Oscillation and Variation Inequalities for Convolution Powers.

By

Roger L. Jones[1]
Department of Mathematics
DePaul University
2219 N. Kenmore
Chicago IL 60614

and

Karin Reinhold
Department of Mathematics
University at Albany, SUNY
1400 Washington Ave.
Albany, NY 12222

Abstract We prove L^2 variation inequalities for operators defined by the convolution powers of probability measures on locally compact Abelian groups. In some cases we also obtain L^p results for $1 < p < \infty$. These inequalities imply the pointwise convergence of these operators and give an estimate on the number of upcrossings.

---

[1] Partially supported by NSF Grant DMS—9531526

1. INTRODUCTION

Variation norms estimate the fluctuations of a given sequence of operators. In the ergodic theory context they were introduced by Bourgain [4] for L^2, and extended to L^p, $1 \leq p < \infty$, by Jones, Kaufman, Rosenblatt and Wierdl in "Oscillations in ergodic theory" [10], where fluctuations of ergodic averages and differentiation operators were studied.

In this paper we study the variation norms for operators given by the convolution powers of probability measures on locally compact Abelian groups.

The variation norms are related to the square functions as in Littlewood–Paley theory, but give stronger results. Square functions for the convolution operators were studied by Jones, Ostrovskii and Rosenblatt in "Square functions in ergodic theory" [9].

To handle the variation inequalities for convolution powers operators we will also use some of the techniques we developed in the paper "Oscillation and variation for the Hilbert transform" by Campbell, Jones, Reinhold and Wierdl in [7]. However, the main technique is the association of the convolution powers with a reverse martingale in a larger space.

The paper is structured as follows. Section 2 states the basic definitions. Section 3 studies the case of symmetric measures. Here, the results are obtained by viewing the operators as a reverse martingale in a much larger space. Through this representation, we can obtain more complete results than for the general case.

Sections 4 studies the general case of convolution measures on locally compact Abelian groups. The main result is obtained by comparing the given measure to its symmetrized version and by developing a tool handle the short step variations. Section 5 extends the result for the short step variation of Section 4 to the continuous parameter case, and gives applications to well known continuous parameter kernels.

The results obtained in Section 4 can also be extended to measures on σ–compact metric groups via direct integral decomposition of the group representation. This result will be presented in a subsequent paper.

2. OSCILLATION AND VARIATION NORMS

Definitions 2.1. Given an increasing sequence of positive integers $\{n_k\}_{k=1}^{\infty}$, define the *oscillation norm* of a sequence of numbers $\{x_n\}_{n=1}^{\infty}$

by

$$\|x_n\|_{o(s)} = \Big(\sum_{k=1}^{\infty} \max_{n_k \le n \le n_{k+1}} |x_n - x_{n_k}|^s\Big)^{1/s}.$$

Remark 2.2. We will be interested in the case $s = 2$. If $s < 2$ the operators associated with this norm are often not well behaved (see [1]), and for $s > 2$ we have $\|x_n\|_{o(s)} \le \|x_n\|_{o(2)}$, so understanding $s = 2$ is usually the general case. Operators based on these norms were studied by Gaposhkin, who showed that if $A_n f(x) = \frac{1}{n}\sum_{j=0}^{n-1} T^j f$, for T a normal contraction, then with $n_k = 2^k$, $\|\|A_n f\|_{o(2)}\|_2 \le c\|f\|_2$. With this result, to prove a.e. convergence, it is sufficient to prove a.e. convergence along only the sequence $\{A_{n_k}\}$.

A second very useful norm on a sequence of numbers is the variation norm, denoted by $\|\cdot\|_{v(\varrho)}$.

Definitions 2.3. Given a set of real numbers $\{x_n\}_{n\in I}$, where I is a countable index set, define its variation ϱ norm by:

$$\|x_n\|_{v(\varrho)} = \sup_{n(j)} \Big(\sum_{j=1}^{\infty} |x_{n(j)} - x_{n(j+1)}|^{\varrho}\Big)^{1/\varrho},$$

where the supremum is taken over all possible sequences $n(1) \le n(2) \le \cdots$, and each $n(j) \in I$.

Often we will take $I = \mathbb{Z}^+$, but sometimes we will be interested in other countable sets. Unless otherwise noted, we will assume $I = \mathbb{Z}^+$. When it is important to specify the index set, we will write $\|x_n : n \in I\|_{v_\varrho}$. In particular, we will write $\|x_n : n_k \le n < n_{k+1}\|_{v_\varrho}$ for $I = \{n : n_k \le n < n_{k+1}\}$.

This norm has been used in martingale theory, and more recently in ergodic theory by Bourgain in [4]; and Jones, Kaufman, Rosenblatt and Wierdl in [10].

We can also define the continuous version of the $v(\varrho)$ norm as follows:

Definitions 2.4. Given a continuously indexed system, $\{x_t\}_{t\in\mathbb{R}}$, define its variation ϱ norm by:

$$\|x_t\|_{v(\varrho)} =$$

$$\sup\Big\{\Big(\sum_{j=1}^{J} |x_{t_j} - x_{t_{j-1}}|^{\varrho}\Big)^{1/\varrho}\Big|\ J = 1, 2, \dots\ ;\ t_0 \le t_1 \le t_2 \le \cdots \le t_J\Big\}.$$

The continuous version of the $o(\varrho)$ norm is defined in the analogous way.

Proposition 2.5 (Properties of the $v(\varrho)$–norms).

1. *For each* ϱ, $1 \leq \varrho < \infty$, $\|.\|_{v(\varrho)}$ *is a semi–norm.*
2. *If* $\{x_n\}$ *is a sequence such that* $x_{n_k} = 0$ *for each* k, *then* $\|x_n\|_{v(\varrho)} \leq 2\Big(\sum_k \|x_n : n_k \leq n < n_{k+1}\|_{v(\varrho)}^{\varrho}\Big)^{1/\varrho}$.
3. $\|x_n\|_{v(\varrho)} \leq 2\Big(\sum_n |x_n|^{\varrho}\Big)^{1/\varrho}$.

We refer the reader to [10] for a proof of these properties, as well as some discussion of the applications of the variation and oscillation norms to ergodic theory.

Although neither $\|.\|_{v(\varrho)}$ and $\|.\|_{o(\varrho)}$ are true norms, by abuse of notation we will continue to refer to them as norms throughout the paper.

Definitions 2.6. Let $\{x_n\}$ denote a sequence of real numbers. We define

$$N(x_n, \lambda) = \sup\{N : \text{there exists } s_1 < t_1 \leq s_2 < t_2 \cdots \leq s_N < t_N \text{ such that } |x_{t_i} - x_{s_i}| \geq \lambda \text{ for } i = 1, 2 \ldots, N\}.$$

The operator $N(x_n, \lambda)$ gives the number of times the sequence $\{x_n\}$ changes by a distance λ. It is related to the number of upcrossings of the sequence, but can be much larger. For a continuously indexed system, $\{x_t\}_{t\in\mathbb{R}}$, $N(x_t, \lambda)$ is defined in the same way.

3. Symmetric measures

We first study the case of symmetric measures. There are three reasons why we begin with this case. First, the results are easier to obtain, second, the results we obtain are useful to obtain later results in the more general case, and third, the results in this case are more complete.

Square functions, oscillation results and variation results for symmetric measures can be studied directly by lifting the operators to a reverse martingale in a larger space. This technique was used by Kakutani [12], and further developed by Oseledets [15] who studied pointwise convergence of operators induced by convolution powers of a symmetric measure. We begin by stating the main proposition, Proposition 3.1, and some of its applications.

Proposition 3.1. *Let* $\{P_n\}$ *be a sequence of doubly stochastic operators, that is*

1. *for* $1 \leq p \leq \infty$, *we have* $\|P_n f\|_p \leq \|f\|_p$,
2. *the adjoint operator* P_n^* *is also bounded in* L^1 *and* L^∞,

3. *$P_n f \geq 0$ whenever $f \geq 0$, and*
4. *$P_n(1) = 1$.*

Then

1. *For any p, $1 < p < \infty$ and any ϱ, $2 < \varrho < \infty$, we have $\| \| P_1 P_2 \dots P_n P_n^* \dots P_2^* P_1^* f \|_{v(\varrho)} \|_p \leq c(\varrho, p) \|f\|_p$.*
2. *For any p, $1 < p < \infty$, and any increasing sequence $\{n_k\}$, we have $\| \| \| P_1 P_2 \dots P_n P_n^* \dots P_2^* P_1^* f \|_{o(2)} \|_p \leq c(p) \|f\|_p$, where we use the sequence $\{n_k\}$ to compute the $o(2)$ norm.*

We will postpone the proof of this proposition until later in the section. As a consequence of Proposition 3.1 we can state several results. First, we need the following definition.

Definitions 3.2. (See Stein [21]) Let (X, Σ, m) denote a positive measure space. Let $\{T^t\}_{0 \leq t < \infty}$ denote a family of operators, each T^t mapping functions on X to functions on X. Assume:

- The operators T^t satisfy the semigroup axioms $T^t T^s = T^{t+s}$, and $T^0 = I$, the identity operator;
- The T^t are continuous in the sense that $\lim_{t \to 0} T^t f = f$ in L^2.

Further, assume

1. $\|T^t f\|_p \leq \|f\|_p$ for $1 \leq p \leq \infty$.
2. Each T^t is a self-adjoint operator on $L^2(X)$.
3. Each $T^t f \geq 0$ if $f \geq 0$.
4. For each t, $T^t(1) = 1$.

Then $\{T^t\}_{0 \leq t < \infty}$ is called a *symmetric diffusion semigroup.*

We can now state the following theorem.

Theorem 3.3. *Let T^t denote a symmetric diffusion semigroup. Then*

1. *For any p, $1 < p < \infty$ and any ϱ, $2 < \varrho < \infty$, we have $\| \| T^t f \|_{v(\varrho)} \|_p \leq c(\varrho, p) \|f\|_p$.*
2. *For any p, $1 < p < \infty$, and any increasing sequence $\{n_k\}$, we have $\| \| T^t f \|_{o(2)} \|_p \leq c(p) \|f\|_p$.*

Proof. We want to apply Proposition 3.1. First fix L and let $P = T^{\frac{1}{2^{L+1}}}$. Clearly, $P_n = P$ satisfies the hypothesis of Proposition 3.1. Since $P^* = P$, applying the Proposition 3.1, we get $\| \| P^{2n} f \|_{v(\varrho)} \|_p \leq c(\varrho, p) \|f\|_p$. Hence $\| \| T^{\frac{n}{2^L}} f \|_{v(\varrho)} \|_p \leq c(\varrho, p) \|f\|_p$. Let $D_L = \{\frac{n}{2^L} : n \geq 0\}$. We have

$$\Big\| \Big(\sup_{\substack{t_1 \leq t_2 \leq \dots \\ t_i \in D_L}} \sum_{j=1}^{\infty} |T^{t_j} f - T^{t_{j+1}} f|^{\varrho} \Big)^{\frac{1}{\varrho}} \Big\|_p \leq c(\varrho, p) \|f\|_p.$$

Taking the limit as $L \to \infty$, by the monotone convergence theorem, we get

$$\Big\| \Big(\sup_{\substack{t_1 \leq t_2 \leq \dots \\ t_j \in D}} \sum_{j=1}^{\infty} |T^{t_j} f - T^{t_{j+1}} f|^{\varrho} \Big)^{\frac{1}{\varrho}} \Big\|_p \leq c(\varrho, p) \|f\|_p,$$

where D is the set of positive dyadic rationals. Now the continuity of the semi-group implies the first conclusion of the Theorem. The $o(2)$ norm is handled in the same way. □

Corollary 3.4. *Let* $P_t(x) = \frac{1}{\pi}\frac{t}{x^2+t^2}$ *denote the Poisson kernel on* $\mathbb{R}$*, or more generally, let* P_t *denote the Poisson kernel on* $\mathbb{R}^d$*.*

1. *For* $1 < p < \infty$ *and* $2 < \varrho < \infty$ *we have* $\| \| P_t f \|_{v(\varrho)} \|_p \leq c(\varrho, p) \|f\|_p$.
2. *For any increasing sequence* $\{n_k\}$*, and* $1 < p < \infty$*, we have* $\| \| P_t \|_{o(2)} \|_p \leq c(p) \|f\|_p$.

Proof. This follows from Theorem 3.3. We just need to check that P_t satisfies the hypothesis, which is an easy exercise. □

Corollary 3.5. *Let* $Q_t(x) = \frac{1}{\pi}\frac{x}{x^2+t^2}$ *denote the Conjugate Poisson kernel on* $\mathbb{R}$*.*

1. *For* $1 < p < \infty$ *and* $2 < \varrho < \infty$ *we have* $\| \| Q_t f \|_{v(\varrho)} \|_p \leq c(\varrho, p) \|f\|_p$.
2. *For any increasing sequence* $\{n_k\}$*, and* $1 < p < \infty$*, we have* $\| \| Q_t \|_{o(2)} \|_p \leq c(p) \|f\|_p$.

Proof. Let $f \in L^p$, and let Hf denote its Hilbert transform. Then $Hf \in L^p$ and $Q_\epsilon * f(x) = P_\epsilon * Hf(x)$. The result then follows from Corollary 3.4. □

Remark 3.6. Using a different technique, we can also show that the variation and oscillation norms for the Poisson Kernel satisfies a weak type (1,1) inequality as well. This result can be found in [7].

Let G denote a locally compact Abelian group, and Γ its dual group. Let (X, β, m) be a probability space, $\{T_g\}_{\{g \in G\}}$ an action of G by invertible measure preserving transformations on (X, β, m). Given a probability measure μ on G, we define $\mu f(x) = \int_G f(T_{g^{-1}} x) d\mu(g)$, and for $n > 1$, $\mu^n f(x) = \mu(\mu^{n-1} f)(x)$.

Definitions 3.7. We say that a measure μ is *strictly aperiodic* if its support is not contained in a proper left coset of G.

For Abelian groups, we have that μ is a strictly aperiodic probability measure if and only if $|\hat{\mu}(\gamma)| < 1$ for all $\gamma \in \Gamma$, $\gamma \neq 0$. See Rosenblatt [17] and Glasner [8]. For discrete Abelian groups there is a stronger characterization.

Theorem 3.8. *Let G be a discrete Abelian group and let μ be a strictly aperiodic probability measure on G. Then*

$$I(\mu) = \sup_{\gamma \in \Gamma \setminus \{0\}} \frac{|1 - \hat{\mu}(\gamma)|^2}{1 - |\hat{\mu}(\gamma)|^2}$$

is bounded.

This result was proved in Jones, Ostrovskii and Rosenblatt [9] for probability measures on $\mathbb{Z}$, but their proof extends to discrete Abelian groups. This result can be extended to other cases as well provided extra conditions on μ. However it may fail in general. Thus the finiteness of $I(\mu)$ is not always equivalent to the measure being strictly aperiodic.

Lemma 3.9. *Let μ be a strictly aperiodic measure on G satisfying $I(\mu) < \infty$. Then*

$$\Big\| \Big(\sum_{n=1}^{\infty} |\mu^{n+1} f - \mu^n f|^2 \Big)^{1/2} \Big\|_2 \leq C \|f\|_2.$$

This lemma was proved in Jones, Ostrovskii and Rosenblatt [9] for $G = \mathbb{Z}$. The proof given there just involves spectral properties of the measure μ, and consequently also applies to any locally compact Abelian group.

Theorem 3.10. *Let μ be a strictly aperiodic symmetric measure on G such that $I(\mu) < \infty$.*

1. *For any $2 < \varrho < \infty$ we have $\| \| \mu^n f \|_{v(\varrho)} \|_2 \leq C \|f\|_2$.*
2. *For any increasing sequence $\{n_k\}$ we have $\| \| \mu^n f \|_{o(2)} \|_2 \leq C \|f\|_2$.*

Proof. With the above lemma, we can easily complete the proof of part 1. Let $B_n f = \mu^n f$ if n is even and $B_n = \mu^{n+1} f$ if n is odd. Then

$$\begin{aligned}
\|\mu^n f\|_{v(s)} &\le \|\mu^n f - B_n f\|_{v(s)} + \|B_n\|_{v(s)} \\
&\le 2\Big(\sum_{k=1}^{\infty} \| |\mu^n f - B_n f|, n = 2k-1, 2k\|_{v(s)}^{s}\Big)^{1/s} + \|\mu^{2n} f\|_{v(s)} \\
&\le 2\Big(\sum_{k=1}^{\infty} |\mu^{2k} f - \mu^{2k-1} f|^2\Big)^{1/2} + \|\mu^{2n} f\|_{v(s)}.
\end{aligned}$$

Now, part 1 of Theorem 3.10 follows by applying Lemma 3.9 to the first term, and then taking $P_n = \mu$, and applying Proposition 3.1 to the second term.

We now want to prove part 2 of Theorem 3.10.

Given a sequence $\{n_k\}$ of positive number, consider the sequences

$$e_k = \begin{cases} n_k & \text{if } k \text{ is even,} \\ n_k + 1 & \text{if } k \text{ is odd} \end{cases} \quad \text{and} \quad o_k = \begin{cases} n_k & \text{if } k \text{ is odd,} \\ n_k + 1 & \text{if } k \text{ is even} \end{cases}.$$

Then, for any sequence of numbers $\{x_n\}$, we have

$$\begin{aligned}
\max_{n_k \le n \le n_{k+1}} |x_n - x_{n_k}|^2 &\le \max_{e_k \le n \le e_{k+1}, n \text{ even}} |x_n - x_{n_k}|^2 \\
&+ \max_{o_k \le n \le o_{k+1}, n \text{ odd}} |x_n - x_{n_k}|^2 \\
&\le c\Big(\max_{e_k \le n \le e_{k+1}, n \text{ even}} |x_n - x_{e_k}|^2 \\
&+ \max_{o_k \le n \le o_{k+1}, n \text{ odd}} |x_n - x_{o_k}|^2 \\
&+ |x_{n_k} - x_{n_k+1}|^2\Big).
\end{aligned}$$

Thus, by Proposition 3.1, taking $P_n = \mu$, and applying Lemma 3.9, we have

$$\begin{aligned}
\|\|\mu^n f\|_{o(2)}\|_2^2 &\le c\Big(\|\|\mu^{2n} f\|_{o(2)}\|_2^2 + \|\|\mu^{2n}(\mu f)\|_{o(2)}\|_2^2 \\
&+ \|(\sum_{k=1}^{\infty} |\mu^{n_k} f - \mu^{n_k+1} f|^2)^{\frac{1}{2}}\|_2^2\Big) \\
&\le C\|f\|_2^2.
\end{aligned}$$

□

Remark 3.11. If we could prove an L^p version of Lemma 3.9, then we would be able to prove an L^p version $(1 < p < \infty)$ of Theorem 3.10.

Such a result is probably true, but we have not yet been able to prove it. The case $p = 1$, that is, showing that $\|\mu^n f\|_{v(\varrho)}$ is a weak type (1,1) operator is of considerable interest, but it would require different techniques, since the weak type (1,1) analog of Proposition 3.1 is not true. To see this, note that if $\|Q^{2n} f\|_{v(s)}$ were weak type (1,1), then we would have convergence for $Q^{2n} f$ for all $f \in L^1(X)$, and Ornstein, [14], has shown this is false. Hence, special properties of μ^n, beyond those in the hypothesis of Proposition 3.1 would have to be used. In a large class of cases, in particular, if μ is a probability measure on $\mathbb{Z}$, with mean zero and finite second moment [2], it known that $\mu^n f$ converges for all $f \in L^1(X)$, so it is possible that the variation operators is weak (1,1) for such μ.

We now state the main tool that is used in the proof of Proposition 3.1. As mentioned before, the idea for powers of one operator goes back to Kakutani, but here we present a simplified version by Rota [19]. For powers of one operator see also Stein [21]. Using this result and a subsequent lemma, we can then prove Proposition 3.1.

Theorem 3.12 (Rota). *Let $\{P_n\}$ be a sequence of doubly stochastic operators in $L^p(X, \mathcal{F}, m)$ ($p \geq 1$), that is*

1. *P_n is an L^1-L^∞ contraction,*
2. *The adjoint operator P_n^* is also defined on L^1 and L^∞,*
3. *$P_n f \geq 0$ if $f \geq 0$,*
4. *$P_n 1 = 1$.*

Then there exists a huge probability space (Ω, β, v), a collection of sub–σ fields of β, $\mathcal{F}_{n+1} \subset \mathcal{F}_n \subset \ldots \subset \mathcal{F}_0$, and another sub–$\sigma$ field $\hat{\mathcal{F}}_0$ of β such that
(a) the probability space $(X, \mathcal{F}, m)$ and $(\Omega, \hat{\mathcal{F}}_0, v)$ are isomorphic under the normal projection map $\pi : \Omega \to X$. Let π also denote the induced isomorphism from $L^p(\Omega, \hat{\mathcal{F}}_0, v)$ to $L^p(X, \mathcal{F}, m)$.
(b) If $F \in L^p(\Omega, \hat{\mathcal{F}}_0, v)$ then

$$P_1 P_2 \ldots P_n P_n^* \ldots P_2^* P_1^*(\pi(F)) = E(E(F|\mathcal{F}_n)|\hat{\mathcal{F}}_0).$$

The proof of this result can be found in Rota's paper [19]. The proof when $P_n = P$ and $P^* = P$ is also in Stein's book on Littlewood Paley Theory [21], where he uses the result to prove a certain g-function is a bounded operator on L^p.

For the remainder of this section, G denotes a locally compact group. We will use the following notation:

1. If $H \in L^p(\Omega)$, $\hat{E}(H) = E(H|\hat{\mathcal{F}}_0)$, $\qquad E_n(H) = E(H|\mathcal{F}_n)$.

2. Given $f \in L^p(X)$, F denotes the lifting of f to $L^p(\Omega)$, $F = \pi^{-1}f$, and $X_n = E_n(F)$. Note that $\{X_n\}$ is a reverse martingale.

The following version of Minkowski's inequality for conditional expectations is easy to verify.

Lemma 3.13. *Let $(X, \mathcal{F}, m)$ and $(Y, \mathcal{F}', v)$ be two σ–finite measure spaces, and consider their product space $X \times Y$ endowed with the product σ–algebra and the product measure. Let $F(x,y) \geq 0$ be an integrable function on $X \times Y$, and $\mathcal{G}$ a sub–σ algebra of $\mathcal{F}'$. Then*

1.

$$\int_X E\big(F(x,y)|\mathcal{G}\big)dm(x) = E\Big(\int_X F(x,y)dm(x)\Big|\mathcal{G}\Big)$$

2.

$$\Big(\int_X \big|E(F(x,y)|\mathcal{G})\big|^p dm(x)\Big)^{1/p} \leq E\Big(\big(\int_X F(x,y)^p dm(x)\big)^{1/p}\Big|\mathcal{G}\Big)$$

Proof of Proposition 3.1. We begin with a proof of part 1 of Proposition 3.1.

We have

$$\|P_1 P_2 \dots P_n P_n^* \dots P_2^* P_1^* f\|_{v(s)} = \|\hat{E}(X_n)\|_{v(s)} \leq \hat{E}(\|X_n\|_{v(s)})$$

Hence

$$\|\ \|P_1 P_2 \dots P_n P_n^* \dots P_2^* P_1^* f\|_{v(s)}\ \|_{L^p(X)} \leq \|\ \|X_n\|_{v(s)}\ \|_{L^p(\Omega)}$$

and the lemma follows by Lepingle's Theorem. (For a proof of Lepingle's theorem see [13, 16].)

For the second part of Proposition 3.1, we argue as follows. For $f \in L^2(X)$,

$$
\begin{aligned}
\| \, \|P_1 P_2 \dots P_n P_n^* \dots P_2^* P_1^* f\|_{o(2)} \, \|_2^2 \\
&= \int_\Omega \sum_{k=1}^{\infty} \max_{n_k \le n < n_{k+1}} |\hat{E}(X_n)(\bar{x}) - \hat{E}(X_{n_k})(\bar{x})|^2 \; d\bar{x} \\
&\le \int_\Omega \hat{E}\big(\sum_{k=1}^{\infty} \max_{n_k \le n < n_{k+1}} |X_n - X_{n_k}|^2\big)(\bar{x}) \; d\bar{x} \\
&= \int_\Omega \sum_{k=1}^{\infty} \max_{n_k \le n < n_{k+1}} |X_n - X_{n_k}|^2(\bar{x}) \; d\bar{x} \\
&= \sum_{k=1}^{\infty} \int_\Omega \max_{n_k \le n < n_{k+1}} |X_n - X_{n_k}|^2(\bar{x}) \; d\bar{x} \\
&\le \sum_{k=1}^{\infty} \|X_{n_{k+1}} - X_{n_k}\|_2^2 \\
&\le C \; \|F\|_2^2 = C \; \|f\|_2^2,
\end{aligned}
$$

where we have used the fact that $\{X_n\}$ is a reverse martingale, and the fact that for martingales, the oscillation norm gives a bounded operator on all L^p. (For a proof of the L^p boundedness of the oscillation norm for martingales, see [10].) □

Before we state our next result on convolution powers, we need to recall a result from martingale theory. Recall that $N(x_n, \lambda)$ is the number of λ-jumps by the sequence x_n, as defined earlier.

Theorem 3.14. *Let (f_n) denote a uniformly integrable martingale sequence. Then $\|\sqrt{N(f_n, \lambda)}\|_p \le \frac{c}{\lambda}\|f\|_p$ for $1 < p < \infty$.*

Proof. The proof follows from a stopping time argument that relates $\sqrt{N(f_n, \lambda)}$ to a square function of a new martingale, and then observing that this new square function has norm bounded by the norm of the original martingale. See Lemma 6.7 of [10], Theorem 6.8 of [10], and the related remarks for details. □

We can now state the following theorem.

Theorem 3.15. *Let T^t denote a symmetric diffusion semigroup. Then we have the following.*

1. *For* $1 < p < \infty$, $\varrho > 2$, *and* $\lambda > 0$, *we have*

$$\int_X (N(T^t f(x), \lambda))^{p/\varrho} dx \le \frac{c(p, \varrho)}{\lambda^p} \|f\|_p^p.$$

2. *For* $1 < p < \infty$, $\varrho > 2$ *and* $\lambda > 0$, *we have*

$$m\{x : N(T^t f(x), \lambda > a\} \le \frac{c(p, \varrho}{\lambda^p a^{p/\varrho}} \|f\|_p^p.$$

3. *For* $\lambda > 0$, *we have* $\int N(T^t f(x), \lambda) dx \le \frac{c}{\lambda^2} \|f\|_2^2$.

Proof. To prove the first part, we just use the observation that $\lambda N(T^t f(x), \lambda)^{1/\varrho} \le \|T^t f(x)\|_{v(\varrho)}$, and the result from Theorem 3.3 which says $\|\|T^t f\|_{v(\varrho)}\|_p \le c(p, \varrho)\|f\|_p$. The second part is just a consequence of the first part and an application of Chebyshev's inequality.

To prove the third part, we need the martingale result. As in the proof of Theorem 3.3, fix L and let $P = T^{1/2^{L+1}}$. Define $Q = P^2$. Let $f \in L^p$. By Rota's Theorem, Theorem 3.12, we can find a function F, so that $f = \pi F$, and a sequence of expectation operators so that we have $Q^n f(x) = P^{2n} f(x) = E(E(F|\mathcal{F}_n)|\mathcal{F}_0)$. Let $X_n = E(F|\mathcal{F}_n)$. Notice that $|Q^n f(x) - Q^m f(x)| = |E(X_n - X_m|\mathcal{F}_0)| \le E(|X_n - X_m|\,|\mathcal{F}_0)$. Hence

$$N(Q^n f(x), \lambda) = N(E(X_n|\mathcal{F}_0), \lambda) \le E(N(X_n, \lambda)|\mathcal{F}_0).$$

Consequently,

$$\begin{aligned}
\int_X N(Q^n f(x), \lambda)\, dx &\le \int_\Omega E(N(X_n, \lambda)|\mathcal{F}_0) d\omega \\
&\le E\Big(\int_\Omega (\sqrt{N(X_n, \lambda)})^2\, d\omega \Big| \mathcal{F}_0\Big) \\
&\le E(\frac{c}{\lambda^2}\|F\|_2^2|\mathcal{F}_0) \\
&= \frac{c}{\lambda^2}\|f\|_2.
\end{aligned}$$

□

We also have the following theorem.

Theorem 3.16. *Let* μ *be a symmetric measure on* G *satisfying* $I(\mu) < \infty$. *Then for* $\lambda > 0$, *we have* $\int N(\mu^n f(x), \lambda) dx \le \frac{c}{\lambda^2}\|f\|_2^2$.

Proof. First, as in the proof of Theorem 3.15, we see that for $\lambda > 0$, we have $\int N(\mu^{2n} f(x), \lambda) dx \le \frac{c}{\lambda^2}\|f\|_2^2$, and $\int N(\mu^{2n+1} f(x), \lambda) dx \le \frac{c}{\lambda^2}\|f\|_2^2$.

Now just observe that

$$\begin{aligned} N(\mu^n f, 3\lambda) \leq & N(\mu^{2n} f, \lambda) + N(\mu^{2n+1} f, \lambda) \\ & + \frac{1}{\lambda^2} \sum_k |\mu^k f(x) - \mu^{k+1} f(x)|^2. \end{aligned}$$

Integrating both sides, and using the above observations, it is enough to show that $\int_X \frac{1}{\lambda^2} \sum_k |\mu^k f(x) - \mu^{k+1} f(x)|^2 dx \leq \frac{c}{\lambda^2} \|f\|_2^2$. This is just the conclusion of Lemma 3.9, that is, the square function is a bounded operator on L^2. □

4. General measures on locally compact Abelian groups

The proofs of the main results in this section are based on Proposition 4.1 below, and the results for symmetric measures in Section 3. Proposition 4.1 is a slight generalization to a similar result in the work of Jones, Kaufman, Rosenblatt and Wierdl [10], which is in turn a slight generalization of a result by Bourgain in [4].

Given a sequence $\{K_n\}_{n=1}^\infty$ of finite measures on G, we consider the operators they define on $L^p(X)$, $1 \leq p \leq \infty$,

$$K_n f(x) = \int_G f(T_g x) dK_n(g).$$

Proposition 4.1. *Let φ be a non–negative function on Γ. If there is a constant $c > 0$ such that $|\hat{K}_n(\lambda)| \leq c \min(n\varphi(\lambda), \frac{1}{n\varphi(\lambda)})$ and $|\hat{K}_n(\lambda) - \hat{K}_{n+1}(\lambda)| \leq \frac{c}{n}$, for all $\lambda \in \Gamma$, for all $n \neq 2^k - 1$ for some k, then*

$$\Big\| \Big(\sum_{k=1}^\infty \|\{K_t f | 2^{k-1} \leq t < 2^k\}\|_{v(2)}^2 \Big)^{\frac{1}{2}} \Big\|_2 \leq c \|f\|_2.$$

We postpone the proof of this proposition until the end of the section, and first give some additional theorems, and some applications. Our first theorem allows us to relate dyadic powers of a quite general measure to the dyadic powers of an appropriate symmetric measure. This allows us to use some of the results from Section 3.

Theorem 4.2. *Let μ be a probability measure on G which satisfies the Stoltz condition $|\hat{\mu}(\gamma) - 1| \leq C(1 - |\hat{\mu}(\gamma)|)$, and let $\nu = \mu \star \tilde{\mu}$, where $\tilde{\mu}(g) = \mu(-g)$. Then*

$$\Big\| \Big(\sum_{n=1}^\infty |\mu^{2^n} f - \nu^{2^n} f|^2 \Big)^{\frac{1}{2}} \Big\|_2 \leq C \|f\|_2.$$

Proof. First note that $\hat{\hat{\mu}}(\gamma) = \hat{\mu}(-\gamma) = \overline{\hat{\mu}(\gamma)}$. Hence $|1 - \hat{\hat{\mu}}(\gamma)| = |1 - \hat{\mu}(\gamma)|$.

Using Fourier transforms we get the estimate

$$\begin{aligned}
\Big\| \Big(\sum_{k=1}^{\infty} |\mu^{2^k} f - \nu^{2^k} f|^2 \Big)^{\frac{1}{2}} \Big\|_2^2 &= \int \sum_{k=1}^{\infty} |\hat{\mu}^{2^k}(\gamma) - \hat{\nu}^{2^k}(\gamma)|^2 |\hat{f}(\gamma)|^2 d\gamma \\
&\leq \int |1 - \hat{\mu}(\gamma)|^2 \sum_{k=1}^{\infty} 2^{2k} |\hat{\mu}^{2^k}(\gamma)|^2 |\hat{f}(\gamma)|^2 d\gamma \\
&\leq \int |1 - \hat{\mu}(\gamma)|^2 \sum_{k=1}^{\infty} 2^{k+1} \sum_{n=2^{k-1}+1}^{2^k} |\hat{\mu}^n(\gamma)|^2 |\hat{f}(\gamma)|^2 d\gamma \\
&\leq 4 \int |1 - \hat{\mu}(\gamma)|^2 \sum_{n=1}^{\infty} n |\hat{\mu}^n(\gamma)|^2 |\hat{f}(\gamma)|^2 d\gamma \\
&\leq c \int \frac{|1 - \hat{\mu}(\gamma)|^2}{(1 - |\hat{\mu}(\gamma)|^2)^2} |\hat{f}(\gamma)|^2 d\gamma \\
&\leq C \int |\hat{f}(\gamma)|^2 d\gamma = C \|f\|_2^2,
\end{aligned}$$

where the estimate in the second step is due to the fact that $\hat{\nu}(t) = \hat{\mu}(-t) = \overline{\hat{\mu}(t)}$. In the 4th step we used $|\hat{\mu}(t)| \leq 1$. The last step is due to the Stoltz condition. □

Theorem 4.3. *Let μ denote a probability measure on G which satisfies the Stoltz condition $|\hat{\mu}(\gamma) - 1| \leq C(1 - |\hat{\mu}(\gamma)|)$. Then*

$$\Big\| \Big(\sum_{k=1}^{\infty} \| \mu^n f | 2^{k-1} \leq n < 2^k \|_{v(2)}^2 \Big)^{\frac{1}{2}} \Big\|_2 \leq C \|f\|_2.$$

Proof. Our plan is to apply Proposition 4.1 with $\varphi(\gamma) = |1 - \hat{\mu}(\gamma)|$ and for $2^{k-1} \leq n < 2^k$, with $K_n = \mu^n - \mu^{2^{k-1}}$. We will be done if for these choices, the necessary estimates hold.

For the following computations, assume $2^{k-1} \leq n < 2^k$.

First, we have the trivial estimate,

$$\begin{aligned}
|\hat{K}_n(\gamma)| &\leq |\hat{\mu}^{2^{k-1}}(\gamma)| \, |1 - \hat{\mu}^{n-2^{k-1}}(\gamma)| \\
&\leq |1 - \hat{\mu}^{n-2^{k-1}}(\gamma)| \\
&\leq |1 - \hat{\mu}(\gamma)| \, | \sum_{j=0}^{n-2^{k-1}-1} |\hat{\mu}^j(\gamma)| \\
&\leq n |1 - \hat{\mu}(\gamma)| = n \varphi(\gamma).
\end{aligned}$$

Before we can make the remaining estimates, we need to establish a preliminary estimate. Since μ satisfies the Stoltz condition, $|1-\hat{\mu}(\gamma)| \leq C(1-|\hat{\mu}(\gamma)|)$, we can write

$$|\hat{\mu}(\gamma)| \leq 1 - \frac{1}{C}|1-\hat{\mu}(\gamma)| \leq e^{-\frac{1}{C}|1-\hat{\mu}(\gamma)|}.$$

We now have

$$\begin{aligned}
|\hat{K}_n(\gamma)| &\leq |\hat{\mu}^n(\gamma)| + |\hat{\mu}^{2^{k-1}}(\gamma)| \\
&\leq e^{-\frac{n}{C}|1-\hat{\mu}(\gamma)|} + e^{-\frac{2^{k-1}}{C}|1-\hat{\mu}(\gamma)|} \\
&\leq \frac{C}{n|1-\hat{\mu}(\gamma)|} + \frac{C}{2^{k-1}|1-\hat{\mu}(\gamma)|} \\
&\leq \frac{c}{n|1-\hat{\mu}(\gamma)|} \leq \frac{c}{n\varphi(\gamma)}.
\end{aligned}$$

For the final estimate, we have

$$\begin{aligned}
|\hat{K}_n(\gamma) - \hat{K}_{n+1}(\gamma)| &\leq |1-\hat{\mu}(\gamma)||\hat{\mu}^n(\gamma)| \\
&\leq |1-\hat{\mu}(\gamma)|e^{-\frac{n}{C}|1-\hat{\mu}(\gamma)|} \\
&\leq |1-\hat{\mu}(\gamma)|\frac{C}{n|1-\hat{\mu}(\gamma)|} \\
&\leq \frac{C}{n}.
\end{aligned}$$

□

We are now ready for the main result.

Theorem 4.4. *Let μ denote a probability measure on G which satisfies the Stoltz condition $|\hat{\mu}(t)-1| \leq C(1-|\hat{\mu}(t)|)$.*

1. *For $2 < s < \infty$ we have the variation inequality*

$$\Big\| \|\mu^n f\|_{v(s)} \Big\|_2 \leq C\|f\|_2.$$

2. *For any increasing sequence $\{n_k\}$, we have*

$$\Big\| \|\mu^n f\|_{o(2)}\| \Big\|_2 \leq c(\mu)\|f\|_2.$$

Proof. We first prove part 1 of the theorem. Write $B_n f = \mu^{2^{k-1}} f$ for $2^{k-1} \leq n < 2^k$. Then note that

$$\|\mu^n f\|_{v(s)} \leq \|\mu^n f - B_n f\|_{v(s)} + \|B_n f\|_{v(s)}.$$

For the first term just note that

$$\|\mu^n f - B_n f\|_{v(s)} \leq \Big(\sum_{k=1}^{\infty} \|\mu^n f|2^{k-1} \leq n < 2^k\|_{v(s)}^s\Big)^{\frac{1}{s}},$$

and by Theorem 4.3 we know this is a bounded operator on L^2.

For the second term, letting $\nu = \mu \star \tilde{\mu}$, with $\tilde{\mu}(g) = \mu(-g)$, we note that

$$\begin{aligned} \|B_n f\|_{v(s)} &= \|\mu^{2^k} f\|_{v(s)} \\ &\le \|\mu^{2^k} f - \nu^{2^k} f\|_{v(s)} + \|\nu^{2^k} f\|_{v(s)} \\ &\le \Big(\sum_{k=1}^{\infty} |\mu^{2^k} f - \nu^{2^k} f|^2 \Big)^{\frac{1}{2}} + \|\nu^{2^k} f\|_{v(s)}. \end{aligned}$$

We have

$$\Big(\sum_{k=1}^{\infty} |\mu^{2^k} f - \nu^{2^k} f|^2 \Big)^{\frac{1}{2}}$$

is a bounded operator on L^2 by Theorem 4.2. We clearly have

$$\|\nu^{2^k} f\|_{v(s)} \le \|\nu^{2n} f\|_{v(s)}.$$

Since ν is a symmetric measure, we can apply Proposition 3.1, with $P_n = \nu$, to obtain the boundedness of this term too.

We now prove part 2 of the theorem.

Let $B_n f = \mu^{2^k} f$ if $2^k \le n < 2^{k+1}$. Then

$$\|\mu^n f\|_{o(2)} \le \|B_n f\|_{o(2)} + \|\mu^n f - B_n f\|_{o(2)}.$$

Now

$$\begin{aligned} \|\mu^n f - B_n f\|_{o(2)} &\le \|\mu^n f - B_n f\|_{v(2)} \\ &\le 2 \Big(\sum_k \|\mu^n f - B_n f | 2^k \le n < 2^{k+1} \|_{v(2)}^2 \Big)^{1/2}, \end{aligned}$$

which by Theorem 4.3, is a bounded operator on L^2.

Again letting $\nu = \mu * \tilde{\mu}$, we have

$$\begin{aligned} \|B_n f\|_{o(2)} &\le \|\mu^{2^k} f\|_{o(2)} \\ &\le \|\nu^{2^k} f\|_{o(2)} + \Big(\sum_k |\mu^{2^k} f - \nu^{2^k} f|^2 \Big)^{1/2}. \end{aligned}$$

We have already seen that both of these terms are bounded operators in L^2.

$\square$

Corollary 4.5. *Let μ be as in Theorem 4.4. For each $2 < \varrho < \infty$ and each $\lambda > 0$, we have*

$$\int_X N(\mu^n f(x), \lambda)^{\frac{2}{\varrho}} dx \le \frac{c(\varrho)}{\lambda^2}\|f\|_2^2.$$

Proof. Note that as in the proof of Theorem 3.15, we have

$$\lambda N(\mu^n f(x), \lambda)^{1/\varrho} \le \|\mu^n f(x)\|_{v(\varrho)}.$$

Now just take the L^2 norm of both sides, and apply Theorem 4.4. □

Proposition 4.1 and Theorem 4.4 basically utilize Fourier transform techniques and the Stoltz condition of the measure. Thus, they can be extended for measures on non–Abelian groups which admit "nice" representations. That is, Theorem 4.4 and Corollary 4.5 can be extended to measures on σ–compact metric groups which satisfy a condition analogous to the Stoltz condition for the Abelian case. This extension will be discussed in a subsequent paper.

Proof of 4.1. By standard transfer methods (see [6], [18]), it suffices to show the theorem for functions on $\ell^2(G)$, where the action is the group addition, $T_g x = g + x$.

Let η be a function supported in $[\frac{1}{2}, 2] \cup [-2, -\frac{1}{2}]$, with $0 \le \eta \le 1$, $|\theta\eta'(\theta)| < C$ and such that $\sum_{j=-\infty}^{\infty} \eta(2^j|\theta|) = 1$ for all θ. Define $K_{j,n}$ by $\hat{K}_{j,n}(\lambda) = \hat{K}_n(\lambda)\eta(2^j n\varphi(\lambda))$. Then we have $K_n = \sum_{j=-\infty}^{\infty} K_{j,n}$.

Using the fact that $\|.\|_{v(2)}$ is a semi–norm, we now estimate as follows:

$$\begin{aligned}
&\Big\|\Big(\sum_{k=1}^{\infty} \|\{K_n f | 2^{k-1} \le n < 2^k\}\|_{v(2)}^2\Big)^{\frac{1}{2}}\Big\|_2 \\
&\qquad = \Big\|\Big(\sum_{k=1}^{\infty} \|\{\sum_{j=-\infty}^{\infty} K_{j,n} f | 2^{k-1} \le n < 2^k\}\|_{v(2)}^2\Big)^{\frac{1}{2}}\Big\|_2 \\
&\qquad \le \Big\|\Big(\sum_{k=1}^{\infty} \Big(\sum_{j=-\infty}^{\infty} \|\{K_{j,n} f | 2^{k-1} \le n < 2^k\}\|_{v(2)}\Big)^2\Big)^{\frac{1}{2}}\Big\|_2 \\
&\qquad \le \Big\|\sum_{j=-\infty}^{\infty} \Big(\sum_{k=1}^{\infty} \|\{K_{j,n} f | 2^{k-1} \le n < 2^k\}\|_{v(2)}^2\Big)^{\frac{1}{2}}\Big\|_2 \\
&\qquad \le \sum_{j=-\infty}^{\infty} \Big\|\Big(\sum_{k=1}^{\infty} \|\{K_{j,n} f | 2^{k-1} \le n < 2^k\}\|_{v(2)}^2\Big)^{\frac{1}{2}}\Big\|_2 .
\end{aligned}$$

We will now think of j as fixed, and obtain for each j the estimate

$$(1) \qquad \Big\|\Big(\sum_{k=1}^{\infty} \|\{K_{j,n}f|2^{k-1} \le n < 2^k\}\|_{v(2)}^2\Big)^{\frac{1}{2}}\Big\|_2^2 \le C^2 2^{-|j|}\|f\|_2^2.$$

If we can establish this, then we can sum over j, and the proof will be complete.

Consider the equidistributed points $2^{k-1} = a_0 < a_1 < \cdots < a_N = 2^k$ where $N = \min\left(2^{|j|}, 2^k\right)$.

For each fixed j define $A_n = K_{j,a_{m-1}}$ where $a_{m-1} \le n < a_m$. Then we have the estimate

$$\begin{aligned}(2) \quad \|\{K_{j,n}f|2^{k-1} \le n < 2^k\}\|_{v(2)} &\le \|\{A_nf|2^{k-1} \le t < 2^k\}\|_{v(2)} \\ &\quad + \|\{K_{j,n} - A_nf|2^{k-1} \le n < 2^k\}\|_{v(2)}.\end{aligned}$$

To estimate the first term, we note that A_n is constant on intervals of the form $a_{m-1} \le n < a_m$. Hence

$$\|\{A_nf|2^{k-1} \le n < 2^k\}\|_{v(2)} = \|\{K_{j,a_{m-1}}f|1 \le m \le N\}\|_{v(2)},$$

which allows us to estimate the L^2 norm

$$\begin{aligned}&\Big\|\Big(\sum_{k=1}^{\infty}\|\{A_nf|2^{k-1} \le n < 2^k\}\|_{v(2)}^2\Big)^{\frac{1}{2}}\Big\|_2^2 \\ &\qquad = \int_X \sum_{k=1}^{\infty} \|\{A_nf(x)|2^{k-1} \le n < 2^k\}\|_{v(2)}^2 dx \\ &\qquad = \int_X \sum_{k=1}^{\infty} \|\{K_{j,a_m}f(x)|m \le N\}\|_{v(2)}^2 dx \\ &\qquad \le \int_X \sum_{k=1}^{\infty} 2^2 \sum_{m=1}^{N(k)} |K_{j,a_m}f(x)|^2 dx \\ &\qquad = c \sum_{k=1}^{\infty}\sum_{m=1}^{N(k)} \int_\Gamma |\hat{K}_{j,a_m}(\gamma)|^2 \, |\hat{f}(\gamma)|^2 d\gamma,\end{aligned}$$

where between the second and third steps we used property (3) of the v(2)–norm, and in the last step we used Minkowski inequality for integrals.

Note that for each fixed j we have the following facts.

1. For each m we have $\hat{K}_{j,a_m}$ supported in $I_{j,k} = \{\lambda | \frac{1}{2^{j+k+1}} < \varphi(\lambda) < \frac{1}{2^{j+k-1}}\}$.
2. For each m we have $|\hat{K}_{j,a_m}(\lambda)| \leq c2^{-|j|}$ for $\lambda \in I_{j,k}$.
3. By definition, $N = N(k) = \min\left(2^{|j|}, 2^k\right)$.
4. For fixed j, no λ is in more than 2 of the $I_{j,k}$.

Thus we have

$$\begin{aligned}\sum_{k=1}^{\infty} \int_{\Gamma} \sum_{m=1}^{N(k)} |\hat{K}_{j,a_m}(\lambda)|^2 \ |\hat{f}(\lambda)|^2 d\lambda &\leq \sum_{k=1}^{\infty} \int_{I_{j,k}} |\frac{c}{2^{|j|}}|^2 \ N(k) \ |\hat{f}(\lambda)|^2 d\lambda \\ &\leq \frac{c}{2^{|j|}} \sum_{k=1}^{\infty} \int_{I_{j,k}} |\hat{f}(\lambda)|^2 d\lambda \\ &\leq \frac{c}{2^{|j|}} \int_{\Gamma} |\hat{f}(\lambda)|^2 d\lambda \\ &= \frac{c}{2^{|j|}} \|f\|_2^2.\end{aligned}$$

For the second term in (2) we use the second property of the v(2)–norm,

$$\begin{aligned}&\|\{K_{j,n} - A_n f | 2^{k-1} \leq n < 2^k\}\|_{v(2)} \\ &\qquad \leq 2\Big(\sum_{m=1}^{N(k)} \|\{K_{j,n} - A_n f | a_{m-1} \leq n < a_m\}\|_{v(2)}^2\Big)^{\frac{1}{2}} \\ &\qquad = 2\Big(\sum_{m=1}^{N(k)} \|\{K_{j,n} f | a_{m-1} \leq n < a_m\}\|_{v(2)}^2\Big)^{\frac{1}{2}}.\end{aligned}$$

For fixed m we have

$$\|\{K_{j,n} f | a_{m-1} \leq n < a_m\}\|_{v(2)} \leq \|\{K_{j,n} f | a_{m-1} \leq n < a_m\}\|_{v(1)}$$

$$= \sup_{a_{m-1} \leq b_1 < b_2 \cdots < b_\ell < a_m} \sum_{r=1}^{\ell} |K_{j,b_r} f - K_{j,b_{r+1}} f|, \tag{3}$$

where as usual, the supremum is taken over all possible increasing sequences contained in $[a_{m-1}, a_m)$.

We have

$$\sum_{r=1}^{\ell} |K_{j,b_r} f - K_{j,b_{r+1}} f(x)| \leq \sum_{r=a_m}^{a_{m+1}} |K_{j,r} f(x) - K_{j,r+1} f(x)|.$$

Using this and Holder's inequality, we see that (3) can be estimated by

$$\Big(a_m - a_{m-1}\Big)^{\frac{1}{2}}\Big(\sum_{r=a_{m-1}}^{a_m} |K_{j,r}f(x) - K_{j,r+1}f(x)|^2\Big)^{\frac{1}{2}}.$$

Hence we have

$$\begin{aligned}
&\|\{K_{j,n} - A_nf|2^{k-1} \le n < 2^k\}\|_{v(2)}^2 \\
&\qquad \le 4\sum_{m=1}^{N(k)} \Big(a_m - a_{m-1}\Big) \sum_{r=a_{m-1}}^{a_m} |K_{j,r}f(x) - K_{j,r+1}f(x)|^2 \\
(4) \qquad &\qquad \le 4\sum_{m=1}^{N(k)} \max\Big(\frac{2^k}{2^{|j|}}, 1\Big) \sum_{r=a_{m-1}}^{a_m} |K_{j,r}f(x) - K_{j,r+1}f(x)|^2 \\
&\qquad \le 4\max\Big(\frac{2^k}{2^{|j|}}, 1\Big) \sum_{r=2^{k-1}}^{2^k} |K_{j,r}f(x) - K_{j,r+1}f(x)|^2.
\end{aligned}$$

We now sum over k, and take the L^2 norm of (4) to obtain

(5)

$$\begin{aligned}
&\Big\| \sum_{k=1}^{\infty} \|\{K_{j,n} - A_n|2^{k-1} \le n < 2^k\}\|_{v(2)}^2 \Big\|_2^2 \\
&\quad \le \sum_{k=1}^{\infty} 4 \max\Big(\frac{2^k}{2^{|j|}}, 1\Big) \int \sum_{r=2^{k-1}}^{2^k} |K_{j,r}f(x) - K_{j,r+1}f(x)|^2 dx \\
&\quad \le 4 \sum_{k=1}^{\infty} \max\Big(\frac{2^k}{2^{|j|}}, 1\Big) \sum_{r=2^{k-1}}^{2^k} \int_{\Gamma} |\hat{K}_{j,r}(\gamma) - \hat{K}_{j,r+1}(\gamma)|^2 \ |\hat{f}(\gamma)|^2 \ d\gamma.
\end{aligned}$$

We now estimate the inner integral of (5). First note that we have

$$\begin{aligned}
|\hat{K}_{j,r}(\lambda) - \hat{K}_{j,r+1}(\lambda)| &= |\hat{K}_r(\lambda)\eta(2^j r\varphi(\lambda)) - \hat{K}_{r+1}(\lambda)\eta(2^j(r+1)\varphi(\lambda))| \\
&\le |\hat{K}_r(\lambda)\eta(2^j r\varphi(\lambda)) - \hat{K}_{r+1}(\lambda)\eta(2^j r\varphi(\lambda))| \\
&\quad + |\hat{K}_{r+1}(\lambda)\eta(2^j r\varphi(\lambda)) - \hat{K}_{r+1}(\lambda)\eta(2^j(r+1)\varphi(\lambda))| \\
&= |\hat{K}_r(\lambda) - \hat{K}_{r+1}(\lambda)| \ |\eta(2^j r\varphi(\lambda))| \\
&\quad + |\hat{K}_{r+1}(\lambda)| \ |\eta(2^j r\varphi(\lambda)) - \eta(2^j r|\lambda|^\delta + 2^j\varphi(\lambda))|.
\end{aligned}$$

For the first term we use the estimate

$$|\hat{K}_r(\lambda) - \hat{K}_{r+1}(\lambda)| \le \frac{1}{r},$$

and the fact that $\eta(2^j r\varphi(\lambda))$ is non–zero only if $\lambda \in I_{j,k}$, to conclude that

$$|\hat{K}_r(\lambda) - \hat{K}_{r+1}(\lambda)|\ |\eta(2^j r\varphi(\lambda))| \leq \frac{c}{2^k}\chi_{I_{jk}}(\lambda),$$

for $2^{k-1} \leq r < 2^k$.

For the second term we note that by the mean value theorem,

$$|\eta(2^j r\varphi(\lambda)) - \eta(2^j r\varphi(\lambda) + 2^j\varphi(\lambda))| \leq 2^j\varphi(\lambda)\eta'(\theta)|,$$

for some $\theta \in [2^j r\varphi(\lambda), 2^j r\varphi(\lambda) + 2^j\varphi(\lambda))]$.

Using the property of η that $\theta\eta'(\theta) < C$, we see

$$|\eta(2^j r\varphi(\lambda)) - \eta(2^j r\varphi(\lambda) + 2^j\varphi(\lambda))| \leq \frac{1}{r+1}(2^j(r+1)\varphi(\lambda))\ \eta'(\theta)| \leq \frac{c}{r}.$$

Consequently the second term is controlled by $\frac{1}{r}|\hat{K}_{r+1}(\lambda)|$ for $\lambda \in I_{jk}$ and is zero elsewhere.

In either case we see that for $2^{k-1} \leq r < 2^k$, we have

$$|\hat{K}_{j,r}(\lambda) - \hat{K}_{j,r+1}(\lambda)| \leq \frac{C}{2^k}\chi_{I_{jk}}(\lambda).$$

We also have the easy estimate

$$|\hat{K}_{j,r}(\lambda) - \hat{K}_{j,r+1}(\lambda)| \leq |\hat{K}_{j,r}(\lambda)| + |\hat{K}_{j,r+1}(\lambda)| \leq \frac{C}{2^{|j|}}\chi_{I_{jk}}(\lambda).$$

Hence we can estimate the inner integral by

$$\begin{aligned} &\int_\Gamma \sum_{r=2^{k-1}}^{2^k} |\hat{K}_{j,r}(\lambda) - \hat{K}_{j,r+1}(\lambda)|^2 |\hat{f}(\lambda)|^2 d\lambda \\ (6) \qquad &\leq c\min(\frac{c}{2^k}, \frac{c}{2^{|j|}})^2 2^k \int_{I_{jk}} |\hat{f}(\lambda)|^2 d\lambda \end{aligned}$$

Hence, we have that (6) can be estimated by

$$
\begin{aligned}
\Big\| \sum_{k=1}^{\infty} \|\{K_{j,n}f - A_n f | 2^{k-1} \le n < 2^k\}\|_{v(2)}^2 \Big\|_2^2 \\
&\le 4 \sum_{k=0}^{\infty} \max(\frac{2^k}{2^{|j|}}, 1) c \min(\frac{1}{2^k}, \frac{1}{2^{|j|}})^2 2^k \int_{I_{jk}} |\hat{f}(\lambda)|^2 d\lambda \\
&\le c \sum_{k=1}^{|j|} (\frac{1}{2^{|j|}})^2 2^k \int_{I_{j,k}} |\hat{f}(\lambda)|^2 d\lambda \\
&\quad + c \sum_{k=|j|}^{\infty} (\frac{2^k}{2^{|j|}}) \frac{1}{2^k} \int_{I_{j,k}} |\hat{f}(\lambda)|^2 d\lambda \\
&\le c \frac{1}{2^{|j|}} \sum_{k=1}^{\infty} \int_{I_{j,k}} |\hat{f}(\lambda)|^2 d\lambda \\
&\le c \frac{1}{2^{|j|}} \|f\|_2^2,
\end{aligned}
$$

as required. □

5. Convolutions and singular integrals in the continuous case.

In this section we extend the theory to the continuous parameter case. To do this we need a continuous analog of Proposition 4.1. We will restrict our attention to convolution operators on $\mathbb{R}^d$. In the case $d = 1$ the version we need for the applications which follow is contained in Bourgain's paper [4]. Here we state a slightly more general version. The proof is the same as the proof of Proposition 4.1, with the obvious replacement of estimates of differences by estimates involving derivatives.

Given a sequence $\{K_t\}_{t\in\mathbb{R}}$ of convolution kernels on $\mathbb{R}^d$, we consider the operators they define on $L^p(X)$, $1 \le p \le \infty$,

$$K_t f(x) = K_t \star f(x).$$

Proposition 5.1. *Let φ be a non–negative function on $\mathbb{R}^d$. If there is a constant $c > 0$ such that (i)$|\hat{K}_t(\lambda)| \le c \min(t\varphi(\lambda), \frac{1}{t\varphi(\lambda)})$ and (ii) $|\lambda \frac{\partial}{\partial t} \hat{K}_t(\lambda)| \le \frac{c}{t}$, for all $\lambda \in \mathbb{R}$, for all $t \in \mathbb{R}^+$, then*

$$\Big\| \Big(\sum_{k=-\infty}^{\infty} \|\{K_t f | 2^{k-1} \le t < 2^k\}\|_{v(2)}^2 \Big)^{\frac{1}{2}} \Big\|_2 \le c\|f\|_2.$$

With this theorem we can obtain several results about convolution operators. In each of the following cases we can take $\varphi(\lambda) = |\lambda|$. In what follows, $\Phi_t(x) = \Phi(x/t)/t^d$ where $x \in \mathbb{R}^d$.

Theorem 5.2. *Let Φ be a function on $\mathbb{R}^d$ such that for $\lambda \neq 0$, we have (a) $|\hat{\Phi}(\lambda)| \leq c/|\lambda|$, (b) $|\nabla\hat{\Phi}(\lambda)| \leq c$, and (c) $|\lambda \cdot \nabla\hat{\Phi}(\lambda)| \leq c$. Then*

$$\|\Big(\sum_{k=-\infty}^{\infty} \|\Phi_t f|2^k \leq t < 2^{k+1}\|^2_{v(2)}\Big)^{1/2}\|_2 \leq c\|f\|_2.$$

Proof. Let $K_t(x) = \Phi_t(x) - \Phi_{2^k}(x)$ if $2^k \leq t < 2^{k+1}$. Then

$$\hat{K}_t(\lambda) = \hat{\Phi}_t(\lambda) - \hat{\Phi}_{2^k}(\lambda) = \hat{\Phi}(\lambda t) - \hat{\Phi}(\lambda 2^k).$$

By (a) we can estimate

$$|\hat{K}_t(\lambda)| \leq |\hat{\Phi}(\lambda t)| + |\hat{\Phi}(\lambda 2^k)| \leq c\frac{1}{|\lambda t|},$$

with (b),

$$|\hat{K}_t(\lambda)| \leq |\lambda t - \lambda 2^k||\nabla\Phi(\lambda^\star t^\star)| \leq c|\lambda t|,$$

and by (c),

$$|\frac{\partial}{\partial t}\hat{K}_t(\lambda)| \leq |\lambda \cdot \nabla\hat{\Phi}(\lambda t)| \leq \frac{|\lambda t \cdot \nabla\hat{\Phi}(\lambda t)|}{t} \leq \frac{c}{t}.$$

Now apply Proposition 5.1. ☐

As a simple example of the kind of Φ that satisfies the above Theorem, we mention the following.

Corollary 5.3 (Differentiation operator). *Let $\Phi = \chi_{[0,1]^d}$. Then $D_t f(x) = \Phi_t f(x)$ is the differentiation operator, and we have*

$$\Big\|\Big(\sum_{k=-\infty}^{\infty} \|D_t f|2^k \leq t < 2^{k+1}\|^2_{v(2)}\Big)^{1/2}\Big\|_2 \leq c\|f\|_2.$$

Proof. We just need to see that the hypothesis of the Lemma 5.4, are satisfied. We make the following estimates.

We have

$$\hat{\Phi}(\lambda) = \prod_{k=1}^{d} \int_0^1 e^{-2\pi i\lambda_k x_k} dx_k = \prod_{1 \leq k \leq d, \lambda_k \neq 0} \frac{e^{-2\pi i\lambda_k} - 1}{-2\pi i\lambda_k}.$$

Now,

$$\sum_{k=1}^{d} \lambda_k^2 \leq \sum_{|\lambda_k|<1} + \sum_{|\lambda_k|\geq 1} \leq d + d \prod_{|\lambda_k|\geq 1} \lambda_k^2 \leq 2d \prod_{|\lambda_k|\geq 1} \lambda_k^2.$$

Hence

$$|\hat{\Phi}(\lambda)| \leq \prod_{1\leq k\leq d, |\lambda_k|\geq 1} \frac{|e^{-2\pi i\lambda_k} - 1|}{|-2\pi i\lambda_k|} \leq c\sqrt{d}\frac{1}{|\lambda|},$$

showing (a).

For (b), we compute

$$\frac{\partial}{\partial\lambda_j}\hat{\Phi}(\lambda) = \prod_{k\neq j}\int_0^1 e^{-2\pi i\lambda_k x_k}\, dx_k \int_0^1 -2\pi i x_j e^{-2\pi i\lambda_j x_j}\, dx_j.$$

Clearly the product, with $k \neq j$ is bounded by 1, and the last integral is also clearly bounded by 1. Now just sum over j.

For (c), the argument is the same except this time we need to estimate

$$\int_0^1 -2\pi i\lambda_j x_j e^{-2\pi i\lambda_j x_j}\, dx_j,$$

and an easy integration by parts shows this to be bounded too. $\square$

For the applications that follow, we also need the following lemma involving the differentiation operator.

Lemma 5.4. *Let Φ be a function on $\mathbb{R}^d$ such that for $\lambda \neq 0$, we have (a) $|\hat{\Phi}(\lambda)| \leq c\min(1/|\lambda|, 1)$, (b) $|\nabla\hat{\Phi}(\lambda)| \leq c$, (c) $|\lambda\cdot\nabla\hat{\Phi}(\lambda)| \leq c$ and (d) there exists Φ_0 such that $\lim_{s\to 0}\hat{\Phi}_s(t) = \hat{\Phi}_0(t)$ for all $t \in R^d$. Then we have*

$$\Big\|\Big(\sum_{k=-\infty}^{\infty} |\Phi_{2^k} * f - D_{2^k} * \Phi_0 * f|^2\Big)^{1/2}\Big\|_2 \leq c\|f\|_2.$$

Proof. By the assumptions about Φ and the properties of D_t established in the proof of Corollary 5.3, we have

$$|\hat{\Phi}_{2^k}(t) - \hat{D}_{2^k}(t)\hat{\Phi}_0(y)| \leq |\hat{\Phi}_{2^k}(t)| + |\hat{D}_{2^k}(t)||\hat{\Phi}_0(t)| \leq c_1\frac{1}{2^k|t|},$$

and

$$\begin{aligned}|\hat{\Phi}_{2^k}(t) - \hat{D}_{2^k}(t)\hat{\Phi}_0(t)| \leq & |\hat{\Phi}_{2^k}(t) - \hat{\Phi}_0(t)| \\ & + |\hat{D}_{2^k}(t) - 1||\hat{\Phi}_0(t)| \leq c_2|t|2^k.\end{aligned}$$

Hence, letting $c = \max c_1, c_2$,

$$
\begin{aligned}
\int \sum_k |\Phi_{2^k} * f - D_{2^k} * \Phi_0 * f|^2 dx \\
&= \int \sum_k |\hat{\Phi}_{2^k}(t) - \hat{D}_{2^k}(t)\hat{\Phi}_0(t)|^2 |f(t)|^2 dt \\
&= \int \sum_{k \geq 0} |\hat{\Phi}_{2^k}(t) - \hat{D}_{2^k}(t)\hat{\Phi}_0(t)|^2 |f(t)|^2 dt \\
&\quad + \int \sum_{k \geq 1} |\hat{\Phi}_{1/2^k}(t) - \hat{D}_{1/2^k}(t)\hat{\Phi}_0(t)|^2 |f(t)|^2 dt \\
&\leq c^2 \int \Big[\sum_{0 \leq k \leq \log_2 |t|^{-1}} 2^{2k}|t|^2 + \sum_{k > \log_2 |t|^{-1}} \frac{1}{|t|^2 2^{2k}} \Big] |f(t)|^2 dt \\
&\quad + c^2 \int \Big[\sum_{1 \leq k \leq \log_2 |t|} \frac{2^{2k}}{|t|^2} + \sum_{k > \log_2 |t|} \frac{|t|^2}{2^{2k}} \Big] |f(t)|^2 dt \\
&\leq C \|f\|_2^2,
\end{aligned}
$$

proving the lemma. □

For the differentiation operator considered in Corollary 5.3, more is true than what was established there. To prove the variation inequalities that follow, we need the stronger statement from Jones, Rosenblatt and Wierdl [11]. A discrete version in $\mathbb{R}$ can also be found in [10].

Theorem 5.5. *Let D_t denote the differentiation operator defined by $D_t f(x) = \frac{1}{t^d} \int \chi_{[0,1)^d}(\frac{x}{t}) f(x-t) dt$. Then for $\varrho > 2$,*

$$\| \|D_t f\|_{v(\varrho)} \|_2 \leq c(\varrho) \|f\|_2,$$

and

$$\| \| D_t f \|_{o(2)} \|_2 \leq c \|f\|_2.$$

Corollary 5.6. *Let Φ as in Lemma 5.4, then*

$$\| \|\Phi_t * f\|_{v(\varrho)} \|_2 \leq c(\varrho) \|f\|_2,$$

and

$$\| \| \Phi_t * f \|_{o(2)} \|_2 \leq c \|f\|_2.$$

Proof. The proof is an application of Theorem 5.2, Lemma 5.4 and Theorem 5.5, together with the observation that $\Phi_0 * f \in L^2$ if $f \in L^2$, and $\|\Phi_0 * f\|_2 \leq c\|f\|_2$. □

The Poisson kernel has been studied earlier in Corollary 3.4 with L^p results for $p > 1$. For illustration purposes we now apply the results of this section to the Poisson kernel, obtaining L^2 results.

Corollary 5.7 (Poisson kernel). *Let*

$$P_\epsilon(x) = c_d \frac{\epsilon}{(|x|^2 + \epsilon^2)^{(d+1)/2}}.$$

Let $2 < \varrho < \infty$. *Then*

1. $\| \| P_\epsilon f \|_{v(\varrho)} \|_2 \le c(\varrho) \|f\|_2$;
2. $\| \| P_\epsilon f \|_{o(2)} \|_2 \le c \|f\|_2$;
3. $\displaystyle\int_X N(P_\epsilon f(x), \lambda)^{\frac{2}{\varrho}} dx \le \frac{c(\varrho)}{\lambda^2} \|f\|_2^2$.

Proof. For the Poisson kernel, we have $\hat{\Phi}(\lambda) = e^{-2\pi|\lambda|}$. From this we easily get the estimate $|\hat{\Phi}(\lambda)| \le \min(1, \frac{1}{2\pi|\lambda|})$. For (b) we have

$$\frac{\partial}{\partial \lambda_j} \hat{\Phi}(\lambda) = e^{-2\pi|\lambda|} \frac{\lambda_j}{|\lambda|}.$$

Now just sum over j, and the result follows. For (c) the argument is almost the same. In this case we just need to note that $|\lambda| e^{-2\pi|\lambda|}$ is bounded.

For the third statement, note that as we have observed before, we have $\lambda N(P_\epsilon f(x), \lambda)^{1/\varrho} \le \|\mu^n f(x)\|_{v(\varrho)}$. Now just take the L^2 norm of both sides, and apply the first statement of the corollary.

□

Corollary 5.8 (Conjugate Poisson kernel). *The conjugate Poisson kernel on* $\mathbb{R}$ *is defined by*

$$Q_\epsilon(x) = \frac{1}{\pi} \frac{x}{x^2 + \epsilon^2}.$$

Let $2 < \varrho < \infty$. *Then*

1. $\| \| Q_\epsilon f \|_{v(\varrho)} \|_2 \le c(\varrho) \|f\|_2$;
2. $\| \| Q_\epsilon f \|_{o(2)} \|_2 \le c \|f\|_2$;
3. $\displaystyle\int_X N(Q_\epsilon f(x), \lambda)^{\frac{2}{\varrho}} dx \le \frac{c(\varrho)}{\lambda^2} \|f\|_2^2$.

Proof. Let $f \in L^2$, and let Hf denote its Hilbert transform. Then $Hf \in L^2$ and $Q_\epsilon * f(x) = P_\epsilon * Hf(x)$. The result then follows from Corollary 5.7. □

This last corollary and Corollary 3.5 give indication that the corresponding variation, oscillation and jump inequalities should also hold for the Hilbert transform. This is indeed the case. The complete study of the variation and oscillation inequalities for the Hilbert transform can be found in Campbell, Jones, Reinhold and Wierdl [7].

REFERENCES

[1] M. Akcoglu, R. L. Jones and P. Schwartz. Variation in probability, ergodic theory and analysis, *Il. J. of Math.* **42** (1998), 154-177.

[2] A. Bellow and A. P. Calderón. A weak type inequality for convolution products, in *Harmonic Analysis and Partial Differential Equations*, M. Christ, C. Kenig and C. Sadosky ed., Chicago Lectures in Mathematics, University of Chicago Press, Chicago (1999) pages 41-48.

[3] A. Bellow, R. L. Jones, and J. Rosenblatt. Almost everywhere convergence of convolution powers, *Erg. Th. and Dyn. Sys.* **14** (1994) no. 3, 415–432.

[4] J. Bourgain. Pointwise ergodic theorems for arithmetic sets, *Publications Mathematiques, Institut des Hautes Etudes Scientifique* **69** (1989), 5-45.

[5] D. L. Burkholder. Martingale transforms, *Ann. Math. Statist.* **37** (1966), 1494-1504.

[6] A. P. Calderón. Ergodic theory and translation invariant operators, *Proc. Nat. Acad. of Sciences, USA* **59** (1968), 349-353.

[7] J. Campbell, R. L. Jones, K. Reinhold and M. Wierdl. Oscillation and variation for the Hilbert transform, to appear in *Duke Math. J.* (2000).

[8] S. Glasner. On Choquet–Deny measures, *Ann. Inst. Henri Poincare* **12** (1976), 1–10.

[9] R. L. Jones, I. Ostrovskii and J. Rosenblatt. Square functions in ergodic theory, *Erg. Th. & Dyn. Sys.* **16** (1996), 267–305.

[10] R. L. Jones, R. Kaufman, J. Rosenblatt, and M. Wierdl. Oscillation in ergodic theory, *Erg. Th. & Dyn. Sys.* **18** No. 4 (1998), 889–936.

[11] R. L. Jones, J. Rosenblatt, and M. Wierdl. Oscillation in ergodic theory: Higher dimensional results, preprint.

[12] S. Kakutani. Random ergodic theorem and Markoff processes with a stable distribution, *Proc. of the second Berkeley Symposium on Math. Stats. and Prob.* (1950), 247–261. Univ. of California Press, Berkeley and Los Angeles, 1951.

[13] D. Lepingle. La variation d'order p des semi-martingales, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **36** (1976), 295-316.

[14] D. Ornstein. On the pointwise behavior of iterates of a self-adjoint operator, *J. of Math. & Mech.* **18** (1968), 473-477.

[15] V. I. Oseledets. Markov chains, skew products and ergodic theorems for "general" dynamical systems, *Teor. Verojanost. i Primenen* **10** (1965), 551–557 (Russian); English trans. *Annals of Prob. and Appl.* **10** (1965), 499–504.

[16] G. Pisier and Q. Xu. The strong p–variation of martingales and orthogonal series, *Prob. Theory* **77** (1988), 497-514.

[17] J. Rosenblatt. Ergodic and mixing random walks on locally compact groups, *Math. Ann.* **257** (1981), 31–42.

[18] J. Rosenblatt and M. Wierdl. Pointwise ergodic theorems via harmonic analysis, *Ergodic Theory and its connections to Harmonic Analysis: Proc. of the 1993 Alexandria Conference*, Cambridge Univ. Press, NY (1995), 3–151.

[19] G. C. Rota. An "Alternieven de Verfahren" for general positive operators, *Bull. AMS* **68** (1962), 95–102.

[20] E. M. Stein. *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, N.J., 1970.

[21] E. M. Stein. *Topics in Harmonic analysis related to the Littlewood–Paley theory*, Princeton Univ. Press 1970, Princeton, New Jersey.

[22] E. M. Stein. On the maximal ergodic theorem, *Proc. Nat. Acad. Sci.* **47** (1961), 1894–1897.