Beyond coherence: Recovering structured time-frequency represenatations

Borup, Lasse; Gribonval, Rémi; Nielsen, Morten

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

Beyond coherence: Recovering structured time-frequency representations

by

Lasse Borup, Rémi Gribonval and Morten Nielsen

R-2007-21 September 2007

Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7 G ▪ DK - 9220 Aalborg Øst ▪ Denmark
Phone: +45 96 35 80 80 ▪ Telefax: +45 98 15 81 29
URL: http://www.math.aau.dk

ISSN 1399–2503 ▪ On-line version ISSN 1601–7811

BEYOND COHERENCE: RECOVERING STRUCTURED TIME-FREQUENCY REPRESENTATIONS

LASSE BORUP, RÉMI GRIBONVAL†, AND MORTEN NIELSEN

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.

1. Introduction

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

$$X = \mathbf{\Phi}(S), \qquad S \in \ell^2(F), \tag{1}$$

of a signal $X \in \mathcal{H}$ relative to a besselian dictionary $\mathbf{\Phi}$ 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 [15], 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 [7, 8, 9, 19, 12, 14] 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 $\mathbf{\Phi}$

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

Then for signals X with a representation $X = \mathbf{\Phi}(S)$ satisfying $|\operatorname{supp}(S)| < \lfloor \frac{1}{2}(1 + 1/\mu)\rfloor$, Basis Pursuit will recover the representation S. One serious problem with this type of results

Date: August 27, 2007.

Key words and phrases. Sparse representation, basis pursuit, matching pursuit, time-frequency dictionaries, wavelets, Wilson basis.

†Partially supported by the European Union's Human Potential Programme, under contract HPRN-CT-2002-00285 (HASSIP).

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 [19]

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

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 shortcommings 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. [3, 6, 2, 1, 20]. 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 [13] as follows. For $1 \leq p < \infty$ and a set $I \subseteq F$ we define

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

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

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

Notice that we allow infinite dictionaries $\mathbf{\Phi}$ so it may happen that $\mu_p(\mathbf{\Phi}, I) = +\infty$. The structured 1-Babel function $\mu_1(\mathbf{\Phi}, \mathcal{S})$ generalizes the Babel function $\mu_1(\mathbf{\Phi}, m)$. In fact, let $\mathcal{S}_m = \{I \subseteq F : |I| = m\}$, $m = 1, 2, \ldots, |F|$. Then $\mu_1(\mathbf{\Phi}, m) = \mu_1(\mathbf{\Phi}, \mathcal{S}_m)$. The case $p = 1$ is especially interesting due to the following result considered by Tropp [19] for the Babel function. The proof of Lemma 1 is a straightforward generalization of the proof in [19] and was pointed out in [13]. It involves the sub-dictionary $\mathbf{\Phi}_I = [\mathbf{g}_i]_{i \in I}$ made of atoms from the set $I \subseteq F$.

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

$$\mu_1(\mathbf{\Phi}, I) + \max_{\ell \in I} \mu_1(\mathbf{\Phi}_I, I \backslash \{\ell\}) < 1. \tag{4}$$

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 $|c_k| \approx const$, the condition (4) is also sufficient to ensure that simple thresholding will recover S [16, 18]. A similar condition involving the 2-Babel function instead of the 1-Babel function was recently shown to be related [11] to the probabilty of recovery with thresholding for simultaneous (multichannel) sparse approximation. All recovery results, which are expressed here for noiseless models $X = \mathbf{\Phi}(S)$, 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 $\mathbf{\Phi}_{FD}$ 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 $\mu(\mathbf{\Phi}_{FD}) = 1/\sqrt{N}$. 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 $\mathbf{\Phi}_{HW}$ defined as the union of the Haar system (see e.g. [17]) $\mathcal{B}_1 = \{h_n\}_{n=0}^{\infty}$ and the Walsh system (see [10]) $\mathcal{B}_2 = \{W_n\}_{n=0}^{\infty}$ on $[0,1]$. The Haar system and the Walsh system both form an orthonormal basis for $L^2([0,1])$ so $\mathbf{\Phi}_{HW}$ is a tight frame for $L^2([0,1])$. The point we wish to make is the following; it is easy to check that $\mu(\mathbf{\Phi}_{HW}) = 1$, so *naively* one would expect that no decent recovery result is possible. However, $\mathbf{\Phi}_{HW}$ 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 $2^j \le n < 2^{j+1}$ 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 $W_j := \operatorname{span} Q(j) = \operatorname{span} Q'(j)$ with $W_j \perp W_{j'}$ for $j \neq j'$, see [10]. Moreover, the 2^j-dimensional subdictionary $\mathbf{\Phi}_{HW}(j) := Q(j) \cup Q'(j)$ is perfectly incoherent with $\mu(\mathbf{\Phi}_{HW}(j)) = 2^{-j/2}$. Hence, if a signal x has a representation

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

with $|\operatorname{supp}(\{c_{j,k}\}_k)| + |\operatorname{supp}(\{d_{j,k}\}_k)| < \frac{1}{2}(1+2^{j/2})$ for $j \ge 0$ (the notation 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 $\mathbf{\Phi}_{HW}$ 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 Torrésani [5]. For any pair $\mathbf{c} := c_{j,n}$ and $\mathbf{d} := d_{n,m}$ of coefficient sequences we define

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

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 $x=\sum_{j=0}^{\infty}\sum_{n\in\mathbb{Z}}\sum_{\ell=0}^{2^j-1}\left(c_{j,2^jn+\ell}\psi_{j,2^jn+\ell}+d_{n,2^j+\ell}g_{n,2^j+\ell}\right)$.

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. [17]. 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

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

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

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

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 [17, Sec. 1.4]. The besselian dictionary considered is the tight frame $\mathbf{\Phi} = \mathcal{B}^1\cup\mathcal{B}^2$. 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

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

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 $\{H_{j,n}\}_{n \in \mathbb{Z}, j \in \mathbb{N}_0}$ essentially[1] 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 $n \in \mathbb{Z}$

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

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

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

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

$$N_j(I) := \sup_{n \in \mathbb{Z}} \max \left(|I^\psi_{j,n}|, |I^g_{j,n}| \right) \tag{7}$$

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

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

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.

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

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*

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

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

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

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

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

and

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

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* $\mathrm{supp}(\phi), \mathrm{supp}(\psi) \subseteq [-\lambda, \lambda]$ *for some $\lambda < \infty$, and* $\mathrm{supp}(g) \subseteq [-1/2, 3/2]$. *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,

$$\mu_p^p(\mathbf{\Phi}, I) \leq (3+2\lambda) \cdot \tilde{C}^p \cdot \sum_{j\geq 0} N_j(I) \cdot 2^{-jp/2}. \tag{12}$$

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 $k \in \mathbb{Z}$

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

Fix $j \geq 0$. By Lemma 3 the only indices k which may yield a nonzero inner product $|\langle g', \psi_{j,k}\rangle|$ are contained in the interval $\big(2^j(n'-\frac{1}{2})-\lambda, 2^j(n'+\frac{3}{2})+\lambda\big)$. Notice that this interval covers at most $3+2^{1-j}\lambda \leq 3+2\lambda$ intervals of the form $[2^j n, 2^j(n+1))$. Therefore, we may take the p-th power and sum up to get

$$\sum_{n\in\mathbb{Z}}\sum_{\psi_{j,k}\in I^\psi_{j,n}}|\langle g_{n',m'},\psi_{j,k}\rangle|^p \leq (3+2\lambda)\cdot N_j(I)\cdot\tilde{C}^p\cdot 2^{-jp/2}\cdot(1+2^{-j}|m'|)^{-Ap}.$$

Since we assume I^ψ is empty, we now take the sum over j yielding

$$\sum_{\psi\in I^\phi\cup I^\psi}|\langle g',\psi\rangle|^p \leq (3+2\lambda)\cdot\tilde{C}^p\cdot\sum_{j\geq 0}N_j(I)\cdot 2^{-jp/2}\cdot(1+2^{-j}|m'|)^{-Ap},$$

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

$$\sup_{g'\in F^g}\sum_{\psi\in I^\psi\cup I^\phi}|\langle g',\psi\rangle|^p \leq (3+2\lambda)\cdot\tilde{C}^p\cdot\sum_{j\geq 0}2^{-jp/2}\cdot N_j(I). \tag{13}$$

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 are at most $\lambda 2^{1-j'}+2 \leq 2\lambda+3$ of them. Therefore, taking the p-th power and summing we get

$$\sum_{n\in\mathbb{Z}}\sum_{g_{n,m}\in I^g_{j,n}}|\langle\psi', g_{n,m}\rangle|^p \leq (3+2\lambda)\cdot N_j(I)\cdot\tilde{C}^p\cdot 2^{-j'p/2}\cdot(1+2^{j-j'})^{-Ap}.$$

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

$$\sum_{g\in I^g}|\langle\psi',g\rangle|^p \leq (3+2\lambda)\cdot\sum_{j\geq 0}N_j(I)\cdot\tilde{C}^p\cdot 2^{-j'p/2}\cdot(1+2^{j-j'})^{-Ap} \tag{14}$$

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

$$\sum_{g\in I^g}|\langle\psi',g\rangle|^p \leq (3+2\lambda)\cdot\sum_{j\geq 0}N_j(I)\cdot\tilde{C}^p\cdot(1+2^j)^{-Ap}$$

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

$$\sup_{\psi' \in F^{\phi} \cup F^{\psi}} \sum_{g \in I^g} |\langle \psi', g \rangle|^p \leq (3+2\lambda) \cdot \tilde{C}^p \cdot \sup_{j' \geq 0} \sum_{j \geq 0} \left(2^{-j'p/2} \cdot (1+2^{j-j'})^{-Ap} \cdot N_j(I) \right). \tag{15}$$

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). □

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*

$$\sum_{j \geq 0} N_j(I) \cdot 2^{-j/2} < \frac{1}{(6+4\lambda)\tilde{C}} \tag{16}$$

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 $J_\ell = I \backslash \{\ell\}$ 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 (17) that

$$\mu_1(\mathbf{\Phi}, I) + \sup_{\ell \in I} \mu_1(\mathbf{\Phi}_I, J_\ell) \leq \mu_1(\mathbf{\Phi}, I) + \sup_{\ell \in I} \mu_1(\mathbf{\Phi}, J_\ell) \leq (6+4\lambda)\tilde{C} \cdot \sum_{j \geq 0} N_j(I) \cdot 2^{-j/2} < 1.$$

□

Our theorem can also be combined with the main theorem in [11] to prove that, under a white Gaussian model on the coefficients S, if

$$\sum_{j \geq 0} N_j(I) \cdot 2^{-j} < \frac{1}{(3+2\lambda)\tilde{C}^2} \tag{17}$$

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* $\operatorname{supp}(\phi^N), \operatorname{supp}(\psi^N) \subseteq [-N, N]$ *with*

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

with $\mu \approx 0.1887$, see [4, 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* $\operatorname{supp}(g) \subseteq [-\frac{1}{2}, \frac{3}{2}]$.

3. Conclusion

In this note we have derived sufficient conditions for the Basis Pursuit and Matching Pursuit algorithms to recover structured representations of admissible signals 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 Torrésani [5] in recovering meaningful signal representations in a union of a wavelet and a local Fourier basis

Appendix 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 [17, Chapter 6]. Let us nevertheless give the proof.

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

$$|\widehat{\psi_{j,k}}(\xi)| = 2^{-j/2}|\widehat{\psi}(2^{-j}\xi)| \leq C2^{-j/2}(1+|2^{-j}\xi|)^{-A},$$

and likewise

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

Thus

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

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

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

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 $j \geq 0$

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

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

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

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

□

References

[1] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007.

[2] E. J. Candès and J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Found. Comput. Math.*, 6(2):227–254, 2006.

[3] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math*, 59:1207–1223, 2006.

[4] I. Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[5] L. Daudet and B. Torrésani. Hybrid representations for audiophonic signal encoding. *Signal Processing, special issue on Image and Video Coding Beyond Standards*, 82(11):1595–1617, 2002.

[6] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[7] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization. *Proc. Natl. Acad. Sci. USA*, 100(5):2197–2202 (electronic), 2003.

[8] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory*, 47(7):2845–2862, 2001.

[9] M. Elad and A. M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Trans. Inform. Theory*, 48(9):2558–2567, 2002.

[10] B. Golubov, A. Efimov, and V. Skvortsov. *Walsh series and transforms*, volume 64 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1991. Theory and applications, Translated from the 1987 Russian original by W. R. Wade.

[11] R. Gribonval, B. Mailhe, H. Rauhut, K. Schnass, and P. Vandergheynst. Average case analysis of multichannel thresholding. In *Proc. IEEE Intl. Conf. Acoust. Speech Signal Process (ICASSP'07)*, 2007.

[12] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49(12):3320–3325, 2003.

[13] R. Gribonval and M. Nielsen. Beyond sparsity : recovering structured representations by ℓ^1-minimization and greedy algorithms. *Advances in Computational Mathematics*, 2006. accepted.

[14] R. Gribonval and M. Nielsen. Highly sparse representations from dictionaries are unique and independent of the sparseness measure. *Appl. Comput. Harm. Anal. (in press)*, 2006.

[15] R. Gribonval and M. Nielsen, editors. *Sparse Approximations in Signal and Image Processing*, Special issue of EURASIP Signal Processing Journal, March 2006.

[16] R. Gribonval, M. Nielsen, and P. Vandergheynst. Towards an adaptive computational strategy for sparse signal approximation. preprint of the Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), 2007.

[17] E. Hernández and G. Weiss. *A first course on wavelets*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1996.

[18] K. K. Herrity, A. C. Gilbert, and J. Tropp. Sparse approximation with iterative thresholding. In *Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006.

[19] J. A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, 2004.

[20] J. A. Tropp. Random subdictionaries of general dictionaries. *Appl. Comp. Harmonic Anal. (to appear)*, 2007.

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

IRISA-INRIA, Campus de Beaulieu, F-35042 Rennes cedex, France
E-mail address: remi.gribonval@inria.fr

Department of Mathematics, Washington University, St. Louis, MO 63130 and Department of Mathematical Sciences, Aalborg University, Fr. Bajers Vej 7G, DK-9220 Aalborg East, Denmark
E-mail address: mnielsen@math.wustl.edu