \in Y$, so $\sum_i \lambda_i |A_r\phi_i(x)|^2 \leq \sum_i \lambda_i (M\phi_i(x))^2 < \infty$. Hence for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $|\sum_{i=m+1}^{n} \lambda_i |A_r\phi_i(x)|^2| < \epsilon$ for all $m, n > N$ with $n > m$ and all $x \in Y$. This says that for each $x \in Y$ the series $\sum_i \lambda_i |A_r\phi_i(x)|^2$ is uniformly and absolutely convergent with respect to $r \in [0,\infty)$. Since $\langle 1, \phi_i\rangle = \overline{\langle \phi_i, 1\rangle}$ we obtain the identity:

$$\begin{aligned}
A_r\phi_i(x)A_r\overline{\phi_i(x)} &= \left(\frac{1}{m(Q_r(x))}\int_{Q_r(x)} \phi_i(t)dt\right)\left(\frac{1}{m(Q_r(x))}\int_{Q_r(x)} \overline{\phi_i(s)}ds\right) \\
&= \frac{1}{m(Q_r(x))^2}\int_{Q_r(x)} \phi_i(t)dt\overline{\int_{Q_r(x)} \phi_i(s)ds} \\
&= \frac{1}{m(Q_r(x))^2}|\int_{Q_r(x)} \phi_i(t)dt|^2 \\
&= \left|A_r\phi_i(x)\right|^2
\end{aligned}$$

which gives rise to

$$\begin{aligned}
A_r^{(2n)}P(x,x) &= A_r^{(2n)}\sum_i \lambda_i\phi_i(x)\overline{\phi_i(x)} \\
&= \sum_i \lambda_i A_r^{(n)}\phi_i(x)A_r^{(n)}\overline{\phi_i(x)} \\
&= \sum_i \lambda_i |A_r\phi_i(x)|^2
\end{aligned}$$

for $r > 0$. Therefore

$$\begin{aligned}
\tilde{P}(x,x) &= \lim_{r\to 0} A_r^{(2n)}P(x,x) \\
&= \lim_{r\to 0}\sum_i \lambda_i |A_r\phi_i(x)|^2 \\
&= \sum_i \lambda_i \lim_{r\to 0}|A_r\phi_i(x)|^2 \\
&= \sum_i \lambda_i |\phi_i(x)|^2
\end{aligned}$$

for all $x \in Y$ and it is $L^1(\mathbb{R}^n)$ convergent by (3.2)-(3.4). Consequently, we have arrived at a result that we wish to illustrate:

$$\begin{aligned} \mathrm{tr}P &= \sum_i \langle\, P\phi_i, \phi_i \rangle \\ &= \sum_i \int P\phi_i(x)\overline{\phi_i(x)}dx \\ &= \sum_i \int \lambda_i \phi_i(x)\overline{\phi_i(x)}dx \\ &= \int \sum_i \lambda_i |\phi_i(x)|^2 dx \\ &= \int \tilde{P}(x,x)dx \end{aligned}$$

that proves the theorem for any trace class operator with a given kernel. □

In the light of the statement of the preceding theorem, the following example shows the insufficiency of the integrability of $\tilde{K}(x,x)$ and $MK(x,x)$ to make the related operator traceable.

Example 1. Volterra integral operator $V : L^2[0,1] \to L^2[0,1]$ is defined by

$$Vf(t) = \int_0^t f(s)ds,$$

$t \in (0,1)$, or equivalently

$$Vf(x) = \int_0^1 V(x,y)f(y)dy$$

with the kernel

$$V(x,y) = \begin{cases} 1, & \text{if } y < x \\ 0, & \text{if } y \geq x \end{cases}$$

V is Hilbert-Schmidt operator on $L^2[0,1]$, yet it has no non-zero eigenvalues in its spectrum. By definitions of $\tilde{V}(x,x)$ and $MV(x,x)$ it is not so hard to see

$$\tilde{V}(x,x) = MV(x,x) = \frac{1}{2}. \tag{3.5}$$

Then let us show (3.5). Since $V(x,y) = 0$ if $y \geq x$, integration taken over the below half of the cubes of radius $r > 0$ centered at (x,x) on the diagonal of $[0,1]^2$ merely contributes to the total resulting value of the integration (in this case cubes are ordinary squares whose diagonals lie on the line $y = x$). And this contribution is nothing but the exactly half of the value of integration over the same cube that is centered at (x,x) with same radius $r > 0$. We can see this by assuming (3.6)

$$V(x,y) = 1 \text{ for } y \geq x. \tag{3.6}$$

Since then $V(x,y) = 1$ on $[0,1]^2$ we have

$$\begin{aligned} \frac{1}{m(Q_r)^2} \int_{Q_r} \int_{Q_r} V(x+t, x+s) dtds &= \frac{1}{m(Q_r)^2} \int_{Q_r} \int_{Q_r} 1 dtds \\ &= \frac{1}{m(Q_r)^2} m(Q_r)^2 \\ &= 1. \end{aligned}$$

Therefore by definition of $\tilde{V}(x,x)$ and $MV(x,x)$, for the kernel

$$V(x,y) = \begin{cases} 1, & \text{if } y < x \\ 0, & \text{if } y \geq x \end{cases}$$

we obtain (3.5).

Now let's prove that V is not traceable. Since the singular values of V are eigenvalues of $\sqrt{V^*V} = |V|$ we must find V^* first. Let $f, g \in L^2[0,1]$. By Fubini's

Theorem we have

$$\begin{aligned}\langle\, Vf, g\rangle &= \int_0^1 \int_0^1 V(x,y) f(y)\overline{g(x)}dydx \\ &= \int_0^1 f(y)\overline{\int_0^1 \overline{V(y,x)}g(x)dxdy} \\ &= \langle\, f, V^* g\rangle.\end{aligned}$$

Therefore, $V^*(x,y) = \overline{V(y,x)} \Rightarrow$

$$V^*(x,y) = \begin{cases} 1 & \text{if } y \geq x \\ 0 & \text{if } y < x \end{cases}$$

and

$$\begin{aligned}V^* f(x) &= \int_0^1 V^*(x,y) f(y)dy \\ &= \int_x^1 f(y)dy.\end{aligned}$$

Therefore,

$$\begin{aligned}(V^* V f)(x) &= \int_x^1 Vf(y)dy \\ &= \int_x^1 \int_0^y f(t)dtdy.\end{aligned}$$

Next set $V^*V = W$. Compactness of V results in the compactness of W which is self-adjoint as well, thus by the Spectral Theorem we know the existence of the set of eigenvalues of W with respect to the set of corresponding eigenfunctions of W. Subsequently, let λ and f be non-zero eigenvalue and eigenfunction of W respectively.

Then,

$$\begin{aligned} Wf(x) &= \int_x^1 \int_0^y f(t)dtdy \\ &= \lambda f(x). \end{aligned}$$

Let

$$\int_0^y f(t)dt = g(y), \tag{3.7}$$

so we have

$$\int_x^1 g(y)dy = \lambda f(x). \tag{3.8}$$

Since the left side of (3.8) is continuous so is $f(x)$ which makes $g(y)$ continuously differentiable by (3.7). Obviously this results in continuously differentiability of

$$\int_x^1 g(y)dy.$$

Let us denote the following for convenience by

$$G(x) = -\int_1^x g(y)dy. \tag{3.9}$$

And by using Fundamental Theorem of Calculus twice in (3.9) we obtain

$$G'(x) = -g(x) \;\Rightarrow\; G''(x) = -f(x) \text{ for all } x \in (0,1).$$

In other words we have:

$$\begin{aligned} \frac{d^2}{dx^2}(Wf(x)) &= \frac{d^2}{dx^2}(\lambda f(x)) \\ &= -f(x) \end{aligned}$$

or,

$$-f(x) = \lambda f''(x). \tag{3.10}$$

Solving the differential equation (3.10) we find $f(x) = \text{a}(e^{ikx} + e^{-ikx})$ for some $a \in \mathbb{C}$ and $k = \frac{1}{\sqrt{\lambda}}$. It follows $f(x) = 2a\cos(kx)$ and hence

$$\frac{1}{\lambda} Wf(x) = 2a\cos(kx). \tag{3.11}$$

Putting $x = 1$ into the left hand side of (3.11) it yields $2a\cos(k) = 0$ by the definition of $Wf(x)$. Moreover, by the assumption $f \neq 0 \Rightarrow a \neq 0$. Therefore, we obtain $k = \frac{(2n+1)\pi}{2} \Rightarrow \lambda_n = \frac{4}{(2n+1)^2\pi^2}$, $n \in \mathbb{N}$. Taking the square root of this gives the singular values of V which are $\frac{2}{(2n+1)\pi}$. Consequently we get

$$\text{tr}|V| = \sum_{n=1}^{\infty} \frac{2}{(2n+1)\pi}$$

as divergent series revealing that V is not traceable.

Next let's try to understand the alteration of the trace formula for trace class operator in Theorem 3.2 once writing it in terms of product of Hilbert-Schmidt operators. It may easily be verified that any trace class operator can be written as a product of two Hilbert-Schmidt operators, so how could we write this formula in this case ? Gaal [10] answers this question when the associated kernel K is operator valued that is to say once $K : G \times G \to \mathscr{L}(\mathscr{H})$ where G is an arbitrary locally compact group and $\mathscr{L}(\mathscr{H})$ is the linear function space with $\mathscr{H}$ being separable Hilbert space. Since our kernel $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{C}$ is a $\mathbb{C}-$valued function we will find slightly different formula than that of Gaal found in [10, Theorem 6.7.17]. Before proving forthcoming proposition, we shall define the convolution of kernels L and J in $L^2(\mathbb{R}^{2n})$, which is almost everywhere equal to $K(x, y)$, corresponding to the trace-class operator $T_K = T_L T_J$ where T_L and T_J are Hilbert-Schmidt operators with associated kernels L and J respectively.

Thus

$K(x,y) = L * J(x,y)$ a.e. $[dxdy]$ where the convolution is defined as follows:

$$L * J(x,y) = \int L(x,z)J(z,y)dz.$$

Proposition 3.3. *If $T_K : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is a trace-class operator, $T_K = T_L T_J$, then*

$$\text{tr}T_K = \int L * J(x,x)dx. \tag{3.12}$$

Proof. We will prove the proposition step by step depending on how the kernel $K(x,y)$ is written in terms of the convolution of kernels. Firstly let T_K be a Hilbert-Schmidt operator such that the associated kernel $K(x,y)$ is of the form $K = L^* * L$, $L \in L^2(\mathbb{R}^{2n})$, and define $L^*(x,y) = L(y,x)^*$. According to [10, Theorem 6.7.14], there is an isomorphism between the space of Hilbert-Schmidt operators and the space $L^2(\mathbb{R}^{2n})$ which sets up correspondence between Hilbert-Schmidt operator T_K and $K(x,y)$. Using this fact and the next two results

$$\begin{aligned}
\langle\ (T_L)^* f, g\rangle &= \langle\ f, T_L g\rangle \\
&= \int f(x)\overline{T_L g(x)}dx \\
&= \int f(x) \int \overline{L(x,y)g(y)}dydx \\
&= \iint \overline{L(x,y)}f(x)\overline{g(y)}dydx \\
&= \iint L^*(y,x)f(x)\overline{g(y)}dxdy \\
&= \langle\ T_{L^*} f, g\rangle
\end{aligned}$$

and

$$
\begin{aligned}
\langle\, T_{L^*}T_Lf, g\rangle &= \langle\, (T_L)^*T_Lf, g\rangle \\
&= \langle\, T_Lf, T_Lg\rangle \\
&= \iiint L(x,y)f(y)\overline{L(x,z)g(z)}dydzdx \\
&= \iiint L^*(z,x)L(x,y)f(y)\overline{g(z)}dxdydz \\
&= \iint L^* * L(z,y)f(y)\overline{g(z)}dydz \\
&= \langle\, T_{L^* * L}f, g\rangle
\end{aligned}
$$

for some f and $g \in L^2(\mathbb{R}^n)$, we may write

$$
\begin{aligned}
T_K &= T_{L^* * L} \\
&= T_{L^*}T_L \\
&= (T_L)^*T_L.
\end{aligned}
$$

So we have

$$
\begin{aligned}
\mathrm{tr}T_K &= \sum_i \langle\, T_K e_i, e_i\rangle \\
&= \sum_i \langle\, (T_L)^*T_L e_i, e_i\rangle \\
&= \sum_i \langle\, T_L e_i, T_L e_i\rangle \\
&= \sum_i \|T_L e_i\|^2 \\
&= \|T_L\|_{HS}^2
\end{aligned}
$$

where $\{e_i\}_i$ is an orthonormal basis of $L^2(\mathbb{R}^n)$. Then

$$\begin{aligned}\int L^* * L(x,x)dx &= \iint L^*(x,z)L(z,x)dzdx \\ &= \iint L(z,x)^* L(z,x)dzdx \\ &= \iint |L(z,x)|^2 dzdx \\ &= \|T_L\|_{HS}^2 \\ &= \mathrm{tr} T_K.\end{aligned}$$

Secondly we want to extend this formula to arbitrary trace-class operators $T_K = T_L T_J$ with corresponding kernel $L * J$ where T_L and T_J are Hilbert-Schmidt operators with kernels L and J respectively. To do so we shall write the identity $L * J$ as a linear combination of the kernels which are of the form M^*M for some kernel M which is in $L^2(\mathbb{R}^n \times \mathbb{R}^n)$. It is easy but tedious to check the correctness of the following

$$\begin{aligned}L * J = \frac{1}{4}\{&(L^* + J)^* * (L^* + J) + i(L^* - iJ)^* * (L^* - iJ) - (L^* - J)^* * (L^* - J) \\ &- i(L^* + iJ)^* * (L^* + iJ)\}.\end{aligned}$$

Letting

$$(L^* + J) = M_1,\ (L^* - iJ) = M_2,\ (L^* - J) = M_3,\ (L^* + iJ) = M_4$$

gives more apparent one

$$L * J = \frac{1}{4}\{M_1^* * M_1 + iM_2^* * M_2 - M_3^* * M_3 - iM_4^* * M_4\}.$$

Since trace is a linear functional and

$T_{\frac{1}{4}\{M_1^* * M_1 + iM_2^* * M_2 - M_3^* * M_3 - iM_4^* * M_4\}} = \frac{1}{4}\{T_{M_1^* * M_1} + iT_{M_2^* * M_2} - T_{M_3^* * M_3} - iT_{M_4^* * M_4}\}$ which can be seen by the definition of the operator associated with these kernels, we obtain

$$
\begin{aligned}
\mathrm{tr}T_K &= \mathrm{tr}T_{L*J} \\
&= \frac{1}{4}\left(\mathrm{tr}T_{M_1^* * M_1} + i\mathrm{tr}T_{M_2^* * M_2} - \mathrm{tr}T_{M_3^* * M_3} - i\mathrm{tr}T_{M_4^* * M_4}\right) \\
&= \frac{1}{4}(\int M_1^* * M_1(x,x)dx + i\int M_2^* * M_2(x,x)dx - \int M_3^* * M_3(x,x)dx \\
&- i\int M_4^* * M_4(x,x)dx) \\
&= \int \frac{1}{4}(M_1^* * M_1 + iM_2^* * M_2 - M_3^* * M_3 - iM_4^* * M_4)(x,x)dx \\
&= \int L * J(x,x)dx.
\end{aligned}
$$

Therefore, we are done. □

The next theorem will show that the trace formula in Proposition 3.3 agrees with the trace formula in Theorem 3.2 whatever the factorization of the trace class operator is. Before dealing with the proof let us define the n-dimensional averaging and maximal operators acting on the first or second variables of a kernel and we shall prove a lemma which has the result that will be used in the following theorem. Let $L(x, y)$ be the pointwise representative for the kernel of the Hilbert-Schmidt operator L. Then

$$
\begin{aligned}
A_{r(n)}L(x,y) &= \frac{1}{m(Q_r)}\int_{Q_r} L(x+s,y)ds \\
A^{r(n)}L(x,y) &= \frac{1}{m(Q_r)}\int_{Q_r} L(x,y+t)dt \\
M_{r(n)}L(x,y) &= \sup_{r>0}\frac{1}{m(Q_r)}\int_{Q_r} |L(x+s,y)|ds \\
M^{r(n)}L(x,y) &= \sup_{r>0}\frac{1}{m(Q_r)}\int_{Q_r} |L(x,y+t)|dt.
\end{aligned}
$$

Lemma 3.4. *If $L(x, y) \in L^2(\mathbb{R}^{2n})$, then $M_{r(n)}L(x, y)$ and $M^{r(n)}L(x, y) \in L^2(\mathbb{R}^{2n})$.*

Proof. For almost every x and y, $L(., y)$ and $L(x, .) \in L^2(\mathbb{R}^n)$. Maximal Theorem implies

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |M_{r(n)}L(x,y)|^2 dxdy = \left\|M_{r(n)}L\right\|_2^2$$
$$\leq C_2^2\|L\|_2^2$$
$$= C_2^2 \int_{\mathbb{R}^n}\left[\int_{\mathbb{R}^n} |L(x,y)|^2 dx\right] dy < \infty,$$

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |M^{r(n)}L(x,y)|^2 dxdy = \left\|M^{r(n)}L\right\|_2^2$$
$$\leq C_2^2\|L\|_2^2$$
$$= C_2^2 \int_{\mathbb{R}^n}\left[\int_{\mathbb{R}^n} |L(x,y)|^2 dy\right] dx < \infty.$$

□

Theorem 3.5. *If $K = LJ$ is a factorization of the trace class operator into a product of Hilbert-Schmidt operators L and J, then $\tilde{K}(x,x) = L * J(x,x)$ a.e. $[dx]$.*

Proof. Let $L(x,y)$ and $J(x,y)$ be the pointwise representatives for the kernels corresponding to L and J respectively. $L(x+s,.)$, $J(.,y+t) \in L^2(\mathbb{R}^n)$ for almost every x and y. Thus,

$$\left(\int_{\mathbb{R}^n} |L(x+s,z)J(z,y+t)|dz\right)^2 \leq \int_{\mathbb{R}^n} |L(x+s,z)|^2 dz \int_{\mathbb{R}^n} |J(z,y+t)|^2 dz < \infty.$$

Let

$$f_t(s,z) := L(x+s,z)J(z,y+t) \text{ and } g_s(z,t) := L(x+s,z)J(z,y+t).$$

Then

$$\left(\iint_{\mathbb{R}^{2n}} |f_t(s,z)|dzds\right)^2 \leq \iint_{\mathbb{R}^{2n}} |L(x+s,z)|^2 dzds \iint_{\mathbb{R}^{2n}} |J(z,y+t)|^2 dzds < \infty,$$

$$\left(\iint_{\mathbb{R}^{2n}} |g_s(z,t)|dzdt\right)^2 \leq \iint_{\mathbb{R}^{2n}} |L(x+s,z)|^2 dzdt \iint_{\mathbb{R}^{2n}} |J(z,y+t)|^2 dzdt < \infty$$

$\Rightarrow$ $f_t(s,z)$ and $g_s(z,t) \in L^1(\mathbb{R}^{2n})$. Then applying the Fubini Theorem twice in the following yields:

$$\begin{aligned}
A_r^{(2n)}K(x,y) &= \frac{1}{(m(Q_r))^2}\int_{Q_r}\int_{Q_r} K(x+s,y+t)dsdt \\
&= \frac{1}{(m(Q_r))^2}\int_{Q_r}\int_{Q_r} L*J(x+s,y+t)dsdt \\
&= \frac{1}{(m(Q_r))^2}\int_{Q_r}\int_{Q_r}\int_{\mathbb{R}^n} L(x+s,z)J(z,y+t)dzdsdt \\
&= \frac{1}{(m(Q_r))^2}\int_{Q_r}\int_{Q_r}\int_{\mathbb{R}^n} f_t(s,z)dzdsdt \\
&= \frac{1}{(m(Q_r))^2}\int_{Q_r}\int_{\mathbb{R}^n}\int_{Q_r} g_s(z,t)dsdzdt \\
&= \frac{1}{(m(Q_r))^2}\int_{\mathbb{R}^n}\int_{Q_r}\int_{Q_r} L(x+s,z)J(z,y+t)dsdtdz \\
&= \int_{\mathbb{R}^n}\left(\frac{1}{m(Q_r)}\int_{Q_r} L(x+s,z)ds\right)\left(\frac{1}{m(Q_r)}\int_{Q_r} J(z,y+t)dt\right)dz \\
&= \int_{\mathbb{R}^n} A_{r(n)}L(x,z)A^{r(n)}J(z,y)dz
\end{aligned}$$

for almost every x and y. We have the following inequality:

$$|A_{r(n)}L(x,z)A^{r(n)}J(z,y)| \leq M_{r(n)}L(x,z)M^{r(n)}J(z,y) \tag{3.13}$$

for $r > 0$. Therefore by the inequality (3.13) and the Lemma 3.4 we obtain

$$\begin{aligned}
\left(\int |A_{r(n)}L(x,z)A^{r(n)}J(z,y)|dz\right)^2 &\leq \left(\int |M_{r(n)}L(x,z)M^{r(n)}J(z,y)|dz\right)^2 \\
&\leq \int |M_{r(n)}L(x,z)|^2dz \int |M^{r(n)}J(z,y)|^2dz \\
&< \infty.
\end{aligned}$$

Thus,

$$A_{r(n)}L(x,z)A^{r(n)}J(z,y) =: f_r \in L^1$$

and $M_{r(n)}L(x,z)M^{r(n)}J(z,y) =: g$ is nonnegative L^1 function such that $|f_r| \leq g$ a.e. for all $r > 0$. Also for almost every z,

$$\begin{aligned}
&\lim_{r\to 0} A_{r(n)}L(x,z) = L(x,z) \quad \text{a.e. [dx]} \\
&\lim_{r\to 0} A^{r(n)}J(z,y) = J(z,y) \quad \text{a.e. [dy]} \\
&\Rightarrow \lim_{r\to 0} f_r = L(x,z)J(z,y) \quad \text{a.e. [dxdy]}.
\end{aligned}$$

Hence by the Dominated Convergence Theorem we have obtained

$$\begin{aligned}
\tilde{K}(x,x) &= \lim_{r\to 0} A_r^{(2n)}K(x,x) \\
&= \lim_{r\to 0} \int_{\mathbb{R}^n} A_{r(n)}L(x,z)A^{r(n)}J(z,x)dz \\
&= \int_{\mathbb{R}^n} \lim_{r\to 0} A_{r(n)}L(x,z)A^{r(n)}J(z,x)dz \\
&= \int_{\mathbb{R}^n} L(x,z)J(z,x)dz \\
&= L * J(x,x) \quad \text{a.e. [dx]}.
\end{aligned}$$

□

Up to now almost all the results have been established for the measure space $(X, \mathcal{B}_X, m)$ where $X \subseteq \mathbb{R}^n$, $\mathcal{B}_X$ Borel σ-algebra on X and m is a Lebesgue measure. The next chapter demonstrates that these results also hold once the measure space $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, m)$ is replaced with more general one: countably generated measurable space $(X, \mathcal{M})$ with σ-finite measure μ on $\mathcal{M}$. Furthermore we will introduce martingales in order to define averaging and maximal functions of martingales generated by a single $\mathcal{M}$-measurable function. Subsequently analogous theorems to Theorem 2.3 and Corollory 2.4 will be put forward to be used in the next chapter.

4. TRACE CLASS OPERATORS ON MORE GENERAL HILBERT SPACES

To present the martingales it is essential to figure out the conditional expectations first.

4.1. Conditional Expectations

Let $(X, \mathcal{M}, \mu)$ be a σ- finite measure space and $\{\mathcal{M}_n\}_{n\in\mathbb{N}}$ be a sequence of sub-σ-algebras of $\mathcal{M}$ such that

- $\mathcal{M}_j \subseteq \mathcal{M}_{j+1}$ for all j.
- $\mathcal{M} = \bigvee_{j\in\mathbb{N}} \mathcal{M}_j$.

To define the conditional expectations of $f \in L^p(X)$, $1 \leq p \leq \infty$, we need the following lemma:

Lemma 4.1. *Let $f \in L^p(X, \mathcal{M}, \mu)$ and $n \in \mathbb{N}$. There is a unique locally integrable $\mathcal{M}_n$-measurable function g_n vanishing off a σ-finite set such that*

$$\int_A g_n d\mu = \int_A f d\mu$$

for all $A \in \mathcal{M}_n$ with $\mu(A) < \infty$ and $1 \leq p \leq \infty$.

Proof. See [11, Lemma 5.2.1]. □

One should note that the proof of Lemma 4.1 makes use of the Radon-Nikodym Theorem so as to define the function g_n and prove the uniqueness of it.

Definition 4.2. The function g_n in Lemma 4.1 is called the conditional expectation of f relative to the sub-σ-algebra $\mathcal{M}_n$ and it is denoted by $E(f|\mathcal{M}_n)$ or E_nf.

Therefore $E(f|\mathcal{M}_n)$ is $\mathcal{M}_n$-measurable and

$$\int_A E(f|\mathcal{M}_n)d\mu = \int_A f d\mu$$

for all $A \in \mathcal{M}_n$ with $\mu(A) < \infty$. Some basic properties of conditional expectations are collected in the appendix since we have sufficient information to continue so far.

4.2. Martingales

Definition 4.3. A sequence $\{f_j\}_{j\in\mathbb{N}}$ of functions on a measurable space $(X, \mathcal{M})$ is a martingale if

(i) Each f_n is $\mathcal{M}_n$-measurable.
(ii) $E(f_n|\mathcal{M}_j) = f_j$ whenever $j < n$.

$\{\mathcal{M}_j\}_{j\in\mathbb{N}}$ is the same as stated in section 4.1. Note that second requirement is a consequence of some elementary properties of conditional expectations. If $f \in L^p(X, \mathcal{M}, \mu)$ is an $\mathcal{M}$-measurable function, $1 \leq p \leq \infty$, then the maximal function of a martingale which is generated by f is defined as follows :

$$Mf(x) = \sup_{n\in\mathbb{N}} E(|f||\mathcal{M}_n)(x) \tag{4.1}$$

The analogue of the Theorem 2.3 comes at once:

Theorem 4.4 (Doob's Martingale Maximal Theorem). *(i) If M is the maximal operator as defined in* (4.1)*, then* $\mu(\{x \in X : Mf(x) > \alpha\}) \leq \frac{1}{\alpha}\|f\|_1$ *for all* $f \in L^1(X)$ *and* $\alpha > 0$.

(ii) If $1 < p \leq \infty$*, then* $\|Mf\|_p \leq C_p\|f\|_p$ *where* C_p *depends only on* p *for all* $f \in L^p(X)$.

The proof of Theorem 4.4 is virtually the same as that of Theorem 2.3. A good reference for the proof is [12, Theorem IV.6]. The next theorem which is analogous to

Corollory 2.4 will be proved in the appendix in the case where the measure space is a probability space $(\Omega, \mathcal{F}, P)$ because J. L. Doob originally establishes such results in [13] for $(\Omega, \mathcal{F}, P)$. Yet here we state the theorem for $f \in L^p(X)$ on the σ-finite measure space $(X, \mathcal{M}, \mu)$.

Theorem 4.5 (Doob's Martingale Convergence Theorem)**.** *Let $\{\mathcal{M}_n\}_{n\in\mathbb{N}}$ be the sequence of sub-σ-algebras of $\mathcal{M}$. Then for each $f \in L^p(X)$, $1 \leq p \leq \infty$, the sequence $E(f|\mathcal{M}_n) \to f$ a.e. $[\mu]$ as $n \to \infty$.*

Now we shall clarify the meaning of conditional expectations. Since $(X, \mathcal{M}, \mu)$ is a σ-finite measure space we may partition X into sequence $\{P_i^n\}_{i\in\mathbb{N}}$ of pairwise disjoint sets each of which has finite measure. Let $\{P_i^n\}_{i\in\mathbb{N}} = \mathcal{P}_n$ and $\mathcal{P}_j \prec \mathcal{P}_k$ mean that $\mathcal{P}_j$ is a refinement of $\mathcal{P}_k$. Take a sequence $\{\mathcal{P}_j\}_{j\in\mathbb{N}}$ of partitions of X into sets of finite measure such that $\mathcal{P}_{n+1} \prec \mathcal{P}_n$. Assume that $\mathcal{M}_n = \sigma(\mathcal{P}_n)$ such that $\mathcal{M} = \bigvee_n \mathcal{M}_n$. So $\mathcal{M}_n \subset \mathcal{M}_{n+1}$. For each $x \in X$ and each $n \in \mathbb{N}$, $\mathcal{P}_n$ has some set P_k^n of finite measure such that $x \in P_k^n$ for some $k \in \mathbb{N}$. This set is of course unique for each $n \in \mathbb{N}$. Hence for simplicity let $P_k^n = Q_n(x)$. Whenever $\mu(Q_n(x)) \neq 0$ we set

$$E(f|\mathcal{M}_n)(x) = \frac{1}{\mu(Q_n(x))} \int_{Q_n(x)} f d\mu. \tag{4.2}$$

Let $\mathcal{N} = \{x \in X : \mu(Q_n(x)) = 0 \text{ for some n} \in \mathbb{N}\}$. It is clear that for each $x \in \mathcal{N}^c$, we define the averaging operator as in chapter 1:

$$A_n f(x) = E(f|\mathcal{M}_n)(x) \tag{4.3}$$

$$= \frac{1}{\mu(Q_n(x))} \int_{Q_n(x)} f d\mu. \tag{4.4}$$

To define the averaging operator on $X \times X$ we must surely exclude the points on $\mathcal{N} \times \mathcal{N}$. Since the definition of $\mathcal{N}$ points that $\mu(\mathcal{N}) = 0$, $\mathcal{N}^c \times \mathcal{N}^c$ is conull with respect to the product measure. Therefore we would like to define

$$A_n^{(2)} f(x,y) = \frac{1}{\mu(Q_n(x))\mu(Q_n(y))} \int_{Q_n(x)} \int_{Q_n(y)} f(s,t) d\mu(t) d\mu(s) \tag{4.5}$$

$$= E(f|\mathcal{M}_n \otimes \mathcal{M}_n)(x,y) \tag{4.6}$$

for all $(x,y) \in \mathcal{N}^c \times \mathcal{N}^c$ where $\mathcal{M}_n \otimes \mathcal{M}_n$ is the product σ-algebra of sub-σ-algebra $\mathcal{M}_n$ of $\mathcal{M}$ with itself for some $n \in \mathbb{N}$. Note that $\mathcal{M}_n \otimes \mathcal{M}_n = \sigma(\mathscr{A}_n)$ where

$$\mathscr{A}_n = \{\bigcup_{i=1}^{m} P_i \times R_i : P_i, R_i \in \mathcal{M}_n; (P_i \times R_i) \cap (P_j \times R_j) = \emptyset \text{ for } i \neq j \text{ and } m \in \mathbb{N}\}$$

is an algebra. Consequently for all $(x,y) \in \mathcal{N}^c \times \mathcal{N}^c$

- $A_n^{(2n)}(\phi \otimes \psi)(x,y) = A_n\phi(x) A_n\psi(y)$
- $\sup_{n\in\mathbb{N}} E(|\phi \otimes \psi||\mathcal{M}_n \otimes \mathcal{M}_n) = M^{(2)}(\phi \otimes \psi)(x,y) \leq M\phi(x) M\psi(y)$
- $M(\phi + \psi)(x) \leq M\phi(x) + M\psi(x)$

In chapter 2 we defined the regular point for the expansion of the kernel $P(x,y)$ of the positive Hilbert-Schmidt operator T_P. Later we will define it in a slightly different approach for the kernel of arbitrary trace class operator on $L^2(X,\mu)$ where $(X,\mathcal{M},\mu)$ is countably generated σ-finite measure space. For the time being for any function $f \in L^p(X)$, call $x \in \mathcal{N}^c$ a regular point for f if

$$\lim_{n\to\infty} A_n f(x) = \tilde{f}(x) \text{ exists} \tag{4.7}$$

and by Theorem 4.5, $\tilde{f}(x) = f(x)$ a.e. $[\mu]$. Moreover the set

$$Y_f = \{x \in \mathcal{N}^c : \lim_{n\to\infty} A_n f(x) = \tilde{f}(x) \text{ exists}\}$$

is conull in X.

Theorem 4.6. *Let X be a second countable space with σ-finite Borel measure μ. If $f \in L^p(X,\mu)$, $1 \leq p \leq \infty$, then each point of continuity $x \in \mathcal{N}^c$ of f is a regular point* *for f with $\tilde{f}(x) = f(x)$.*

Proof. Assume $(X, \mathcal{M})$ is a measurable space and $\mathcal{M}$ is generated by the countable base $\mathcal{U} = \{U_n\}_{n\in\mathbb{N}}$ of the topology $\mathscr{T}_X$ of X. Since μ is σ-finite assume X is partitioned inductively by using $\mathcal{U}$:

$$\mathcal{P}_n = \{U_n \cap \mathcal{P}_{n-1},\ U_n^c \cap \mathcal{P}_{n-1}\}, n \in \mathbb{N}$$

and $\mathcal{P}_0$ is a countable partition of X into sets of finite measure. As one can see that $\mathcal{P}_{j+1} \prec \mathcal{P}_j$ for all j we deduce that $\sigma(\mathcal{P}_j) = \mathcal{M}_j \subset \mathcal{M}_i$ for $i > j$. Thus by the inductive definition of $\mathcal{P}_j$'s we observe that $\bigvee_{n\in\mathbb{N}} \mathcal{M}_n = \mathcal{M}$ also since the Borel σ-algebra $\mathcal{B}_X$ is generated by the family of open sets in X we write $\mathcal{M} = \mathcal{B}_X$. The reason why we have shown this is based on the fact that the definition (4.2) is defined by making use of the unique sets $Q_n(x)$ which are elements of $\mathcal{P}_n \subset \mathcal{B}_X$ for $x \in \mathcal{N}^c$. Now let $x \in \mathcal{N}^c$ be the point on which f is continuous. So for every $\epsilon > 0$ there exists $n > 0$ and $y \in X$ such that $|y - x| < \frac{1}{n} \Rightarrow\ |f(y) - f(x)| < \epsilon$. Therefore

$$|A_n f(x) - f(x)| \leq \frac{1}{\mu(Q_n(x))} \int_{Q_n(x)} |f(y) - f(x)| d\mu(y) < \epsilon$$

which shows us that for every such x, $A_n f(x) \to f(x)$ as $n \to \infty$. As a result, $\tilde{f}(x) = f(x)$ by the definition of regular point for $f \in L^p(X)$. □

4.3. Trace of a Trace Class Operator

Let $(X, \mathcal{M}, \mu)$ be a countably generated, σ-finite measure space and K be a trace class operator on $L^2(X, \mu)$. As K^*K is positive and compact, by the Spectral Theorem for compact self adjoint operators it can be diagonalized by an orthonormal sequence $\{\psi_i\}_{i\in\mathbb{N}}$ of eigenvectors with corresponding eigenvalues $\{\mu_i\}_{i\in\mathbb{N}}$, $\mu_i > 0$ for all $i \in \mathbb{N}$.

Let us define $\lambda_i^{-1} K\psi_i = \phi_i$, $\lambda_i = \sqrt{\mu_i}$. The sequence $\{\phi_i\}_{i\in\mathbb{N}}$ is orthonormal as well:

$$\begin{aligned}\langle \phi_i, \phi_j\rangle &= \langle \lambda_i^{-1} K\psi_i, \lambda_j^{-1} K\psi_j\rangle \\ &= \lambda_i^{-1}\lambda_j^{-1}\langle K^* K\psi_i, \psi_j\rangle \\ &= \lambda_i^{-1}\lambda_j^{-1}\mu_i\langle \psi_i, \psi_j\rangle \\ &= 0\end{aligned}$$

for all $i, j \in \mathbb{N}$. We know

$$\begin{aligned}\text{clspan}\{\psi_i\}_{i\in\mathbb{N}} &= \ker(K^* K)^{\perp} \\ &= \ker(K)^{\perp}.\end{aligned}$$

Thus any $v \in \ker(K)^{\perp}$ can be expanded in terms of $\{\psi_i\}_{i\in\mathbb{N}}$ and with this in the hand we claim that K has the canonical expansion $K = \sum_{i=1}^{\infty} \lambda_i \phi_i \otimes \overline{\psi_i}$. Consider the rank 1 operator $\phi \otimes \overline{\psi} : v \mapsto \langle v, \psi\rangle\phi$. For $v \in \ker(K)^{\perp}$,

$$\begin{aligned}\left(\sum_{i=1}^{\infty} \lambda_i \phi_i \otimes \overline{\psi}_i\right)(v) &= \sum_{i=1}^{\infty} \lambda_i \phi_i \langle v, \psi_i\rangle \\ &= \sum_{i=1}^{\infty} K\psi_i \langle v, \psi_i\rangle \\ &= K\left(\sum_{i=1}^{\infty} \langle v, \psi_i\rangle \psi_i\right) \\ &= Kv.\end{aligned}$$

Thus this demonstrates $K = \sum_{i=1}^{\infty} \lambda_i \phi_i \otimes \overline{\psi}_i$ which corresponds to an orthonormal expansion

$$\sum_{i=1}^{\infty} \lambda_i \phi_i(x) \otimes \overline{\psi_i(y)} \in L^2(X \times X, \mu \times \mu). \tag{4.8}$$

Definition 4.7. If the limits $\lim_{n\to\infty} A_n \phi_i(x) = \tilde{\phi}_i(x)$ and $\lim_{n\to\infty} A_n \psi_i(x) = \tilde{\psi}_i(x)$ exist at x for all $i \in \mathbb{N}$, then $x \in \mathcal{N}^c$ is called regular point for the expansion (4.8).

Let the set of regular points for (4.8) be denoted by $Y_{\phi_i \cap \psi_i} = \bigcap_{i\in\mathbb{N}}(Y_{\phi_i} \cap Y_{\psi_i})$. Theorem 4.5 implies that almost every point in X belongs to $Y_{\phi_i \cap \psi_i}$ so we obtain $\tilde{\phi}_i(x) = \phi_i(x)$ a.e. $[\mu]$; $\tilde{\psi}_i(x) = \psi_i(x)$ a.e. $[\mu]$. Now for all $(x, y) \in \mathcal{N}^c \times \mathcal{N}^c$ define

$$\tilde{K}(x, y) = \lim_{n\to\infty} A_n^{(2)} K(x, y) \text{ pointwise a.e. } [\mu \times \mu] \tag{4.9}$$

where

$$A_n^{(2)} K(x, y) = \sum_i \lambda_i A_n \phi_i(x) A_n \overline{\psi}_i(y) \tag{4.10}$$

by earlier results.

Theorem 4.8. *Let K be a trace class operator on $L^2(X, \mu)$; then*

(i) The canonical expansion for the kernel $K(x, y)$ converges absolutely a.e. $[\mu \times \mu]$ and it converges absolutely a.e. $[\mu]$ on the diagonal of $X \times X$.
(ii) $M^{(2)} K(x, x) \in L^1(X)$.
(iii) $\tilde{K}(x, x)$ exist a.e. $[\mu]$ and hence $\mathrm{tr}K = \int \tilde{K}(x, x) d\mu(x)$.

Proof. (i) Since K is trace class operator we have $\sum_i \lambda_i < \infty$, where λ_i is a singular value of K for each $i \in \mathbb{N}$. The associated kernel expansion for K is of the form as that of (4.8): $K(x, y) = \sum_i \lambda_i \phi_i(x) \overline{\psi_i(y)}$ with orthonormal sequences $\{\phi_i\}_{i\in\mathbb{N}}$ and $\{\psi_i\}_{i\in\mathbb{N}}$. Thus

$$\begin{aligned} \int \sum_i \lambda_i |\phi_i(x)|^2 d\mu(x) &= \int \sum_i \lambda_i |\psi_i(y)|^2 d\mu(y) \\ &= \sum_i \lambda_i \\ &< \infty \end{aligned}$$

which shows that $\sum_i \lambda_i |\phi_i(x)|^2 < \infty$ and $\sum_i \lambda_i |\psi_i(y)|^2 < \infty$ almost everywhere

$[\mu]$. Moreover,

$$\sum_i |\lambda_i \phi_i(x)\overline{\psi_i(y)}| \leq \left(\sum_i \lambda_i |\phi_i(x)|^2\right)^{\frac{1}{2}} \left(\sum_i \lambda_i |\psi_i(y)|^2\right)^{\frac{1}{2}}. \tag{4.11}$$

By the above results and (4.11) the series $\sum_i \lambda_i \phi_i(x)\overline{\psi_i(y)}$ converges absolutely a.e. $[\mu \times \mu]$ to $K(x,y)$ and $\sum_i \lambda_i \phi_i(x)\overline{\psi_i(x)}$ converges absolutely a.e. $[\mu]$ to $K(x,x)$.

(ii) By submultiplicativity, subadditivity and boundedness of martingale maximal operator the following is obtained:

$$\begin{aligned}
\int M^{(2)}K(x,x)d\mu(x) &\leq \int \sum_i \lambda_i M^{(2)}(\phi_i \otimes \overline{\psi_i})(x,x)d\mu(x) \\
&\leq \int \sum_i \lambda_i M\phi_i(x)M\overline{\psi_i}(x)d\mu(x) \\
&= \sum_i \lambda_i \int M\phi_i(x)M\overline{\psi_i}(x)d\mu(x) \\
&\leq \sum_i \lambda_i \left(\int (M\phi_i(x))^2 d\mu(x)\right)^{\frac{1}{2}} \left(\int (M\overline{\psi_i}(x))^2 d\mu(x)\right)^{\frac{1}{2}} \\
&= \sum_i \lambda_i \|M\phi_i\|_2 \left\|M\overline{\psi_i}\right\|_2 \\
&\leq C_2^2 \sum_i \lambda_i \|\phi_i\|_2 \left\|\overline{\psi_i}\right\|_2 \\
&= C_2^2 \sum_i \lambda_i \\
&< \infty.
\end{aligned}$$

Therefore $M^{(2)}K(x,x) \in L^1(X)$ and remark that $\sum_i \lambda_i M_i \phi_i(x) M\overline{\psi_i}(x) < \infty$ a.e. $[\mu]$ which is going to play a significant role in (iii).

(iii) Let

$$Y = \{x \in Y_{\phi_i \cap \psi_i} : \sum_i \lambda_i |\phi_i(x)|^2,\ \sum_i \lambda_i |\psi_i(x)|^2,\ \sum_i \lambda_i M\phi_i(x) M\overline{\psi_i}(x) < \infty\}$$

be a conull set in $Y_{\phi_i \cap \psi_i}$. By earlier definitions of maximal and averaging operators for martingales we know

$$|A_n\phi_i(x)A_n\overline{\psi_i}(x)| \leq M\phi_i(x)M\overline{\psi_i}(x).$$

Thus for each $x \in Y$, $\sum_i A_n\phi_i(x)A_n\overline{\psi_i}(x)$ converges absolutely and by (4.10) $A_n^{(2)}K(x,y) = \sum_i \lambda_i A_n\phi_i(x)A_n\overline{\psi}_i(y)$. Furthermore

$$\lim_{n\to\infty} A_n\phi_i(x)A_n\overline{\psi_i}(x) = \phi_i(x)\overline{\psi_i(x)}$$

for all $x \in Y$. Then by the dominated convergence theorem for sums we get

$$\begin{aligned}\tilde{K}(x,x) &= \lim_{n\to\infty} A_n^{(2)}K(x,x)\\ &= \lim_{n\to\infty}\sum_i \lambda_i A_n\phi_i(x)A_n\overline{\psi_i}(x)\\ &= \sum_i \lambda_i \lim_{n\to\infty} A_n\phi_i(x)A_n\overline{\psi_i}(x)\\ &= \sum_i \lambda_i\phi_i(x)\overline{\psi_i(x)}\end{aligned}$$

for each $x \in Y$. And this series converges absolutely in $L^1(X)$ by (4.11). Using this result in [7, Theorem 2.25] and applying it in the following calculation yields:

$$\begin{aligned}\mathrm{tr}K &= \sum_i \langle K\psi_i, \psi_i\rangle\\ &= \sum_i \int K\psi_i(x)\overline{\psi_i(x)}d\mu(x)\\ &= \sum_i \int \lambda_i\phi_i(x)\overline{\psi_i(x)}d\mu(x)\\ &= \int \sum_i \lambda_i\phi_i(x)\overline{\psi_i(x)}d\mu(x)\\ &= \int \tilde{K}(x,x)d\mu(x).\end{aligned}$$

□

Corollary 4.9. *Let X be a second countable space with σ-finite Borel measure μ and K be a trace class operator on $L^2(X, \mu)$. If the kernel $K(x, y)$ is continuous at (x, x) for almost every x then*

$$\mathrm{tr}K = \int K(x, x) d\mu(x). \tag{4.12}$$

Proof. Combine Theorem 4.8 and Theorem 4.6 to obtain (4.12). □

REFERENCES

1. Brislawn, C., "Kernels of Trace Class Operators", *Proc. Amer. Math. Soc.*, pp. 1181-1190, 1988.

2. Brislawn, C., "Traceable Integral Kernels on Countably Generated Measure Space", *Pacific Journal of Mathematics*, pp. 229-240, 1991.

3. Duflo, M., "Généralités sur les représentations induites, Représentations des Groupes de Lie Résolubles", *Monographies de la Soc. Math de France*, Vol. 4, pp. 93-119, 1972.

4. Weidmann, J., "Integraloperatoren der spurklasse", *Math. Ann.*, pp. 340-345, 1966.

5. Riesz, F. and B. Sz.-Nagy, *Functional Analysis*, Frederick Ungar Publishing Co., New York, 1955.

6. Tricomi, F. G., *Integral Equations*, Dover Publications, New York, 1985

7. Folland, G. B., *Real Analysis: Modern Techniques and Their Applications*, Wiley, New York, 1984.

8. Stein, E. M., *Singular Integrals and Differentiability Properties of Functions*, Princeton Univ. Press, Princeton, N.J., 1970.

9. Conway, J. B., *A Course in Functional Analysis*, New York, 1997.

10. Gaal, S., *Linear Analysis and Representation Theory*, Grundlehren Math. Wiss., Springer-Verlag, Berlin, Heidelberg, New York, 1973.

11. Edwards, R. and G. Gaudry, "Littlewood-Paley and Multiplier Theory", *Ergebnisse Math. und Grenz.*, Vol. 90, 1977.

12. Stein, E. M., “Topics in Harmonic Analysis”, *Annals of Math. Studies*, Vol. 63, Princeton Univ. Press, Princeton, NJ, 1970.

13. Doob, J. L., *Stochastic Process*, John Wiley & Sons, New York, 1953.

APPENDIX A: CONDITIONAL EXPECTATIONS AND DOOB'S MARTINGALE CONVERGENCE THEOREM

A.1. Some Basic Properties of Conditional Expectation

(i) $E(f|\mathcal{M}) \geq 0$ if $f \geq 0$.

(ii) $E(f+g|\mathcal{M}) = E(f|\mathcal{M}) + E(g|\mathcal{M})$ and $E(\alpha f|\mathcal{M}) = \alpha E(f|\mathcal{M})$ for some α.

(iii) $|E(f|\mathcal{M})| \leq E(|f||\mathcal{M})$ (Jensen's Inequality).

(iv) $E(E(f|\mathcal{M}_m)|\mathcal{M}_n) = E(f|\mathcal{M}_n)$ if $f \in L^p(X)$ and $m \geq n$.

(v) $\left\|E(f|\mathcal{M}_n)\right\|_p \leq \|f\|_p$ if $f \in L^p(X)$ and $1 \leq p \leq \infty$.

(vi) If f is $\mathcal{M}_n$-measurable then $E(f|\mathcal{M}_n) = f$, $1 \leq p \leq \infty$. In particular, $E(1|\mathcal{M}_n) = 1$ whenever $\mu(X) < \infty$.

(vii) If $f \in L^p(X)$, $1 \leq p \leq \infty$, which is $\mathcal{M}_n$-measurable, then $E(fg|\mathcal{M}_n) = gE(f|\mathcal{M}_n)$.

(viii) $\overline{E(f|\mathcal{M}_n)} = E(\overline{f}|\mathcal{M}_n)$,

$$\begin{aligned}\int_X E(f|\mathcal{M}_n)\overline{g}d\mu &= \int_X E(f|\mathcal{M}_n)\overline{E(g|\mathcal{M}_n)}d\mu \\ &= \int_X f\overline{E(g|\mathcal{M}_n)}d\mu.\end{aligned}$$

Proof. Most of the proofs are readily seen by the very definition of the conditional expectation. Here we will prove (iv), (vii), (viii).

(iv) Since $m \geq n$ we have $\mathcal{M}_n \subseteq \mathcal{M}_m$. For $A \in \mathcal{M}_n$ with $\mu(A) < \infty$

$$\int_A E(E(f|\mathcal{M}_m)|\mathcal{M}_n)d\mu = \int_A E(f|\mathcal{M}_m)d\mu \tag{A.1}$$

$$= \int_A f d\mu \tag{A.2}$$

$$= \int_A E(f|\mathcal{M}_n)d\mu \tag{A.3}$$

because $A \in \mathcal{M}_m$ as well. $(A.1)$ and $(A.3)$ imply that $E(E(f|\mathcal{M}_m)|\mathcal{M}_n)$ is $\mathcal{M}_n$-

measurable. Thus $E(E(f|\mathcal{M}_m)|\mathcal{M}_n) = E(f|\mathcal{M}_n)$.

(vii) If $g \in L^+$, there is a sequence $\{\phi_j\}_j$ of integrable $\mathcal{M}_n$-measurable functions such that $0 \leq |\phi_1| \leq |\phi_2| \leq ... \leq |g|$, $\phi_n \to g$ pointwise. For not necessarily non-negative function g we may write $g = g_1 - g_2$ where

$$g_1(x) = \begin{cases} g(x), & \text{if } g(x) > 0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$g_2(x) = \begin{cases} -g(x), & \text{if } g(x) < 0 \\ 0, & \text{otherwise} \end{cases}$$

Since g_1 and g_2 are L^+ functions they can both be approximated by simple functions. In this case for any $\mathcal{M}$-measurable g it is sufficient to prove (vii) for simple $\mathcal{M}_n$-measurable integrable function $\phi = \sum_{i=1}^{n} c_i 1_{A_i}$. Let $A \in \mathcal{M}_n$ and $\mu(A) < \infty$,

$$\begin{aligned}
\int_A E(f\phi|\mathcal{M}_n) d\mu &= \int_A f\phi d\mu \\
&= \int_A f \sum_{i=1}^{n} c_i 1_{A_i} d\mu \\
&= \sum_{i=1}^{n} c_i \int_{A\cap A_i} f d\mu \\
&= \sum_{i=1}^{n} c_i \int_{A\cap A_i} E(f|\mathcal{M}_n) d\mu \\
&= \sum_{i=1}^{n} c_i \int_A E(f|\mathcal{M}_n) 1_{A_i} d\mu \\
&= \int_A E(f|\mathcal{M}_n) \sum_{i=1}^{n} c_i 1_{A_i} d\mu \\
&= \int_A E(f|\mathcal{M}) \phi d\mu.
\end{aligned}$$

(viii) Let $A \in \mathcal{M}_n$ with finite measure. Then

$$\begin{aligned}\int_A \overline{E(f|\mathcal{M}_n)} d\mu &= \overline{\int_A E(f|\mathcal{M}_n) d\mu} \\ &= \overline{\int_A f d\mu} \\ &= \int_A \overline{f} d\mu \\ &= \int_A E(\overline{f}|\mathcal{M}_n) d\mu\end{aligned}$$

and

$$\begin{aligned}\int_X E(f|\mathcal{M}_n)\overline{g} d\mu &= \int_X E(E(f|\mathcal{M}_n)\overline{g}|\mathcal{M}_n) d\mu \\ &= \int_X E(f|\mathcal{M}_n)E(\overline{g}|\mathcal{M}_n) d\mu \\ &= \int_X E(f|\mathcal{M}_n)\overline{E(g|\mathcal{M}_n)} d\mu\end{aligned}$$

where we have used (vii) in the second equality.

□

A.2. Doob's Martingale Convergence Theorem

We will give some theorems without proofs for in use before handling the Doob's Martingale Convergence Theorem. Below let $(\Omega, \mathcal{F}, P)$ be a probability sample space and $\{X_n\}$ be a sequence of random variables.

Theorem A.1 (Doob's Inequality). *Let $\{X_n\}$ be a sub-martingale. Then*

$$\begin{aligned}P(\sup_{1\leq k\leq n} X_k \geq \lambda) &\leq E[X_n 1_{\{w:\sup_{1\leq k\leq n} X_k \geq \lambda\}}] \\ &\leq E[X_n].\end{aligned}$$

Theorem A.2 (Jensen's Inequality). *Let P be a probability measure. If $\psi : \mathbb{R} \to \mathbb{R}$ is* *convex and each* $\psi(f)$, f *are integrable then* $\int \psi(f)dP \geq \psi\left(\int f dP\right)$.

Example 2. The function $|.|$ is obviously convex and satisfies Jensen's inequality. Moreover there are versions of Jensen's inequality for expectations and conditional expectations each of which is same as Theorem A.2 yet has notational difference. Let $\psi = |.|$ and $X_n = E(X|\mathcal{F}_n)$, X random variable, then $E[|X_n|] \leq E[|X|]$.

Theorem A.3. *Let* $X_n = E(X|\mathcal{F}_n)$, $n \in \mathbb{N}$. *If* $X \in L^p$, $1 \leq p \leq \infty$, *then* $\|X_n - X\|_p \to 0$ *as* $n \to \infty$.

A standard reference for the proof of Theorem A.3 is [11, Theorem 5.2.6].

Theorem A.4 (Doob's Martingale Convergence Theorem)**.** *For each random variable* $X \in L^p(\Omega)$, $1 \leq p \leq \infty$, *the sequence* $X_n = E(X|\mathcal{F}_n)$ *converges to* X *almost surely* P.

Proof. Let $(\Omega, \mathcal{F})$ be a probability space with probability measure P. Since $P(\Omega) = 1 < \infty$, $L^1(\Omega) \supseteq L^p(\Omega)$ by Proposition 6.12 in [7]. Thus it is enough to prove the theorem for $p = 1$. Let

$$S = \{X \in L^1(\Omega) : X \text{ satisfies the statement of the theorem}\} \subset L^1(\Omega)$$

and

$$S_n = \{X \in L^1(\Omega) : X \text{ is } \mathcal{F}_n\text{-measurable}\}.$$

Clearly $S_n \subset S$ for every $n \in \mathbb{N}$, because $E(X|\mathcal{F}_n) = X$ for each $\mathcal{F}_n$-measurable $X \in L^1(\Omega)$. What is more is that $\bigcup_n S_n$ is dense in $L^1(\Omega)$ since $\bigvee_n \mathcal{F}_n = \mathcal{F}$ and $\mathcal{F}_i \subset \mathcal{F}_{i+1}$ for all i. Therefore each $\mathcal{F}$-measurable $\tilde{X} \in L^1(\Omega)$ is approximated by sequence $\{X_n\}_{n\in\mathbb{N}} \subset \bigcup_{n\in\mathbb{N}} S_n$ of random variables where $X_n \in S_n$ for each $n \in \mathbb{N}$. Consequently S is dense in $L^1(\Omega)$. Let $X \in L^1$ and pick a sequence $\{Y_j\}_j \subset S$ such that $\lim_{j\to\infty} Y_j = X$ in $L^1(\Omega)$. Define $Y_n^j = E(Y_j|\mathcal{F}_n)$ and $X_n = E(X|\mathcal{F}_n)$. Also $\|X_n - X\|_1 \to 0$ and $\{X_n\}_{n\in\mathbb{N}}$ is uniformly bounded in $L^1(\Omega)$ by Theorem A.3 and property (v) in section A.1 respectively. Since $X_n \to X$ in L^1 it is a fact that there is a subsequence $\{X_{n_j}\}$ of $\{X_n\}$ such that $X_{n_j} \to X$ almost surely P by [7, Corollory 2.32]. Let

$$F = \{\omega \in \Omega : \limsup_n X_n(\omega) - \liminf_n X_n(\omega) \geq \epsilon\}.$$

X_n converges some random variable a.s. P iff $P(F) = 0$ and by the above fact this random variable is exactly X which ends up the proof. Therefore let us show that $P(F) = 0$.

We shall write $X = (X - Y_j) + Y_j$ and $X_n = (X_n - Y_n^j) + Y_n^j$. Since $Y_j \in S$ we have $Y_n^j \to Y_j$ a.s. P. As a result,

$$\limsup_n Y_n^j = \liminf_n Y_n^j \quad \text{a.s. } P \tag{A.4}$$

$$\begin{aligned}
\limsup_n X_n - \liminf_n X_n &\leq \limsup_n (X_n - Y_n^j) - \liminf_n (X_n - Y_n^j) + \\
&\quad \limsup_n Y_n^j - \liminf_n Y_n^j \\
&\leq 2 \sup_n |X_n - Y_n^j|
\end{aligned}$$

by the result (A.4). After taking advantage of Theorems A.1 and A.2 in the following inequality we are done.

$$\begin{aligned}
P(F) &\leq P(\sup_n |X_n - Y_n^j| \geq \frac{\epsilon}{2}) \\
&\leq \frac{2}{\epsilon} E[|X - Y_j|]
\end{aligned}$$

as F is independent of j, letting $j \to \infty$ we get $P(F) = 0$ since $Y_j \to X$ in $L^1(\Omega)$. $\square$