Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients

Matthieu Kowalski, Bruno Torrésani

Signal, Image and Video Processing manuscript No.
(will be inserted by the editor)

Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients

Matthieu Kowalski · Bruno Torrésani

Received: October 2007 / Accepted:

Abstract Sparse regression often uses ℓ_p norm priors (with $p < 2$). This paper demonstrates that the introduction of mixed-norms in such contexts allows one to go one step beyond in signal models, and promote some different, structured, forms of sparsity. It is shown that the particular case of $\ell_{1,2}$ and $\ell_{2,1}$ norms lead to new *group shrinkage* operators. Two different problems are considered, that illustrate the relevance of the proposed approach, in the context of audio signals. Mixed norm priors are shown to be particularly efficient for multichannel audio denoising, in a generalized basis pursuit denoising approach. Mixed norm priors are also used in a context of *morphological component analysis* of time-frequency sound representations, for which an adapted version of Block Coordinate Relaxation algorithm is derived. This yields a new approach for sparse regression in time-frequency dictionaries.

Keywords Mixed-norms · Time-frequency decompositions · Sparse representations

1 Introduction

Sparse approximation approaches have enjoyed considerable popularity in recent signal processing applications. Sparsity seems to be a particularly efficient guiding principle in view of a number of tasks such as signal compression, denoising, image de-blurring, blind source separation,... The guiding principle may be summarized as follows: for most signal classes, it is possible to find a basis or a dictionary of elementary building blocks (or atoms) with respect to which all (or most) signals in the class may be expanded, so that when the expansion is truncated in a suitable way, high precision

M. Kowalski
LATP, CMI, 39 rue Joliot-Curie, 13453 Marseille Cedex 13, France
Tel.: +33-4-91054740
Fax: +33-4-91054742
E-mail: kowalki@cmi.univ-mrs.fr

B. Torrésani
LATP, CMI, 39 rue Joliot-Curie, 13453 Marseille Cedex 13, France
Tel.: +33-4-91054678
Fax: +33-4-91054742
E-mail: Bruno.Torresani@cmi.univ-mrs.fr

approximations are obtained even when a small number of terms are retained. A large number of signal and image processing "success stories" may be described in such a way, including image compression and denoising using wavelets, curvelets, or more sophisticated *-lets, audio coding using MDCT bases, and so forth. Several efficient sparse expansion algorithms have been proposed, including among others simple expansion with respect to a fixed basis followed by soft or hard coefficient thresholding, iterative thresholding strategies in redundant dictionaries, greedy (pursuit) algorithms, or more elaborate approaches such as sparse regression in Bayesian frameworks. Thresholding and iterative thresholding strategies are particularly interesting, mainly because thresholding automatically generates sparsity, and corresponding algorithms are easy to implement and generally exhibit fast convergence properties.

A main strength of these thresholding approaches is that they process the signal representation coefficientwise, which results in low complexity algorithms. However, this may become a weakness when it comes to applications to real signals. Indeed, the assumption of independence of coefficients is generally not realistic. For example, when using wavelet or local cosine bases for expanding 1D signals, abrupt changes manifest themselves by groups of time-localized large coefficients, and frequency modulated signals exhibit *ridges* of frequency localized large coefficients. The same remark applies to edges and regular textures in wavelet or local cosine representations of images. Several different approaches have been considered to handle such dependencies between coefficients, including *structured* versions of matching pursuit (for example, harmonic or molecular versions of matching pursuit), coefficient domain modelling, or construction of suitable bases. We propose here to keep the coefficient modeling approach. However, rather than introducing explicit models for coefficients, we follow the thresholding and iterative thresholding approaches and design new *group thresholding* methods, associated with mixed norms in the coefficient domain.

More precisely, we consider here with the following problem. Let $y \in \mathbb{R}^M$ be a noisy observation of a signal $s \in \mathbb{R}^M$. Let $\mathcal{D}$ denote a fixed dictionary for $\mathbb{R}^M$, and denote by $A \in \mathbb{R}^{M\times N}$ be the matrix whose columns are the vectors from the dictionary $\mathcal{D}$. We assume that s has a sparse expansion in $\mathcal{D}$, and we want to estimate s from y. A classical estimate is given by the basis pursuit denoising [1] introduced by Donoho and coworkers, also known as the lasso estimate [2] of Tibshirani, and is obtained by the following optimization:

$$\hat{x} = \underset{x\in\mathbb{R}^N}{\operatorname{argmin}} \|y - Ax\|_2^2 + \lambda\|x\|_1 \tag{1}$$

where $\lambda \in \mathbb{R}$ is a fixed parameter, so that, $A\hat{x}$ is the estimate of s. The ℓ_1 norm directly leads to soft thresholding strategies. Similar algorithms may be derived using more general ℓ_p norms, i.e. replacing $\|x\|_1$ with $\|x\|_p^p$. That estimate treats all coefficients independently. Dependencies between selected subsets of coefficients may be introduced as soon as the latter may be labelled using a double index (for example, a time-frequency index), say $x = \{x_{ab}, a = 1, \dots N_a,\, b = 1, \dots N_b\}$. Then a new estimate is obtained via

$$\hat{x} = \underset{x\in\mathbb{R}^N}{\operatorname{argmin}} \|y - Ax\|_2^2 + \lambda\|x\|_{p,q}^q \ , \tag{2}$$

where $\|x\|_{p,q}$ is the mixed norm defined by

$$\|x\|_{p,q} = \left(\sum_{a=1}^{N_a} \left(\sum_{b=1}^{N_b} |x_{ab}|^p \right)^{q/p} \right)^{1/q} \ . \tag{3}$$

Here, the roles of indices a and b is purely conventional, and a and b can be interchanged, which corresponds to a different problem.

It is worth noticing that like the lasso method and ℓ_p generalizations, the mixed norm approach admits a simple Bayesian interpretation, assuming Gaussian white noise (which justifies the choice of ℓ_2 norm for the data fidelity term), and a coefficient prior of the form

$$f(x) \;\propto\; \exp\left\{-\lambda \|x\|_{p,q}^q\right\} \;,$$

which explicitely introduces couplings between coefficients.

As a simple application of this approach, we will consider the case of multichannel audio signal denoising in a MDCT (Modified Discrete Cosine Transform) basis. We show that multichannel denoising based upon an appropriate $\ell_{2,1}$ norm, that implements across-channel persistence, and within-channel sparsity, significantly outperforms multichannel basis pursuit denoising.

Mixed norms can also be implemented into multilayered type signal expansions, such as the ones used in [3–5] for audio signals, or in the Morphological Component Analysis (MCA for short) for images [6]. The goal of MCA is to minimize functionals of the type

$$\Phi(x_1, x_2) = \|x_1\|_1 + \|x_2\|_1 + \lambda \|y - A_1 x_1 - A_2 x_2\|_2^2 \tag{4}$$

where A_1 and A_2 are the matrices corresponding to two dictionaries, chosen to be able to describe sparsely edges and textures respectively. A similar approach may be taken to separate transient and tonal layers in audio signals. According to the discussion above, we shall show that the two ℓ_1 norms in the latter expression can be conveniently replaced with suitable mixed norms, to enforce relevant dependencies between coefficients.

We show in this paper that multilayered signal and image expansions can be combined with coefficient dependencies thanks to adequate mixed norms. For the sake of simplicity, we stick to variational formulations, and derive a corresponding extension of the MCA algorithm. These mixed norms are used in a block coordinate relaxation algorithm to minimize an appropriate functional in section 3, and some illustrations of the obtained results to show the relevance of mixed norms are described in section 4.

Audio signals and time-frequency representation are so natural illustrations and applications of mixed norms and will be the breadcum trail of this paper.

2 Mixed norms

We give in this section the definition of the mixed norms we shall be interested in. For the sake of simplicity, we shall stick to the case of two indices, even though extensions are clearly possible.

The reader may think of these two indices as the indices of a time-frequency signal expansion, as follows. Let $\mathbf{x} \in \mathbb{R}^N$ be a time-frequency representation of a given signal y. We suppose that $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_K)$, where for all $k \in \{1, \ldots, K\}$, $\mathbf{x}_k = \big(x_{k,1}, \ldots, x_{k,\nu}, \ldots, x_{k,F}\big)$. So that, (k, ν) denote a time-frequency index, and $N = K \times F$. With these notations, $x_{k,\nu}$ denote the coefficient at time k and frequency ν, the vector $\mathbf{x}_k$ represent all the frequency coefficients at time k. We will denote by $\mathbf{x}_{\cdot,\nu}$ the vector which contain all the time coefficients at frequency ν.

However, let us stress that the developments below are not specific at all to time-frequency signal representations, and apply to any situation where signals are expanded with respect to a dictionary with two indices. Another simple example of that is multichannel signals, where a first index labels (scalar) dictionary elements and a second one labels channels. In an even more general situation, any discrete signal expansion may be re-labelled so as to be processed by our approach.

Now that the notations are introduced, we are ready to define the mixed norms.

Definition 1 Let $\mathbf{x} \in \mathbb{R}^N$, labelled by a double index (k, ν) . Let $p \geq 1$ and $q \geq 1$, then one can define two mixed norms $\ell_{1;p,q}$ and $\ell_{2;p,q}$ on x

$$\|\mathbf{x}\|_{1;p,q} = \left(\sum_{k=1}^{K} \left(\sum_{\nu=1}^{F} |x_{k,\nu}|^p \right)^{q/p} \right)^{1/q} , \tag{5}$$

$$\|\mathbf{x}\|_{2;p,q} = \left(\sum_{\nu=1}^{F} \left(\sum_{k=1}^{\nu} |x_{k,\nu}|^p \right)^{q/p} \right)^{1/q} . \tag{6}$$

The cases $p = +\infty$ and $q = +\infty$ are obtained by replacing the corresponding norm by the supremum.

Mixed norms have been used extensively by mathematicians in functional analysis (see for example [7] and references therein). We limit ourselves here to the finite dimensional case, and focuss on the particular cases $\ell_{\bullet;1,2}$ and $\ell_{\bullet;2,1}$. For the sake of simplicity, we will use the $\ell_{1;p,q}$ norm for the theoretical study, and then denote it simply by $\ell_{p,q}$. The second case is obtained by simply switching the roles of k and ν. In the numerical applications described in section 4 the choices will be specified precisely.

It is interesting to stress that a $\ell_{p,q}$ mixed norm can be seen as a "composition" of ℓ_p and ℓ_q norms. With the above notations,

$$\|\mathbf{x}\|_{p,q} = \left(\sum_{k=1}^{K} \|\mathbf{x}_k\|_p^q \right)^{1/q} = \|(\|\mathbf{x}_1\|_p, \ldots, \|\mathbf{x}_K\|_p)\|_q \ . \tag{7}$$

For $p < 2$, ℓ_p norms are often used as *diversity* measures, and minimizing the ℓ_p norm of a coefficient sequence of a signal generally aims at promoting sparsity for the expansion. The case $p = 1$ has a particular status, since the ℓ_1 norm promotes sparsity and remains convex. The situation with mixed norms is a bit more tricky, since two exponents have to be taken into account. However, we shall see below that values of p (or q) smaller than 2 still yield some form of sparsity, in a somewhat *structured* way. More precisely, depending on the choice of p and q, sparsity will be promoted on each indivual variable $x_{k,\nu}$ if p is close to 1, and on an entire group of variables if q is close to 1.

3 Regression with mixed norm

To show the relevance of mixed-norms in signal processing contexts, we consider a couple of examples borrowed from audio signal processing. The first one concerns multichannel audio denoising, and the second one is the multilayer audio signal decomposition. Let us stress that the application range of this approach is not at all restricted to audio signals, and that numerous other domains can be addressed.

We first introduce generalized shrinkage operators, extending lasso and group lasso estimators, before turning to extensions to the multilayered case.

3.1 Introduction of new group-shrinkage operators

Our aim in this subsection is to solve the following optimisation problem, in the particular case where A is an orthogonal matrix:

$$\min_{x\in\mathbb{R}^N} \frac{1}{2}\|y - Ax\|_2^2 + \frac{\lambda}{q}\|x\|_{p,q}^q \tag{8}$$

which can also be written

$$\min_{x\in\mathbb{R}^N} \frac{1}{2}\|\bar{y} - x\|_2^2 + \frac{\lambda}{q}\sum_{k=1}^{K}\left(\sum_{\nu=1}^{F}|x_{k,\nu}|^p\right)^{q/p} \tag{9}$$

where $\bar{y} = A^T y$. The solution in the cases where the mixed norm are $\ell_{1,2}$ or $\ell_{2,1}$ is given by the following proposition.

Proposition 1 *Let A be an orthogonal matrix. Then, the minimimum $\hat{x}$ of*

(a) $\frac{1}{2}\|y - Ax\|_2^2 + \frac{\lambda}{2}\sum_{k=1}^{K}\left(\sum_{\nu=1}^{F}|x_{k,\nu}|^1\right)^2$ *is given by, for all k,ν*

$$\hat{x}_{k,\nu} = \begin{cases} \bar{y}_{k,\nu} - \frac{\lambda\|\bar{y}_k\|_1}{1+F\lambda} & \text{if } \bar{y}_{k,\nu} > \frac{\lambda\|\bar{y}_k\|_1}{1+F\lambda} \\ \bar{y}_{k,\nu} + \frac{\lambda\|\bar{y}_k\|_1}{1+F\lambda} & \text{if } -\bar{y}_{k,\nu} < -\frac{\lambda\|\bar{y}_k\|_1}{1+F\lambda} \\ 0 & \text{if } |\bar{y}_{k,\nu}| < \frac{\lambda\|\bar{y}_k\|_1}{1+F\lambda} \end{cases}$$

(b) $\frac{1}{2}\|y - Ax\|_2^2 + \lambda\sum_{k=1}^{K}\left(\sum_{\nu=1}^{F}|x_{k,\nu}|^2\right)^{1/2}$ *is given by, for all k,ν*

$$\hat{x}_{k,\nu} = \bar{y}_{k,\nu}\left(1 - \frac{\lambda}{\|\bar{y}_k\|_2}\right)^+ .$$

Remark 1 The $\ell_{2,1}$ case is known in the statistical community as the group-lasso estimate, and the result was given in [8]. The $\ell_{1,2}$ result is proven in [9] as a part of a more general result. Notice that in both cases, the result is a generalized soft thresholding, or shrinkage, that is applied to a group of coefficients rather than single coefficients. Hence, coefficients are not processed independently any more.

Remark 2 It is interesting to notice the striking difference between the two new shrinkage operators. In the second case (the group-lasso case), a 1D group of coefficients is either globally retained or discarded. In the first case, each coefficient is shrunk individually, but the corresponding threshold depends on its 1D neighborhood. The difference between these two situations will appear clearly in the numerical results below.

Now we are able to find the solution of the simpler problem (8) where A is an orthogonal matrix (corresponding to an orthonormal basis), we can turn to the more complex functional (10) in the particular case where A is a concatenation of orthogonal matrices (corresponding to unions of orthonormal bases).

3.2 Multichannel denoising

Let us consider multichannel signals $\mathbf{y} = \{y_{nc},\, n = 1, \dots N,\, c = 1, \dots C\}$, n denoting the time index and c the channel index. Consider an orthonormal basis $\mathbf{U} = \{u_m, m = 1, \dots N\}$ (here, m index the atoms of the basis) for the single channel signal space. We are interested in expansions of the form $\mathbf{y} = \sum_m \mathbf{x}_m u_m$, where multichannel vectors are denoted with bold symbols, in cases where the observations are noisy, and the basis $\mathbf{U}$ has been chosen in such a way that the coefficient sequences x are sparse in the n direction, and persistent across channels.

Sparse approximation techniques have been extended recently to multichannel signals (see [10,11] and references therein). We address such a problem directly using a generalized basis pursuit denoising approach, using ℓ_1 norm in the n direction, and ℓ_2 norm across channels. In this case, the A matrix is built as a block diagonal matrix, the blocks being equal to the orthogonal matrix U associated with basis $\mathbf{U}$, and the optimization problem is formulated as before:

$$\min_{\{y_{nc}\}} \left(\|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_{p,q}^q \right) .$$

The results above may then be applied directly. Since we aim at privileging lines of coefficients, we choose here the group-lasso norm, i.e. $p = 2$ and $q = 1$, and use the estimator provided in Proposition 1-(b). Numerical results showing the superiority of this approach over standard multichannel version of basis pursuit denoising, are given in section 4.1.

3.3 Multilayered expansion: application to *tonal* + *transients* + *noise* separation

Our aim is to decompose an audio signal into three layers *transient*+*tonal*+*noise*. This problem was already studied and some algorithm are already available. The approach outlined here has the advantage of being extremely simple.

We start from an optimization problem similar to the one given by MCA, but, instead of using two ℓ_1 norm to estimate the tonal and transient layer, we will use adapted mixed-norm. So that, we define the following functional we will minimize

$$\Phi(x, \tilde{x}) = \|y - A(x, \tilde{x})^T\|_2^2 + \lambda \|x\|_{p,q}^q + \mu \|\tilde{x}\|_{\tilde{p},\tilde{q}}^{\tilde{q}} \tag{10}$$

where the $\ell_{p,q}$ and $\ell_{\tilde{p},\tilde{q}}$ will be chosen adequately.

To decompose a signal into several layer, one chooses an adapted dictionary for each layer. For audio signals, the transient layer is known to be sparsely represented in wavelets dictionaries, or time-frequency dictionaries (like Gabor or MDCT) with a narrow window. At the opposite, tonal layer is known to be sparsely represented in time-frequency dictionaries with a wide window.

We choose here the special case where each dictionary is an orthonormal basis, for example, two MDCT bases with two differents sizes for the windows, in order to apply the Block Coordinate Relaxation method [12] (BCR for short) which inspired the MCA algorithms [13]. BCR is specially adapted for the union of orthogonal bases, and is known to converge to a minimum of the basis-pursuit objective functional.

Let us introduce the following notations. We denote by $\mathbf{U}$ and $\mathbf{V}$ the two bases under consideration, and by U and V the corresponding matrices. In the multilayered audio signal expansion example, $\mathbf{U}$ may be the basis adapted for the tonal layer, and $\mathbf{V}$

the one adapted to the transient layer. We denote by $\mathbf{x_U}$ the coefficients corresponding to the basis $\mathbf{U}$ and $\mathbf{x_V}$ the coefficients corresponding to the basis $\mathbf{V}$. So that, $U\mathbf{x_U}$ will correspond to the tonal layer and $V\mathbf{x_V}$ to the transient layer.

The functional chosen to minimize and obtain an estimate of the two layers is the following

$$\Phi(x_{\mathbf{U}}, x_{\mathbf{V}}) = \frac{1}{2}\|y - Ux_{\mathbf{U}} - Vx_{\mathbf{V}}\|_2^2 + \frac{\lambda}{q}\|x_{\mathbf{U}}\|_{p,q}^q + \frac{\mu}{\tilde{q}}\|x_{\mathbf{V}}\|_{\tilde{p},\tilde{q}}^{\tilde{q}} \tag{11}$$

The BCR algorithm is then slightly modified in order to yield a minimizer of (11):

Algorithm 1

- **Let** $x_{\mathbf{U}}^{(0)} \in \mathbb{R}^N$ *and* $x_{\mathbf{V}}^{(0)} \in \mathbb{R}^N$
- **Do**
  1. $r_{\mathbf{U}}^{(m)} = y - Vx_{\mathbf{V}}^{(m)}$
  2. *Find an estimate* $x_{\mathbf{U}}^{(m+1)}$ *by solving*

$$x_{\mathbf{U}}^{(m+1)} = \underset{x\in\mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|y - Ux\|_2^2 + \frac{\lambda}{q}\|x\|_{p,q}^q$$

  using proposition 1
  3. $r_{\mathbf{V}}^{(m)} = y - Ux_{\mathbf{U}}^{(m+1)}$
  4. *Find an estimate* $x_{\mathbf{V}}^{(m+1)}$ *by solving*

$$x_{\mathbf{V}}^{(m+1)} = \underset{x\in\mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|y - Vx\|_2^2 + \frac{\mu}{\tilde{q}}\|x\|_{\tilde{p},\tilde{q}}^{\tilde{q}}$$

  using proposition 1

  Until *convergence*

Theorem 1 *Let U and V be two orthogonal matrices of $\mathbb{R}^{N\times N}$. Let $y \in \mathbb{R}^N$ and $p \geq 1$, $q \geq 1$ $\tilde{p} \geq 1$, and $\tilde{q} \geq 1$. Then the algorithm 1 converges to a minimum of* (11).

Sketch of the proof: The considered algorithm may be seen as an extension of the MCA version of the BCR algorithm, whose convergence was proved in [13], using arguments from [12] and [14].

The main arguments of the proof are the fact the terms in (11) are convex functionals, which is still true here, and that the objective function (11) has a separable form, which is also true in our case. Notice that now the decoupled variables are not coefficients any more, but groups of coefficients, since the considered mixed norms only partially decouple coefficients.

Thus, following [12], we may conclude that the convergence follows from the convergence of the Dual Block Coordinate Ascent (DBCA) algorithm of [14]. □

4 Results

We illustrate here the interest of mixed norm formulations with a couple of problems, namely denoising of multichannel signals, and multilayered signal decomposition.

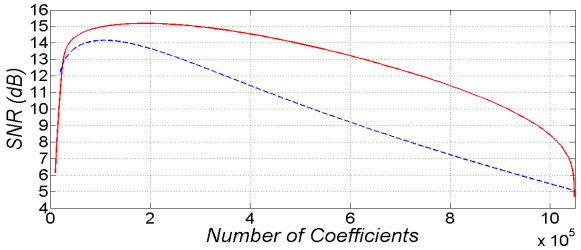

Fig. 1 SNR as a function of the number of retained coefficients; dashed curve: ℓ_1 norm; full curve: $\ell_{2,1}$ norm.

4.1 Denoising of multichannel signals

We consider here the framework of section 3.2, and illustrate it with a sound example recorded in a running train. Let us stress that the same approach may be developed in many other multichannel signal denoising contexts, such as color image denoising, multispectral imaging,...

The considered signal features low frequency noise, phone ring, voice, clicks and additional transient components. The signal is a four channels signal, recorded using three directional and one omnidirectional microphones. Gaussian white noise was added to the four channels, yielding input SNR equal to 5.07 dB. The signal was denoised by applying soft thresholding (corresponding to ℓ_1 norm prior on the set of coefficients, i.e. lasso estimate), and group soft thresholding corresponding to $\ell_{2,1}$ norm prior on coefficients (group lasso estimate). As stressed before, this choice is motivated by the desire of using the same significance map (i.e. the set of labels of nonzero coefficients) for all channels. Simulations were run with various values of the threshold (i.e. the Lagrange parameter). Corresponding SNR curves are displayed in figure 1.

The mixed norms approach clearly outperforms the classical approach significantly. Similar results (not shown here) were also obtained on different multichannel audio signals. The improvement appears to increase with the number of channels, as may be expected.

4.2 Multilayered audio signal expansion

This section illustrates the influence of the mixed-norm in the regression problem (11), in comparison to the usual ℓ_1 norm used in the MCA regression problem (4). For that, we choose the difficult problem of a single sensor source separation and we consider the mixture of two signals: a "tonal" one, namely a song of trumpet and a "transient" one played by castanets. The two signals are then simply added up to obtain the mixture of about 1.5s long (2^{16} samples). The two signals and the mixture are represented in figure 2.

One then expects to obtain an estimate $V\hat{x}_{\mathbf{V}}$ of the castanet signal, and an estimate $U\hat{x}_{\mathbf{U}}$ of the trumpet signal. We will compare the estimates given by choosing two ℓ_1 norm (like the MCA), and several mixed norms specified below. We choose for $\mathbf{U}$ a MDCT basis with a 4096 samples length window, and for $\mathbf{V}$ a MDCT basis with a 128 samples length window. The representations of the MDCT coefficients of the

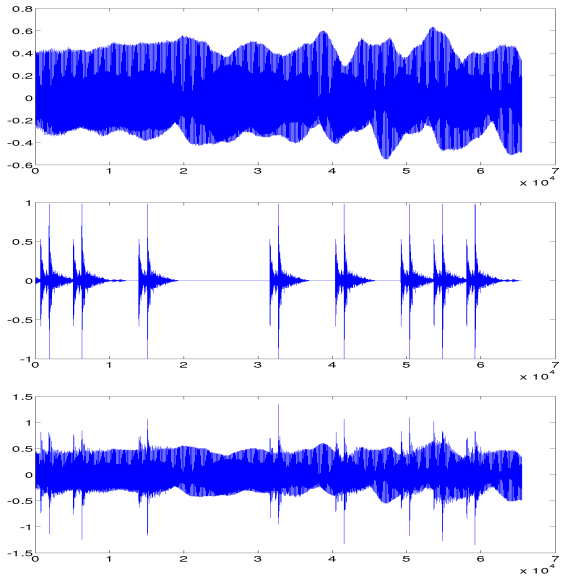

Fig. 2 From top to bottom: trumpet signal, castanet signal, mixture.

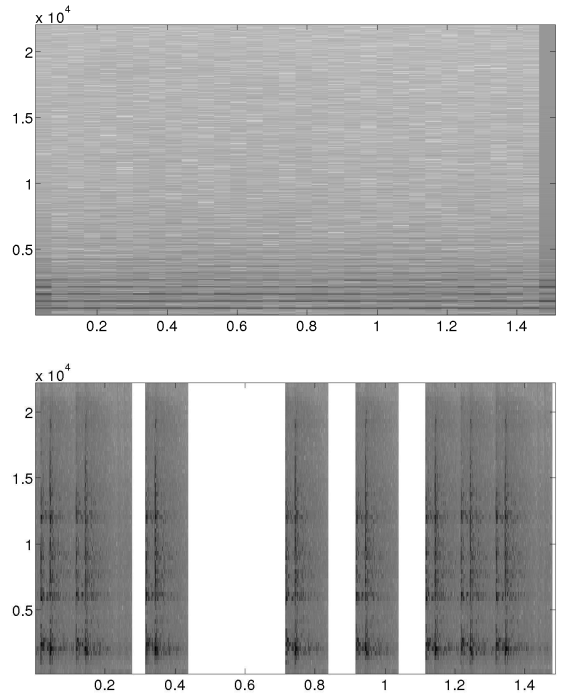

Fig. 3 MDCT coefficients of the two signals. Top: the trumpet signal, bottom: the castanet signal.

norms	ℓ_1 / ℓ_1	$\ell_{1;1,2}$ / $\ell_{2;1,2}$	$\ell_{1;1,2}$ / $\ell_{1;2,1}$
nbcoeff $x_{\mathbf{U}}$	18 665	17 915	17 841
nbcoeff $x_{\mathbf{V}}$	17 098	16 925	18 880
SNR $x_{\mathbf{U}}$	14.8814	15.4417	15.0803
SNR $x_{\mathbf{V}}$	3.9619	4.3732	3.7965

Table 1 Results obtained for three different choices of mixed norms: number of retained coefficients in each layer, and corresponding SNR.

trumpet (resp. castanet) signal in $\mathbf{U}$ (res. $\mathbf{V}$) are shown in figure 3. Then one can see the particular structure of each layers.

Definition 1 provides several choices for mixed-norms defined in the time-frequency domain. The tonal layer is expected to be sparsely represented in the frequency domain, with emergent frequencies that may evolve slowly with time. Sticking to combinations of ℓ_1 and ℓ_2 norms, possible choices are then $\ell_{1;1,2}$ and $\ell_{2;2,1}$ mixed norm (we recall that these are defined in Definition 1). In a similar spirit, the transient layer is expected to be sparse in time, but wide in the frequency domain. The possible choices are then $\ell_{2;1,2}$ or the $\ell_{1;2,1}$ mixed norm.

Choosing the $\ell_{2;2,1}$ mixed-norm for the tonal layer is actually not a good strategy because of the slow evolution in time of the frequencies. So we prefer the $\ell_{1;1,2}$ for the tonal layer. On the other hand, for the transient layer, if $\ell_{2;1,2}$ is still a good choice, $\ell_{1;2,1}$ can also be interesting because of the particular structures of transients, which are generally sharply time-localized.

Table 1 summarizes the different results obtained using the three possible functionals. The λ and μ parameters were tuned to obtain approximately the same numbers of coefficients for each functional, but we did not seek the best SNR (regard to the size of the parameter space). The choice made give acceptable results for the source separation, and illustrate well the behavior of the different norms. The first column of the table contains the norms that were chosen for the tonal layer and the transient layer, the second and the third columns contain respectively the number of retained coefficient for $x_{\mathbf{U}}$ and $x_{\mathbf{V}}$, and the last two columns contain the signal to noise ratio of the estimates.

This table and the figures 4 and 5 clearly show the different behaviors of the norms. The best results in term of SNR are obtained for the second functional. The choice of the $\ell_{1;1,2}$ (resp. $\ell_{2;1,2}$) norm for the tonal (resp. transient) layer clearly promotes time (resp. frequency) persistence. For the tonal layer, and compared to the ℓ_1 norm, less hight frequencies are catched and the persistence of a frequency during time is supported. For the transient layer, the vertical structures are better preserved compared to the ℓ_1 norm which catches a lot of low frequency components, which would be better represented in the tonal layer.

It is interesting to stress that the $\ell_{1;2,1}$ norm for the transient layer, despite a lower SNR, preserves straight lines as expected, and then makes this norm a good choice if one wants to preserve this particular structure. The price to pay is a significant increase of the number of retained coefficients.

5 Conclusion

We have shown in this paper the relevance of mixed norm priors in the framework of sparse regression problems. Such mixed norms have been extensively used in the

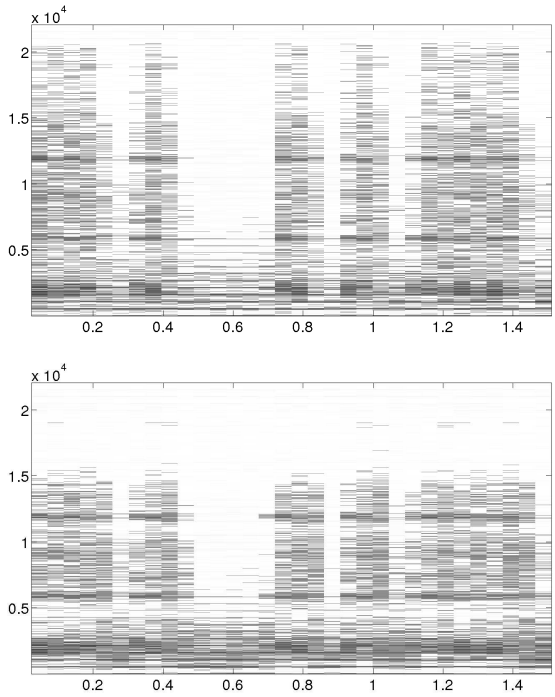

Fig. 4 MDCT coefficients of two estimates of the trumpet signal. Top: the ℓ_1 estimate, bottom: the $\ell_{1;1,2}$ estimate.

mathematical analysis literature [7], but their use in practical situations has been quite limited so far, except for a few particular ones (such as the group lasso algorithm). For the sake of simplicity, the mixed norms discussed here are $\ell_{1,2}$ and $\ell_{2,1}$ norms, but similar results may be obtained using more general $\ell_{p,q}$ norms, and several standard sparse approximation algorithms may be extended to that situation. We refer to the forthcoming paper [9] for a thorough analysis of the latter.

We have only emphasized here a couple of applications, in the domain of audio signal processing, where the results were quite spectacular. Let us stress that in both cases, our point was not to compare to state of the art approaches, but rather to show what can be done using very simple techniques, that can be refined further. We would also like to point out that this approach is not at all specific to audio signals, and may be applied *mutatis mutandis* to image decomposition, for example in the framework of the MCA approach of [6].

It is worth coming back to the behavior of mixed norms in the present context. The rationale of our approach is to use a combination of ℓ_1 and ℓ_2 norms, to promote sparsity in the direction of one of the two indices, and persistence in the direction of the other. Now, as we have stressed at the beginning of this paper, a doubly labelled coefficient sequences can be obtained by arbitrary relabelling of a given coefficient sequence. Therefore, mixed norm approaches can be used to introduce models for coefficients involving a small number of clusters of significant coefficients. Such a representation

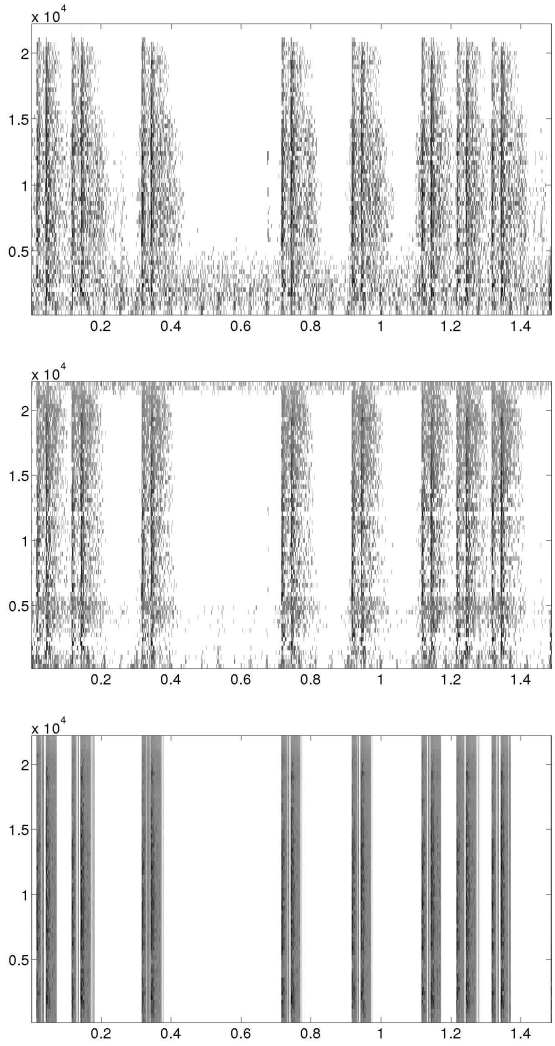

Fig. 5 MDCT coefficients of three estimates of the castanet signal. From top to bottom: the ℓ_1 estimate, the $\ell_{2;1,2}$ estimate, the $\ell_{1;2,1}$ estimate.

features both sparsity (in the domain of coefficient groups) and persistence (within a group). We believe that the potential of such approaches is extremely important.

Acknowledgements

We wish to thank Stéphane Molla for kindly providing the train sound example.

References

1. S. S. Chen, D. L. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
2. R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Serie B*, vol. 58, no. 1, pp. 267–288, 1996.
3. J. Berger, R. Coifman, and M. Goldberg, "Removing noise from music using local trigonometric bases and wavelet packets," *J. Audio Eng. Soc.*, vol. 42, no. 10, pp. 808–818, 1994.
4. L. Daudet and B. Torrésani, "Hybrid representations for audiophonic signal encoding," *Signal Processing*, vol. 82, no. 11, pp. 1595–1617, 2002, special issue on Image and Video Coding Beyond Standards. [Online]. Available: http/::www.cmi.univ-mrs.fr/ torresan/papers/SigPro.ps.gz
5. L. Daudet, S. Molla, and B. Torrésani, "Towards a hybrid audio coder," in *International Conference Wavelet analysis and Applications*, J. P. Li, Ed., Chongqing, China, 2004, pp. 13–24.
6. M. Elad, J.-L. Starck, D. L. Donoho, and P. Querre, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Journal on Applied and Computational Harmonic Analysis*, vol. 19, pp. 340–358, November 2005.
7. S. Samarah, S. Obeidat, and R. Salman, "A shur test for weighted mixed-norm spaces," *Analysis Mathematica*, vol. 31, pp. 277–289, 2005.
8. M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society Serie B*, vol. 68, no. 1, pp. 49–67, 2006.
9. M. Kowalski, "Sparse regression using mixed norms," in preparation.
10. J. Bobin, Y. Moudden, J. Fadili, and J.-L. Starck, "Morphological component analysis for sparse multichannel data: Application to inpainting," 2007, preprint, submitted.
11. R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! average case analysis of multi-channel sparse recovery using greedy algorithms," 2007, iNRIA technical report PI 1848, submitted.
12. A. G. Bruce, S. Sardy, and P. Tseng, "Block coordinate relaxation methods for nonparametric signal denoising," in *Proceedings of the SPIE - The International Society for Optical Engineering*, no. 3391, 1998, pp. 75–86.
13. J.-L. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Transactions on Image Processing*, vol. 14, no. 10, 2004.
14. P. Tseng, "Dual coordinate ascent methods for non-stricly convex minimization," *Mathematical Programming*, vol. 59, pp. 231–247, 1993.