



HAL
open science

Beyond coherence: recovering structured time-frequency representations

Lasse Borup, Rémi Gribonval, Morten Nielsen

► **To cite this version:**

Lasse Borup, Rémi Gribonval, Morten Nielsen. Beyond coherence: recovering structured time-frequency representations. [Research Report] 2007, pp.13. inria-00130123v1

HAL Id: inria-00130123

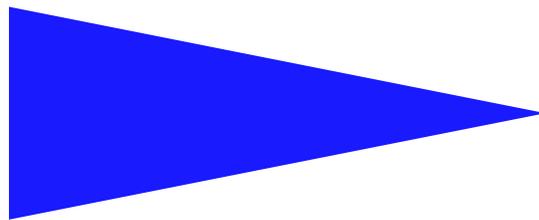
<https://inria.hal.science/inria-00130123v1>

Submitted on 9 Feb 2007 (v1), last revised 22 Jun 2007 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PUBLICATION
INTERNE
N° 1833



BEYOND COHERENCE : RECOVERING STRUCTURED
TIME-FREQUENCY REPRESENTATIONS.

LASSE BORUP , RÉMI GRIBONVAL , MORTEN NIELSEN

Beyond coherence : recovering structured time-frequency representations.

Lasse Borup^{*} , Rémi Gribonval^{**} , Morten Nielsen^{***}

Systemes cognitifs
Projet Metiss

Publication interne n° 1833 — Février 2007 — 13 pages

Abstract: We consider the problem of recovering a structured sparse representation of a signal in an overcomplete time-frequency dictionary with a particular structure. For infinite dictionaries that are the union of a nice wavelet basis and a Wilson basis, sufficient conditions are given for the Basis Pursuit and (Orthogonal) Matching Pursuit algorithms to recover a structured representation of an admissible signal. The sufficient conditions take into account the structure of the wavelet/Wilson dictionary and allow very large (even infinite) support sets to be recovered even though the dictionary is highly coherent.

Key-words: sparse representation, overcomplete dictionary, matching pursuit, basis pursuit, time-frequency dictionaries, wavelets, Wilson basis, linear programming, greedy algorithm, , identifiability, inverse problem.

(Résumé : *tsvp*)

This paper has been submitted for possible publication to Applied and Computational Harmonic Analysis. This work is supported in part by the European Union's Human Potential Programme, under contract HPRN-CT-2002-00285 (HASSIP).

* Department of Mathematical Sciences, Aalborg University, Fr. Bajers Vej 7G, DK-9220 Aalborg East, Denmark, lasse@math.aau.dk

** remi.gribonval@irisa.fr.

*** Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7 G, DK-9220 Aalborg East, Denmark, mnielsen@math.wustl.edu

Sur l'identification de représentations temps-fréquence parcimonieuses structurées

Résumé : Nous nous intéressons au comportement des principaux algorithmes d'approximation parcimonieuse sur des signaux admettant une représentation à la fois parcimonieuse et structurée dans des dictionnaires temps-fréquence particuliers constitués de l'union d'une base d'ondelette et d'une base de Wilson, qui sont utilisés pour certains schémas de codage audio dits hybrides. Il s'agit de déterminer des conditions de "structure" suffisantes garantissant qu'une représentations d'un signal dans ce type de dictionnaire soit retrouvée par les algorithmes de type *Matching Pursuit* (MP) et *Basis Pursuit*. Les conditions suffisantes les plus classiques, basées sur la cohérence (cumulative) du dictionnaire, sont en effet inopérantes dans ce contexte car elles amènent à se restreindre à des signaux qui coïncident avec un atome du dictionnaire, et les résultats probabilistes valables pour des dictionnaires aléatoires ne peuvent être directement appliqués sur un tel dictionnaire très structuré. En supposant que le dictionnaire est généré à partir de fonctions (ondelette mère et fenêtre d'analyse) suffisamment régulières, nous exprimons ici des conditions suffisantes beaucoup moins pessimistes qui prennent en compte la structure particulière du dictionnaire hybride ondelettes/Wilson. Les conditions obtenues reposent sur le fait que ce dictionnaire peut être décomposé en sous-ensembles d'ondelettes et de fonctions de base de Wilson approximativement associés à des "blocs" temps-fréquence disjoints, et qui forment essentiellement l'union de deux bases maximale-ment incohérentes des fonctions ayant ces blocs pour support.

Mots clés : représentation parcimonieuse, dictionnaire redondant, matching pursuit, basis pursuit, dictionnaire temps-fréquence, ondelette, base de Wilson, programmation linéaire, algorithme glouton, séparation de sources, traitement du signal multicanal, identifiabilité, problème inverse.

Contents

1	Introduction	4
2	Wavelet and local cosine dictionary	6
2.1	Basis functions	6
2.2	Cumulative coherence	7
2.3	Sketch with time-frequency blocks	7
2.4	Inner products between wavelet and local cosine basis functions	8
2.5	Upper bound on the cumulative coherence	9
3	Conclusion	11
A	Time-frequency estimates	12

1 Introduction

Let $\Phi = [\mathbf{g}_i]_{i \in F}$ be an at most countable collection of normalized elements in a Hilbert space \mathcal{H} . We say that Φ is a *besselian dictionary* if the associated linear map $\Phi : \ell^2(F) \rightarrow \mathcal{H}$ given by $\Phi[(c_k)_k] = \sum_{i \in F} c_k \mathbf{g}_i$ is bounded. In this note we consider the problem of recovering a sparse representation S

$$X = \Phi(S), \quad S \in \ell^2(F), \quad (1)$$

of a signal $X \in \mathcal{H}$ relative to a besselian dictionary Φ with a specific structure. Sparse representations provide a very useful tool to solve many problems in signal processing including blind source separation, feature extraction and classification, denoising, and detection, to name only a few (see also [?], and references therein). Several algorithms, such as Basis Pursuit (ℓ^1 -minimization) and Matching Pursuits (also known as greedy algorithms), have been introduced to compute sparse representations/approximations of signals. The problem we face is that such algorithms *a priori* only provide sub-optimal solutions. That is, we do get a representation of the type (1), but we may not recover the sparse representation S of X .

Several recent papers [?, ?, ?, ?, ?, ?] have identified situations where algorithms such as Basis Pursuit actually compute an optimal representation of a given signal, in the sense that they solve the best approximation problem under a constraint on the size of the support of the signal. Typically, one calculates the coherence of Φ

$$\mu(\Phi) = \sup_{i \neq j} |\langle \mathbf{g}_i, \mathbf{g}_j \rangle|.$$

Then for signals X with a representation $X = \Phi(S)$ satisfying $|\text{supp}(S)| < \lfloor \frac{1}{2}(1 + 1/\mu) \rfloor$, Basis Pursuit will recover the representation S . One serious problem with this type of results using the coherence is that they represent *worst case* estimates. For example, the coherence is close to one as soon as we have one pair of atoms that are approximately colinear while the rest of the dictionary may be much nicer. A more refined type of result can be obtained by considering the cumulative coherence introduced by Tropp [?]

$$\mu_1(\Phi, m) := \sup_{|\Lambda|=m} \sup_{j \notin \Lambda} \sum_{i \in \Lambda} |\langle \mathbf{g}_i, \mathbf{g}_j \rangle|.$$

However, the cumulative coherence also gives a *worst case* estimate that does not take into account the finer structure of the dictionary, and the mentioned bounds are too weak for many applications.

One way to overcome these shortcomings is by shifting to a probabilistic viewpoint and consider *random dictionaries*. The probabilistic approach has been considered in a number of recent papers, see e.g. [?, ?]. Random dictionaries are typically created by picking a number of unit vectors randomly from some larger ensemble. The results on sparse representations using random dictionaries are typically much better than the corresponding deterministic results. One problem is that the results are difficult to interpret when we consider a specific dictionary.

We follow a different deterministic path in this note. The goal is to give more optimistic results for some concrete dictionaries that are often used in signal processing and harmonic analysis. The idea is to take into account some of the internal structure of the dictionary.

The typical example of an admissible dictionary is the union of a nice wavelet basis and a Wilson basis.

The main tool to extend the classical estimates is to consider the *setwise p -Babel function*, which trivially extends the setwise Babel function defined in [?] as follows. For $1 \leq p < \infty$ and a set $I \subseteq F$ we define

$$\mu_p(\Phi, I) := \left(\sup_{i \notin I} \sum_{j \in I} |\langle \mathbf{g}_i, \mathbf{g}_j \rangle|^p \right)^{1/p}. \quad (2)$$

For \mathcal{S} a family of subsets of F , we define the *structured p -Babel function* as

$$\mu_p(\Phi, \mathcal{S}) := \sup_{I \in \mathcal{S}} \mu_p(\Phi, I). \quad (3)$$

Notice that we allow infinite dictionaries Φ so it may happen that $\mu_p(\Phi, I) = +\infty$. The structured 1-Babel function $\mu_1(\Phi, \mathcal{S})$ generalizes the Babel function $\mu_1(\Phi, m)$. In fact, let $\mathcal{S}_m = \{I \subseteq F : |I| = m\}$, $m = 1, 2, \dots, |F|$. Then $\mu_1(\Phi, m) = \mu_1(\Phi, \mathcal{S}_m)$. The case $p = 1$ is especially interesting due to the following result considered by Tropp [?] for the Babel function. The proof of Lemma 1 is a straightforward generalization of the proof in [?] and was pointed out in [?]. It involves the sub-dictionary $\Phi_I = [\mathbf{g}_i]_{i \in I}$ made of atoms from the set $I \subseteq F$.

Lemma 1. *Let $X = \Phi(S)$ and suppose that $\text{supp}(S) = I$ is such that*

$$\mu_1(\Phi, I) + \max_{\ell \in I} \mu_1(\Phi_I, I \setminus \{\ell\}) < 1. \quad (4)$$

Then Basis Pursuit and Orthogonal Matching Pursuit exactly recover the representation S of X .

In the special case where the nonzero coefficients in the representation S have similar magnitudes $|c_k| \approx cst$, the condition (4) is also sufficient to ensure that simple thresholding will recover S [?, ?]. A similar condition involving the 2-Babel function instead of the 1-Babel function was recently shown to be related [?] to the probability of recovery with thresholding for simultaneous (multichannel) sparse approximation. All recovery results, which are expressed here for noiseless models $X = \Phi(S)$, have been shown to be stable to noise.

To illustrate how we can use dictionary structure to get improved recovery results, let us consider two examples. The first example is finite dimensional. In \mathbb{C}^N we consider Φ_{FD} given as the union of the Dirac and the Fourier orthonormal basis for \mathbb{C}^N . One easily checks that the dictionary is maximally incoherent with $\mu(\Phi_{FD}) = 1/\sqrt{N}$. Thus, the classical result shows that Basis pursuit recovers sparse signals having a representation with support less than $\lfloor \frac{1}{2}(1 + \sqrt{N}) \rfloor$ atoms. Here we only estimate the size of the support.

The second example is infinite dimensional. We consider Φ_{HW} defined as the union of the Haar system (see e.g. [?]) $\mathcal{B}_1 = \{h_n\}_{n=0}^{\infty}$ and the Walsh system (see [?]) $\mathcal{B}_2 = \{W_n\}_{n=0}^{\infty}$ on $[0, 1]$. The Haar system and the Walsh system both form an orthonormal basis for $L^2([0, 1])$ so Φ_{HW} is a tight frame for $L^2([0, 1])$. The point we wish to make is the following; it is easy to check that $\mu(\Phi_{HW}) = 1$, so *naively* one would expect that no decent recovery result is possible. However, Φ_{HW} has a lot of *structure* we can exploit. In the time-frequency plane, the Walsh function W_n is supported in the block $[0, 1] \times [n, n + 1]$ while

the Haar function h_n with $2^j \leq n < 2^{j+1}$ is supported on $H_{j,0} := [0, 1] \times [2^j, 2^{j+1} - 1]$. In fact, let $Q(j) = \{h_k\}_{k=2^j}^{2^{j+1}-1}$ and $Q'(j) = \{W_k\}_{k=2^j}^{2^{j+1}-1}$. Then $W_j := \text{span } Q(j) = \text{span } Q'(j)$ with $W_j \perp W_{j'}$ for $j \neq j'$, see [?]. Moreover, the 2^j -dimensional subdictionary $\Phi_{HW}(j) := Q(j) \cup Q'(j)$ is perfectly incoherent with $\mu(\Phi_{HW}(j)) = 2^{-j/2}$. Hence, if a signal x has a representation

$$x = c_0 + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} (c_{j,k} h_{2^j+k} + d_{j,k} W_{2^j+k}),$$

with $|\text{supp}(\{c_{j,k}\}_k)| + |\text{supp}(\{d_{j,k}\}_k)| < \frac{1}{2}(1 + 2^{j/2})$ for $j \geq 0$ (the notation $\text{supp}()$ stands for the support set where a sequence is nonzero, and $|\cdot|$ denotes the cardinality of such a set), then we can use the simple finite dimensional estimate using the coherence (and the fact that $W_j \perp W_{j'}$ for $j \neq j'$) to conclude that MP and BP recover this representation of x . Notice how this estimate takes into account the structure of the dictionary and not only the size of the support of signals.

The main result of this paper is to extend the straightforward considerations for Φ_{HW} to other dictionaries with the same type of underlying structure. However, we will not assume that the dictionary can be decomposed into orthogonal finite dimensional dictionaries which will give rise to some added technicalities in the estimates. Our result holds for unions of an orthonormal wavelet basis $\{\psi_{j,k}\}$ and a Wilson basis $\{g_{n,m}\}$ with sufficient smoothness, a type of dictionary which was proposed for audio signal modeling and compression by Daudet and Torr sani [?]. For any pair $\mathbf{c} := c_{j,n}$ and $\mathbf{d} := d_{n,m}$ of coefficient sequences we define

$$N_j(\mathbf{c}, \mathbf{d}) := \sup_{n \in \mathbb{Z}} \max (|\text{supp}(\{c_{j,2^j n + \ell}\}_{\ell=0}^{2^j-1})|, |\text{supp}(\{d_{n,2^j + \ell}\}_{\ell=0}^{2^j-1})|).$$

Theorem 1. *There is a constant K (which depends on the support size and smoothness of the mother wavelet ψ and Wilson window function g) such that any pair of sequences satisfying*

$$\sum_{j \geq 0} N_j(\mathbf{c}, \mathbf{d}) \cdot 2^{-j/2} < K$$

will be recovered by both Basis Pursuit and (Orthonormal) Matching Pursuit performed on the signal $x = \sum_{j=0}^{\infty} \sum_{n \in \mathbb{Z}} \sum_{\ell=0}^{2^j-1} (c_{j,2^j n + \ell} \psi_{j,2^j n + \ell} + d_{n,2^j + \ell} g_{n,2^j + \ell})$.

2 Wavelet and local cosine dictionary

In this section we introduce the main function dictionary considered in this paper. The dictionary is the union of an orthonormal wavelet and local cosine basis and is consequently a tight frame with frame constant 2. We will not discuss the details involved in the construction of these bases here, but just refer the reader to e.g. [?]. To avoid unnecessary technicalities, we only consider the univariate case.

2.1 Basis functions

Let ϕ and ψ be a scaling function and a wavelet, both with compact support, such that $\mathcal{B}^1 := \{\phi_k\}_{k \in \mathbb{Z}} \cup \{\psi_{j,k}\}_{j \geq 0, k \in \mathbb{Z}}$, is an orthonormal wavelet basis for $L^2(\mathbb{R})$, where

$$\psi_{j,k}(x) := 2^{j/2} \psi(2^j x - k) \quad \text{and} \quad \phi_k(x) := \phi(x - k).$$

Suppose that g is a smooth compactly supported “cut-off” function, and let

$$g_{n,m}(x) := \sqrt{2}g(x-n) \cos(\pi(m + \frac{1}{2})(x-n))$$

for $n \in \mathbb{Z}$ and $m \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For a suitable choice of g , $\mathcal{B}^2 := \{g_{n,m}\}_{n \in \mathbb{Z}, m \in \mathbb{N}_0}$ is an orthonormal basis for $L^2(\mathbb{R})$, see [?, Sec. 1.4]. The besselian dictionary considered is the tight frame $\Phi = \mathcal{B}^1 \cup \mathcal{B}^2$. It is indexed by $F = F^\phi \cup F^\psi \cup F^g$ where F^ϕ (resp. F^ψ , F^g) indexes scaling functions (respectively wavelets, local cosines).

2.2 Cumulative coherence

We may partition any index set $I \subset F$ into scaling function indices $I^\phi = I \cap F^\phi$, wavelet indices $I^\psi = I \cap F^\psi$ and local cosine indices $I^g = I \cap F^g$. Since each basis is orthogonal, the p -cumulative coherence of I is given by

$$\begin{aligned} \mu_p^p(\Phi, I) &= \max \left\{ \sup_{\psi' \in (F^\phi \cup F^\psi) \setminus I} \left(\sum_{\psi \in I^\phi \cup I^\psi} |\langle \psi', \psi \rangle|^p + \sum_{g \in I^g} |\langle \psi', g \rangle|^p \right), \right. \\ &\quad \left. \sup_{g' \in F^g \setminus I} \left(\sum_{\psi \in I^\phi \cup I^\psi} |\langle g', \psi \rangle|^p + \sum_{g \in I^g} |\langle g', g \rangle|^p \right) \right\} \\ &\leq \max \left\{ \sup_{\psi' \in F^\phi \cup F^\psi} \sum_{g \in I^g} |\langle \psi', g \rangle|^p, \sup_{g' \in F^g} \sum_{\psi \in I^\phi \cup I^\psi} |\langle g', \psi \rangle|^p \right\} \end{aligned} \quad (5)$$

where we slightly abused notations by confusing basis functions with their indices, e.g., in the notation $g \in I^g$.

2.3 Sketch with time-frequency blocks

To estimate each of the two terms which appear in the maximum (5) we will partition further the index sets I^ϕ , I^ψ and I^g . Given $j \in \mathbb{N}_0$ and $n \in \mathbb{Z}$, it is easy to see that for a nice mother wavelet ψ , the 2^j functions $\{\psi_{j,k}\}_{2^j n \leq k < 2^j(n+1)}$ are essentially localized in time in the neighborhood of the interval $[n, n+1]$, and essentially localized in frequency in the neighborhood of the interval $[2^j, 2^{j+1}]$. In other words, they are localized around the “time-frequency block” $H_{j,n} := [n, n+1] \times [2^j, 2^{j+1}]$. The same goes for the 2^j local cosine functions $\{g_{n,m}\}_{2^j \leq m < 2^{j+1}}$, if g is well-localized in time and frequency. The coherence between any wavelet and Wilson function “living” on such a block is of the order $2^{-j/2}$. The distinct regions $\{H_{j,n}\}_{n \in \mathbb{Z}, j \in \mathbb{N}_0}$ essentially¹ tile the time-frequency plane, and in contrast to the relatively large coherence between functions from the same tile of the partition, the coherence between any two functions in two different pieces is (much) smaller. Thus, we may cut the sets I^ψ and I^g into pieces in parallel to the tiling of the time-frequency plane, and define for $j \geq 0$ and $n \in \mathbb{Z}$

$$\begin{aligned} I_{j,n}^\psi &:= \{\psi_{j,k} \in I^\psi, 2^j n \leq k < 2^j(n+1)\} \\ I_{j,n}^g &:= \{g_{n,m} \in I^g, 2^j \leq m < 2^{j+1}\}. \end{aligned}$$

¹For a complete tiling one would also need to include low frequency regions $[n, n+1] \times [0, 1]$

For a given wavelet $\psi_{j,k} \notin I^\psi$, letting $n := \lfloor 2^{-j}k \rfloor$ be such that $2^j n \leq k < 2^j(n+1)$ we have

$$\sum_{g \in I^g} |\langle \psi_{j,k}, g \rangle|^p = \sum_{j' \geq 0} \sum_{n' \in \mathbb{Z}} \sum_{g \in I_{j',n'}^g} |\langle \psi_{j,k}, g \rangle|^p \approx \sum_{g \in I_{j,n}^g} |\langle \psi_{j,k}, g \rangle|^p \approx |I_{j,n}^g| \cdot 2^{-jp/2} \quad (6)$$

provided that the above sketchy analysis is valid. A similar estimate holds if the role of the wavelet and local cosine bases is exchanged, and the numbers

$$N_j(I) := \sup_{n \in \mathbb{Z}} \max \left(|I_{j,n}^\psi|, |I_{j,n}^g| \right) \quad (7)$$

are therefore involved in the estimation of the coherence. Indeed, if we assume that the index set I does not contain any scaling function (i.e., $I^\phi = \emptyset$) and no Wilson function $g_{n,0}$ either, we get the following estimate of the p -cumulative coherence

$$\mu_p(\Phi, I) \approx \sup_{j \geq 0} \left(2^{-j/2} N_j(I)^{1/p} \right). \quad (8)$$

The above approach is only a sketch: in practice the inner products between wavelets and local cosine functions do not depend as sharply as depicted here on the time-frequency block to which their indices belong, and we will see below how the above approach should be corrected.

Note that we assumed in this sketch that the index set I did not contain any low-frequency atom (i.e., no scaling function and no Wilson function of the type $g_{n,0}$). This restriction is only natural since scaling functions and Wilson functions of this type have very similar shapes, and have a very large coherence, so if they were to be included in I then the p -cumulative coherence would almost certainly exceed one.

2.4 Inner products between wavelet and local cosine basis functions

To estimate the p -coherence from above we will need to control the inner products $|\langle \psi_{j,k}, g_{n,m} \rangle|$ between wavelets and local cosine functions. The following lemmata give the fundamental estimates and are proved in Appendix A.

Lemma 2. *Let ϕ , ψ , and g be three univariate functions such that for some $C < \infty$ and $A > 1$ we have*

$$\max(|\widehat{\phi}(\xi)|, |\widehat{\psi}(\xi)|, |\widehat{g}(\xi)|) \leq C (1 + |\xi|)^{-A}. \quad (9)$$

Then, for all $k, n \in \mathbb{Z}$, and $m, j \in \mathbb{N}_0$, we have

$$\begin{aligned} |\langle \psi_{j,k}, g_{n,m} \rangle| &\leq \tilde{C} \cdot 2^{-j/2} (1 + 2^{-j}|m|)^{-A}, \\ |\langle \phi_k, g_{n,m} \rangle| &\leq \tilde{C} \cdot (1 + |m|)^{-A}, \end{aligned}$$

with

$$\tilde{C} := (1 - A^{-1})^{-1} \cdot 2^{3/2} (2^A + 3^A) C^2. \quad (10)$$

Lemma 2 is proved using the frequency localization of the wavelet and Wilson basis. The next simple lemma uses the time localization to obtain other estimates of the inner products.

Lemma 3. *Suppose that $\text{supp}(\phi), \text{supp}(\psi) \subseteq [-\lambda, \lambda]$ for some $\lambda < \infty$, and $\text{supp}(g) \subseteq [-1/2, 3/2]$. Then we have the estimates*

$$|\langle \psi_{j,k}, g_{n,m} \rangle| \leq \begin{cases} 2^{-j/2+1} \min\{\lambda, 2^j\}, & \text{if } 2^j(n - 1/2) - \lambda < k < 2^j(n + 3/2) + \lambda \\ 0, & \text{else,} \end{cases}$$

and

$$|\langle \phi_k, g_{n,m} \rangle| \leq \begin{cases} 2\lambda, & \text{if } n - 1/2 - \lambda < k < n + 3/2 + \lambda \\ 0, & \text{else.} \end{cases}$$

for all $j, k, n \in \mathbb{Z}$, and $m \in \mathbb{N}$.

2.5 Upper bound on the cumulative coherence

Using the estimates above we can now upper bound the cumulative coherence as expressed in the following result.

Theorem 2. *Let ϕ , ψ , and g be three univariate functions with $\text{supp}(\phi), \text{supp}(\psi) \subseteq [-\lambda, \lambda]$ for some $\lambda < \infty$, and $\text{supp}(g) \subseteq [-1/2, 3/2]$. Assume that for some $C < \infty$ and $A > 1$ we have*

$$\max(|\widehat{\phi}(\xi)|, |\widehat{\psi}(\xi)|, |\widehat{g}(\xi)|) \leq C(1 + |\xi|)^{-A}. \tag{11}$$

Consider a support set I which does not contain any scaling function, and no Wilson function $g_{n,0}$ of frequency index 0 either. Then, for any p ,

$$\mu_p^p(\Phi, I) \leq 2(\lambda + 1) \cdot \tilde{C}^p \cdot \sum_{j \geq 0} N_j(I) \cdot 2^{-jp/2}. \tag{12}$$

with \tilde{C} given by (10).

Comparing this result with the sketch (8) we notice that, in addition to a constant factor, the supremum over j has been replaced by a sum, which is quite strong but does not fundamentally change the rate at which $N_j(I)$ can grow with j . One can compare it to going from a weak ℓ_1 norm to a strong ℓ_1 -norm.

Proof. First we estimate $\sup_{g' \in F^g} \sum_{\psi \in I \phi \cup I \psi} |\langle g', \psi \rangle|^p$. For that we consider a given Wilson basis function $g' = g_{n',m'} \in F^g$. By Lemma 2 we have, for all $j \geq 0$ and $k \in \mathbb{Z}$

$$|\langle g', \psi_{j,k} \rangle| \leq \tilde{C} \cdot 2^{-j/2} (1 + 2^{-j}|m'|)^{-A}.$$

By Lemma 3 the only indices k which may yield a nonzero inner product are in the interval $(2^j(n' - \frac{1}{2}) - \lambda, 2^j(n' + \frac{3}{2}) + \lambda)$, which covers at most $2 + 2^{1-j}\lambda \leq 2(\lambda + 1)$ intervals of the form $[2^n \ell, 2^{j'}(n + 1))$. Therefore, we may take the p -th power and sum up to get

$$\sum_{n \in \mathbb{Z}} \sum_{\psi_{j,k} \in I_{j,n}^\psi} |\langle g_{n,m}, \psi_{j,k} \rangle|^p \leq 2(\lambda + 1) \cdot N_j(I) \cdot \tilde{C}^p \cdot 2^{-jp/2} \cdot (1 + 2^{-j}|m'|)^{-Ap}.$$

Since we assume I^ψ is empty, summing up over j gives

$$\sum_{\psi \in I^\phi \cup I^\psi} |\langle g', \psi \rangle|^p \leq 2(\lambda + 1) \cdot \tilde{C}^p \cdot \sum_{j \geq 0} N_j(I) \cdot 2^{-jp/2} \cdot (1 + 2^{-j}|m'|)^{-Ap}$$

and by taking the supremum over m , which is achieved at $m = 0$, it follows that

$$\sup_{g' \in F^g} \sum_{\psi \in I^\psi \cup I^\phi} |\langle g', \psi \rangle|^p \leq 2(\lambda + 1) \cdot \tilde{C}^p \cdot \sum_{j \geq 0} 2^{-jp/2} \cdot N_j(I) \quad (13)$$

Reversing now the roles between Wilson basis functions and wavelets we now want to estimate $\sup_{\psi' \in F^\phi \cup F^\psi} \sum_{g \in I^g} |\langle \psi', g \rangle|^p$. We consider a given wavelet $\psi' = \psi_{j', k'}$. By Lemma 2 we have for any $n \in \mathbb{Z}$, $m \geq 1$ and $j \geq 0$ such that $2^j \leq m < 2^{j+1}$

$$|\langle \psi', g_{n,m} \rangle| \leq \tilde{C} \cdot 2^{-j'/2} (1 + 2^{-j'}|m|)^{-A} \leq \tilde{C} \cdot 2^{-j'/2} (1 + 2^{-j'}2^j)^{-A}.$$

Moreover, by Lemma 3, the only indices n for which this inner product can be nonzero satisfy $-\frac{1}{2} - 2^{-j'}\lambda < n - 2^{-j'}k' < \frac{3}{2} + 2^{-j'}\lambda$, so there is at most $\lambda 2^{1-j'} + 1 \leq 2(\lambda + 1)$ of them. Therefore, taking the p -th power and summing we get

$$\sum_{n \in \mathbb{Z}} \sum_{g_{n,m} \in I_{j,n}^g} |\langle \psi', g_{n,m} \rangle|^p \leq 2(\lambda + 1) \cdot N_j(I) \cdot \tilde{C}^p \cdot 2^{-j'p/2} \cdot (1 + 2^{j-j'})^{-Ap}.$$

Since we assume I^g does not contain any $g_{n,m}$ with $m = 0$, summing up over j gives

$$\sum_{g \in I^g} |\langle \psi', g \rangle|^p \leq 2(\lambda + 1) \cdot \sum_{j \geq 0} N_j(I) \cdot \tilde{C}^p \cdot 2^{-j'p/2} \cdot (1 + 2^{j-j'})^{-Ap} \quad (14)$$

Similarly, for any scaling function $\psi' = \phi_k$ we obtain

$$\sum_{g \in I^g} |\langle \psi', g \rangle|^p \leq 2(\lambda + 1) \cdot \sum_{j \geq 0} N_j(I) \cdot \tilde{C}^p \cdot (1 + 2^j)^{-Ap}$$

and we notice that the right hand side is exactly that of (14) for $j' = 0$. Therefore we obtain

$$\sup_{\psi' \in F^\phi \cup F^\psi} \sum_{g \in I^g} |\langle \psi', g \rangle|^p \leq 2(\lambda + 1) \cdot \tilde{C}^p \cdot \sup_{j' \geq 0} \sum_{j \geq 0} \left(2^{-j'p/2} \cdot (1 + 2^{j-j'})^{-Ap} \cdot N_j(I) \right). \quad (15)$$

Since $A > \frac{1}{2}$ it follows that for any $\ell \in \mathbb{Z}$, $2^{\ell p/2} (1 + 2^\ell)^{-Ap} \leq 1$. Therefore, for any $j, j' \geq 0$ we have

$$2^{-j'p/2} \cdot (1 + 2^{j-j'})^{-Ap} = 2^{(j-j')p/2} \cdot (1 + 2^{j-j'})^{-Ap} \cdot 2^{-jp/2} \leq 2^{-jp/2}.$$

Combining these facts with (13) and (15) we get the desired result (12). \square

We have the following corollary.

Corollary 1. *Assume that a wavelet basis and a Wilson basis satisfy the decay conditions of Theorem 2, and consider a support set I which does not contain any scaling function, and no Wilson function $g_{n,0}$ of frequency index 0 either. If*

$$\sum_{j \geq 0} N_j(I) \cdot 2^{-j/2} < \frac{1}{4(\lambda + 1)\tilde{C}} \quad (16)$$

then all standard pursuit algorithms will (stably) recover the support of any combination of atoms from the support set I .

Proof. For any $\ell \in I$ we consider the subset $J_\ell = I \setminus \{\ell\}$ and notice that, for all j , $N_j(J_\ell) \leq N_j(I)$, therefore, applying Theorem 2 with $p = 1$ we get under the condition (16) that

$$\mu_1(\Phi, I) + \sup_{\ell \in I} \mu_1(\Phi_I, J_\ell) \leq \mu_1(\Phi, I) + \sup_{\ell \in I} \mu_1(\Phi, J_\ell) \leq 4(\lambda + 1)\tilde{C} \cdot \sum_{j \geq 0} N_j(I) \cdot 2^{-j/2} < 1.$$

□

Our theorem can also be combined with the main theorem in [?] to prove that, under a white Gaussian model on the coefficients S , if

$$\sum_{j \geq 0} N_j(I) \cdot 2^{-j} < \frac{1}{2(\lambda + 1)\tilde{C}^2} \quad (17)$$

then the probability that multichannel thresholding fails to recover the support set I decays exponentially fast with the number of channels.

Example 1. *The compactly supported Daubechies wavelets $\{\phi^N\}$, $\{\psi^N\}$ (filter length $2N$) satisfy $\text{supp}(\phi^N), \text{supp}(\psi^N) \subseteq [-N, N]$ with*

$$\max(|\widehat{\phi^N}(\xi)|, |\widehat{\psi^N}(\xi)|) \leq C(1 + |\xi|)^{-\mu N - 1},$$

with $\mu \approx 0.1887$, see [?, Chap. 7]. Thus we can apply Theorem 2 and its corollary with $A = \mu N + 1$ with any infinitely differentiable cut-off function g with $\text{supp}(g) \subseteq [-\frac{1}{2}, \frac{3}{2}]$.

3 Conclusion

In this note we have derived sufficient conditions for the Basis Pursuit and Matching Pursuit algorithms to recover structured representations of admissible signal with respect to an infinite dictionary given as the union of a nice wavelet basis and a Wilson basis. The sufficient conditions, although quite natural given the known coherence results for finite dictionaries, take into account the time-frequency structure of the dictionary and are thus much more optimistic than estimates taking only into account the overall dictionary coherence or its cumulative coherence. The conditions allow very large (even infinite) support sets to be recovered. These results somehow explain the success of audio signal processing techniques such as those proposed by Daudet and Torr sani [?] in recovering meaningful signal representations in a union of a wavelet and a local Fourier basis

A Time-frequency estimates

This appendix contains estimates of the inner product between a wavelet $\psi_{j,k}$ and a Wilson atom $g_{n,m}$ using the time and frequency localization of the respective systems. First we give a proof of Lemma 2. The result is similar to Lemma 3.12 in [?, Chapter 6]. Let us nevertheless give the proof.

Proof. (Proof of Lemma 2) Suppose $m \neq 0$. Notice that

$$|\widehat{\psi_{j,k}}(\xi)| = 2^{-j/2} |\widehat{\psi}(2^{-j}\xi)| \leq C 2^{-j/2} (1 + |2^{-j}\xi|)^{-A},$$

and likewise

$$|\widehat{g_{n,m}}(\xi)| \leq 2^{-1/2} [|\widehat{g}(\xi - m)| + |\widehat{g}(\xi + m)|] \leq 2^{-1/2} C [(1 + |\xi - m|)^{-A} + (1 + |\xi + m|)^{-A}].$$

Thus

$$\begin{aligned} |\langle \psi_{j,k}, g_{n,m} \rangle| &= |\langle \widehat{\psi_{j,k}}, \widehat{g_{n,m}} \rangle| \\ &\leq 2^{-1/2} C^2 \cdot 2^{-j/2} \int_{\mathbb{R}} (1 + |2^{-j}\xi|)^{-A} [(1 + |\xi - m|)^{-A} + (1 + |\xi + m|)^{-A}] d\xi \\ &= 2^{-1/2} C^2 \cdot 2^{j/2} \int_{\mathbb{R}} (1 + |\xi|)^{-A} [(1 + 2^j|\xi - \xi_0|)^{-A} + (1 + 2^j|\xi + \xi_0|)^{-A}] d\xi, \end{aligned} \tag{18}$$

where $\xi_0 := 2^{-j}m$. Define

$$\begin{aligned} E_1 &= \{\xi \in \mathbb{R} : |\xi - \xi_0| \leq 1\}, \\ E_2 &= \{\xi \in \mathbb{R} : |\xi - \xi_0| > 1 \text{ and } |\xi| > \frac{1}{2}|\xi_0|\}, \\ E_3 &= \{\xi \in \mathbb{R} : |\xi - \xi_0| > 1 \text{ and } |\xi| \leq \frac{1}{2}|\xi_0|\}. \end{aligned}$$

For $\xi \in E_1$ we have $1 + |\xi_0| \leq 1 + |\xi - \xi_0| + |\xi| \leq 2 + |\xi|$. If $\xi \in E_2$ we have $1 + |\xi_0| < 1 + 2|\xi|$. Thus, for $\xi \in E_1 \cup E_2$, we have $1 + |\xi_0| \leq 2(1 + |\xi|)$, and obtain for $j \geq 0$

$$\begin{aligned} \int_{E_1 \cup E_2} (1 + |\xi|)^{-A} (1 + 2^j|\xi - \xi_0|)^{-A} d\xi &\leq 2^A (1 + |\xi_0|)^{-A} \int_{\mathbb{R}} (1 + 2^j|\xi - \xi_0|)^{-A} d\xi \\ &\leq 2^A 2^{-j} (1 + |\xi_0|)^{-A} \int_{\mathbb{R}} (1 + |\xi|)^{-A} d\xi. \end{aligned}$$

If $\xi \in E_3$, $|\xi_0 - \xi| \geq \frac{1}{2}|\xi_0|$ and $3|\xi - \xi_0| = |\xi - \xi_0| + 2|\xi_0 - \xi| > 1 + |\xi_0|$. Thus, $1 + 2^j|\xi - \xi_0| > 2^j|\xi - \xi_0| > 2^j(1 + |\xi_0|)/3$. Therefore,

$$\int_{E_3} (1 + |\xi|)^{-A} (1 + 2^j|\xi - \xi_0|)^{-A} d\xi \leq 3^A 2^{-Aj} (1 + |\xi_0|)^{-A} \int_{\mathbb{R}} (1 + |\xi|)^{-A} d\xi.$$

Since $A > 1$, combining the above estimates we get for $j \geq 0$

$$\begin{aligned} \int_{\mathbb{R}} (1 + |\xi|)^{-A} (1 + 2^j|\xi - \xi_0|)^{-A} d\xi &\leq (2^A + 3^A) 2^{-j} (1 + |\xi_0|)^{-A} \int_{\mathbb{R}} (1 + |\xi|)^{-A} d\xi \\ &= \frac{(2^A + 3^A)}{1 - A^{-1}} 2^{-j} (1 + |\xi_0|)^{-A}. \end{aligned}$$

Since this estimate is independent of the sign of ξ_0 , we can conclude by combining the estimate with Eq. (18). The other inequalities are proved similarly. \square

Proof. Proof of Lemma 3 The result follows from the fact that

$$\text{supp}(\psi_{j,k}) \subseteq [2^{-j}(k - \lambda), 2^{-j}(k + \lambda)] \quad \text{and} \quad \text{supp}(g_{n,m}) \subseteq [n - 1/2, n + 3/2).$$

□