TIME-FREQUENCY ANALYSIS IN THE DISCRETE PHASE PLANE

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

in Candidacy for the Degree of

Doctor of Philosophy

by

Christoph Martin Thiele

May 1995

ABSTRACT

Time-Frequency Analysis in the Discrete Phase Plane

Christoph Martin Thiele
Yale University

1995

This thesis is divided into two independent parts.

First Part: The bilinear Hilbert transform is given by

$$H(f,g) = \int 1/t f(x-t) g(x+t) dt \ .$$

We introduce a Walsh model H_W, called quartile operator, for this bilinear operator and prove that the model operator satisfies

$$\|H_W(f,g)\|_r \leq C(s,t)\|f\|_s\|g\|_t$$

for $1 \leq r < \infty$, $1 < s,t \leq \infty$ and $1/r = 1/s + 1/t$. It is open whether the bilinear Hilbert transform satisfies the corresponding estimates.

Then we apply the techniques used for proving boundedness of the quartile operator to obtain a conceptual proof of the well known fact that the Walsh Fourier series of a function in $L^p[0,1)$, with $1 < p$, converges pointwise almost everywhere.

Second Part: Performing the fast Fourier transform on a signal of length 2^n, one calculates $(n+1)2^n$ coefficients of the signal. It suffices to keep certain subsets of size 2^n of these coefficients to recover the full information about the signal. We classify all these subsets and give a fast algorithm to pick out the best subset with respect to a class of cost functions. This algorithm is completely analogous to an algorithm by Lars Villemoes for Walsh bases, which is also presented in this part of the thesis.

ACKNOWLEDGEMENTS

It is my pleasure to thank my advisor, Ronald R. Coifman, for his guidance through the work of this thesis. I am grateful to him for enthusiastically introducing me into the world of pure and applied harmonic analysis, and for his generous support throughout my years at Yale.

My special thanks to Lars Villemoes. This work had been impossible without the ideas that came out of our collaboration. I am very happy to include a joint paper with Lars in this thesis.

There are many people inside and outside the Yale Mathematics Department, who deserved being mentioned here for supporting my research, encouraging me or teaching me useful mathematics. I am grateful to all these people, but unfortunately it would not fit onto this page to list them all by name.

Finally, I would like to dedicate this thesis to my parents, my family, and to my friend Andrea, who is a true master of applied harmonic analysis.

Contents

List of Figures

Chapter 0

Introduction

Time-frequency analysis deals with decompositions of functions or operators into pieces which are localized in time and frequency. This localization can be defined in several ways. The most restrictive definition is that a function in one real parameter, which we call time, is localized at a bounded time interval I, if it is supported on this interval, and it is localized at a bounded frequency interval ω, if its Fourier transform is supported on this interval. Unfortunately, with this definition, there is no function which is simultanously localized in time and frequency. Hence one often considers a weaker notion of localization, which is defined only relative to one fixed function. One fixes a mother function f and defines it to be localized at the rectangle $I \times \omega$ in the time-frequency plane. Then the translated, rescaled and modulated function $\tilde{f}(x) = f(2^k x - x_0)e^{2\pi i \xi_0 (2^k x - x_0)}$ is localized at the rectangle $2^{-k}(I + x_0) \times 2^k(\omega + \xi_0)$. The time-frequency plane is also called the phase plane.

The recent development of multiresolution analysis has shown that there are many orthonormal bases of $L^2(\mathbf{R})$, the wavelet bases, which consist of translated and rescaled copies of one mother wavelet. If the mother wavelet is chosen appropriately, then the weak definition of time-frequency localization reflects the properties of the wavelet basis very well. These bases are used to obtain an elegant approach to many problems of Littlewood Paley theory and the theory of Calderon Zygmund singular integral operators, as well as to numerical calculations.

However, for some applications, one needs a larger variety of time-frequency localized functions than those given by wavelet bases. In addition to translating and rescaling one mother function, one also needs to modulate it, in order to shift the localization in frequency. We call a set of functions a *complete collection* of time-frequency localized functions, if it is the set of functions obtained from one mother function by applying an appropriate discrete group of translations, rescalings, and modulations.

Complete collections of time-frequency localized functions are frequently used in data compression and noise reduction. For example, the acoustic signal of a musical piece is well described as a sum of notes, each having a well defined time, duration and frequency.

A more theoretical application of complete collections of time-frequency localized functions appears in the various proofs of the classical theorem of Carleson [3], which states that the Fourier series of a function in $L^2[0,1)$ converges almost everywhere in $[0,1)$. Our point of view of time-frequency analysis is especially apparent in the proof by C. Fefferman [4]. More recently, M. Lacey [10] has tried to use complete collections of time-frequency localized functions to prove a bound on the bilinear Hilbert transform. Such a bound has been conjectured by A. Calderon in the early 1970's. Those two applications have been the main source of inspiration for much of the work in this thesis.

One of the main ideas of time-frequency localization is that two functions are almost orthogonal, if they are localized at regions of the phase plane which are far away from each other. For wavelet bases one can even achieve exact orthogonality of the whole basis by choosing the mother wavelet appropriately. Unfortunately for complete collections of time-frequency localized functions the situation is much worse. Here one cannot choose a mother function such that two functions in the complete collection, which are localized at distant regions of the phase plane, are orthogonal. This causes many technical difficulties in applying complete collections of functions, which is reflected in the fact that the proofs of the Carleson theorem are technically very complicated and the fact that the bilinear Hilbert transform is still not well understood.

However, one can find discrete models of time-frequency analysis, in which this difficulty does not appear. Probably the most important one is the Walsh model, which is obtained by taking the characteristic function of the unit interval $[0,1)$ as mother function, and by modulating with Walsh functions W_n instead of exponentials $e^{2\pi i \xi_0 x}$. Here the Walsh functions are defined recursively on the unit interval $[0,1)$ by

$W_0 \equiv 1$ and, for every nonnegative integer n,

$$W_{2n}(x) = \begin{cases} W_n(2x) & \text{if } x < 0.5 \\ W_n(2x-1) & \text{if } x \geq 0.5 \end{cases},$$

$$W_{2n+1}(x) = \begin{cases} W_n(2x) & \text{if } x < 0.5 \\ -W_n(2x-1) & \text{if } x \geq 0.5 \end{cases}.$$

The complete collection of time-frequency localized functions in the Walsh model contains all functions

$$2^{-j/2}\, W_n(2^{-j}x - l) \ ,$$

with integers j, l, n and $n > 0$, and we define such a function to be localized at the rectangle

$$[2^j l, 2^j(l+1)) \times [2^{-j}n, 2^{-j}(n+1)) \ .$$

The Walsh model is a discrete model, because all functions in the complete collection are finite step functions. It has the property, that two functions in the complete collection are orthogonal, whenever they are localized at disjoint rectangles. An important subset of this complete collection is the well known Haar basis of $L^2(\mathbf{R})$. It plays the role of a standard wavelet basis in the Walsh model.

The aim of this thesis is to develop the theory of time-frequency analysis in the Walsh model and other discrete models. The main example for a discrete model of time-frequency analysis besides the Walsh model is related to the Fast Fourier Transform. The functions of the complete collection of time-frequency localized functions are most conveniently described by their inner products with an arbitrary signal:

these inner products are obtained as intermediate coefficients while performing a Fast Fourier Transform.

It is clear, that the complete collection of time-frequency localized functions in the Walsh model is too large to form a basis of $L^2(\mathbf{R})$. For example the Haar functions, a small subset of this collection, form a basis, and there are many other bases obtained by modulating the Haar basis with a fixed Walsh function. It is a natural question, to classify all those subsets of the complete collection which form bases of $L^2(\mathbf{R})$. This is important for applications like data compression and noise reduction, in which one wants to expand signals into bases which are adapted to the signals. This classification is done in terms of tilings of the phase plane into rectangles, which are associated to the time-frequency localized functions. L. Villemoes has first given a fast algorithm which selects the best basis with respect to certain cost functions. This algorithm is described in Part II of this thesis, together with its generalizations to other models of time-frequency analysis.

In Part I of this thesis, we discuss the analogues in the Walsh model of Carleson's theorem and Calderon's conjecture. We will prove almost everywhere convergence of Walsh Fourier series of L^p functions by proving a bound on the maximal partial sum operator of Walsh Fourier series. We will also define an operator which is a Walsh model for the bilinear Hilbert transform, and prove boundedness of this operator. Both operators can be naturally decomposed into sums of time-frequency localized pieces, and these sums are essentially indexed by the complete collection of time-frequency localized functions in the Walsh model.

There are three main themes in the proofs of boundedness of the two operators. The first theme is to study clusters of time-frequency localized pieces in the phase plane, i.e., sets of time-frequency localized pieces which are localized at a region in the phase plane. Such clusters we call *trees*. The region, where a tree is localized, corresponds to a subspace of $L^2(\mathbf{R})$, and part of the theory is to restrict the attention to the projections of functions onto these subspaces. Inside the region of a tree, we will detect a subset of the Haar basis or another basis, which is obtained by modulating the Haar basis by a fixed Walsh function. This means we can do Littlewood Paley theory and Calderon Zygmund theory to study the trees.

The second theme is to study the union of trees, which are localized at distant regions of the phase plane. Such a union is called a *forest*. The main tool is here orthogonality of the functions or operators associated to the trees. We will apply the Parseval inequality to obtain several estimates, which we call Carleson measure estimates.

The third theme is to collect all forests together and to obtain an estimate on the sum over the operators associated to the forests. Here the convergence is obtained by making use of two effects. For some forests, the operator norm of the associated operator is small. For other forest, the operator norm is large, but then the operator is supported only on a small exceptional set. These two effects are measured by a quantity called density, which is associated to the forests.

Part I

A.E. Convergence of Walsh Fourier Series and a Walsh Model for the Bilinear Hilbert Transform

Chapter 1

The Walsh Phase Plane

1.1 Introduction

The *Walsh functions* W_0, W_1, W_2, ... are defined recursively on the unit interval $[0, 1)$ by $W_0 \equiv 1$ and, for every nonnegative integer n,

$$W_{2n}(x) = \begin{cases} W_n(2x) & \text{if } x < 0.5 \\ W_n(2x-1) & \text{if } x \geq 0.5 \end{cases},$$

$$W_{2n+1}(x) = \begin{cases} W_n(2x) & \text{if } x < 0.5 \\ -W_n(2x-1) & \text{if } x \geq 0.5 \end{cases}.$$

They form an orthonormal basis of the Hilbert space $L^2[0, 1)$, (see [15]). The *Walsh Fourier transform* of a function f is its expansion

$$\sum_{n=0}^{\infty} (f, W_n) W_n$$

into this basis. With the natural pairing,

$$(f, W_n) = \int f(x) W_n(x)\, dx$$

the Walsh Fourier transform is formally defined for every integrable function. We have the following analogue of the famous Carleson Hunt theorem for a.e. convergence of Fourier series.

Theorem 1.1 *For every function $f \in L^p[0,1)$ with $1 < p \leq \infty$, the partial sums*

$$\sum_{n=0}^{N} (f, W_n) W_n(x)$$

converge as $N \to \infty$ for almost every x in $[0,1)$.

We extend the Walsh functions to the whole real line by defining them to be zero outside the unit interval. Then we define for each triple j, l, n of integers with $n \geq 0$ the *Walsh wave packet*

$$w_{j,l,n}(x) = 2^{-j/2}\, W_n(2^{-j}x - l)$$

The *quartile operator* H_W is a bilinear operator which formally assigns to two locally integrable functions f and g on $\mathbf{R}$ the function

$$H_W(f,g) = \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{n=0}^{\infty} 2^{-j/2}\, (f, w_{j,l,4n})\, (g, w_{j,l,4n+1})\, w_{j,l,4n+2}$$

As usual we denote the norm of the Lebesgue space $L^p[\mathbf{R}]$ by $\|f\|_p$. Then we have the following bound on the quartile operator.

Theorem 1.2 *Let $1 < s, t \leq \infty$ and $1 \leq r < \infty$, such that $1/r = 1/s + 1/t$. Then the above series for $H_W(f,g)$ converges absolutely in L^r, whenever $f \in L^s$ and $g \in L^t$, and we have the bound*

$$\|H_W(f,g)\|_r \leq C \|f\|_s \|g\|_t$$

with a constant C depending only on s and t.

In this paper we prove Theorem 1.2. and we give a new proof of Theorem 1.1

Theorems 1.1 and 1.2 are very close in nature, which is reflected by the fact that the proofs we present are largely identical. The main feature of these proofs is a decomposition of operators, in the case of Theorem 1.1 the linearized maximal partial sum operator, and in the case of Theorem 1.2 the quartile operator, into pieces which are localized in time and frequency. For the quartile operator this decomposition is already apparent from its definition.

We will represent time-frequency localized pieces graphically by rectangles in the phase plane. As an instance of the Heisenberg uncertainty, these rectangles have area one. Hence they are given by three parameters, a time , a frequency and a scale, which relates the width in time to the width in frequency. The main objective of this paper is to describe the interplay between the geometric structure of these rectangles in the phase plane and the structure of the operators which are decomposed into time-frequency localized pieces.

The notions *time* and *frequency* in the above are understood in terms of the Walsh Fourier transform. The Walsh Fourier transform is a discrete model for the Fourier transform, which in nature relates time and frequency. In the Walsh model, time-frequency localization is very sharp and allows a straightforward translation of geometric properties of the phase plane into properties of functions which are built of time frequency localized pieces. The corresponding theory for the Fourier transform, which should reflect many of the aspects of the model theory, is much more complicated because strict localization of functions can be achieved only either in time or in frequency. The theory in the Fourier case is not yet understood satisfactorily.

The analogue of Theorem 1.1 in the case of the Fourier transform is clearly the famous Carleson Hunt Theorem about almost everywhere convergence of Fourier series. The basic ideas of time-frequency localization, which are worked out in this thesis, appear already in the two classical proofs of this theorem by L.Carleson in [3] and C.Fefferman in [4]. The proof for the Walsh case has first been done by Billard for $L^2(\mathbf{R})$ in [2] and Sjölin in [13] for $L^p(\mathbf{R})$. The proof we present here is slightly refined such that it gives the L^p estimate without using Hunt's trick [7] with characteristic functions.

The Fourier analogue of the quartile operator is the bilinear Hilbert transform, given by

$$H(f,g)(x) = \int_R f(x-t)g(x+t)\frac{1}{t}\,dt$$

It is conjectured, that the bilinear Hilbert transform satisfies the same estimates as the quartile operator in Theorem 1.2. This conjecture was first formulated by A.P. Calderon in the early 1970's for the special case $L^2 \times L^\infty \mapsto L^2$. Recently M. Lacey [10] has tried to apply the techniques of time-frequency localization to prove Calderon's conjecture, an approach which has also been suggested by R. Coifman and Y. Meyer. Some of the ideas in the present paper have been inspired by Lacey's work. Unfortunately we don't see a method to apply the results for the quartile operator to obtain any information about the bilinear Hilbert transform.

The rest of this chapter is devoted to developing the geometry of the Walsh phase plane and to discussing how it translates into a theory of time-frequency localization of functions.

In Chapter 2 we discuss the quartile operator and prove Theorem 1.2. This is done in two steps. First we pass to the trilinear form

$$T(f_1, f_2, f_3) = (H(f_1, f_2), f_3) \quad .$$

By duality Theorem 1.2 translates into boundedness properties of the trilinear form. There is a special symmetry in the case that all three arguments in the trilinear form are in Lebesgue spaces with exponents between 2 and ∞, allowing an especially elegant proof of boundedness. To get the remaining estimates, we then work directly on the quartile operator, which will also give another proof for the symmetric estimates of the trilinear form. Finally we will give a heuristic argument, which shows the connection between the quartile operator and the bilinear Hilbert transform.

In Chapter 3 we will formulate a proof of Theorem 1.1 using the techniques that have been introduced in the previous Chapters.

1.2 Geometric properties of the Walsh phase plane

Recall that a *dyadic interval* is of the form $[2^j n, 2^j(n+1))$ with integers j, n. Given two dyadic intervals, either they have empty intersection, or one is contained in the other.

The Walsh phase plane is the closed upper half plane $\mathbf{R} \times \mathbf{R}_0^+$. We think of the first coordinate as time and the second coordinate as frequency. A *tile* p is a rectangle $I_p \times \omega_p$ in the Walsh phase plane, such that $|I_p||\omega_p| = 1$ and both I_p and ω_p are dyadic intervals. We call I_p the *time interval* and ω_p the *frequency interval* of p.

The dyadic structure forces many interesting properties of the set of tiles, and the aim of this section is to summarize those which are needed later.

A *bitile* is a dyadic rectangle $P = I_P \times \omega_P$ in the Walsh phase plane of area two, i.e., $|I_P||\omega_P| = 2$ and both I_P and ω_P are dyadic intervals. Similarly we define *quartiles* as dyadic rectangles of area four. If $P = [x_0, x_1) \times [\xi_0, \xi_1)$ is a bitile, we call the tile $[x_0, x_1) \times [\xi_0, (\xi_0 + \xi_1)/2)$ the *upper son* of P. Similarly we define the *lower*, *left*, and *right sons* to be the tiles $[x_0, x_1) \times [(\xi_0 + \xi_1)/2, \xi_1)$, $[x_0, (x_0 + x_1)/2) \times [\xi_0, \xi_1)$, and $[(x_0 + x_1)/2, x_1) \times [\xi_0, \xi_1)$. We call the upper and lower sons *frequency brothers* of each other, and the left and right sons *time brothers*. In this way each tile has a unique frequency brother and a unique time brother.

Lemma 1.3 *Let P be a bitile with upper and lower son p and p', and let Q be a different bitile with upper and lower sons q and q'. Then at least one of the intersections $p \cap q$ and $p' \cap q'$ is empty.*

To prove the lemma, assume that both intersections are nonempty. Pick two points $(x, \xi) \in p \cap q$ and $(x', \xi') \in p' \cap q'$. Then the frequency interval of the bitile P is the smallest dyadic interval which contains both ξ and ξ'. The same holds for the frequency interval of Q, hence the two frequency intervals are equal. This implies that the time intervals of both bitiles have the same length, and as the bitiles have nonempty intersection, the time intervals must be equal too. This is a contradiction to the assumption that P and Q are different.

We define a tile p to be *less than* a tile p', and we write $p \leq p'$, if $I_p \subset I_{p'}$ and $\omega_{p'} \subset \omega_p$. This defines a partial order on the set of all tiles. Observe that two tiles are *comparable*, i.e., $p \leq p'$ or $p' \leq p$, if and only if they have nonempty intersection. If $\mathbf{p}$ is a set of tiles, we denote by $\mathbf{p}^{\min}$ the set of minimal tiles in $\mathbf{p}$ and by $\mathbf{p}^{\max}$ the set of maximal tiles in $\mathbf{p}$.

Lemma 1.4 *Let $\mathbf{p}$ be a finite set of pairwise disjoint tiles. Let $\tilde{\mathbf{p}}$ be the set of all tiles $\tilde{p}$ with the property*

$$\tilde{p} \subset \bigcup_{p \in \mathbf{p}} p$$

Then either $\mathbf{p} = \tilde{\mathbf{p}}^{\min}$ or $\mathbf{p}$ contains a pair of frequency brothers.

Proof: The tiles in $\tilde{\mathbf{p}}^{\min}$ are minimal and therefore pairwise disjoint. They are all covered by tiles of $\mathbf{p}$. Hence if $\mathbf{p} \subset \tilde{\mathbf{p}}^{\min}$ we have already $\mathbf{p} = \tilde{\mathbf{p}}^{\min}$ and we are done. Therefore we can assume that there is a tile in $\mathbf{p}$ which is not in $\tilde{\mathbf{p}}^{\min}$. We pick such a tile p with maximal time interval I_P. There is a tile $\tilde{p} \in \tilde{\mathbf{p}}$ which is strictly less than p. This $\tilde{p}$ is also strictly less than the frequency brother q of p. The tile $\tilde{p}$ is covered by $\mathbf{p}$, hence there is a tile $p' \in \mathbf{p}$ which intersects q nontrivially. This tile does not intersect p, hence we conclude $q \leq p'$. But then p' is also a tile in $\tilde{\mathbf{p}}$ which is not minimal, and by choice of p we must have $|I_{p'}| \leq |I_p|$. This implies $q = p'$, and we have shown that $\mathbf{p}$ contains a pair of frequency brothers.

Corollary 1.5 *Let $\mathbf{p}$ and $\tilde{\mathbf{p}}$ be as in Lemma 1.4. Then*

$$\bigcup_{p \in \mathbf{p}} p = \bigcup_{\tilde{p} \in \tilde{\mathbf{p}}^{\min}} \tilde{p}$$

As long as $\mathbf{p} \neq \tilde{\mathbf{p}}^{\min}$, we can replace a pair of frequency brothers in $\mathbf{p}$ by the pair of time brothers which covers the same bitile, and it suffices to show the corollary for the new set of tiles. Doing this successively, the procedure must terminate, because changing from frequency brothers to time brothers gives tiles with larger frequency intervals, but there is an upper bound on the size of frequency intervals of tiles in $\tilde{p}$. Hence this procedure terminates with $\mathbf{p} = \tilde{\mathbf{p}}^{\min}$, and the corollary follows.

Corollary 1.6 *Let* $\mathbf{p}$ *and* $\tilde{\mathbf{p}}$ *be as in Lemma* 1.4*. Then for each* $\tilde{p}$ *in* $\tilde{\mathbf{p}}$ *there is a set* $\mathbf{p}'$ *of pairwise disjoint tiles which contains* $\tilde{p}$ *and satisfies*

$$\bigcup_{p \in \mathbf{p}'} p = \bigcup_{p \in \mathbf{p}} p$$

By the previous corollary we can assume that $\mathbf{p} = \tilde{\mathbf{p}}^{\min}$. We can also assume that each tile in $\tilde{\mathbf{p}}^{\min}$ intersects $\tilde{p}$ nontrivially, by removing all those, which don't. Then $\tilde{p}$ is maximal in $\tilde{\mathbf{p}}$. By the dual version to the previous corollary the set of maximal tiles in $\tilde{\mathbf{p}}$ satisfies the requirements we need.

The following lemma gives a criterion for a subset of the phase plane to be decomposable into a disjoint union of tiles. We define for bitiles the same partial order as for tiles. Then a set $\mathbf{P}$ of bitiles is called *convex*, if for all ordered triples $P \leq P' \leq P''$ of bitiles we have the property that if P and P'' are in $\mathbf{P}$, then P' is also in $\mathbf{P}$.

Lemma 1.7 *The union of a finite convex set of bitiles can be decomposed into a disjoint union of tiles.*

We use induction on the number of bitiles. Clearly the lemma is true for the empty set of bitiles. Given a nonempty convex set we pick a minimal bitile P in the set.

Let p be one of the two frequency brothers which constitute P. Let P' be the bitile which is the union of p and its time brother. If there is any bitile $P'' \in \mathbf{P}$ besides P which intersects p nontrivially, then we have $P'' \leq P' \leq P$, and by convexity we conclude $P' \in \mathbf{P}$. This means that p is either contained in or disjoint from the union of bitiles in $\mathbf{P} \setminus \{P\}$. As this is true for both frequency brothers in P, we can apply the induction hypothesis to conclude the statement of the lemma.

1.3 Basic algebraic properties of the Walsh phase plane

We relate the geometric properties of tiles in the Walsh phase plane to algebraic properties of wave packets via the following correspondence. A tile p is determined by three integer parameters j, l, n with $n \geq 0$, if we write

$$p = [2^j l, 2^j(l+1)) \times [2^{-j}n, 2^{-j}(n+1))$$

We associate to this tile the wave packet

$$w_p(x) = w_{j,l,n}(x) = 2^{-j/2}\, W_n(2^{-j}x - l)$$

Observe that the wave packet w_p is supported on the time interval I_p of the tile p, and that its absolute value is constant on this interval.

Lemma 1.8 *Let P be a bitile and let w_{up},w_{lo},w_l, and w_r be the wave packets corresponding to upper, lower, left and right son of P. Then we have*

$$w_l = \frac{1}{\sqrt{2}}(w_{lo} + w_{up})$$

$$w_l = \frac{1}{\sqrt{2}}(w_{lo} - w_{up})$$

After rewinding the definitions this is nothing but a reformulation of the recursion relations for the Walsh functions.

Corollary 1.9 *Let* $\mathbf{p}$ *and* $\mathbf{p}'$ *be two finite sets of pairwise disjoint tiles, which cover the same area in the phase plane. Then the corresponding two sets of wave packets span the same vector subspace of* $L^2(\mathbf{R})$.

As in Corollary 1.5, we can replace in each set successively pairs of frequency brothers by pairs of time brothers without changing the spanned vector space, until both sets become $\tilde{\mathbf{p}}^{\min}$. This proves Corollary 1.9.

This Corollary gives for each subset S of the Walsh phase plane, which is a disjoint union of tiles, a unique vector space associated to the subset. We denote by Π_S the orthogonal projection onto this subspace in $L^2(\mathbf{R})$, and we say that the set S *defines a projection.*

Corollary 1.10 *Let* $\mathbf{p}$ *be a finite set of pairwise disjoint tiles and let* A *denote their union. Then the wave packet of any tile* $p \subset A$ *is contained in the vector space spanned by the wave packets of the tiles in* $\mathbf{p}$.

This is an immediate consequence of Corollary 1.6 and the previous corollary.

Lemma 1.11 *If two tiles* p *and* p' *are disjoint, then the corresponding wave packets are orthogonal in* $L^2(\mathbf{R})$.

If the time intervals of the two tiles are disjoint, the two wave packets have disjoint support, which implies orthogonality. If the time intervals are not disjoint, then the frequency intervals must be disjoint. By rescaling the time axis we can assume that both time intervals are subsets of $[0,1)$. Let $\mathbf{p}$ be the set of tiles of the form $[0,1)\times\omega$ which intersect p and let $\mathbf{p}'$ be the set of tiles of this form which intersect p'. These two sets are disjoint. The wave packet corresponding to a tile $[0,1)\times\omega$ is a Walsh function restricted to $[0,1)$, and two different Walsh functions are known to be orthogonal on $[0,1)$. Hence by the previous Corollary the two wave packets in question are contained in orthogonal subspaces of $L^2(\mathbf{R})$, which proves the lemma.

1.4 Carleson measure estimates

Assume we are given a function $f\in L^2$ and a set $\mathbf{p}$ of pairwise disjoint tiles in the phase plane. Then we can use orthogonality of the wave packets to obtain the Parseval inequality

$$\sum_{p\in\mathbf{p}}|(f,w_p)|^2\leq\|f\|_2^2$$

We will be concerned with sets of tiles for which $|(f,w_p)|/\sqrt{|I_p|}$ is roughly of a fixed size 2^k. Then the Parseval inequality becomes

$$\sum_{p\in\mathbf{p}}|I_p|\leq C\frac{\|f\|_2^2}{2^k}$$

We can rewrite the left hand side of this inequality as

$$\sum_{p\in\mathbf{p}}|I_p|=\int_R N_{\mathbf{p}}(x)\,dx$$

where $N_{\mathbf{p}}$ is the function which assigns to each time x the number of tiles in $\mathbf{p}$ whose time intervals contain x.

We call a bound as above on the sum of time intervals or the function $N_{\mathbf{p}}$ a *Carleson measure estimate*. The following lemma describes Carleson measure estimates when the function f is merely in some L^r space.

Lemma 1.12 *Let f be a function in L^r. Denote the conjugate exponent of r by r', i.e., $1/r' + 1/r = 1$. For each integer k let $\mathbf{p}_k$ be a set of pairwise disjoint tiles p such that $\|\Pi_p f\|_\infty \geq 2^k$. Then*

1. *if $1 < r < 2$,*
$$\|N_{\mathbf{p}_k}^{1/(r'+\epsilon)}\|_r \leq C(\epsilon, r)\frac{\|f\|_r}{2^k}$$
for any small $\epsilon > 0$.

2. *if $r = 2$,*
$$\sum_{p \in \mathbf{p}_k} |I_p| \leq C\frac{\|f\|_2^2}{2^k}$$

3. *if $2 < r < \infty$,*
$$\sum_{k=-\infty}^{\infty} 2^{rk} \sum_{p \in \mathbf{p}_k} |I_p| \leq C(r)\|f\|_r^r$$

Given the fact that $\|\Pi_p f\|_\infty = |(f, w_p)|/\sqrt{|I_p|}$, we have already shown the lemma for $r = 2$. For $r > 2$ we use that the function f is locally square integrable. Denote by $M_2 f$ the maximal function

$$M_2 f(x) = \sup_{I \text{ dyadic}; x \in I} \left(\frac{1}{|I|}\int_I |f(y)|^2\, dy\right)^{1/2}$$

Clearly the time interval of a tile in $\mathbf{p}_k$ must be contained in the set

$$\{x : M_2 f(x) \geq 2^k\}$$

Hence we can apply our L^2 estimate to the restriction of f to this set. We claim that

$$\int_{\{x:M_2 f(x)\geq 2^k\}} |f(y)|^2 \, dy \leq C 2^{2k} |\{x : M_2 f(x) \geq 2^k\}|$$

Multiplying by 2^{rk-2k} and summing over k proves the lemma in the case $r > 2$, because the right hand side becomes $\|M_2 f\|_r^r$, which by the maximal theorem is bounded by $\|f\|_r^r$.

To see the claim, observe that on each maximal dyadic subinterval I of the set $\{x : |M_2 f(x)| > 2^{k-1}\}$ we have

$$\frac{1}{2|I|} \int_I |f(y)|^2 \, dy \leq C 2^{2k}$$

because the next larger dyadic interval contains a point x where the maximal function $M_2 f(x)$ is less than 2^k. Multiplying by $|I|$ and summing over all maximal dyadic subintervals shows the claim.

To show the lemma for $r < 2$ we need to invoke interpolation of vector valued functions. Let p be any collection of pairwise disjoint tiles. For each function g on $\mathbf{R}$ we can form the vector valued function G on $\mathbf{R}$ with values in $\mathbf{R}^{\mathbf{p}}$ as follows

$$G(x) = (\Pi_p g(x))_{p\in\mathbf{p}}$$

The $L^s(l^t)$-norm of the function G is defined as

$$\left\|(G(x))_{p\in\mathbf{p}}\right\|_{L^s(l^t)} := \left(\int \left(\sum_{p\in\mathbf{p}'} |G(x)|^t\right)^{s/t} dx\right)^{1/s}$$

We want to use interpolation for the linear operator which maps g to G. The endpoints of the interpolation are the obvious bounds

$$\left|\left|(\Pi_p g(x))_{p\in\mathbf{p}}\right|\right|^2_{L^2(l^2)} = \sum_{p\in\mathbf{p}} ||\Pi_p g(x)||_2^2 \leq \|g\|_2^2$$

and for some some small $\epsilon > 0$

$$\left|\left|(\Pi_p g(x))_{p\in\mathbf{p}}\right|\right|^{1+\epsilon}_{L^{1+\epsilon}(l^\infty)} = \int \left(\sup_{p\in\mathbf{p}} |\Pi_p g(x)|\right)^{1+\epsilon} dx \leq$$

$$C\int Mg(x)^{1+\epsilon}\, dx \leq C\|g\|_{1+\epsilon}^{1+\epsilon}$$

Complex interpolation (see [8]) gives for $1 < r < 2$

$$\left|\left|(\Pi_p g(x))_{p\in\mathbf{p}}\right|\right|_{L^r(l^{r'+\epsilon})} \leq C\|g\|_r$$

with a new $\epsilon > 0$ which can be made arbitrarily small.

Now we apply this estimate to the function f and the set of tiles $\mathbf{p}_k$. Together with the observation

$$\|N_{\mathbf{p}_k}^{1/(r'+\epsilon)}\|_r^r = \int N_{\mathbf{p}_k}(x)^{r/(r'+\epsilon)}\, dx \leq \frac{1}{2^r k}\int \left(\sum_{p\in\mathbf{p}_k} |\Pi_p f(x)|^{r'+\epsilon}\right)^{r/(r'+\epsilon)} dx$$

this finishes the proof of the lemma.

Corollary 1.13 *The Carleson measure estimates of Lemma 1.2 are also true, if the sets $\mathbf{p}_k$ are replaced by sets $\mathbf{P}_k$ of pairwise disjoint bitiles (quartiles).*

We prove the corollary for bitiles, the case of quartiles being similar. If P is a bitile such that $\|\Pi_P f\|_\infty \geq 2^k$, then one of the frequency brothers in the bitile, p, satisfies $\|\Pi_p(2f)\|_\infty \geq 2^k$. We pick for each bitile in the set $\mathbf{P}_k$ such a tile and call this set of tiles $\mathbf{p}_k$. Then we apply the lemma to the sets $\mathbf{p}_k$ and the function $2f$. This proves the corollary.

Chapter 2

The Quartile Operator

2.1 Introduction

In this chapter $\mathbf{P}$ denotes the set of all quartiles. Each quartile P is the union of four tiles a_P, b_P, c_P, and d_P, which have the same time interval as the quartile. They are shown in Figure 2.1.

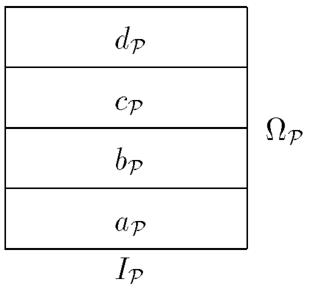

Figure 2.1: Subdivision of quartiles

The quartile operator is

$$H_W(f,g) = \sum_{P \in \mathbf{P}} \frac{1}{\sqrt{|I_P|}} (w_{a_P}, f)(w_{b_P}, g) w_{c_P}$$

It has no significant meaning that we have broken the symmetry between the four tiles in writing down the quartile operator. For example the operator

$$H'_W(f,g) = \sum_{P\in\mathbf{P}} \frac{1}{\sqrt{|I_P|}}(w_{d_P}, f)(w_{c_P}, g)w_{a_P}$$

has the same boundedness properties as the quartile operator. It is important though, that the three quartiles we have picked in the definition of the quartile operator are different.

One makes sense of the infinite sum in the definition of the quartile operator as follows. Fix an $f \in L^s$ with $1 \le s < \infty$. Then the sum is uniformly absolutely convergent if g is the characteristic function of the interval $[0,1]$. To see this, observe that in this case a quartile P contributes to the sum only if b_P intersects the square $[0,1) \times [0,1)$. We find 2^n such quartiles with time interval $[0, 2^{n+1})$, if n is a positive integer. This gives the trivial estimate:

$$\sum_{P\in\mathbf{P}} \left\| \frac{1}{\sqrt{|I_P|}}(w_{a_P}, f)(w_{b_P}, g)w_{c_P} \right\|_\infty \le C\sum_{n=0}^{\infty} 2^{-n/s}\|f\|_s \le C(s)\|f\|_s$$

By translating and dilating g dyadically, and taking linear combinations, one obtains uniform absolute convergence for all dyadic step functions g. Hence the quartile operator extends to all $g \in L^t$ with $1 \le t < \infty$ and is bounded, as soon as the bound of Theorem 1.2 is shown for all step functions g. In the case when f or g is in L^∞, one uses similar arguments for the appropriate dual of the quartile operator.

We will present two different approaches to Theorem 1.2, the first one proving only part of the theorem, and the second proving the full theorem. The reason that we present also the first approach is, that it makes explicit use of the symmetry of the quartile operator and therefore results in a more elegant and straight forward proof of this part of the theorem.

The symmetry is expressed by pairing the quartile operator with a third function and thereby passing to the trilinear form

$$T(f_1, f_2, f_3) = \sum_{P \in \mathbf{P}} \frac{1}{\sqrt{|I_P|}} (w_{a_P}, f_1)(w_{b_P}, f_2)(w_{c_P}, f_3) \tag{2.1}$$

We will show the bound

$$|T(f_1, f_2, f_3)| \leq C \|f_1\|_{r_1} \|f_2\|_{r_2} \|f_3\|_{r_3}$$

for $2 < r_1, r_2, r_3 < \infty$ and $1/r_1 + 1/r_2 + 1/r_3 = 1$. By duality this proves the theorem for $1 < r < 2$ and $2 < s, t < \infty$.

To prove the remaining estimates of the theorem we will argue directly for the quartile operator and prove the weak type estimates

$$|\{x : |H_W(f, g)(x)| > \lambda\}| \leq C \frac{(\|f\|_s \|g\|_t)^r}{\lambda^r}$$

whith $1/r = 1/s + 1/t$, $1 < s < 2$, $2 \leq t < \infty$. By real interpolation we obtain the theorem for triples (r, s, t) of the form $(1, t', t)$ with $2 < t < \infty$. If we fix g and view the quartile operator as linear operator in f, the adjoint operator is essentially of the same form as the quartile operator and satisfies the same estimates. By duality we obtain the theorem for triples of the form (t, ∞, t) with $2 < t < \infty$. Interpolation gives the result for all triples having $2 < t < \infty$ as third entry. By symmetry we also get the estimate for all triples having $2 < s < \infty$ as second entry. Then the only estimate missing is $(1, 2, 2)$, but this is proved by an appropriate interpolation with one of the weak type estimates above.

2.2 The decomposition of the trilinear form

We organize the set of quartiles into a hierarchy of subsets. We need the partial order of the set of quartiles given by:

$$P \leq P' \text{ if } \quad P \cap P' \neq \emptyset \text{ and } I_P \subset I_{P'}$$

We define the following density of the quartile P with respect to the function f,

$$\delta(P, f) = \sup_{P' \geq P} \|\Pi_{P'} f\|_\infty$$

where Π_P denotes the orthogonal projection onto the space spanned by w_{a_P}, w_{b_P}, w_{c_P}, and w_{d_P}. Then we define for all integers k the sets of homogenous density

$$\mathbf{P}_k(f) = \{P \in \mathbf{P} : 2^k \leq \delta(P, f) < 2^{k+1}\}$$

The top level in our hierarchy of subsets are the subsets with constant density with respect to all three functions that are involved in the trilinear form:

$$\mathbf{P}_{k_1,k_2,k_3} = \bigcap_{i=1,2,3} \mathbf{P}_{k_i}(f_i)$$

This gives a partition of all quartiles with nonvanishing density, but the quartiles with vanishing density don't contribute to the trilinear form anyway. We point out that if one of the functions f_i is a linear combination of characteristic functions of dyadic intervals, and we can assume that this is the case, then the sets $\mathbf{P}_{k_1,k_2,k_3}$ are finite.

Let $\mathbf{P}^{\max}_{k_1,k_2,k_3}$ denote the set of maximal quartiles in $\mathbf{P}_{k_1,k_2,k_3}$ with respect to the ordering $\leq$ of quartiles. We enumerate the quartiles in $\mathbf{P}^{\max}_{k_1,k_2,k_3}$ as $P^1_{k_1,k_2,k_3}$, $P^2_{k_1,k_2,k_3}$, and so on, and we define recursively $\mathbf{P}^i_{k_1,k_2,k_3}$ to be the set of elements in $\mathbf{P}_{k_1,k_2,k_3}$ which are less than $P^i_{k_1,k_2,k_3}$ but not less than any of the quartiles $P^{i'}_{k_1,k_2,k_3}$ with $i' < i$. The sets $\mathbf{P}^i_{k_1,k_2,k_3}$ are called *trees*, and the maximal quartiles $P^i_{k_1,k_2,k_3}$ *tree tops.*

The trees are the smallest subsets in our hierarchy. The rest of this section is devoted to proving the following estimate on a single tree.

$$\left| \sum_{P \in \mathbf{P}^i_{k_1,k_2,k_3}} \frac{1}{\sqrt{|I_P|}} (w_{a_P}, f_1)(w_{b_P}, f_2)(w_{c_P}, f_3) \right| \leq 2^{k_1+k_2+k_3} |I_{P^i_{k_1,k_2,k_3}}| \tag{2.2}$$

For boundedness of the trilinear form it will then suffice to show

$$\sum_{k_1,k_2,k_3} 2^{k_1+k_2+k_3} \sum_i |I_{P^i_{k_1,k_2,k_3}}| \leq C \|f_1\|_{r_1} \|f_2\|_{r_2} \|f_3\|_{r_3} \tag{2.3}$$

which is done in the next section using Carleson measure estimates with respect to the functions f_1, f_2, and f_3.

We denote the tree $\mathbf{P}^i_{k_1,k_2,k_3}$ for short by $\mathbf{Q}$ and the tree top $P^i_{k_1,k_2,k_3}$ by Q. Fix a frequency ξ in the frequency interval of the tree top and split the tree into four subsets, calles *subtrees*, $\mathbf{Q}_a$, $\mathbf{Q}_b$, $\mathbf{Q}_c$, and $\mathbf{Q}_d$, where $\mathbf{Q}_a$ contains all quartiles P in $\mathbf{Q}$ such that ξ lies in the frequency interval of a_P, etc. The tree estimate follows by the corresponding estimate for each of the four subtrees, and as the four subtrees are handled in similar fashion, we discuss only the tree $\mathbf{Q}_a$.

The first crucial observation is, that the tiles b_P and $b_{P'}$ are disjoint, whenever P and P' are two different quartiles in $\mathbf{Q}_a$. This follows from Lemma 1.3, because whenever the time intervals of P and P' have nonempty intersection, the tiles a_P and a'_P have nonempty intersection.

Hence as P runs through the quartiles of $\mathbf{Q}_a$, the wave packets corresponding to the tiles b_P form an orthogonal set.The same is true by a similar argument for the tiles c_P. Therefore we can apply the Cauchy Schwarz inequality to get

$$\left|\sum_{P\in\mathbf{Q}} \frac{(w_{a_P}, f_1)}{\sqrt{|I_P|}}(w_{b_P}, f_2)(w_{c_P}, f_3)\right| \leq$$

$$2^{k_1} \sum_{P\in\mathbf{Q}} |(w_{b_P}, f_2)||(w_{c_P}, f_3)| \leq 2^{k_1}\|f_2\|_2\|f_3\|_2$$

The second crucial observation is that we can replace in this estimate the functions f_2 and f_3 by their projections onto the area $A_{\mathbf{Q}} = \bigcup_{P\in\mathbf{Q}} P$ of the tree. This projection $\Pi_{A_{\mathbf{Q}}}$ is defined, because by construction the tree $\mathbf{Q}$ is a convex set of quartiles, and we can apply Lemma 1.7. The left hand side of the above estimate doesn't change if we replace f_2 and f_3 by their projections onto the area of the tree, because by Corollary 1.6 and Corollary 1.9 the functions w_{b_P} and w_{c_P} are contained in the image of the orthogonal projection $\Pi_{A_{\mathbf{Q}}}$.

We estimate the norms of the functions $\Pi_{A_{\mathbf{Q}}} f_i$ for $i = 2, 3$. Clearly they are supported on I_Q . Furthermore if $x \in I_Q$, and if P is the minimal quartile in the tree whose time interval contains x, then $\Pi_{A_{\mathbf{Q}}} f_i(x) = \Pi_P f_i(x)$, because the difference of the two projections is supported outside x. But $\Pi_P f_i(x)$ is bounded by 2^{k_i} by construction of the tree, hence $\Pi_{A_{\mathbf{Q}}} f_i$ is also a bounded function. Now we can estimate

$$2^{k_1}\|\Pi_{A_{\mathbf{Q}}} f_2\|_2\|\Pi_{A_{\mathbf{Q}}} f_3\|_2 \leq 2^{k_1}2^{k_2}\sqrt{|I_Q|}2^{k_3}\sqrt{|I_Q|} = 2^{k_1+k_2+k_3}|I_Q|$$

which finishes the proof of the tree estimate.

2.3 An integration lemma

We first reduce Inequality 2.3 to a slightly simpler inequality by an easy geometric argument. Fix a quartile P in $\mathbf{P}_{k_1}^{\max}(f_1)$ and consider the collection $\mathbf{P}'$ of quartiles in $\mathbf{P}_{k_1,k_2,k_3}^{\max}$ which are less than P. The quartiles in $\mathbf{P}'$ are pairwise incomparable. But the frequency intervals of all these quartiles contain the interval ω_P, hence the time intervals of the quartiles in $\mathbf{P}'$ are pairwise disjoint. This implies the estimate

$$\sum_{P' \in \mathbf{P}'} |I_{P'}| \leq |I_P|$$

If we sum these inequalities for all $P \in \mathbf{P}_{k_1}^{\max(f_1)}$, each quartile in $\mathbf{P}_{k_1,k_2,k_3}^{\max}$ occurs at least once in the sum on the left hand side, because it is majorized by at least one quartile in $\mathbf{P}_{k_1}^{\max}(f_1)$. This gives

$$\sum_{P' \in \mathbf{P}_{k_1,k_2,k_3}^{\max}} |I_{P'}| \leq \sum_{P \in \mathbf{P}_{k_1}^{\max}(f_1)} |I_P|$$

This holds also with the sets $\mathbf{P}_{k_2}^{\max}(f_2)$ and $\mathbf{P}_{k_3}^{\max}(f_3)$, on the right hand side. To finish the proof of boundedness of the trilinear form it therefore suffices to show

$$\sum_{k_1,k_2,k_3} 2^{k_1+k_2+k_3} \inf_{i=1,2,3} \sum_{P \in \mathbf{P}_{k_i}^{\max}(f_i)} |I_P| \leq C \|f_1\|_{r_1} \|f_2\|_{r_2} \|f_3\|_{r_3}$$

This follows by a purely algebraic bootstrapping argument using the Carleson measure estimates

$$\sum_k 2^{r_i k} \sum_{P \in \mathbf{P}_{k_i}^{\max}(f_i)} |I_P| \leq C \|f_i\|_{r_i}$$

for $i = 1, 2, 3$ of Lemma 1.12.

At this point it is convenient to pass from discrete sums over the integers to integrals over the real line. Define for $i = 1, 2, 3$ and any real number x:

$$S_i(x) = \sum_{P \in \mathbf{P}^{\max}_{[x/\log 2]}(f_i)} |I_P|$$

where $[x/\log 2]$ denotes the largest integer which does not exceed $x/\log 2$. With this definition the above Carleson measure estimates are equivalent to

$$\int_R e^{r_i x} S_i(x)\, dx \leq C \|f_i\|_{r_i}$$

and the inequality we want to prove is equivalent to

$$\int_{\mathbf{R}^3} e^{x_1+x_2+x_3} \inf_{i=1,2,3} S_i(x_i)\, dx_1 dx_2 dx_3$$

It is now a direct application of the following lemma.

Lemma 2.1 *Let S_i for $i = 1, \ldots, n$ be positive measurable functions on the real line such that*

$$\int_R e^{r_i x} S_i(x)\, dx \leq F_i^{r_i}$$

for some constants r_i , F_i with $1 < r_i < \infty$ and $\sum_i \frac{1}{r_i} = 1$. Then

$$I := \int_{R^n} e^{x_1+\ldots+x_n} \left(\inf_{i=1,\ldots,n} S_i(x_i) \right) dx_1 \ldots dx_n \leq n \prod_{i=1}^{n} F_i$$

Proof: The interesting case is $n \geq 2$. Let r_i' denote the conjugate exponent of r_i, i.e.

$$1/r_i + 1/r_i' = 1.$$

We write each $x \in R^n$ uniquely as

$$x = u + tv + w$$

where t is a real number, the coordinates of u and v are for $i = 1, \dots n$

$$v_i = 1/r_i,$$

$$u_i = \log\left(F_i r_i^{1/r_i}\right),$$

and w is a vector orthogonal to $(1, 1, \dots, 1)$. Then we can change variables as follows

$$\begin{aligned} I &= \frac{1}{\sqrt{n}} \int_{R^{n-1}} \int_R e^{\sum_j u_j + tv_j + w_j} \inf_i \left(S_i(u_i + tv_i + w_i)\right) \, dt \, dw \\ &= \frac{1}{\sqrt{n}} \prod_i F_i r_i^{1/r_i} \int_{R^{n-1}} \int_R e^t \inf_i \left(S_i(u_i + tv_i + w_i)\right) \, dt \, dw \end{aligned} \tag{2.4}$$

where dw is the standard normalized Lebesgue measure on the orthogonal complement of the vector $(1, 1, \dots, 1)$ and the factor $1/\sqrt{n}$ comes from the change of variables.

We estimate the inner integral I_w for a fixed w. Let j be the index such that $r_j w_j$ is maximal. It is clear that w_j is positive.

$$I_w = \int_R e^t \inf_i \left(S_i(u_i + tv_i + w_i)\right) \, dt$$

$$= \int_R \left(e^t \inf_i S_i(u_i + tv_i + w_i)\right)^{1/r_j'} \left(e^t S_j(u_j + tv_j + w_j)\right)^{1/r_j} \, dt$$

$$\leq \left(\int_R e^t \inf_i S_i(u_i + tv_i + w_i) \, dt\right)^{1/r_j'} \left(\int_R e^t S_j(u_j + tv_j + w_j) \, dt\right)^{1/r_j}$$

$$= I_w^{1/r_j'} \left(e^{-u_j - w_j} \left(\int_R e^{r_j(u_j + tv_j + w_j)} \left(S_j(u_j + tv_j + w_j)\right) \, dt\right)^{1/r_j}\right)$$

$$\leq I_w^{1/r_j'} e^{-u_j - w_j} F_j r_j^{1/r_j} = I_w^{1/r_j'} e^{-w_j}$$

Hence

$$I_w \leq e^{-r_j w_j}$$

It is clear now, that the integral (2.1) converges, because w_j is up to a factor of at most n equal to the Euclidean norm of w. The purpose of the rest of this section is to get the right constants. The impatient reader may omit this part of the proof. We have

$$\int_{R^{n-1}} I_w \, dw \leq \int_{R^{n-1}} e^{-r_{j(w)} w_{j(w)}} \, dw = \sum_{j=1}^{n} \int_{C_j} e^{r_j w_j} \, dw$$

where C_j is the cone of those w in the orthogonal complement of $(1, 1, \ldots, 1)$ for which we have picked $j(w) = j$. We consider the case $j = 1$. We cut the cone into $n-2$ dimensional slices with constant height x_1. Then the scaling behaviour of these slices gives a constant C depending only on r_1 and the geometry of the cone, such that

$$\int_{C_1} e^{r_1 w_1} \, dw = C \int_{R+} (r_1 x_1)^{n-2} e^{-r_1 x_1} r_1 \, dx_1 = C(n-2)!$$

To get this constant, we integrate the characteristic function on the truncated cone

$$C \int_0^{1/r_1} (r_1 x_1)^{n-2} r_1 \, dx_j = C \frac{1}{(n-1)}$$

and observe that this integral is the volume of the convex hull of the origin and the vertices

$$(\frac{1}{r_1}, \frac{1}{r_2}, \frac{1}{r_3}, \ldots \frac{1}{r_{i-1}}, \frac{1}{r_i} - 1, \frac{1}{r_{i+1}}, \ldots, \frac{1}{r_n})$$

where i runs from 2 to n. This volume is equal to

$$\frac{1}{(n-1)!} |\det(M)|$$

where M is the matrix

$$\begin{pmatrix} 1/\sqrt{n} & 1/r_1 & 1/r_1 & \dots & 1/r_1 \\ 1/\sqrt{n} & 1/r_2 - 1 & 1/r_2 & \dots & 1/r_2 \\ 1/\sqrt{n} & 1/r_3 & 1/r_3 - 1 & \dots & 1/r_3 \\ \dots & \dots & \dots & \dots & \dots \\ 1/\sqrt{n} & 1/r_n & 1/r_n & \dots & 1/r_n - 1 \end{pmatrix}$$

The first column is a unit vector chosen to be orthogonal to the others, and the rest of the matrix maps the convex hull of the vectors $(0,1,0,\dots 0)$, ..., $(0,0,0,\dots 1)$, which has volume $\frac{1}{(n-1)!}$ to the convex hull which is to be investigated. The matrix acts like minus the identity on all vectors orthogonal to $(1,0,\dots,0)$ and $(0,1,\dots,1)$ and acts like

$$\begin{pmatrix} 1/\sqrt{n} & (n-1)/r_1 \\ 1/\sqrt{n} & -1/r_1 \end{pmatrix}$$

on those two vectors modulo their orthogonal complement. Hence the determinant of the matrix has absolute value $\sqrt{n}/r_1$ and we use this to get

$$\int_{C_1} e^{r_1 w_1}\, dw = \frac{\sqrt{n}}{r_1}$$

Resubstituting everything gives

$$I \leq \prod_{i=1}^{n} F_i r_i^{1/r_i}$$

It is well known that the entropy $\sum_{i=1}^{n} \frac{1}{r_i} \log r_i$ is maximal subject to the constraint $\sum_{i=1}^{n} \frac{1}{r_i} = 1$ if all r_i are equal n. This proves

$$I \leq n \prod_{i=1}^{n} F_i$$

which was claimed in the lemma.

2.4 The decomposition of the bilinear operator

We fix a function $f \in L^s$ with $1 < s < 2$ and view $H_W(f,g)$ as a linear operator in the second argument.

We decompose the operator similarly, but this time we consider only the density of a quartile with respect to the function$f$. We write for simplicity $\mathbf{P}_k := \mathbf{P}_k(f)$, the latter being defined as before. The maximal quartiles in $\mathbf{P}_k$ are again denoted by $\mathbf{P}_k^{\max}$. For each quartile in $\mathbf{P}_k$ we count how many quartiles of $\mathbf{P}_k^{\max}$ are above the quartile, and we define for each nonnegative integer i:

$$\mathbf{P}_{k,i} = \{P \in \mathbf{P}_k : 2^i \leq |\{P' \in \mathbf{P}_k^{\max} : P' \geq P\}| < 2^{i+1}\}$$

The sets $\mathbf{P}_{k,i}$ are called forests.

We define the *forest operator* $T_{\mathbf{P}_{k,i}}$ as

$$T_{\mathbf{P}_{k,i}} g = \sum_{P \in \mathbf{P}_{k,i}} \frac{1}{|I_P|} (w_{a_P}, f)(w_{b_P}, g) w_{c_P}$$

Then the basic decomposition of the quartile operator is

$$H(f,g) = \sum_{k=-\infty}^{0} \sum_{i=0}^{\infty} T_{\mathbf{P}_{k,i}}(g)$$

We will prove in the next Section that the forest operator is bounded on the Hilbert space $L^2(\mathbf{R})$:

$$\|T_{\mathbf{P}_{k,i}} g\|_2 \leq C 2^k \|g\|_2 \tag{2.5}$$

With this estimate, the rest of the proof consists of putting the pieces together in the appropriate way, using a Carleson measure estimate. Fix a large constant $m > s'$, and let k_0 be an integer to be chosen conveniently later. Let A be the set of all forests

$\mathbf{P}_{k,i}$ such that $k < k_0$ and $i < m(k_0 - k)$, and let B be the set of all other forests. This situation is shown in Figure 2.2.

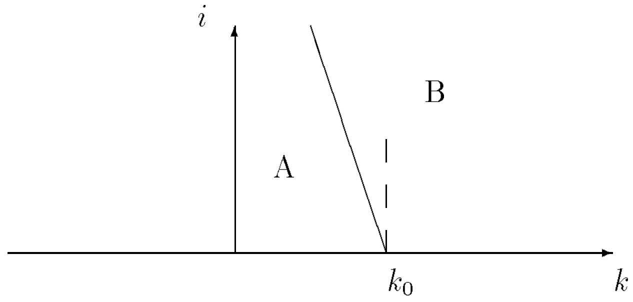

Figure 2.2: Decomposition of the set of forests

For a time x to be contained in the time interval of a quartile of the forest $\mathbf{P}_{k,i}$, it must be contained in the time interval of at least 2^i quartiles of $\mathbf{P}_k^{\max}$. Hence if $\mathbf{P}_{k,i}$ is a forest in B, then the support of $T_{\mathbf{P}_{k,i}} g$ is a subset of

$$E_k = \begin{cases} \{x : N_k(x) \geq 2^{m(k_0-k)}\} & \text{if} \quad k \leq k_0 \\ \{x : N_k(x) \geq 1\} & \text{if} \quad k > k_0 \end{cases}$$

The size of the set E_k is controlled by the Carleson measure estimate of lemma 1.12.

$$|E_k| \leq \begin{cases} C\|f\|_s^s/(2^{sk}2^{sm(k_0-k)(1/s'-\epsilon)}) & \text{if} \quad k \leq k_0 \\ C\|f\|_s^s/(2^{sk}) & \text{if} \quad k > k_0 \end{cases}$$

The constant M is chosen such that these esimates are summable over k. We obtain for $E = \bigcup_{k=-\infty}^{\infty} E_k$ the bound

$$|E| \leq C\frac{\|f\|_s^s}{2^{sk_0}}$$

Outside the set E we only have to consider the forests in A. Hence if $g \in L^2$ we have for the quartile operator outside the set E:

$$\|H_W(f,g)1_{E^c}\|_2 \leq \| \sum_{\mathbf{P}_{k,i} \in A} T_{\mathbf{P}_{k,i}} g\|_2 \leq$$

$$\sum_{k \leq k_0} C\|g\|_2 2^k m(k_0 - k) \leq C\|g\|_2 2^{k_0}$$

If $g \in L^t$ with $t > 2$ we observe again that the time intervals of all quartiles of the forest $\mathbf{P}_{k,i}$ are contained in the set $\{x : N_k(x) \geq 1\}$. Applying the forest operator to g gives the same as applying it to the restriction of g to this set. For this restriction we have, using the Carleson measure estimate

$$\|g 1_{\{x:N_k(x)\geq 1\}}\|_2 \leq \|g\|_t |\{x : N_k(x) \geq 1\}|^{1/2-1/t} \leq$$

$$\|g\|_t \|f\|_s^{s/2-s/t} / (2^{k(s/2-s/t)})$$

This gives for $t \geq 2$ the weak type estimate

$$|\{H_W(f,g)(x) > \lambda\}| \leq \frac{\|H_W(f,g)1_{E^c}\|_2^2}{\lambda^2} + |E| \leq$$

$$C \frac{\|g\|_t^2 \|f\|_s^{s(1-2/t)} 2^{-sk_0(1-2/t)} 2^{2k_0}}{\lambda^2} + C\frac{\|f\|_s^s}{2^{sk_0}}$$

Now for given λ we pick k_0 to be the closest integer to

$$\log_2 \left(\lambda^{t/(s+t)} \|f\|_s^{s/(s+t)} \|g\|_t^{-t/(s+t)} \right)$$

such that we obtain the desired weak type estimate

$$|\{T(f,g)(x) > \lambda\}| \leq C \frac{(\|f\|_s \|g\|_t)^{\frac{ts}{t+s}}}{\lambda^{\frac{ts}{t+s}}}$$

2.5 The forest estimate

We decompose the forests further. For each quartile Q in $\mathbf{P}_{k,i}^{\max}$ we consider the set of all quartiles in $\mathbf{P}_{k,i}$ which are less than Q and call this set the *tree* with *tree top* Q.

Clearly the union of the trees in the forest gives back the forest. It is a beautiful observation of Fefferman in [4] that two such trees are disjoint. For a proof of this fact assume to the contrary that a quartile $P \in \mathbf{P}_{k,i}$ lies below two maximal quartiles Q and Q'. This implies that Q and Q' have disjoint frequency intervals. By definition of $\mathbf{P}_{k,i}^{\max}$ the quartiles Q and Q' are less than at least 2^i maximal quartiles in $\mathbf{P}_k$, and by the disjointness of frequency intervals these two collections of maximal quartiles are disjoint. Hence P is less than at least 2^{k+1} maximal quartiles in $\mathbf{Q}_k$, a contradiction to $P \in \mathbf{P}_{k,i}$.

We conclude that whenever two quartiles in a forest are comparable, they are both comparable to a common maximal quartile, and therefore they are in the same tree. This means that quartiles from different trees are incomparable. This implies that the *tree operators*

$$\sum_{P \in \mathbf{P}_{k,i}, P \leq Q} \frac{1}{|I_P|}(w_{a_P}, f)(w_{b_P}, g) w_{c_P}$$

are mutually orthogonal as Q runs through $\mathbf{P}_{k,i}^{\max}$. Therefore it suffices to prove the desired L^2 estimate for a single tree operator.

We decompose as before the tree $\mathbf{Q}$ into four subtrees $\mathbf{Q}_a$, $\mathbf{Q}_b$, $\mathbf{Q}_c$, and $\mathbf{Q}_d$ by fixing a frequency in the frequency interval of each tree top. The bound for the trees $\mathbf{Q}_a$ and $\mathbf{Q}_d$ can be done in the same way as we did the tree estimate for the trilinear form. The adjoint of the tree operator of $\mathbf{Q}_c$ is essentially of the same form as the tree operator of $\mathbf{Q}_b$, hence by analogy we are done once we have shown the bound for the tree operator of $\mathbf{Q}_b$

As in the proof of the tree estimate for the trilinear form we can replace f by its restriction to the area $A_{\mathbf{Q}}$ of the tree, and this projection $\Pi_{A_{\mathbf{Q}}} f$ is pointwise bounded by $C2^k$.

We calculate, using that the tiles c_P run over a disjoint collection of tiles

$$\|\sum_{P\in\mathbf{Q}_b} \frac{1}{\sqrt{|I_P|}}(w_{a_P}, f)(w_{b_P}, g)w_{c_P}\|_2$$

$$= \sqrt{\sum_{P\in\mathbf{Q}_b} \frac{1}{|I_P|}|(w_{a_P}, f)|^2|(w_{b_P}, g)|^2}$$

Now we view $\mathbf{Q}_b$ as a measure space, where each quartile has the measure

$$\mu(\{P\}) = |(w_{a_P}, f)|^2$$

Then the previous sum can be written as

$$= \sqrt{\int_0^\infty \mu\left(\left\{P : \frac{1}{|I_P|}|(w_{b_P}, g)|^2 \geq \lambda\right\}\right) d\lambda}$$

We estimate the measure of the set inside the integral. The time intervals of the quartiles occuring in the set are obviously contained in the set

$$E = \{x : Mg(x)^2 \geq \lambda\}$$

where M is the maximal function. The tiles a_P are pairwise disjoint, hence the measure of the above set by the Plancherel theorem is bounded by the square of the

L^2-norm of the function $\Pi_{A_{\mathbf{Q}}} f$, restricted to the set E. This norm is bounded by $C2^k\sqrt{|E|}$. We continue the calculation, using the maximal theorem, to get

$$\surd \leq \sqrt{\int_0^\infty C2^{2k}|\{x : Mg(x)^2 \geq \lambda\}|d\lambda}$$

$$= C2^k\|Mg\|_2 \leq C2^k\|g\|_2$$

This proves the bound for the forest operator.

2.6 The connection between the quartile operator and the bilinear Hilbert transform

A formal calculation using the Fourier transform gives for the bilinear Hilbert transform

$$H(f,g)(x) = \int_R \int_R \hat{f}(\eta)\hat{g}(\xi)e^{i(\eta+\xi)x}(-i)sgn(\eta-\xi)\, d\eta\, d\xi$$

Consider the η, ξ plane and a Whitney decomposition into squares of the area below the diagonal $\eta = \xi$. as in Figure 2.3.

This Whitney decomposition is chosen such that for each dyadic interval $\omega = [2^j n, 2^j(n+1))$ there is a square $\omega \times \omega'$ in the decomposition with $\omega' = [2^j(n-2), 2^j(n-1))$. We fix the notation that ω' denotes the dyadic interval which is two steps below ω. Let Ω be the set of dyadic intervals of the form

$$\left[2^j(n+\frac{2}{4}), 2^j(n+\frac{3}{4})\right)$$

with integers j and n, i.e. those intervals which occur as frequency intervals of third

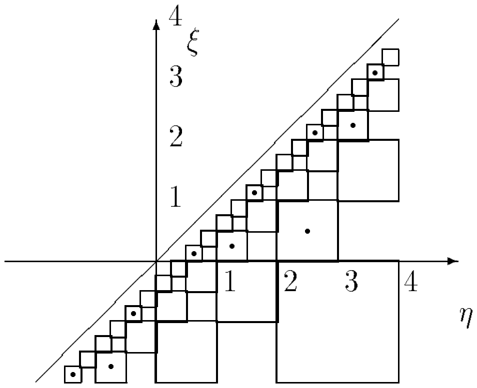

Figure 2.3: Whitney decomposition of the lower triangle

tiles c_P of quartiles P. The squares $\omega \times \omega'$ with $\omega \in \Omega$ are marked with a dot in Figure 2.3.

We consider the following modified bilinear operator, where 1_ω denotes the characteristic function of the interval ω and $\check{1_\omega}$ its inverse Fourier transform.

$$\sum_{\omega\in\Omega} \int_R \int_R \hat{f}(\eta)\hat{g}(\xi)e^{i(\eta+\xi)x} 1_\omega(\eta) 1_{\omega'}(\xi)\, d\eta\, d\xi = \sum_{\omega\in\Omega} (f * \check{1_\omega})(x)(g * \check{1_{\omega'}})(x)$$

This is to be understood in a formal and heuristic sense. If we needed to avoid difficulties with the fact that $\check{1_\omega}$ is not integrable and therefore the convolutions on the right hand side are troublesome, we would choose smooth cutoff functions instead of 1_ω.

To recover the bilinear Hilbert transform from this expression, we have to take a linear combination with the corresponding sum for the area above the diagonal and to average over translated and dilated dyadic grids in order to get rid of the gaps we have left over in the Whitney decomposition. Therefore boundedness properties of the modified operator will translate into boundedness properties of the bilinear

Hilbert transform.

The expression $f * \check{1}_\omega$ can be understood as taking the frequency band ω of the function f. In the Walsh model this corresponds to

$$\sum_{I:\text{dyadic}, \, |I||\omega|=1} (f, w_{I\times\omega}) w_{I\times\omega}$$

In the continuous case, the frequency band ω of f is multiplied to the frequency band ω' of g to get a function with frequency band $\omega+\omega'$. This situation is shown in Figure 2.4. In contrast to that the product

$$\sum_I (f, w_{I\times\omega}) w_{I\times\omega} \sum_I (g, w_{I\times\omega'}) w_{I\times\omega'} =$$

$$\sum_I \frac{1}{\sqrt{|I|}} (f, w_{I\times\omega})(g, w_{I\times\omega'}) w_{I\times\{|\omega|,2|\omega|\}}$$

would give a frequency band which is close to the bottom of the half plane. In order to model the cancellation properties of the bilinear Hilbert transform, one has to shift this frequency band up. We shift to get the frequency band of the continuous case scaled by the factor $1/2$, such that it is the band right in between ω and ω'.

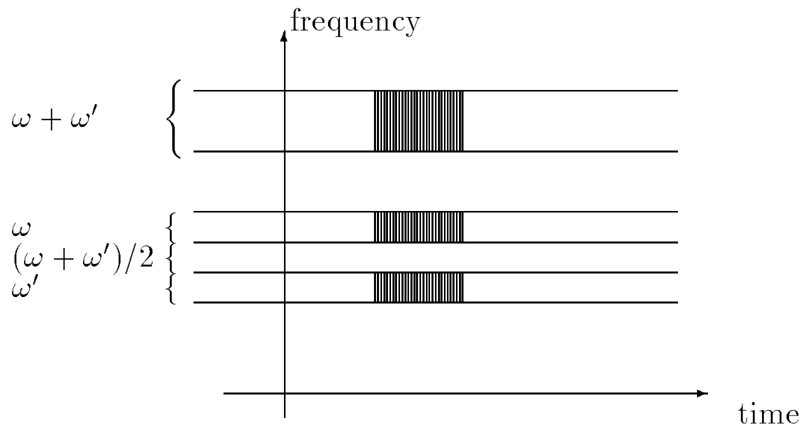

Figure 2.4: Time-frequency diagram for $(f * \check{1}_\omega)(x)(g * \check{1}_{\omega'})$

Then each term

$$\frac{1}{\sqrt{|I|}}(f, w_{I\times\omega})(g, w_{I\times\omega'})w_{I\times(\omega+\omega')/2}$$

equals a summand of the quartile operator except for a permutation of the three frequency bands, which in view of the symmetry of the corresponding trilinear form is inessential to the theory. The double sum over all $\omega \in \Omega$ and all I with $|I||\omega| = 1$ can be indexed as a sum over all quartiles, hence we have arrived at the quartile operator.

Chapter 3

A.E. Convergence of Walsh Fourier series

3.1 Introduction

The method to prove Theorem 1.1 is to prove the boundedness of the maximal partial sum operator

$$S^{\max} f(x) = \sup_n \left| \sum_{k=0}^{n} (f, W_k) W_k(x) \right|$$

We will prove the weak type estimate

$$|\{x : S^{\max} f(x) > \lambda\}| \leq C \frac{\|f\|_r^r}{\lambda^r}$$

for $1 < r \leq 2$. By standard arguments this proves Theorem 1.1.

Fix a very large number N and fix a function $n : [0, 1) \mapsto \mathbf{N}_0$ that is bounded by 2^N and constant on each dyadic interval of length 2^{-N}. The linearized maximal partial sum operator is

$$S^n(x) = \left| \sum_{k=0}^{n(x)} (f, W_k) W_k(x) \right|$$

To prove the above weak type estimate it suffices to show

$$|\{x : S^n f(x) > \lambda\}| \leq C \frac{\|f\|_r^r}{\lambda^r}$$

with a constant C independent of the number N and the function n.

We decompose the linearized maximal partial sum operator into a sum of operators associated to *bitiles*. Throughout this chapter $\mathbf{P}$ will denote the set of bitiles. Recall that each bitile P is the union of a lower son l_P and an upper son u_P. Define for each bitile P the operator T_P by

$$T_P f(x) = \begin{cases} (f, w_{l_P}) w_{l_P}(x) & \text{if} \quad (x, n(x)) \in u_P \\ 0 & \text{if} \quad (x, n(x)) \notin u_P \end{cases}$$

We claim that

$$S^n f(x) = \sum_{P \in \mathbf{P}} T_P f(x)$$

Observe that due to our special choice of the function n there are only finitely many summands on the right hand side of the equation which are nonzero.

Fix an $x_0 \in [0, 1)$. Then the rectangle $R = [0, 1) \times [0, n(x_0))$ defines a projection Π_R. Decomposing the rectangle into tiles of the form $[0, 1) \times [k, k + 1)$ with corresponding wave packets W_k, we see that $S^n f(x_0) = \Pi_R f(x_0)$.

Now we decompose the rectangle R into another collection of pairwise disjoint tiles. For each point (x, ξ) in the rectangle there is a unique bitile whose lower son contains (x, ξ) and whose upper son contains $(x, n(x_0))$. Indeed, the frequency interval of this bitile is the minimal dyadic interval which contains ξ and $n(x_0)$, and this determines the length of the time interval of the bitile, which is therefore unique. The rectangle is therefore the disjoint union of the lower sons of all bitiles whose upper son intersects the set $[0,1) \times \{n(x_0)\}$. We use this decomposition to calculate $\Pi_R f(x_0)$. In doing this, we can remove all tiles from the collection, whose time interval does not contain x_0, or equivalently whose upper son does not contain the point $(x_0, n(x_0))$. But what remains is clearly equal to $\sum_{P\in\mathbf{P}} T_P f(x_0)$.

3.2 Proof of Theorem 1.1

We organize the set of bitiles into a hierarchy of subsets, as we have done in Section 2.4 for the set of quartiles. For any subset $\mathbf{Q} \subset \mathbf{P}$ we write $T_{\mathbf{Q}}$ for the operator $\sum_{P\in\mathbf{Q}} T_P$.

Fix a function $f \in L^r$ with $1 < r \leq 2$. With the partial order of the set of birtiles,

$$P \leq P' \text{ if } \quad P \cap P' \neq \emptyset \text{ and } I_P \subset I_{P'}$$

we define the density of a bitile P with respect to the function f as

$$\delta(P, f) = \sup_{P' \geq P} \|\Pi_{P'} f\|_\infty$$

The sets of homogeneous density are

$$\mathbf{P}_k = \mathbf{P}_k(f) = \{P \in \mathbf{P} : 2^k \leq \delta(P, f) < 2^{k+1}\},$$

and the forests are

$$\mathbf{P}_{k,i} = \{P \in \mathbf{P}_k : 2^i \leq |\{P' \in \mathbf{P}_k^{\max} : P' \geq P\}| < 2^{i+1}\}$$

In the next section we will prove for $1 < s < \infty$ the following estimate on the forest operator

$$\|T_{\mathbf{P}_{k,i}} f\|_s \leq C 2^k \left(\sum_{P \in \mathbf{P}_{k,i}^{\max}} |I_P| \right)^{1/s}$$

for $1 < s < \infty$.

With that we simply proceed as in Section 2.4 and split the set of forests into the set A of all forests $\mathbf{P}_{k,i}$ such that $k < k_0$ and $i < m(k_0 - k)$ and the set B of all other forests, with M large enough and k_0 to be chosen. The forests in B are supported in the small set $E = \bigcup_k E_k$, which was defined in Section 2.4 by If

$$E_k = \begin{cases} \{x : N_k(x) \geq 2^{m(k_0 - k)}\} & \text{if} \quad k \leq k_0 \\ \{x : N_k(x) \geq 1\} & \text{if} \quad k > k_0 \end{cases}$$

and which is bounded in measure by

$$|E| \leq C \frac{\|f\|_r^r}{2^{r k_0}} \ .$$

$\mathbf{P}_{k,i}$ is a forest in the set A, then we can remove all bitiles of this forest whose time interval is contained in the set E, because we will disregard this set anyway. Then we have the following Carleson measure estimate

$$\begin{aligned} \sum_{P \in P_{k,i}^{\max} \text{ s.t } I_P \not\subset E} |I_P| &\leq \int_{\mathbf{R}} \min(N_k(x), 2^i)\, dx \\ &\leq \int_{\mathbf{R}} (N_k(x))^{\frac{r}{r'+\epsilon}} \left(2^{m(k_0-k)}\right)^{1-\frac{r}{r'+\epsilon}} \, dx \\ &\leq C \frac{||f||_r^r}{2^{rk}} 2^{m(k_0-k)(1-\frac{r}{r'+\epsilon})} \end{aligned}$$

We substitute this into the forest estimate and sum over all forests in the set A to get

$$\|T_{\mathbf{P}} f 1_{E^c}\|_s \leq C \sum_{k=-\infty}^{k_0} m(k_0 - k) 2^k \left(\frac{||f||_r^r}{2^{rk}} 2^{m(k_0-k)(1-\frac{r}{r'+\epsilon})} \right)^{1/s}$$

which is for large s summable to

$$\|T_{\mathbf{P}} f 1_{E^c}\|_s \leq C 2^{k_0} \left(\frac{||f||_r^r}{2^{rk_0}} \right)^{1/s}$$

Now we pick $\lambda = 2^{k_0}$, and we obtain the desired weak type estimate:

$$|\{T_{\mathbf{P}} f(x) > \lambda\}| \leq C \frac{||T_{\mathbf{P}}(f)||_{s,E^c}^s}{\lambda^s} + |E| \leq C \frac{||f||_r^r}{\lambda^r}$$

3.3 The forest estimate

As in Section 2.5, each forest $\mathbf{P}_{k,i}$ is the union of convex trees such that two different trees cover subsets of the Walsh phase plane which are disjoint. If $\mathbf{Q}$ is a tree, then the support of the function $T_{\mathbf{Q}} f$ is contained in the set of all x such that $(x, n(x))$ is contained in $\bigcup_{Q \in \mathbf{Q}} Q$. Hence for two different trees $\mathbf{Q}$ and $\mathbf{Q}'$ in the forest $\mathbf{P}_{k,i}$ we conclude that $T_{\mathbf{Q}} f$ and $T_{\mathbf{Q}'} f$ are supported on disjoint sets. It therefore suffices to show for $1 < s < \infty$

$$\|T_{\mathbf{Q}_{k,i}} f\|_s \leq C 2^k |I_Q|$$

for a convex tree $\mathbf{Q} \subset \mathbf{P}_{k,i}$ with tree top Q.

Let the frequency interval of the tree top be $[\xi, \xi')$. We split the tree $\mathbf{Q}$ into two subtrees $\mathbf{Q}_l$ and $\mathbf{Q}_u$, the first one containing all bitiles of the tree such that the frequency interval of the lower son contains ξ, and the second one containing all bitiles of the tree such that the frequency interval of the upper son contains ξ. It suffices to show the estimate for both subtrees separately.

Consider first $\mathbf{Q}_l$. Then the upper sons of the bitiles in $\mathbf{Q}_l$ are pairwise disjoint by Lemma 1.3 . That means for a given x there is at most one bitile in $\mathbf{Q}_l$ whose upper son contains $(x, n(x))$. Hence $T_{\mathbf{Q}_l} f(x)$ is given by $\Pi_{I_P} f(x)$, which is bounded by 2^k because the projection $\Pi_P f(x)$ is bounded by 2^k, by definition of $\mathbf{P}_{k,i}$.

Hence $T_{\mathbf{Q}_l} f(x)$ is pointwise bounded by 2^k and is supported on the time interval of the tree top. This proves the desired estimate in this case.

Now consider the subtree $\mathbf{Q}_u$. Let $I = I_Q$ be the time interval of the tree top Q of $\mathbf{Q}$. All frequency intervals we will mention in this paragraph have the property that their Cartesian product with I is a rectangle which defines a projection. We will show later that for each x there is an interval $\omega(x)$, which is not necessarily dyadic, such that

$$T_{\mathbf{Q}_u} f(x) = \Pi_{I \times \omega(x)} f(x)$$

Moreover we will show that this interval can be written as

$$\omega(x) = (\omega_1(x) \setminus \omega_2(x)) \cap \omega \tag{3.1}$$

where $\omega_1(x)$ and $\omega_2(x)$ are dyadic intervals with $\omega_2(x) \subset \omega_1(x)$ and $\omega = [0, \xi)$ with ξ as above. By slicing everything into tiles with time interval I, we see that this gives the operator identity

$$\Pi_{I \times \omega(x)} = \left(\Pi_{I \times \omega_1(x)} - \Pi_{I \times \omega_2(x)} \right) \Pi_{I \times \omega}$$

The operator $\Pi_{I \times \omega_1(x)}$ is pointwise bounded in absolute value by the maximal operator. This one can see from slicing the rectangle $I \times \omega_1(x)$ into minimal tiles, i.e., those which have frequency interval $\omega_1(x)$. Then at each point the projection operator coincides with the projection onto a single tile, which clearly is bounded by the maximal operator. The same is true for $\Pi_{I\times\omega_2(x)}$. This gives

$$|T_{\mathbf{Q}_u}f(x)| \leq \left|\Pi_{I\times\omega(x)}f(x)\right| \leq 2M\Pi_{I\times\omega}f(x)$$

The operator $\Pi_{I\times\omega}$ is nothing but the ordinary partial sum operator

$$\sum_{k=0}^{n}(f, W_k)W_k$$

with a certain n, rescaled to the interval I. This one can see from slicing the rectangle into maximal tiles, i.e, tiles with time interval I. By Paley's theorem [11] this operator is bounded in L^r for $1 < r < \infty$ with a bound not depending on n. This and the maximal theorem applied to the previous inequality gives the desired estimate for the tree operator.

It remains to show the existence of the frequency interval $\omega(x)$ with the stated properties. Pick $\omega_1(x)$ to be the frequency interval of the minimal bitile in $\mathbf{Q}_u$ which contains the point $(x, n(x))$, and let $\omega_2(x)$ be the frequency interval of the upper son of the maximal bitile in $\mathbf{Q}_u$ which contains the point $(x, n(x))$. If no bitile in $\mathbf{Q}_u$ contains this point, there is nothing to prove because then the operator vanishes at x. By Equation 3.1 this determines the interval $\omega(x)$.

Let P be any bitile in $\mathbf{Q}_u$ which contributes to $T_{Q_u}f(x)$, i.e., whose upper son contains $(x, n(x))$. We show that the lower son l_P is contained in the rectangle $I \times \omega$. The time interval of l_P is certainly contained in the time interval I of the tree top. As to the frequency interval of l_P, we observe that it lies below the frequency ξ by definition of $\mathbf{Q}_u$, and hence is contained in ω. Furthermore it is contained in $\omega_1(x)$, because of the extremality of the latter. But it is disjoint from $\omega_2(x)$, because by extremality the interval $\omega_2(x)$ must be contained in the frequency interval of u_P, which is certainly disjoint from the frequency interval of l_P.

By Lemma 1.3 the lower sons of the bitiles in $\mathbf{Q}_u$ whose upper sons contain $(x, n(x))$ are disjoint. To show that the projection onto the union of these tiles coincides at the point x with the projection onto the rectangle $I \times \omega_x$, it remains to show that those tiles cover each point of the line $\{x\} \times \omega(x)$,bbecause then the difference between the two projections is supported outside x.

So let (x, ξ'') be a point on this line. Then there is a unique bitile whose upper son contains $(x, n(x))$ and whose lower son contains (x, ξ''). By reversing the arguments above we see that it is squeezed between two bitiles of $\mathbf{Q}_u$ w.r.t. the partial order of bitiles. By convexity of $\mathbf{Q}$ the bitile is in $\mathbf{Q}$. Moreover the frequency interval of its upper son contains $\omega_2(x)$ and hence contains the point (x, ξ). Hence the bitile is in $\mathbf{Q}_u$.

Part II

A Fast Algorithm for Adapted Time-Frequency Tilings

Chapter 4

A Fast Algorithm for Adapted Time-Frequency Tilings

We first consider orthonormal bases of $\mathbf{R}^N$ consisting of discretized rescaled Walsh functions, where N is a power of two. Given a vector, the best basis with respect to an additive cost function is found with an algorithm of order $O(N \log N)$. The algorithm operates in the time-frequency plane by constructing a tiling of minimal cost among all possible tilings with dyadic rectangles of area one. Then we discuss generalizations replacing the Walsh group, which controls the structure of the time-frequency plane, by other finite abelian groups. The main example here involves the Fast Fourier Transform.

[0]This chapter is taken from [14], which is joint work with Lars Villemoes.

4.1 Introduction

The goal of time-frequency analysis is to decompose a signal into elementary waveforms, each having a well defined *position, duration*, and *frequency.* For applications like data compression or noise reduction, the decomposition should be realizable by a fast, stable, and adaptive algorithm. The recent best basis algorithm of Coifman and Wickerhauser [12] achieves this goal by starting from a large library of orthonormal bases, containing one basis for each dyadic segmentation of a fixed interval. This interval is the time axis when the library consists of local trigonometric bases and it can be thought of as the frequency axis when the library consists of wavelet packet bases. Given a signal, the best basis is the one which minimizes an additive information cost function defined on the set of expansion coefficients. It can be found by a binary tree search, requiring only $O(N)$ operations for a signal of length N. The whole algorithm for expanding a signal in the best basis is of order $O(N \log N)$ for the wavelet packet library and $O(N(\log N)^2)$ for the local trigonometric library.

Each basis element from these two libraries of bases can be identified with a rectangle of area one in the time-frequency plane, such that orthonormal bases correspond to disjoint coverings, or *tilings* of this plane. This method is applied in the software tool [16] to visualize the time-frequency decomposition of a signal. From that point of view, the local trigonometric bases correspond to tilings that can be obtained by first choosing a dyadic partition of the time axis, thereby dividing the time-frequency plane into vertical sectors, and then tiling each sector with rectangles of maximal time width and area one. For wavelet packet bases, the role of the time and frequency axes are interchanged.

In the double tree algorithm proposed by Herley, Kovačević, Ramchandran and Vetterli [5], a time segmentation is followed by local frequency segmentations. The elementary waveforms are wavelet packets on each time interval, constructed as the interval wavelets of Cohen, Daubechies, and Vial [1]. The number of bases in the double tree library is substantially larger than for global wavelet packets, but there are still many dyadic tilings which do not correspond to orthonormal bases.

On the other hand, for the special wavelet packets defined from the Haar filter, any tiling corresponds to an orthonormal basis. These wavelet packets are just rescaled versions of Walsh functions. The first purpose of this paper is to present a fast algorithm for selecting the best basis among *all* these bases, given a signal and an additive cost function. Instead of relying on dyadic interval segmentations, the algorithm will operate directly in the time-frequency plane.

Of course, since the frequency resolution of Walsh functions is very poor the algorithm should be viewed only as a first step towards a new class of time-frequency decompositions, based on a more symmetric treatment of the time and frequency parameters. A better frequency resolution would require some nontrivial smoothing procedure.

The fast algorithm for selecting the best basis as described in the first part of this paper applies to more general bases than just the one consisting of Walsh functions. Those bases are described by successively applying orthonormal transformations on pairs of basis vectors. However in the case of Walsh functions there is a global understanding of the basis vectors which comes from the fact that the Walsh functions are characters of a group.

In the second part of this paper we specialize to this point of view and introduce libraries of bases attached to other groups, for which there is an analogue algorithm to pick the minimal cost basis. The basis elements in these libraries correspond to subsets of the phase plane hat are no longer rectangles, but spread out over the plane. Therefore they are not localized in time and frequency in the usual sense, so we call them algebraically localized.

However, for the case of a finite cyclic group, the basis elements have a very natural meaning, they are implicitely used in the Fast Fourier Transform. That means in the intermediate steps of the Fast Fourier Transform one calculates the expansions of the signal into all bases of the library. Especially the pure frequency basis is contained in the library. Generally the basis elements are localized both in time and frequency at sampling points with certain sampling rates.

Rather than trying to formulate the algorithm of this paper in most generality, we decided to present it for the cases we find most important and interesting. That is for Walsh functions, here the algorithm was discovered by the second author [1], and for bases attached to finite groups, this application was pointed out by the first author. The library of orthonormal bases of rescaled Walsh functions is defined in Section 4.2 and the selection algorithm is described in Section 4.3. To illustrate the algorithm, a small numerical example is worked out in Section 4.4. The algebraic point of view

[1] L. Villemoes

is developed in Section 4.5, and illustrated by an example related to the Fast Fourier Transform in Section 4.6.

4.2 Walsh tilings

Let $W_0(t) = 1$ for $0 \leq t < 1$ and zero elsewhere, and define $W_1, W_2, \ldots$ recursively by

$$\begin{aligned} W_{2n}(t) &= W_n(2t) + (-1)^n W_n(2t-1), \\ W_{2n+1}(t) &= W_n(2t) - (-1)^n W_n(2t-1). \end{aligned} \tag{4.1}$$

Then $\{W_n\}_{n=0}^{\infty}$ is the Walsh system in sequency order, as constructed in [15]. It is an orthonormal basis for $L^2(0,1)$ and each basis function is piecewise equal to either 1 or -1 on $[0,1[$. The number of sign changes for W_n on this interval is equal to n as it can easily be seen by induction from (4.1). Therefore, we will think of n as a frequency parameter, although the actual frequency localization is very poor. The natural (Paley) ordering is obtained by omitting the factors $(-1)^n$ in (4.1).

We will build bases consisting of rescaled versions of the Walsh functions. Considering $S = [0,1[\times[0,\infty[$ as a time-frequency plane, we will define a *dyadic rectangle* to be a rectangle with dyadic sides, that is, a subset of S of the form

$$I \times \omega = [2^{-j}k, 2^{-j}(k+1)[\times[2^{j'}n, 2^{j'}(n+1)[,$$

where j, j', n, k are nonnegative integers and $k < 2^j$. Let $\mathcal{P}$ be the collection dyadic rectangles of area one, $(j = j')$. These special rectangles will be called *tiles*. To each tile $p \in \mathcal{P}$ we assign the corresponding rescaled Walsh function

$$w_p(t) = 2^{j/2} W_n(2^j t - k). \tag{4.2}$$

Thus w_p is supported on I where it has n sign changes. It is therefore natural to think of p as a Heisenberg box with time interval I and frequency interval ω. We will see that disjoint tiles lead to orthogonal functions w_p. For the moment, just observe that this is true for tiles with disjoint time intervals.

A dyadic rectangle of area two in S can either be split in a left and a right tile $\{l, r\}$ or in a lower and upper tile $\{d, u\}$. All orthogonality properties of rescaled Walsh functions can then be derived from the following simple observation.

Lemma 4.1 *Assume $T = I \times \omega$ be a dyadic rectangle of area two. Define the four tiles $\{l, r, d, u\} \subset \mathcal{P}$ by $T = l \cup r = d \cup u$, $\omega(l) = \omega(r)$, and $I(d) = I(u)$. Then $\{w_l, w_r\}$ and $\{w_d, w_u\}$ are orthonormal bases of the same subspace of $L^2(0,1)$.*

Proof. We can assume that

$$I \times \omega = [2^{-j}k, 2^{-j}(k+1)[\times[2^{j+1}n, 2^{j+1}(n+1)[$$

so that $I(l) = [2^{-j-1}(2k), 2^{-j-1}(2k+1)[$, $I(r) = [2^{-j-1}(2k+1), 2^{-j-1}(2k+2)[$, $\omega(d) = [2^j(2n), 2^j(2n+1)[$, and $\omega(u) = [2^j(2n+1), 2^j(2n+2)[$. From (4.1) it follows that

$$\begin{aligned} 2^{j/2}W_{2n}(2^jt-k) &= 2^{j/2}W_n(2^{j+1}t-2k) + (-1)^n 2^{j/2}W_n(2^{j+1}t-2k-1), \\ 2^{j/2}W_{2n+1}(2^jt-k) &= 2^{j/2}W_n(2^{j+1}t-2k) - (-1)^n 2^{j/2}W_n(2^{j+1}t-2k-1). \end{aligned}$$

Therefore, by (4.2),

$$\begin{pmatrix} w_d \\ w_u \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & (-1)^n \\ 1 & -(-1)^n \end{pmatrix} \begin{pmatrix} w_l \\ w_r \end{pmatrix} \tag{4.3}$$

showing that $\{w_d, w_u\}$ is obtained from $\{w_l, w_r\}$ by an orthogonal transformation. $\square$

Remark The matrix in (4.3) could actually be replaced by any unitary matrix, even depending on the dyadic rectangle T. The algorithm of Section 4.3 would still work. In fact, Lemma 4.1 can be seen as a definition of the kind of libraries of orthonormal bases for which Algorithm 4.7 can be used directly. Obvious variations are possible like changing from dyadic to triadic rectangles, (letting the matrix in (4.3) be of size 3×3), or starting from another basis for $\mathbf{R}^N$ than the standard basis which we use in Section 4.3. Bases that have multiplicative structures like the Walsh functions are discussed in Section 4.5.

The next result is obtained by induction from Lemma 4.1. It states that dyadic rectangles can be identified with subspaces of $L^2(0,1)$.

Corollary 4.2 *Let $T = I \times \omega$ be a dyadic rectangle of area $|T| = |I||\omega| \geq 1$, and assume that $\mathcal{B}$ and $\mathcal{B}'$ are two collections of tiles, both defining disjoint coverings of T. Then $\{w_p \mid p \in \mathcal{B}\}$ and $\{w_p \mid p \in \mathcal{B}'\}$ are orthonormal bases of the same subspace of $L^2(0,1)$.*

Proof. Put $V = \text{span}\{w_p \mid p \in \mathcal{B}\}$ and $L_0 = \max\{I(p) \mid p \in \mathcal{B}\}$. Then the set of tiles in $\mathcal{B}$ with $|I(p)| = L_0$ consists of pairs of the form $\{d, u\}$ as in Lemma 4.1. An application of this lemma to all those pairs changes $\mathcal{B}$ to a collection $\mathcal{B}_\infty$ with maximal time interval length $L_1 = L_0/2$ and such that the rescaled Walsh functions corresponding to $\mathcal{B}_\infty$ also define an orthonormal basis of V.

We can repeat this process until all tiles have time intervals of length $1/|\omega|$. This condition is only satisfied by one tiling $\mathcal{L} \subset \mathcal{P}$ of T. Therefore the same modification of $\mathcal{B}'$ must also lead to $\mathcal{L}$. Hence $\{w_p \mid p \in \mathcal{B}\}$, $\{w_p \mid p \in \mathcal{B}'\}$, and $\{w_p \mid p \in \mathcal{L}\}$ are orthonormal bases of the same subspace of $L^2(0,1)$. $\square$

Corollary 4.3 *The functions w_p and w_q are orthogonal if and only if the tiles p and q are disjoint.*

Proof. Assume p and q are disjoint tiles. If they have disjoint time intervals $I(p) \cap I(q) = \emptyset$ then w_p and w_q are trivially orthogonal. If $I(p)$ and $I(q)$ are not disjoint one of the intervals must be contained in the other, because they are dyadic. Let us assume that $I(p) \subset I(q)$, and let ω be a dyadic frequency interval such that $\omega(p), \omega(q) \subset \omega$. Then it is not hard to construct a tiling of $T = I(q) \times \omega$ containing the tiles p and q. Therefore, by Corollary 4.2, w_p and w_q are different members of an orthonormal set.

Conversely, suppose q intersects p. We can assume that $I(p) \subset I(q)$ and hence $\omega(q) \subset \omega(p)$. The rectangle $I(q) \times \omega(p)$ can now be covered by disjoint tiles congruent with p or by disjoint tiles congruent with q. By Corollary 4.2 the two corresponding sets of rescaled Walsh functions are orthonormal bases for the same subspace V of $L^2(0,1)$, and we see that V contains both w_p and w_q. Expanding w_q in the basis defined from tiles congruent with p it follows that $w_q \perp w_p$ implies $w_q = 0$ on $I(p)$, contradicting the fact that the support of w_q is all of $I(q)$. $\square$

The above results about finite systems of rescaled Walsh functions are sufficient for understanding the best basis algorithm of Section 4.3, and the next theorem concerning infinite systems is only included for aesthetic reasons. Note that Theorem 4.4 contains the Haar and Walsh bases as special cases.

Theorem 4.4 *A collection* $\{w_p \mid p \in \mathcal{B}\}$ *is an orthonormal basis of* $L^2(0,1)$ *if and only if* $\mathcal{B} \subset \mathcal{P}$ *is a disjoint covering of* S *except for a set of measure zero.*

Proof. **1** Let us first note that the complement of the union of a pairwise disjoint set of tiles $\mathcal{B} \subset \mathcal{P}$ is a disjoint union of tiles and a set of the form $E \times [0, \infty[$. To see this, let $S' = \cup\mathcal{B}$ and pick a point in the complement $(t, \xi) \in S \setminus S'$. Now either $(\{t\} \times [0, \infty[) \cap S' = \emptyset$ or there is a largest dyadic frequency interval $\omega(\xi) \subset [0, \infty[$ containing ξ and such that $(\{t\} \times \omega(\xi)) \cap S' = \emptyset$. In the last case, consider the tile $q = I(t) \times \omega(\xi)$ where $I(t)$ is the dyadic time interval of length $|\omega(\xi)|^{-1}$ containing t. Let $\omega'(\xi)$ be the neighbor of $\omega(\xi)$ in the sense that the union of these two intervals is dyadic. By construction, we know that the line segment $\{t\} \times \omega'(\xi)$ must be intersected by some tile $p \in \mathcal{B}$, and we see that this tile must have $I(p) \supset I(t)$. Now it is clear that any tile intersecting q must intersect either p or $\{t\} \times \omega(\xi)$. In particular, such a tile cannot belong to $\mathcal{B}$.

Repeating this procedure yields the claimed decomposition of $S \setminus S'$

2 Let $\mathcal{B} \subset \mathcal{P}$ be a disjoint covering of $S \setminus A$, where $|A| = 0$. Since tiles have area one $\mathcal{B}$ must be countable and we know that $A = E \times [0, \infty[$ from point **1** above. It is clear from Corollary 4.3 that $\{w_p \mid p \in \mathcal{B}\}$ is an orthonormal set. We need to show that this set is complete in $L^2(0,1)$.

For each $J = 0, 1, \ldots$ let ω_J be the frequency interval $\{0, \ldots, 2^J - 1\}$ and define the sub-collection

$$\mathcal{B}_J = \{p \in \mathcal{B} \mid \omega(p) \subset \omega_J\}.$$

The corresponding finite dimensional subspace of $L^2(0,1)$ is $V_J = \mathrm{span}\{w_p \mid p \in \mathcal{B}_{\mathcal{J}}\}$.

We now proceed as in the proof of Corollary 4.2 to obtain a collection of tiles $\mathcal{B}'_{\mathcal{J}}$ all having time intervals of length 2^{-J} and such that the $\{w_p \mid p \in \mathcal{B}'_{\mathcal{J}}\}$ also span V_J. Moreover, if E_J is the complement of the union of intervals $\{I(p) \mid p \in \mathcal{B}'_{\mathcal{J}}\}$, then we conclude that V_J is simply the space of functions piecewise constant on dyadic intervals of length 2^{-J} and vanishing on E_J.

By construction, $S_J = \cup B_J$ is an increasing sequence of subsets of S such that $\cup_J S_J = S \setminus (E \times [0, \infty[)$. This implies that E_J is a decreasing sequence of subsets of $[0, 1[$ with $\cap_J E_J = E$. In particular, $\lim_J |E_J| = |E| = 0$. Let φ be a function piecewise constant on dyadic subintervals of $[0, 1[$ of length 2^{-m}. Then the projection φ_J of φ on V_J is just the restriction of φ to $[0, 1[\setminus E_J$, provided $J \geq m$. Hence $\varphi_J \to \varphi$ for $J \to \infty$ in $L^2(0, 1)$. The density of functions like φ therefore implies that $\{w_p \mid p \in \mathcal{B}\}$ is complete in $L^2(0, 1)$. **3** Conversely, assume $\{w_p \mid p \in \mathcal{B}\}$ is an orthonormal basis of $L^2(0, 1)$. By Corollary 4.3, $\mathcal{B}$ must be a pairwise disjoint set of tiles. By **1** above, the complement of $S' = \cup\mathcal{B}$ in S is the disjoint union of tiles and a set of the form $E \times [0, \infty[$. If there is a tile $q \in S \setminus S'$, then w_q is orthogonal to all $w_p, p \in \mathcal{B}$, contradicting the completeness of the basis. Therefore $S \setminus S' = E \times [0, \infty[$. Because of the simple support properties of Walsh tiles, the characteristic function of E is orthogonal to all $w_p, p \in \mathcal{B}$. By completeness, this function must be zero almost everywhere, hence $|E| = 0$. $\square$

4.3 Best basis algorithm

Let $N = 2^j$ for some nonnegative integer j. We can identify $\mathbf{R}^N$ with the closed subspace V_j of $L^2(0,1)$ spanned by $\{w_p \mid p \in \mathcal{B}_|\}$, where $\mathcal{B}_|$ consists of the tiles $[k/N, (k+1)/N[\times[0, N[$, $k = 0, 1, \ldots, N-1$. In other words, V_j is the space of functions piecewise constant on dyadic intervals of length 2^{-j}, and we identify $\mathbf{R}^N$ with the subset $S_j = [0,1[\times[0,N[$ of the time-frequency plane $S = [0,1[\times[0,\infty[$. It is clear from the results of Section 4.2 that any disjoint covering $\mathcal{B} \subset \mathcal{P}$ of S_j corresponds to an orthonormal basis of V_j, and therefore of $\mathbf{R}^N$.

The above identification is equivalent to associating the function

$$f(t) = \sum_{k=0}^{N-1} x_k \sqrt{N} W_0(Nt - k) \tag{4.4}$$

to the given vector $x \in \mathbf{R}^N$. For each tiling $\mathcal{B}$ of S_j, Corollary 4.2 implies that we also have

$$f = \sum_{p\in\mathcal{B}} \langle f, w_p\rangle w_p.$$

This gives a large number of bases to choose from, and we will use an additive cost function as in [12] to compare these bases.

Definition 4.5 *A map H from the set of finite pairwise disjoint collections of tiles, $\mathcal{B} \subset \mathcal{P}$ and vectors $v = (v_p)_{p\in\mathcal{B}}$ to $\mathbf{R}$ is called an additive cost function if for all such pairs $(v, \mathcal{B})$*

$$H(v, \mathcal{B}) = \sum_{p\in\mathcal{B}} H(v_p, \{p\}).$$

(In fact, this is a slight generalization of the measure from [12].) Typically $H(v, \{p\}) = h(v)$ and standard choices are $h(v) = -v^2 \log v^2$ or $h(v) = |v|^p$ with

$p < 2$. The idea is that an additive cost function should be small when the energy of the vector (v_p) is concentrated in few elements.

The cost of expanding $f \in V_j$ in the basis induced by $\mathcal{B}$ is then

$$H((\langle f, w_p \rangle)_{p \in \mathcal{B}}, \mathcal{B}) = \sum_{p \in \mathcal{B}} H(\langle f, w_p \rangle, p). \tag{4.5}$$

We are looking for a tiling $\mathcal{B}$ of S_j that minimizes (4.5). The corresponding basis of $\mathbf{R}^N$ is then a "best" basis relative to the vector x and the additive cost function H.

Once the vector x is fixed, all tiles $p \in \mathcal{P}$ with $p \subset S_j$ have a fixed cost $c(p) = H(\langle f, w_p \rangle, p)$. There are $(j+1)N$ such tiles, and the problem above is equivalent to the problem of tiling S_j in the cheapest possible way. The fast algorithm for finding such a minimizing tiling relies on the next lemma which reduces the problem to four problems of half size.

Lemma 4.6 *Let $T \subset S$ be a dyadic rectangle of area greater or equal to two, with left half L, right half R, lower half D, and upper half U. Assume each tile $p \subset T$ has the cost $c(p)$. Define*

$$m_T = \min \left\{ \sum_{p \in \mathcal{B}} c(p) \mid \mathcal{B} \subset \mathcal{P} \textit{ is a disjoint covering of } T \right\},$$

and similarly m_L, m_R, m_D, m_U. Then

$$m_T = \min\{m_L + m_R, m_D + m_U\}.$$

Proof. Let $\mathcal{B}$ be any tiling of $T = I \times \omega$. If a tile $p \in \mathcal{B}$ intersects both L and R, then $I(p) = I$ and if it intersects both D and U then $\omega(p) = \omega$. Hence $\mathcal{B}$ cannot have both types of intersecting tiles since these would intersect each other. Either each $p \in \mathcal{B}$ is contained in one of the halves L or R or each $p \in \mathcal{B}$ is contained in one of the halves D or U.

The union of a tiling of L and a tiling of R is a tiling of T, so $m_T \leq m_L + m_R$ holds. With the same argument we get that $m_T \leq m_D + m_U$. On the other hand, if $\mathcal{B}$ is a minimizing tiling of T, we have just seen above that $\mathcal{B}$ is the disjoint union of two tilings, either of L and R or of D and U. Therefore either $m_T \geq m_L + m_R$ or $m_T \geq m_D + m_U$. $\square$

Repeated use of Lemma 4.6 starting with $T = S_j$ gives a recursive procedure for finding a minimizing tiling of S_j in j steps. The actual algorithm will work in the opposite direction.

Algorithm 4.7

1. *Compute the cost $c(p)$ of all tiles contained in S_j, and put $A(p) = \{p\}$.*

2. *For $l = 1, 2, \ldots j$;*

   Consider the dyadic rectangles $P \subset S_j$ of area 2^{l-1} as tiles with costs $c(P)$. For each dyadic rectangle $P' \subset S_j$ of area 2^l, find a minimizing tiling of P' with the rectangles of area 2^{l-1}. There are two possibilities. Then mark each rectangle P' with the resulting cost $c(P')$, and put $A(P') = A(P_1) \cup A(P_2)$ where P_1 and P_2 are the chosen rectangles of area 2^{l-1}. By Lemma 4.6, $A(P')$ is now a minimizing tiling of P'.

3. *Finally $A(S_j)$ is a minimizing tiling of S_j with the minimal cost $c(S_j)$.*

There are $(j-l+1)2^{j-l}$ dyadic sub-rectangles of S_j with area 2^l, so the lth step of the iteration involves $(j-l+1)2^{j-l}$ times two additions and one comparison. This gives a total of

$$3\sum_{l=1}^{j}(j-l+1)2^{j-l} = 3\sum_{\nu=1}^{j}\nu 2^{\nu-1} = 3(1+(j-1)2^j)$$

operations. That is, the search for the best basis takes less than $3N\log_2 N$ operations.

Finding the coefficients $\langle f, w_p\rangle$ for all $(j+1)2^j$ tiles $p \subset S_j$ can be done by using Lemma 4.1 in j steps of N multiplications and additions. Going backwards from the chosen tiles, the same number of operations will reconstruct the vector $x \in \mathbf{R}^N$. Thus the whole process of representing x with the best among all rescaled Walsh bases takes $O(N\log N)$ operations.

The number a_j of such bases for $\mathbf{R}^N$ can be found recursively from the argument in the proof of Lemma 4.6. Indeed, a_j is equal to the number of tilings of a dyadic rectangle T of area 2^j with dyadic tiles of area one. Each such tiling can be split in two sub-tilings horizontally or vertically. There are a_{j-1}^2 of each of those split tilings, giving $2a_{j-1}^2$ possibilities for the tilings of T, except for the fact that we counted the doubly split tilings twice. There are a_{j-2}^4 such tilings, hence

$$\begin{cases} a_0 = 1, \quad a_1 = 2, \\ a_j = 2a_{j-1}^2 - a_{j-2}^4, \quad j \geq 2. \end{cases}$$

The corresponding recursion for the number b_n of wavelet packet bases is [12]

$$\begin{cases} b_0 = 1, \\ b_j = b_{j-1}^2 + 1, \quad j \geq 1. \end{cases}$$

In the double tree algorithm of [5] the bases are obtained by a dyadic time segmentation followed by dyadic frequency segmentation. The number d_j of such bases is given by

$$\begin{cases} d_0 = 1, \\ d_j = d_{j-1}^2 + b_j - b_{j-1}, \quad j \geq 1. \end{cases}$$

For $j = 0, 1, 2$ we have $d_j = a_j$ but there are twelve Walsh tiling bases of $\mathbf{R}^8$ that can not be obtained by the double tree algorithm. All three numbers of bases grow exponentially with N, and asymptotic estimates can be found by comparison with the recursion $z_{j+1} = \mu z_j^2$, which has the solution $z_j = \mu^{-1}[(\mu z_m)^{2^{-m}}]^{2^j}$ for $j \geq m$. For instance, we can define $\mu_j = a_{j+1}/a_j^2$, and observe that $\mu_0 = 2$ and $\mu_{j+1} = 2 - \mu_j^{-1}$, so that $\mu_j \searrow \mu = (1 + \sqrt{5})/2$ for $j \to \infty$. Therefore $\mu^{-1}[(\mu a_m)^{2^{-m}}]^{2^j} \leq a_j \leq \mu_m^{-1}[(\mu_m a_m)^{2^{-m}}]^{2^j}$ for $j \geq m$. Using similar techniques for b_j and d_j we find that for $j \geq 4$,

$$\left\{ \begin{array}{rcl} 0.618(1.84)^{2^j} \leq & a_j & \leq 0.618(1.85)^{2^j} \\ (1.50)^{2^j} \leq & b_j & \leq (1.51)^{2^j} \\ (1.71)^{2^j} \leq & d_j & \leq (1.72)^{2^j} \end{array} \right\}$$

For example, there are approximately 10^{181} wavelet packet bases, 10^{240} double tree bases, and 10^{272} Walsh tiling bases for $\mathbf{R}^{1024}$.

4.4 An example

The algorithm of Section 4.3 is easiest to understand graphically. Therefore we will consider a small numerical example in some detail.

Consider the vector $x = (3,1,2,2,3,1,1,3) \in \mathbf{R}^8$. For $j = 3$, the time-frequency plane is $S_3 = [0,1[\times[0,8[$, and for figures we will scale the frequency axis to get a square.

The collection of all 32 tiles $p \subset S_3$ and the corresponding coefficients $\langle f, w_p \rangle$ are placed on four copies of S_3 in the top row of squares of Figure 4.1.

The first of these four squares contains just the elements of the vector x, and describes the initial expansion (4.4). In the next square, the eight new expansion coefficients are obtained from Lemma 4.1, more specifically (4.3). Up to a factor $\sqrt{2}$ the numbers are just sums and differences of the elements of x. The remaining two squares in the first row are filled with coefficients $\langle f, w_p \rangle$ obtained by successive use of (4.3) In fact, this is just the usual iterated filter bank operations giving the wavelet packet expansion (in sequency order) of x.

To make calculations simple, we use the cost function based on $h(v) = |v|$. This means that the cost of a tile p is simply the absolute value of the corresponding coefficient $c(p) = |\langle f, w_p \rangle|$. This makes step (1) of Algorithm 4.7 particularly easy.

The next step, $l = 1$, is to consider the twelve dyadic sub-rectangles of S_3 of area 2. These are delimited by solid lines in the second row of squares of Figure 4.1. Each of these rectangles can be covered in two ways by tiles from the the first row of squares. We mark each rectangle with the cost of the cheapest tiling and indicate the addresses of the chosen tiles by a dashed line. For example, the very first rectangle can be covered by time neighbors with cost $3 + 1 = 4$ or frequency neighbors with cost $\sqrt{2} + 2\sqrt{2} = 3\sqrt{3}$. The first choice is cheapest. If the costs of two coverings are the same, we choose time neighbors.

We continue the process with the four dyadic sub-rectangles of S_3 of area four on the third row of squares, $l = 2$. The dashed lines delimit the minimizing tilings. For example, the first rectangle can be covered with dyadic rectangles of area two from the row above in two ways. Time neighbors with cost $4+2\sqrt{2}$ or frequency neighbors with cost $4+\sqrt{2}$. The first choice is cheapest, so the corresponding tiling and cost is carried over to the first rectangle of the row $l = 2$.

Finally, since $4+\sqrt{2}+6 > 4\sqrt{2}+2+\sqrt{2}$ the tiling of the right square of the row $l = 2$ is a minimizing tiling of S_3, so we transfer this tiling to the bottom square, $l = 3$, of Figure 4.1. The tiles are marked with the corresponding expansion coefficients taken from the top row of squares. There are only three nonzero coefficients so the algorithm has indeed concentrated the energy of the signal x. Observe that the resulting tiling can neither be obtained by the wavelet packet algorithm nor by the double tree algorithm, where a time segmentation is followed by a frequency segmentation. In the present situation, both of these algorithms lead to a best basis with the higher cost $1+6\sqrt{2}$ and four nonzero coefficients in the resulting representation of x.

4.5 Algebraic time-frequency localization

In order to motivate the definition below, we now view the interval $[0,1]$ as an image of the group G of all sequences of zeroes and ones with binary addition as group multiplication. Each sequence is mapped to its binary number. Then the intervals $[0, 2^{-k}]$ are images of subgroups of G, and an arbitrary dyadic interval is an image of a coset of such a subgroup.

The dual group of G becomes the set of Walsh functions on $[0, 1]$ with pointwise multiplication as group operation. If we consider as above the sequency order of the Walsh functions, then $W_0, \ldots W_{2^k-1}$ form a subgroup, and the Walsh functions of each dyadic interval of integers form a coset of such a subgroup. The same would be true if we chose the natural Paley order.

Hence a dyadic rectangle is a cartesian product of a coset in G with a coset in $\widehat{G}$. We take this property as our new definition of phase space cells, and call this point of view algebraic time-frequency localization. For simplicity we restrict ourselves to finite groups.

Let G be a finite abelian group, let $\widehat{G} = \mathrm{Hom}[G, T]$ be its dual group consisting of the characters of G. Both are naturally subsets of the group algebra $C(G)$, and with the usual invariant scalar product on the group algebra both are orthogonal bases. We call them time and frequency basis. The linear span of a subset S of the group algebra will be denoted by $\mathrm{span}(S)$.

Definition 4.8 *Consider subgroups $H \subset G$ and $X \subset \widehat{G}$, elements $t \in G$ and $\xi \in \widehat{G}$, and the corresponding cosets Ht and $X\xi$. Then the cartesian product*

$$Ht \times X\xi$$

is called a phase space cell. It's corresponding vector space is

$$V_{Ht \times X\xi} = \mathrm{span}(Ht) \cap \mathrm{span}(X\xi).$$

The following theorem relates the area of a phase space cell to the dimension of the corresponding vector space. Therefore it resembles the Heisenberg uncertainty principle.

The Annihilator $\mathrm{Ann}(H)$ of a subgroup $H \subset G$ is the set of all characters of G that are identically 1 on H. As usual $|G|$ denotes the order of G.

Theorem 4.9 *Consider the data of Definition 4.8. If* $\mathrm{Ann}(H) = X$*, then the phase space cell is one dimensional and spanned by the unit vector*

$$\frac{1}{\sqrt{|H|}} \sum_{t' \in Ht} \xi(t')t' = \frac{\xi(t)}{\sqrt{|G||\mathrm{Ann}(H)|}} \sum_{\xi' \in \mathrm{Ann}(H)\xi} \overline{\xi'(t)}\xi'$$

If $\mathrm{Ann}(H) \subset X$*, then*

$$\dim(V_{Ht \times X\xi}) = |Ht||X\xi|/|G|,$$

and finally if $\mathrm{Ann}(H) \not\subset X$*, then* $V_{Ht \times X\xi} = \{0\}$.

We call the number $|Ht||X\xi|/|G|$ the *normalized area* of the phase space cell $Ht \times X\xi$. Observe that the conditions in Theorem 4.9 are actually symmetric in X and H, if we use Pontryagin duality to identify $\mathrm{Ann}(X)$ with a subgroup of G.

Proof of Theorem 4.9

For the first part of the theorem, observe that both vectors for which equality is claimed are unit vectors, and that once the equality is shown, it follows immediately that the vector is contained in the phase space cell. We calculate the scalar product of both sides using the formula $|H||\mathrm{Ann}(H)| = |G|$, which follows from well known facts about finite abelian groups (see [6], [9]).

$$\left\langle \frac{1}{\sqrt{|H|}} \sum_{t' \in Ht} \xi(t')t', \frac{\xi(t)}{\sqrt{|G||\mathrm{Ann}(H)}} \sum_{\xi' \in \mathrm{Ann}(H)\xi} \overline{\xi'(t)}\xi' \right\rangle$$

$$= \frac{1}{\sqrt{|G||H||\mathrm{Ann}(H)|}} \sum_{t' \in Ht, \xi' \in \mathrm{Ann}(H)\xi} \xi(t')\overline{\xi(t)}\xi'(t)\overline{\xi'(t')}$$

$$= \frac{1}{|G|} \sum_{t'' \in H, \xi'' \in \mathrm{Ann}(H)} \xi''(t'') = 1$$

The scalar product of the two unit vectors is one, hence equality holds.

That this vector spans the phase space cell will follow from the second statement of Theorem 4.9, which we prove by a counting argument. First observe that the corresponding vector spaces of two disjoint phase space cells are orthogonal, because the two cells have disjoint support either in time or in frequency, and both the time basis and the frequency basis are orthogonal. Assume $\mathrm{Ann}(H) \subset X$. The coset $X\xi$ is the disjoint union of as many as $|X/\mathrm{Ann}(H)| = |X||H|/|G|$ cosets of $\mathrm{Ann}(H)$, each giving rise to an at least one dimensional subspace of the corresponding vector space of $Ht \times X\xi$, which therefore itself has at least the dimension claimed in the theorem. There are $|G|/|H|$ cosets of H and $|G|/|X|$ cosets of X in G, hence there are $|G|^2/(|H||X|)$ combinations of such cosets, giving the same number of mutually orthogonal vector spaces. As the dimension of the group algebra is $|G|$, it follows that the lower bound on the dimension of each of these vector spaces, that was found above, is the exact dimension.

Finally assume that $\mathrm{Ann}(H) \not\subset X$, and pick a character in $\mathrm{Ann}(H) \setminus X$. Considering its natural action on the group algebra, it acts as a scalar on $\mathrm{span}(Ht)$, but the only vector in $\mathrm{span}(X\xi)$ on which it acts as a scalar is the zero vector. Hence the intersection of both spaces is the zero vector only.

From now on we will only consider phase space cells which satisfy the condition of Theorem 4.9 so as to represent a nontrivial vector space. We have the following analogue of Corollary 4.2

Corollary 4.10 *Let a phase space cell p be the disjoint union of phase space cells p_j, $j \in J$, each having normalized area one. Then picking a unit vector out of each of the one dimensional vector spaces V_{p_j} gives an orthonormal basis of V_p.*

Proof. Comparison of the areas shows by Theorem 4.9, that we have the right number of unit vectors. That they are orthogonal has been shown in the proof of Theorem 4.9.

From Corollary 4.10 we get a whole library of orthonormal bases of the group algebra to choose from, namely one for each partition of $G \times \widehat{G}$ into phase space cells of normalized area one. To get a fast algorithm as before we have to restrict ourselves to a sublibrary. For that we fix a chain of subgroups

$$G_0 = (0) \subset G_1 \subset \ldots \subset G_k \subset \ldots \subset G_n = G$$

of G. An *admissible tile* with respect to this chain is a phase space cell $G_k t \times \mathrm{Ann}(G_k)\xi$ with G_k contained in the chain. Observe that such a tile has normalized area one. A tiling of a phase space cell is a partition of the latter into admissible tiles . We then have the analogue of Lemma 4.6:

Lemma 4.11 *With the notation above, consider a phase space cell $C \times D$ of normalized area greater than one, where C is a coset of G_k and D is a coset of $\mathrm{Ann}(G_l)$. Then a tiling of this phase space cell is either the disjoint union of tilings of the*

$|G_k|/|G_{k-1}|$ *phase space cells* $C_j \times D$*, where the* C_j *denote the different cosets of* G_{k-1} *in* C*, or it is the disjoint union of tilings of the* $|\mathrm{Ann}(G_l)|/|\mathrm{Ann}(G_{l+1})|$ *phase space cells* $C \times D_j$ *, where the* D_j *denote the different cosets of* $\mathrm{Ann}(G_{l+1})$ *inside* D*.*

Proof. If the tiling fails to split into tilings in the first way, then necessarily for one tile $C' \times D'$ of this tiling the first coset C' is not contained in any of the cosets of G_{k+1} inside C, which implies that $C' = C$. Assuming that the tiling does not split in the second way neither, we similarly find a tile $C'' \times D$ in the tiling of $C \times D$. But then those two tiles have nonempty intersection, a contradiction.

This lemma gives a recursive control over the bases of a phase space cell, which gives a fast algorithm to pick the optimal basis with respect to an additive cost function as before in the Walsh case.

4.6 An example related to the FFT

For numerical treatment of the Fourier transform, for example on the circle, one discretizes the latter to a finite cyclic group. This, besides the Walsh group, provides the main example for the results of Section 4.5. Indeed, when doing the Fast Fourier Transform, one implicitely makes use of phase space cells.

As an example we discuss the cyclic group G of order eight. It is represented by the residues $\{0,1,2,3,4,5,6,7\}$ modulo eight. By G_0 , G_1 , and G_2 we denote the subgroups $\{0\}$, $\{0,4\}$, and $\{0,2,4,6\}$, and we also write $G = G_3$. These are all subgroups of G, and as for any cyclic group they form a chain, so that the fast algorithm works for the collection of all phase space cells.

The characters of the group G are of the form

$$\xi_n(t) = e^{\frac{i\pi}{4}tn},$$

whith n an integer. In fact such a character depends only on the residue of n modulo eight, hence we can identify the dual group also with $\{0, 1, 2, 3, 4, 5, 6, 7\}$. It is clear that the Annihilator of G_k is G_{3-k}.

In Figure 4.2 each square represents the phase plane with coordinates as shown for the square in the bottom row. Inside each square (k, l) the black area shows the phase space cell $G_k \times G_l$. The dark gray and the light gray areas, which are obtained from the black areas just by translations, show phase space cells of cosets of the same subgroups.

One can see from Figure 4.2 how a black phase space cell $G_k \times G_l$ splits into cells of smaller dimension: it is the union of the light gray cell and the black cell in the square $(k-1, l)$, and at the same time it is the union of the dark gray cell and the black cell in $(k, l-1)$.

That each cell of normalized area greater than one is a union of exactly two smaller cells is certainly a special property of our situation, which can be achieved only for groups having a power of two as order. It implies that all calculations for the fast algorithm have the same complexity as in the Walsh case.

For each pair (k, l) the phase plane can be partitioned into phase space cells that are products of a coset of G_k with a coset of G_l. This means that for each square in the top row of Figure 4.2 we obtain an orthonormal basis of the group algebra by picking a unit vector out of each of the one dimensional cells. These special bases are called *homogeneous*. Assume that we are given a signal in time basis, that is the basis corresponding to the square $(0, 3)$. To calculate the coefficients of the basis corresponding to the square $(1, 2)$, we use that each phase space cell in this square is contained in a phase space cell of the square $(1, 3)$, which itself is spanned by two phase space cells of the square $(0, 3)$. Hence each coefficient of square $(1, 2)$ is the linear combination of two coefficients of the square $(0, 3)$. The explicit form of the linear combination can be deduced from the formula in Theorem 4.9.

Continuing this way, we can successively calculate all coefficients in the upper row. In order to get the signal in frequency basis, we have to calculate 24 linear combinations. This algorithm for calculating the Fourier Transform is called the Fast Fourier Transform. Observe that in general for a group of order 2^n we need to calculate $n2^n$ linear combinations.

The algorithm for selecting the best basis now goes along the same lines as in the example in Section 4.4: Given an additive cost function, we first select the optimal basis out of the two choices for each of the two dimensional phase space cells in the second row. Then we successively continue with the other rows, until we obtain the optimal basis for the last square, which represents the whole group algebra. For each phase space cell we have to compare only two bases.

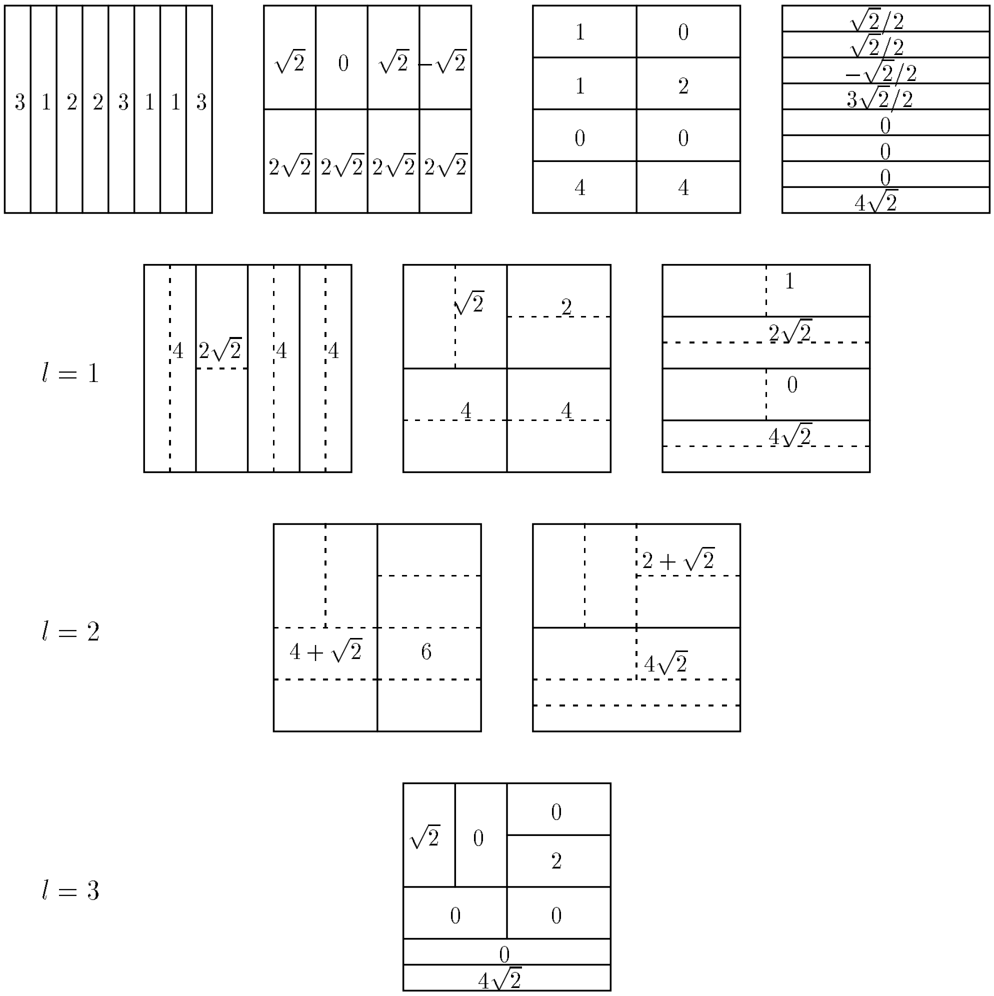

Figure 4.1: Graphical representation of Algorithm 4.7 for a signal of length 8. The cost function is equal to the sum of absolute values of the expansion coefficients. The top row of squares contain all tiles and coefficients. The bottom square represents the best basis expansion.

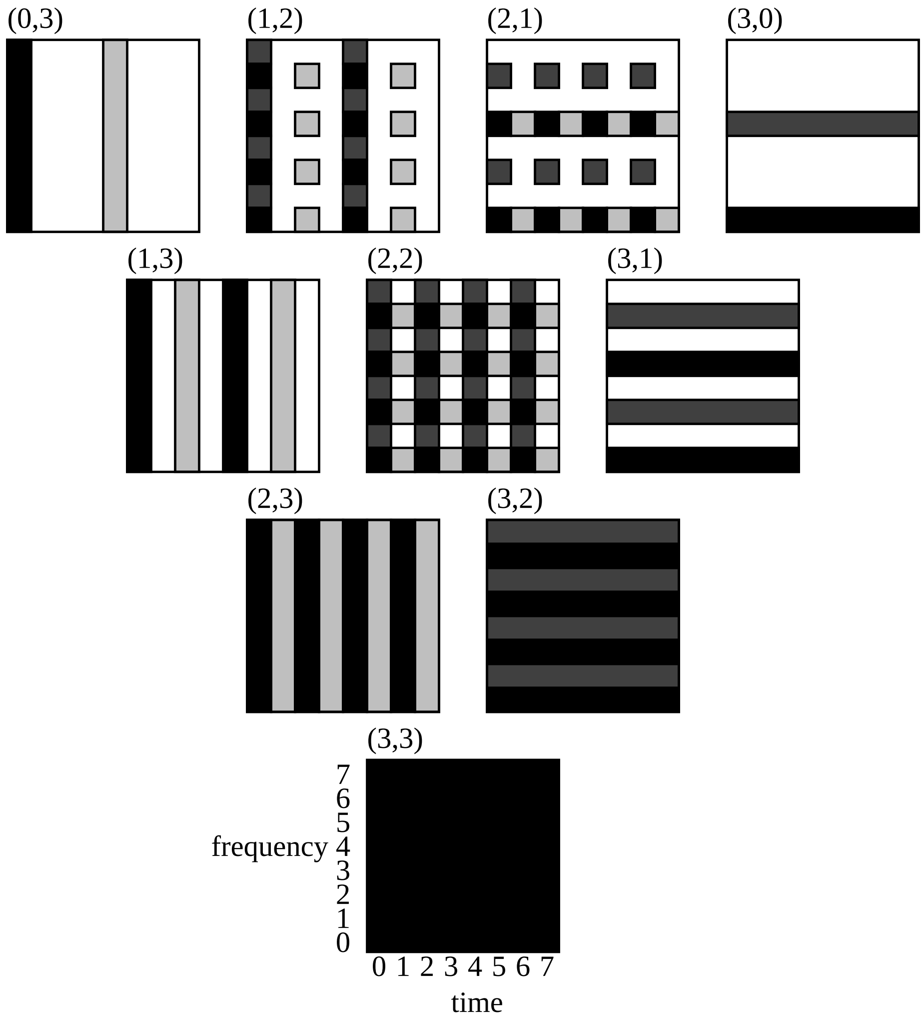

Figure 4.2: Phase space cells for the cyclic group of order eight

Bibliography

[1] A.Cohen, I.Daubechies, and P.Vial. Wavelets on the interval and fast wavelet transforms. *Appl. Comput.Harmonic Anal.*, 1:54–81, 1993.

[2] P. Billard. Sur la convergence presque partout des series de fourier walsh des fonctions de l'espace $l^2(0,1)$. *Studia Math.*, 28:363–388, 1967.

[3] L. Carleson. On convergence and growth of partial sums of fourier series. *Acta Math.*, 116:135–157, 1966.

[4] C.Fefferman. Pointwise convergence of fourier series. *Annals of Math.(2)*, 98:551–571, 1973.

[5] C.Herley, J.Kovačević, K.Ramchandran, and M.Vetterli. Tilings of the time-frequency plane: Construction of arbitrary orthogonal bases and fast tiling algorithms. *IEEE Trans. Signal Proc.*, 41 No 12:3341–3359, 1993.

[6] D.Gorenstein. *Finite Groups.* Chelsea, New York, 1980.

[7] R.A. Hunt. On the convergence of fouries series. In *Orthogonal Expansions and their Continuous Analogues. Proc.Conf.Edwardsville*, pages 235–255, 1967.

[8] J.Bergh and J.Löfström. *Interpolation Spaces, An Introduction.* Springer-Verlag, Berlin Heidelberg New York, 1976.

[9] Y. Katznelson. *An Introduction to Harmonic Analysis.* Dover, New York, 1968.

[10] M. Lacey. Bilinear hilbert transform. several papers in personal communication, 1994.

[11] R.E.A.C. Paley. A remarkable series of orthogonal functions (i). *Proc. London Math. Soc.*, 34:241–264, 1932.

[12] R.R.Coifman and M.V.Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE trans.info.theory*, 38:713–718, 1992.

[13] P. Sjölin. An inequality of paley and convergence a.e. of walsh-fourier series. *Arkiv f. Mat.*, 7 nr 42:551–570, 1968.

[14] C.M. Thiele and L.F. Villemoes. A fast algorithm for adapted time frequency tilings. to appear in Appl.Comp.Harm.An, 1994.

[15] J. L. Walsh. A closed set of normal orthogonal functions. *Amer. J. Math.*, 45:5–24, 1923.

[16] Wavelet packet laboratory for windows. A.K.Peters, 1993. PC software.