Hierarchical Parallel Matching Pursuit

Hans G. Feichtinger, Andreas Türk and Thomas Strohmer
NUHAG – Numerical Harmonic Analysis Group
University of Vienna, Department of Mathematics
Strudlhofgasse 4, A-1090 Vienna, AUSTRIA
email: fei@tyche.mat.univie.ac.at

ABSTRACT

The by now well-known matching pursuit method of S. Mallat[8] as well as the recently proposed orthogonal matching pursuit[10] work purely sequential and are based on the idea of choosing only one single atom at a given time. Pursuing ideas which are related to modifications of the POCS method (as described in[5] and[2]) we suggest a new type of orthogonalization procedure which allows to operate in parallel at different "levels". More precisely, we assume that we have a dictionary which consists of similar "pages", i.e. those pages are collection of functions, generated from a single function by translations along a subgroup which is the same for all such pages. Based on a certain hierarchical structure (preference of pages) we apply an appropriate Gram-Schmidt type orthogonalization procedure which allows to deal with the matching pursuit problem in parallel at different levels. After carrying out appropriate approximations at the different, now orthogonal levels, one comes back to a representation based on the given family of atoms, in a straight-forward way.

Keywords: matching pursuit, coherent frames, wavelet theory, Gram-Schmidt orthogonalization, POCS-method .

1 Introduction

Starting from a general dictionary, i.e. typically a rich collection of potential atoms in a Hilbert space, the so-called matching pursuit method tries to obtain a representation of a given signal in terms of those atoms which show maximal response (hence similarity) to the signal, at least at some stage of an iterative process. This strategy has certain drawbacks, however. Among others, the searching and updating makes the algorithm fairly slow, at least for general dictionaries, and furthermore the same element may occur repeatedly during the process of reconstruction. In order to get rid of this last mentioned problem B.Pati et. al. have proposed a modification of the original concept by S. Mallat,[8] the so-called orthogonal matching pursuit.[10] Both methods work purely sequential and choose only one additional element at a time. In this paper we suggest a well parallelizable way for doing matching pursuit at different levels. Although our approach has been motivated by non-orthogonal wavelet type collections of atoms it is much more general. On the other hand we require (compared to general dictionaries) a certain structure of the dictionary to be used. The detailed explanation is as follows:

We assume that the Hilbert space is $L^2(\mathbb{R}^m)$, and that the dictionary itself can be interpreted as a collection of similar "pages". Those *pages* are defined as closed, H-invariant subspaces of $L^2(\mathbb{R}^m)$, which are generated from a set of H-translates of a given atom φ, H being some discrete subgroup of $\mathbb{R}^m$.

Let us be more precise: Assume we have a spline-type function φ – e.g. a B-spline – which generates a closed linear subspace of $L^2(\mathbb{R}^m)$, either as a Riesz basis or at least as a frame by integer translates. Thus $H = \mathbb{Z}^m \in \mathbb{R}^m$. Assume further that a compressed version (for the sake of discussion let us assume that it is compressed by a factor of 3) of the same function is used in connection with a scaled (by a factor of $\frac{1}{3}$) lattice. In that case we would consider our dictionary as generated by 3 equal lattices (the original one $H = \mathbb{Z}^m$), and 3 compressed version of φ, one in original position, the two others shifted by $\frac{1}{3}$ to the right and to the left respectively. Since we do not assume orthogonality between the different spaces and we want to do matching pursuit we will decide about some "hierarchy", in other words, we have to determine, which of the atoms we prefer in case there is some freedom. It appears to us that one would naturally prefer to use blocks involving the original function φ corresponding to larger scales, as opposed to the compressed bump-functions. For the rest, it appears as a matter of taste if one prefers the compressed atom in its unshifted or shifted version and in which order.

In the main part of this paper we give a description, how a certain Gram-Schmidt type orthogonalization procedure allows to translate this non-orthogonal problem into a more or less equivalent orthogonal one. In the new orthogonal setting any matching pursuit type method or general approximation method may be applied in order to approximate some part of the original signal f (the splitting of f is in terms of the orthogonal decomposition mentioned above) by linear combinations of H-translates of new atoms, obtained in the GS-procedure. Finally, the approximation which is a finite series in those new atoms can be rewritten in terms of the original ones. Just to emphasize the role of the order (which is in some sense up to the "user") we mention, that the order in which the spaces – i.e. the pages – are used makes a difference and that the actual coefficients arising will depend among others on the order in which they are taken.

2 The Non-orthogonal Sum of H-invariant Spaces

Suppose we are given a Hilbert space E and some closed subspace U, which is the not necessarily orthogonal sum of the closed subspaces V_i, $i = 1, \ldots, N$:

$$U = V_1 + \ldots + V_N \,. \tag{1}$$

Generally speaking let $\mathcal{D} = \{\rho_i\}_{i \in I}$ be a collection of vectors (our *dictionary*) with $\text{span}\{\rho_i\} = U$. To each subspace V_j belongs a subdictionary $\mathcal{D}_j \subset \mathcal{D}$, and $\mathcal{D} = \bigcup \mathcal{D}_j$. We want to construct approximations f_a to some signal f in E of the form

$$f_a = \sum_{i \in I} c_i \rho_i \,,$$

by using the elements in those sub-dictionaries $\mathcal{D}_|$ in a certain order of preference ("hierarchy"), with the extra aim of going for an algorithm allowing easy parallelization. Typically one will go for $f_a = P_U f$, where $P_U : E \mapsto U$ is the orthogonal projection operator onto U, and therefore most of our explanation will concern this special case. However, the scope of our approach is more general and can be combined with other approximation methods based on the hierarchy given by the family of V_j's (such as matching pursuit within subspaces, using the smaller dictionaries D_j only).

Typical for what we want to discuss in this paper is the situation that the spaces V_k, $k = 1, \ldots N$ are all very "similar". Compared to a dictionary, where one can find everything, but with no specific preference for the use of certain words, we are thinking of those spaces V_j as "pages" of a book (the subspace U corresponding to this book). Staying with this picture we would like to approximate a given text with pieces from the different pages (use atoms from the subspaces to approximate a signal within U). As we will see, this structure will allow to give a certain preference to the "first" pages, i.e. the procedure described in the sequel will *depend on the order* in which the spaces V_k are labeled. Since however, the spaces V_j are not orthogonal to each other, and need not even be linear independent in whatever sense, there may be many ambiguities in the representation of a signal in terms

of elements from the spaces V_k. It is one aim of this paper to show how one can set up a certain hierarchy among those spaces. This will result in preferences of atoms in the first mentioned spaces in the sum of equation (1).

In order to get a better idea of our approach let us consider the following reasoning: To calculate the orthogonal projection of an element f in E on U, we can of course use the projector on U P_U to get $P_U f$. However this space will be quite large and unstructured in general and therefore for typical applications the calculation of the corresponding huge pseudo-inverse matrix will *not* be feasible. If we make some further assumptions, however, we can reduce the calculation to that of several projections onto small spaces. We require, that not only $U = V_1 + \ldots V_N$ but every sum $V_1, \ldots, V_j$, $j = 1, \ldots, N$ is a closed subspace of E. Now we construct the orthogonal complements of $V_1 + \ldots + V_{j-1}$ in $V_1 + \ldots + V_j$:

$$W_j := (V_1 + \ldots + V_j) \ominus (V_1 + \ldots + V_{j-1}) \qquad \text{for } j = 2, \ldots, N$$

and define $W_1 := V_1$. With the new spaces W_j we can write now U as an orthogonal sum:

$$U = V_1 + \ldots + V_N = W_1 \oplus \ldots \oplus W_N \,. \tag{2}$$

This is due to the fact, that $V_1 + \ldots + V_{j-1} \oplus W_j = V_1 + \ldots + V_j$, because $V_1 + \ldots + V_{j-1}$ is closed. Now the calculation of $P_U f$ can be reduced to the calculation of $P_{W_j} f$, $j = 1, \ldots, N$ and summing up:

$$P_U f = P_{W_1} f + \ldots + P_{W_N} f \tag{3}$$

where P_{W_j} is the orthogonal projection on the space W_j. For convenience of notation we set $Q_j := P_{W_j}$. The point of this reduction will be that for the applications we have in mind the spaces V_k and therefore also the spaces W_k will be nicely structured such that consequently those partial orthogonal projections are computationally much less expensive and can be carried out using FFT-based methods.

An expansion like (3) is only valid, if the larger space U is the *orthogonal* sum of its subspaces. This is the reason why we have to orthogonalize $U = V_1 + \ldots + V_N$ before we can construct $P_U f$ from the projections on the smaller spaces W_j. If we obtain the orthogonal projection on U as result of the orthogonal matching pursuit, this means that we first transform our dictionary, which is made up of dictionaries for the V_j , $j = 1, \ldots, N$, to get dictionaries for W_j , $j = 1, \ldots, N$, which are of course smaller than that of U. Now we can apply the orthogonal matching pursuit to all the W_j , $j = 1, \ldots, N$ seperately, because by (2) approximation from W_i does not affect approximation from W_j for $j \neq i$. Parallelizing this step, one ends up with an algorithm, that is as much faster than the orthogonal matching pursuit for U, as the largest dictionary for the W_j , $j = 1, \ldots, N$ is smaller than the dictionary of U. The proposed strategy makes in particular sense if the same hierarchical sequence V_j is used to represent a larger number of signals f.

REMARK: Much of the subsequent discussion could also be carried out in terms of representation theory of (commutative) groups. In fact, there is no big difference in notation and the proofs. Nevertheless we chose a formulation avoiding those more abstract concepts in order to keep the paper readable to applied scientists interested in more concrete problems.

Thus we specialize the mathematical setting to the situation which is most interesting for applications. Let E be $L^2(\mathbb{R}^m)$ for some $m \geq 1$, and let V_j be the closed linear subspace of $L^2(\mathbb{R}^m)$ spanned by translates $T_h \varphi_j(x) := \varphi_j(x - h)$ of a single function $\varphi_j \in L^2(\mathbb{R}^m)$. Spaces of this form are also know as *principal shift-invariant spaces* in the mathematical literature. Cf.[7] and[4] and the references given there for a rich account on the literature on such spaces.

The lattice H along which we translate the φ_j and the constituting spaces V_j in the above description are given by:

$$V_j = \overline{\langle T_h \varphi_j : h \in H \rangle} \,, \tag{4}$$

with

$$H = \left\{ \sum_{i=1}^{n} z_i v_i : (z_1, \ldots, z_m) \in \mathbb{Z}^m \right\}$$

where $v_i \in \mathbb{R}^m$ $i = 1, \ldots, m$ are the linearly independent generating vectors of the lattice H. Further we shall require, that every $v_j \in V_j$ can not only be approximated by finite linear combinations of the $T_h\varphi_j$, $h \in H$, but has an expansion

$$v_j = \sum_{h \in H} c^h T_h \varphi_j \,, \tag{5}$$

where this sum is understood as an $L^2(\mathbb{R}^m)$-limit. Consider for instance the case that $\{T_h\varphi_j : h \in H\}$ constitutes a Riesz basis, i.e.,

$$A\|(c^h)_{h\in H}\|^2_{l^2(H)} \leq \| \sum_{h \in H} c^h T_h \varphi_j \|_2^2 \leq B\|(c^h)_{h\in H}\|^2_{l^2(H)} \tag{6}$$

holds for every sequence $(c^h)_{h\in H} \in l^2(H)$ and some constants $A > 0, B < \infty$. Then we know, that $(c^h)_{h\in H}$ in (5) is an element of $l^2(H)$ and that this expansion is unique, since the elements of a Riesz basis are linearly independent (but they do not have to be orthogonal). This is due to the fact that a Riesz basis for a Hilbert space is always the image of an orthonormal basis under a linear and invertible mapping of this Hilbert space into itself, (this mapping need not be an isometry). For instance the set of functions $\{N_m(x-k) : k \in \mathbb{Z}\}$ is a Riesz basis, where $N_m(x)$ is the m-th order cardinal B-spline, given inductively by

$$N_m(x) := (N_{m-1} * N_1)(x) \quad = \quad \int_{-\infty}^{\infty} N_{m-1}(x-t)N_1(t)dt \tag{7}$$

$$= \quad \int_0^1 N_{m-1}(x-t)dt, \tag{8}$$

where we have used for the last step that N_1 is defined by

$$N_1(x) = \begin{cases} 0 & 0 \leq x \leq 1 \\ 1 & \text{otherwise}\,. \end{cases}$$

A more general situation occurs, when $\{T_h\varphi_j : h \in H\}$ constitutes a *frame* for V_j. This means, that

$$A\|v_j\|_2^2 \leq \sum_{h \in H} |\langle v_j, T_h\varphi_j\rangle|^2 \leq B\|v_j\|_2^2 \tag{9}$$

for every $v_j \in V_j$, with $A > 0, B < \infty$ (it is more convenient for us to use the same frame bounds for all of the finitely many $V_j's$). For each frame the so called *dual frame* $\{\tilde{\varphi}_j^h : h \in H\}$ can be constructed, having the following property

$$v_j = \sum_{h \in H} \langle v_j, T_h\tilde{\varphi}_j\rangle T_h\varphi_j = \sum_{h \in H} \langle v_j, T_h\varphi_j\rangle \; T_h\tilde{\varphi}_j \,.$$

It is one of the remarkable (and simple facts) about frames generated by shifts from a single function, that the frame operator commutes with translations by elements in H. Consequently the same is true for the inverse frame operator, which in turn implies that the dual frame is given by the translates of some $\tilde{\varphi}_j$. Therefore frame coefficients can be obtained in this case by convolving v_j with $\tilde{\varphi}_j$ and then downsampling to the subgroup H.

3 The H-invariant Orthogonalization Procedure

In this section an H-invariant version of the Gram-Schmidt procedure is proposed. Let us consider the orthogonalization of $U = V_1 + \ldots + V_N$ for the general setting of (4) and (5). It can be shown that the W_j , $j = 1, \ldots, N$ are themselves spanned by the translates of a single function ψ_j. We proceed by induction:
Set $\psi_1 = \varphi_1$ and for $k \geq 2$ consider the function

$$\psi_k := \varphi_k - Q_1\varphi_k - \ldots - Q_{k-1}\varphi_k \,,$$

where $\psi_k \in V_1 + \ldots + V_k$, because ψ_k is a linear combination of elements from $W_k, W_1, \ldots, W_{k-1}$ respectively. This ψ_k is also orthogonal to $V_1 + \ldots + V_{k-1}$, since by construction

$$V_1 + \ldots + V_{k-1} = W_1 \oplus \ldots \oplus W_{k-1}$$

and $Q_1\varphi_k + \ldots + Q_{k-1}\varphi_k$ is therefore just the orthogonal projection of φ_k onto $W_1 \oplus \ldots \oplus W_{k-1}$. Hence $\varphi_k - Q_1\varphi_k \ldots - Q_{k-1}\varphi_k \in (W_1 \oplus \ldots \oplus W_{k-1})^\perp$, by elementary Hilbert space theory. It remains to show, that also the translates of ψ_k are orthogonal to $V_1 + \ldots + V_{k-1}$ and that their span is the orthogonal complement of $V_1 + \ldots + V_{k-1}$ in $V_1 + \ldots + V_k$. The orthogonality of $T_h\psi_k$ to $V_1 + \ldots + V_{k-1}$ is – abstractly speaking – a consequence of the unitarity of the representation $h \mapsto T_h$. In detail,

$$\begin{aligned}
\langle T_{h_1}\varphi_j, T_{h_2}\psi_k\rangle &= \int\limits_{\mathbb{R}^m} \varphi_j(x-h_1)\overline{\psi_k(x-h_2)}dx \\
&= \int\limits_{\mathbb{R}^m} \varphi_j(x-(h_1-h_2))\overline{\psi_k(x)}dx \\
&= \langle T_{(h_1-h_2)}\varphi_j, \psi_k\rangle = 0 \qquad \text{for } j \le k-1 \ .
\end{aligned}$$

In the sequel we show that the translates of ψ_k span W_k. For this, we take an element $f \in V_1 + \ldots + V_k$ that is orthogonal to $V_1 + \ldots + V_{k-1}$. We have therefore by induction an expansion

$$\begin{aligned}
f &= \sum_{h\in H} c_k^h T_h\varphi_k + \sum_{j=1}^{k-1}\sum_{h\in H} c_j^h T_h\psi_j \\
&= \sum_{h\in H} c_k^h T_h(\varphi_k - Q_1\varphi_k - \ldots - Q_{k-1}\varphi_k) + \sum_{j=1}^{k-1}\sum_{h\in H} c_k^h T_h(Q_j\varphi_k) + \sum_{j=1}^{k-1}\sum_{h\in H} c_j^h T_h\psi_j
\end{aligned}$$

where each $\sum_{h\in H} c_k^h T_h(Q_j\varphi_k)$ lies in W_j, $j = 1, \ldots, k-1$ and can therefore be written as

$$\sum_{h\in H} c_k^h T_h(Q_j\varphi_k) = \sum_{h\in H} b_j^h T_h\psi_j \quad .$$

This yields

$$f = \sum_{h\in H} c_k^h T_h\psi_k + \sum_{j=1}^{k-1}\sum_{h\in H}(c_j^h + b_j^h)T_h\psi_j \ . \tag{10}$$

Since $f \in (V_1 + \ldots + V_k) \ominus (V_1 + \ldots + V_{k-1})$, we have

$$\begin{aligned}
0 = Q_l f &= \sum_{h\in H} c_k^h Q_l T_h\psi_k + \sum_{j=1}^{k-1}(c_j^h + b_j^h)Q_l T_h\psi_j \\
&= \sum_{h\in H}(c_l^h + b_l^h)T_h\psi_l \qquad , l = 1, \ldots, k-1 \ .
\end{aligned} \tag{11}$$

because of the orthogonalization of $T_h\psi_j$ to W_l for all $h \in H$ and $j = 1, \ldots, k-1, j \neq l$. From (11) we deduce, that the second sum in (10) is zero. Thus our claim has been verified. Summing up we have obtained following theorem:

THEOREM 3.1. *Let E be a Hilbert space, $U \subseteq E$ a closed subspace of the form $U = V_1 + \ldots + V_N$ where the V_j and the $V_1 + \ldots + V_j$, $j = 2, \ldots, N-1$ themselves are closed subspaces of E. Then U can be decomposed as follows:*

$$U = W_1 \oplus \ldots \oplus W_N$$

where $W_1 = V_1$ *and* $W_j := (V_1 + \ldots + V_j) \ominus (V_1 + \ldots + V_{j-1})$. *Further if* H *is a lattice in* $\mathbb{R}^m$, $E = L^2(\mathbb{R}^m)$ *and* $V_j = \overline{\langle T_h \varphi_j : h \in H \rangle}$, *then*

$$W_j = \overline{\langle T_h \psi_j : h \in H \rangle}$$

with

$$\psi_j := \varphi_j - Q_1 \varphi_j - \ldots - Q_{j-1} \varphi_j$$

where Q_l *is the orthogonal projection onto* W_l.

Consider again the special case of $\{T_h \varphi_j : h \in H\}$ constituting a Riesz basis. From the last theorem we infer that $T_h \psi_j = Q_j T_h \varphi_j$. For the family $\{T_h \psi_j : h \in H\}$ being a Riesz basis it is necessary that Q_j is a linear homeomorphism of V_j onto W_j. If the V_j are linearly independent, Q_j is bijective. But in general it might not be a open mapping. However, if V_j is finite-dimensional this cannot occur. If $\{T_h \psi_j : h \in H\}$ constitutes a Riesz basis, we can directly write down the orthogonal projection of any $f \in L^2(\mathbb{R}^m)$ on W_j. Its Fourier-transform is given by

$$\widehat{P_{W_j} f}(\omega) = \frac{\hat{\psi}_j(\omega)}{\hat{\Psi}_j(\omega)} \sum_{h' \in H'} \hat{f}(\omega + h') \overline{\hat{\psi}_j(\omega + h')}\,, \tag{12}$$

H' being the reciprocal lattice of H and $\hat{\Psi}_j(\omega) := \sum_{h' \in H'} |\hat{\psi}_j(\omega + h')|^2$. This result for general Riesz basis is proved in several papers, for instance in Oakley et al.(see[6]). It is not difficult to adjust it to general translation invariant frames and has some interesting applications in wavelet theory. A similar result – the so called orthogonalization trick – serves to orthogonalize the $N_m(x-k)\, k \in \mathbb{Z}$ from (7) thus giving a orthonormal basis for one multiresolution layer. For a detailed discussion of the concept of multiresolution, especially in connection with B-splines, we refer to.[1] See[3] for examples concerning the application of the orthogonalization trick in wavelet theory.

The assumption that the V_j , $j = 1, \ldots, N$ are linearly independent is no severe restriction, if we allow V_j to be generated by the translates of a finite number of different functions. Then we can put all linearly dependent V_j's into one single V_i, which has a dictionary, that is the union of the dictionaries of these V_j.
One can also apply (12) – substituting f by φ_k – to get the orthogonal projections of φ_k onto W_j $j \leq k-1$, namely $Q_j \varphi_k$, at least, if $\{T_h \psi_j : h \in H\}$ $j \leq k-1$ is a Riesz basis. Thus, if the $Q_j : V_j \to W_j$ are linear homeomorphisms, we can calculate every object in our theorem and all the orthogonal projections onto the spaces W_j , $j = 1, \ldots, N$. Thereby we arrive at a decomposition of $P_U f \in L^2(\mathbb{R}^m)$ in the form:

$$P_U f = \sum_{j=1}^{N} \sum_{h \in H} a_j^h T_h \psi_j\,. \tag{13}$$

This is the case whenever one calculates the orthogonal projection on U from the orthogonal projections on W_j , $j = 1, \ldots, N$. Since we did not start with the ψ_j's but with the φ_j's, it is desirable to get an expansion

$$P_U f = \sum_{j=1}^{N} \sum_{h \in H} c_j^h T_h \varphi_j\ . \tag{14}$$

The coefficients in (14) can easily be calculated from the a_j^h in (13) once an expansion of $Q_l \varphi_j$, $1 \leq l \leq j-1$ in W_l is known. Suppose

$$Q_l \varphi_j = \sum_{h \in H} d_{l,j}^h T_h \psi_l \qquad , j = 2, \ldots, N; \quad 1 \leq l \leq j-1\,.$$

Then

$$P_U f = \sum_{j=1}^{N} \sum_{h \in H} a_j^h T_h \psi_j \quad = \quad \sum_{j=1}^{N-1} \sum_{h \in H} a_j^h T_h \psi_j + \sum_{h \in H} a_N^h T_h (\varphi_N - Q_1 \varphi_N - \ldots - Q_{N-1} \varphi_N)$$

$$= \sum_{j=1}^{N-1} a_j^h T_h \psi_j - \sum_{j=1}^{N-1} \sum_{h\in H} a_N^h T_h \sum_{k\in H} d_{j,N}^k T_k \psi_j + \sum_{h\in H} a_N^h T_h \varphi_N$$

$$= \sum_{j=1}^{N-1} \sum_{h\in H} (a_j^h - \sum_{k\in H} a_N^k d_{j,N}^{h-k}) T_h \psi_j + \sum_{h\in H} a_N^h T_h \varphi_N \ . \tag{15}$$

By this procedure we managed to single out coefficients for the $T_h\varphi_N$ $h \in H$, while manipulating those for $T_h\psi_j$, $h \in H, j \leq N-1$. If we apply now the above formula once more to the first sum in (15) we get coefficients for $T_h\varphi_{N-1}$, $h \in H$. We finally obtain an expansion

$$P_U f = \sum_{h\in H} c_1^h T_h \psi_1 + \sum_{j=2}^{N} \sum_{h\in H} c_j^h T_h \varphi_j \ . \tag{16}$$

Since $\psi_1 = \varphi_1$ equation (16) can be written as

$$P_U f = \sum_{j=1}^{N} \sum_{h\in H} c_j^h T_h \varphi_j \ .$$

So the translation of (13) into (14) is achieved through application of the following algorithm, which calculates the required sequence of coefficients c_j^h for the φ-representation of the best approximation from the corresponding sequence a_j^h based on the ψ-representation.

Algorithm 1 (Backsubstitution from W-expansions to V-expansions):

Initialization	$c_j^h := a_j^h$	for $j = 1, \ldots, N$
First step:	$c_j^h := c_j^h - \sum_{k\in H} c_N^k d_{j,N}^{h-k}$	$h \in H$, $1 \leq j \leq N-1$
Second step:	Apply first step to c_j^h	$1 \leq j \leq N-1$
	redefining c_j^h for	$1 \leq j \leq N-2$
$\vdots$		
Last step:	Apply first step to c_j^h	$j = 1, 2$
	redefining c_j^h for	$j = 1$

Remarks: Typically the family $T_h\varphi_j$, $h \in H$ is not linearly independent, which is the case, if one has frame and not a Riesz basis. In this case the $d_{j,N}^h$ are of course not uniquely determined. Different choices of $d_{j,N}^h$ leed to different coefficients in the above algorithm. Although typically one obtains the minimal-norm coefficients (with respect to the ψ_j) we do not claim that this will result in minimal norm coefficients when translated to sums involving the φ_j. We also point out once more, that the order in which the spaces are given (the "hierarchy") will play a significant role in the calculation of the actual coefficients.

Concerning the original question of matching pursuit we have to make the following remark. Whereas we have been describing in detail the method of converting a *best* approximation it is clear that *any* approximation by linear combinations of the building blocks ψ can be converted into a sum involving the original atoms generating the spaces V_j. Thus an alternative of obtaining good approximations with few terms might be to carry out matching pursuit methods *within the different spaces* W_j. There are two major advantages compared to the *global* matching pursuit using the full dictionary: First of all one has a comparatively small dictionary (the shifted version of the generating atoms ψ), and a lot of additional structure and therefore FFT-methods are applicable in practice to speed up the matching pursuit. Secondly, the matching pursuit can be carried out independently in the different (pairwise orthogonal) subspaces W_j. This is the reason for us to speak of *parallel matching pursuit*. Another feature, which we plan to discuss in more details elsewhere is the possibility to pick out many atoms at once (delivering faster the more relevant coefficients). The point for this feature being the fact that for the case of

good decay of the functions ψ one can pick several building blocks simultaneously, because they are automatically "almost orthogonal" if the are positioned sufficiently far apart from each other. In other words we expect effects similar to orthogonalization without substantially increased computational costs. Of course, this points will require a more detailed explanation and numerical experiments which are to be given elsewhere.

4 Discrete Implementation

For implementations of the above methods it is necessary to go into the (finite) discrete setting. As can be suspected, everything turns out to be less complicated, if we choose the space $l^2((\mathbb{Z}_n)^m)$ for some $n \geq 1$ instead of $L^2(\mathbb{R}^m)$. In the following, we restrict our attention to the case $m = 1$. Then of course, $l^2(\mathbb{Z}_n)$ is naturally identified with the complex n-space $\mathbb{C}^n$. The advantage of interpreting vectors of length n as functions over the unit roots or order n is of course that there is a natural interpretation of *cyclic* rotation instead of ordinary shift of signals in $L^2(R^m)$ and resulting therefrom the obvious connection to FFT-methods. In this interpretation one can think of an element from $l^2(\mathbb{Z}_n)$ as representing n consecutive samples of some n-periodic, discrete signal. The subgroup H of $\mathbb{R}^m$ is substituted by a subgroup of $\mathbb{Z}_n$, which has to be isomorphic to some $\mathbb{Z}_d$, where n is divided by d, i.e. there exists $a \geq 1\, a \in \mathbb{Z}$, such that $n = ad$. In other words, the discrete subgroup is given by $H = a\mathbb{Z}_n$, and the collection of building blocks for a single space V_j or W_j are obtained by cyclic rotation of a given signal by multiples of a.

If one wants to work with many different subgroups in order to imitate the "continuous" situation, we advise the user to choose a natural number $n \in \mathbb{Z}$ that is highly composite (e.g. $n = 360$ or $n = 2520$.

One further advantage if the finite/discrete setting is the fact, that every finite set $\{\xi_1, \ldots, \xi_d\}$ of vectors from $l^2(\mathbb{Z}_n) \cong \mathbb{C}^n$ that generates a subspace S of $\mathbb{C}^n$ is a frame for S.

This follows from linear algebra. The linear mapping $X : f \mapsto (\langle f, \xi_i \rangle)_{i=1,\ldots,d}$ between the finite-dimensional vector spaces S and $\mathbb{C}^d$ is automatically bounded, thus implying that

$$\|(\langle f, \xi_i \rangle)_{i=1,\ldots,d}\|^2 = \sum_{i=1}^{d} |\langle f, \xi_i \rangle|^2 \leq B\|f\|^2 \qquad B < \infty \ .$$

The left inequality in (9) follows from the observation that $X : S \to \mathbb{C}^d$ is one-to-one, since $\{\xi_1, \ldots, \xi_d\}$ span S. Therefore 0 is the only element in S which is orthogonal to all ξ_i $i = 1, \ldots, d$. Hence we can invert $X : S \to Im(X) \subseteq \mathbb{C}^d$, which is also onto $Im(X)$, the image of X. We arrive at a linear mapping $X^{-1} : Im(X) \to \mathbb{C}^n$, which is bounded by the same argument from linear algebra as above. This means

$$\|X^{-1}g\|^2 \leq A'\|g\|^2 \qquad g \in Im(X) \tag{17}$$

with $A' < \infty$. If we apply X^{-1} to $(\langle f, \xi_i \rangle)_{i=1,\ldots,d}$ (17) converts to

$$\begin{aligned} \|f\|^2 = \|X^{-1}(\langle f, \xi_i \rangle)_{i=1,\ldots,d}\|^2 &\leq A'\|(\langle f, \xi_i \rangle)_{i=1,\ldots,d}\|^2 \\ &= A' \sum_{i=1}^{d} |\langle f, \xi_i \rangle|^2 \end{aligned}$$

And since $A' < \infty$ it follows, that

$$A\|f\|^2 \leq \sum_{i=1}^{d} |\langle f, \xi_i \rangle|^2 \ ,$$

where $A = A'^{-1} > 0$.

If we are given a *coherent* frame $(\xi_h)_{h\in H}$ in a Hilbert space E, where the frame is generated from a single vector under the group action of H, the orthogonal projection onto its linear span is given by

$$P_V f = \sum_{h\in H} \langle f, \tilde{\xi}_h \rangle \xi_h \tag{18}$$

or alternatively

$$P_V f = \sum_{h\in H} \langle f, \xi_h \rangle \tilde{\xi}_h \tag{19}$$

see the paper by Oakley et al[6] for a description of the corresponding continuous problem which can be easily adapted to the discrete case. Here $V = \overline{\langle \xi_h : h \in H \rangle}$ and $\{\tilde{\xi}_h : h \in H\}$ is the dual frame of $\{\xi_h : h \in H\}$. Let us consider the finite-dimensional case of $E = \mathbb{C}^n$ with $n = a \cdot d$ and $H \cong Z_d$, which acts through cyclic rotations by multiples of a, $h = a(s-1), 1 \le s \le d$. The elements of the dual frame are simply the rows of the transpose of the pseudo inverse M^+, of the matrix M, whose h-th row is ξ_h. This was pointed out in [CFMSS]. Therefore we can write the orthogonal projection of any $f \in \mathbb{C}^n$ onto the space V spanned by ξ_h, $h = 1, \ldots, d$ as

$$P_V f = \left(f \cdot M^+\right) \cdot M \ .$$

For practical purposes and in order to obtain an efficient implementation one should observe [5] that the formula $M^+ = M^* \cdot (M \cdot M^*)^+$ (where M^* denotes the conjugate transpose of the matrix M) allows to reduce the problem of pseudo-inversion to that of a H-circulant Gram matrix $M \cdot M^*$ whose size corresponds to the number of elements of H and therefore can be obtained quickly using the natural FFT for the (abelian) group H. We will not pursue this (important) aspects further in this presentation, but emphasize at least here that this aspect is crucial for the overall performance of any algorithm of the kind explained here.

Coming back to the algorithm we observe that in the discrete setting ψ_k can be calculated recursively by

$$\psi_k = \varphi_k - \left(\varphi_k \cdot M_{k-1}^+\right) \cdot M_{k-1} - \ldots \left(\varphi_k \cdot M_1^+\right) \cdot M_1 \tag{20}$$

where M_j is the matrix, whose h-th row is $T_h \psi_j$. Remember that $\psi_1 = \varphi_1$. Having obtained all M_j , $j = 1, \ldots, N$ by (20) we can calculate the orthogonal projection of an $f \in \mathbb{C}^n$ on any W_j by

$$\left(f \cdot M_j^+\right) \cdot M_j$$

and thereby the orthogonal projection on $U = V_1 + \ldots + V_N = W_1 \oplus \ldots \oplus W_N$ is

$$\left(f \cdot M_1^+\right) \cdot M_1 + \ldots + \left(f \cdot M_N^+\right) \cdot M_N \ .$$

However we want to obtain not only $P_U f$ but an expansion

$$P_U f = \sum_{j=1}^{N} \sum_{h\in H} c_j^h T_h \varphi_j \ .$$

This can be achieved by Algorithm 1, if we are given the a_j^h and $d_{j,k}^h$, $j = 1, \ldots, N$, $j \le k-1$, $h \in H$. Those can be calculated as:

$$\begin{aligned} (a_j^h)_{h\in H} &= f \cdot M_j^+ \\ (d_{j,k}^h)_{h\in H} &= \varphi_k \cdot M_j^+ \end{aligned}$$

Summing up we have the following algorithm (where all matrices are understood as collections of row vectors carrying the building blocks):

ALGORITHM 2:

First Part	:	Calculation of the ψ_j	(with $n = a \cdot d$)
$j = 1$	:	$\psi_1 = \varphi_1$	$M_1 = (T_{sa}\psi_1)_{0 \le s < d}$
$j = 2$	:	$\psi_2 = \varphi_2 - (\varphi_2 \cdot M_1^+) \cdot M_1$	$M_2 = (T_{sa}\psi_2)_{0 \le s < d}$
		$d_{1,2} = \varphi_2 \cdot M_1^+$	
		$\vdots$	
$j = N$	:	$\psi_N = \varphi_N - (\varphi_N \cdot M_1^+) \cdot M_1$	$M_N = (T_{sa}\psi_N(l))_{0 \le s < d}$
		$- \ldots - (\varphi_N \cdot M_{N-1}^+) \cdot M_{N-1}$	
		$d_{1,N} = \varphi_N \cdot M_1^+$	
		$\vdots$	
		$d_{N-1,N} = \varphi_N \cdot M_{N-1}^+$	
Second Part	:	Calculation of $P_U f$	
		$P_U f = (f \cdot M_1^+) \cdot M_1 + \ldots$	
		$+ (f \cdot M_N^+) \cdot M_N$	
		$a_j = f \cdot M_j^+$	$j = 1, \ldots, N$
Third Part	:	Calculation of c_j	$j = 1, \ldots, N$
		Apply Algorithm 1 to	
		the sequences a_j	$j = 1, \ldots, N$
		and $d_{j,k}$	$k \le j - 1$

Observe that the first part of the above algorithm has to be performed only once, if the family φ_j is unchanged.

Acknowledgements

The authors acknowledge partial support from the project PH 08784 at the Austrian Science foundation FWF.

5 REFERENCES

[1] Chui C. *An Introduction to Wavelets.* Academic Press, 1992.

[2] C. Cenker, H. G. Feichtinger, M. Mayer, H. Steier, and T. Strohmer. New variants of the POCS method using affine subspaces of finite codimension, with applications to irregular sampling. *Conf. SPIE 92 Boston*, pages 299–310, 1992.

[3] I. Daubechies. Ten Lectures on Wavelets. *CBMS-NSF Reg. Conf. Series in Applied Math.*, 1992.

[4] C. DeBoor, R. DeVore, and Ron.A. The structure of finitely generated shift-invariant spaces in $L^2(\mathbb{R}^d)$. *J.Funct.Anal.*, 119:37–78, 1994.

[5] H. G. Feichtinger. Pseudo-inverse matrix methods for signal reconstruction from partial data. In *SPIE-Conf., Visual Comm. and Image Proc., Boston*, pages 766–772, 1991. Int.Soc.Opt.Eng.

[6] J. P. Oakley, M. J. Cunningham, and G. Little. A Fourier-domain formula for the least squares projection of a function on a repititive basis in N-dimensional space. *IEEE Trans. ASSP*, 38(1):114–120, 1990.

[7] A. Ron and Z. Shen. Frames and stable bases for shift-invariant subspaces of $L^2(\mathbb{R}^d)$. *submitted*, 1993.

[8] Mallat S. and Zhang Z. Matching pursuits with time-frequency dictionaries. *IEEE Trans. SSP*, 41/12:3397–3415, 1993.

[9] H. Stark, editor. *Image Recovery: Theory and application*. Acad. Press, 1987.

[10] Pati Y.C., Rezaiifar R., and Krishnaprasad P.S. Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition. In *Proc. 27th Annual Asimolar Conference on Signals, Systems and computers*, 1993.

[11] D. C. Youla and Webb H. Image Restoration by the Method of Convex Projections, Part1, Theory. *IEEE Trans. Med Imaging 1*, pages 81–94, 1982.