Noname manuscript No.
(will be inserted by the editor)

Primal-Dual Algorithms for Audio Decomposition Using Mixed Norms

İlker Bayram and Ö. Deniz Akyıldız

Received: date / Accepted: date

Abstract We consider the problem of decomposing audio into components that have different time frequency characteristics. For this, we model the components using different transforms and mixed norms. While recent works concentrate on the synthesis prior formulation, we consider both the analysis and synthesis prior formulations. We derive algorithms for both formulations through a primal-dual framework. We also discuss how to modify the primal-dual algorithms in order to derive a simpler heuristic scheme.

Keywords Audio decomposition · mixed norms · analysis prior · synthesis prior · primal-dual

1 Introduction

Decomposing an audio signal into components with different time-frequency characteristics can be of interest in a number of different scenarios. For instance, if we wish to modify a certain instrument in a music signal, then separating the instrument from the rest of the signal appears as a plausible preprocessing step. If the time-frequency behavior for the instrument of interest differs from the rest of the signal, then this preprocessing step poses a decomposition problem as described above. Another example may be an audio restoration problem, where the signal contains artifacts (like crackles) due to recording, that are distinct from the signal of interest, from a time-frequency viewpoint. More generally, a closely related problem is the decomposition of audio as

$$\text{Tonal} + \text{Transient} + \text{Residual}.$$

Originally, such a decomposition was proposed as a preprocessing step for audio coding, as well as other modifications [18,28,41]. Although the goal of each example described above is different, they all require a decomposition of audio into components with different time-frequency behavior. In this paper, we consider formulations of the decomposition problem based on 'mixed norms' (see also [27, 37] for related work) and derive algorithms following a primal-dual framework. Based on these new algorithms, we also propose a heuristic scheme that can achieve decompositions with similar characteristics.

Put briefly, mixed norm formulations for decomposition make use of the different structures inherent in the frame coefficients (see [10] for a detailed discussion of frames) of the tonal and the transient component. Specifically, suppose we have a frame that is suitable for analyzing the tonal component and a frame that is suitable for analyzing the transient component. Let A_1, A_2 denote the corresponding frame synthesis operators. Here, we can think of A_1 as a matrix whose columns contain atoms that are well localized in frequency. On the other hand, A_2 may be regarded as a matrix whose columns contain atoms that are well localized in time. Also, let w_1, w_2 denote the frame coefficients of the tonal and transient components. Typically, the distribution of the energy through w_1 and w_2 exhibit different patterns. In w_1, we expect to see thin clusters parallel to the time axis (as in Fig. 7b). In contrast, w_2

This work is supported by a TUBITAK 3501 Project (110E240).

İlker Bayram
Istanbul Technical University
E-mail: ibayram@itu.edu.tr

Ö. Deniz Akyıldız
Bogazici University
E-mail: odakyildiz@gmail.com

is expected to contain thin clusters that lie parallel to the frequency axis (see Fig. 7c). Mixed norms allow to penalize deviations from these patterns.

A description of mixed norms is provided in Section 2.1 (see [25] for a more detailed discussion). For now, suppose that, throughout the set of frame coefficients with unit energy, $\|\cdot\|_a$ is a norm that assumes lower values for the frame coefficients of signals that exhibit tonal behavior. Similarly, suppose that $\|\cdot\|_b$ is a norm that assumes lower values for the frame coefficients of signals that exhibit transient behavior. In this setting, given an audio excerpt y, one possible formulation for decomposition is [27],

$$(w_1, w_2) = \underset{\theta_1,\theta_2}{\operatorname{argmin}} \frac{1}{2}\|y - A_1\,\theta_1 - A_2\,\theta_2\|_2^2 + \lambda_1\,\|\theta_1\|_a + \lambda_2\,\|\theta_2\|_b. \quad (1)$$

This is a *synthesis prior* formulation – the coefficients w_1, w_2 tell us which particular linear combinaton of the corresponding frame atoms should be used to construct the two components. When problems of the form

$$\min_{\theta_1} \frac{1}{2}\|u_1 - \theta_1\|_2^2 + \lambda_1\,\|\theta_1\|_a, \quad (2a)$$

$$\min_{\theta_2} \frac{1}{2}\|u_2 - \theta_2\|_2^2 + \lambda_2\,\|\theta_2\|_b, \quad (2b)$$

are easy to solve (i.e. when the proximal mappings [12] of $\|\cdot\|_a$ and $\|\cdot\|_b$ are easy to realize), the minimization problem in (1) can be solved using iterative shrinkage/thresholding algorithms[1] (ISTA) as in Algorithm 1.

Algorithm 1 An Iterative Shrinkage Algorithm

repeat

$$w_i \leftarrow w_i^k + \alpha A_i^T\,(y - A_1\,w_1^k - A_2\,w_2^k),\ i = 1, 2, \quad (3)$$

$$w_1^{k+1} \leftarrow \underset{\theta_1}{\operatorname{argmin}} \frac{1}{2}\|w_1 - \theta_1\|_2^2 + \lambda_1\,\|\theta_1\|_a, \quad (4)$$

$$w_2^{k+1} \leftarrow \underset{\theta_2}{\operatorname{argmin}} \frac{1}{2}\|w_2 - \theta_2\|_2^2 + \lambda_2\,\|\theta_2\|_b. \quad (5)$$

until convergence

For certain types of mixed norms, (specifically, mixed norms with 'non-overlapping groups' – see Section 2.1), the proximal maps (4), (5) in ISTA type algorithms can be realized in a single step. Unfortunately, this is not always the case – one needs iterative algorithms to solve (4), (5) (see [1,2] for a discussion). This in turn implies nested iterations in an ISTA type algorithm and is not desirable.

An alternative to the synthesis prior formulation is the analysis prior formulation. In this case, the components (x_1, x_2) are estimated as

$$(x_1, x_2) = \underset{t_1,t_2}{\operatorname{argmin}} \frac{1}{2}\|y - t_1 - t_2\|_2^2 + \lambda_1\,\|A_1^T\,t_1\|_a + \lambda_2\,\|A_2^T\,t_2\|_b. \quad (6)$$

Unlike the synthesis prior formulation, where the variables of interest are the frame coefficients of the components, here we directly obtain the components in the time domain.

The formulation in (6) asks that the analysis coefficients of the frame follow the given expected pattern. In view of this, it has been argued [14] that the analysis prior is more demanding since for an overcomplete frame, the analysis coefficients lie on a subspace of smaller dimension than that of the synthesis coefficients – consequently, the analysis coefficients have less freedom (also see [36] for a comparison of the analysis/synthesis formulation and [31] for a recent discussion).

From an algorithmic viewpoint, for the analysis prior formulation, the proximity mappings of $\|A_1^T\cdot\|_a$ and $\|A_2^T\cdot\|_b$ are of interest. These require the solution of

$$\min_{t_1} \frac{1}{2}\|u_1 - t_1\|_2^2 + \lambda_1\,\|A_1^T\,t_1\|_a, \quad (7a)$$

$$\min_{t_2} \frac{1}{2}\|u_2 - t_2\|_2^2 + \lambda_2\,\|A_2^T\,t_2\|_b. \quad (7b)$$

Unfortunately, these proximal mappings are not easy to realize and one has to resort to iterative algorithms [1]. This in turn renders ISTA type algorithms infeasible. To the best of our knowledge, there is no previous work addressing the analysis prior formulation that employs mixed norms for the decomposition problem, as described above.

In this paper, we propose primal-dual algorithms where each substep involves simple operations. The proposed algorithms can handle both the analysis and synthesis formulations. Therefore, they are also of interest in that they allow to compare the results of the two formulations. In addition, we also discuss how to modify the algorithms to obtain a convergent scheme.

In order to derive the algorithms, we first rewrite the minimization problems above as saddle point problems. For this, we first study mixed norms and present a dual description. After transforming the formulation to a saddle point problem, we present a convergent primal-dual algorithm based on recent work [7, 15, 42].

[1] In practice, accelerated versions of ISTA, such as FISTA [5] might also be used. A discussion of FISTA is included in Experiment 2. Here, we employ ISTA in the discussion because of its simpler form.

Related Work on Audio Decomposition

Both the analysis and the synthesis prior formulations described above may be fit into the 'morphological component analysis' (MCA) framework [16,39] – see for example [34, 35]. However, the introduction of the 'tonal + transient + residual' representation of audio predates the work on MCA [18, 28, 41] – see also [29, 38] for earlier work. Different schemes have been proposed to achieve such decompositions. In [13], the authors estimate the tonal and the transient components sequentially (i.e. they first estimate the tonal and then estimate the transient from the residual), using different transforms for the two components and a locally adaptive thresholding strategy. Reference [30] achieves the decomposition by employing a hidden Markov model for the frame coefficients of the components, that takes into account the time-frequency persistence properties of the components. The decomposition is realized sequentially, as in [13]. It is noted in [30] that for such sequential schemes, the performance of the tonal estimation is important as it affects the transient estimate, which is derived from the residual signal. We note that the formulation in this paper *does not* rely on such a scheme – the components are estimated jointly by minimizing a cost function.

A different approach that obtains an adaptive representation of the signal, using different time-frequency frames was proposed in [22]. Roughly, the idea is to consider a tiling of the time-frequency plane and approximate each tile (which needs multiple time-frequency atoms to cover), using atoms from a fixed frame, where the particular frame for a tile is selected from a given set of frames. Given such an approximation of the original signal, the procedure is repeated on the residual. Once a reasonable approximation is at hand, each frame yields a different component/layer.

More recently, mixed norms [27] have been utilized to formulate the multilayer decomposition problem as in (1). A coordinate descent type algorithm was proposed in [27] for this formulation. Albeit effective for the problems in [27], we have found that coordinate descent type algorithms are not very feasible for the analysis prior or when overlapping-group mixed norms are used, since the associated proximal mappings can not be realized in a single step.

A recent work that avoids the use of proximal mappings as in (2), (7) can be found in [8], where the authors consider a denoising formulation based on mixed norms. They derive an algorithm by 'majorizing' [17,20] the mixed norm at each iteration of the algorithm. Although the denoising problem differs from the one considered in this paper, the method in [8] can be extended straightforwardly to a decomposition formulation by majorizing the 'data term' as well.

Remarks on Audio Decomposition

As noted by one of our reviewers, the decomposition of audio as 'tonal+transient+noise' is not necessarily a physically meaningful one. For example, for an audio excerpt with two instruments, such as 'flute + percussion', the work discussed on audio decomposition discussed above does not necessarily aim for the separation of the two instruments – see [9] in this context, which shows that, signals that also contain strong transient components like a glockenspiel signal or a castanets signal can be faithfully represented using sums of AM modulated sinusoids. Nevertheless, for the 'flute + percussion' decomposition problem, it can still be useful to base a decomposition scheme on the difference between the time-frequency characteristics of the two components, which in turn lead to a model akin to 'tonal + transient'. The algorithms discussed in this paper allow to discover the limitations or feasibility of such a model.

The 'transient + tonal' model, when understood in a broad sense as discussed above, may also be useful in restoration tasks, when the undesired artifacts show 'transient' behavior. We present one example in Experiment 1, where the clean signal, which shows strong tonal behavior, is contaminated by artifacts due to recording. The artifacts show quite a different behavior in the time-frequency plane, which, however, also differ from Gaussian noise. In such a situtation, we argue that modelling only the desired signal leads to poorer performance than modelling both the desired signal and the noise. Modelling the two components at the same time leads to a decomposition problem, rather than a denoising problem with a quadratic data discrepancy term.

Outline

Following the Introduction, we discuss how to transform the synthesis/analysis prior problems, in (1), (6) respectively, into saddle point problems in Section 2. For this, we make use of an alternative description of mixed norms. Given the saddle point problem, we describe algorithms for the synthesis and analysis prior formulations in Section 3. We present a heuristic primal-dual scheme in Section 4, by modifying the formally convergent algorithms. Experiments on real audio signals are presented in Section 5, in order to evaluate the

presented algorithms and demonstrate their use. Section 6 is the conclusion.

Notation

From this point on, bold variables as in $\mathbf{x}$, $\mathbf{x}_i$ or $\mathbf{z}^i$ denote vectors. The j^{th} component of the vector $\mathbf{x}$ is denoted by x_j. Throughout the text, $\mathbf{x}_1$ and $\mathbf{x}_2$ denote the 'primary variables' for the tonal and transient component respectively.

2 Saddle Point Formulation

In this section, we rewrite the formulation as a saddle point problem. For this, we will first study the 'dual' representation of mixed norms. Specifically, we will write the mixed norm $\|\mathbf{x}\|$, as '$\sup_{\mathbf{z}\in K}\langle \mathbf{x},\mathbf{z}\rangle$' where K is a set that depends on the norm[2]. This allows us to transform the problem into a saddle point search, which in turn can be solved using primal-dual algorithms.

Remark 1 The function '$\sup_{\mathbf{z}\in K}\langle \mathbf{x},\mathbf{z}\rangle$' is called the 'support function of K' [19].

2.1 Mixed Norms

Definition

Consider a vector $\mathbf{x} = \big(x_1, x_2, \ldots, x_N\big)$. Suppose that $\mathbf{g}^i$'s are a collection of subvectors (groups) where $\mathbf{g}^i = \big(x_{i_1}, x_{i_2}, \ldots, x_{i_M}\big)$. Here, we take equal length $\mathbf{g}^i$'s for simplicity of notation, but in general there are no restrictions on the size of these groups. We note that $\mathbf{g}^i$'s might be sharing elements, i.e. it is possible that $x_{i_n} = x_{j_k}$ for some (i,j,k,n) with $i \neq j$ – in this case we say that the groups overlap [2]. In this setting, the (p,q) - mixed norm of $\mathbf{x}$ is defined as,

$$\|\mathbf{x}\|_{p,q} = \left(\sum_i \|\mathbf{g}^i\|_p^q\right)^{1/q}. \tag{8}$$

Here, $\|\cdot\|_p$ denotes the ℓ_p norm. In this paper, we will be interested in the case $p = 2, q = 1$. For a general discussion, we refer the reader to [25, 27].

For the choice $p = 2$, $q = 1$, a mixed norm, with non-overlapping groups, penalizes deviations from zero of *groups* of coefficients. This is unlike the ℓ_1 norm, which acts on individual coefficients. By allowing the groups to overlap, the goal is to obtain some sort of 'shift invariance', which can be an expected feature for a norm acting on a natural signal like audio. However, we also note at this point that the overlapping group mixed norm definition above is not necessarily the only choice to allow groups to overlap. In particular, as remarked by one of our reviewers, this definition does not force a group to zero, if one of its neighboring groups is non-zero. In this regard, we refer to [21], which introduces an alternative mixed norm. The example below also demonstrates this.

[2] This is indeed possible for any norm, see [19].

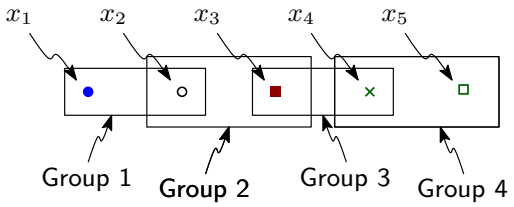

Fig. 1 A length-five vector and four groups for defining a mixed norm.

Example

To give a concrete example, consider the length-five vector in Fig.1. Here, the groups are formed as

$$\mathbf{g}^1 = \big(x_1, x_2\big)\,, \quad \mathbf{g}^2 = \big(x_2, x_3\big)\,,$$
$$\mathbf{g}^3 = \big(x_3, x_4\big)\,, \quad \mathbf{g}^4 = \big(x_4, x_5\big)\,.$$

Given the groups, the mixed norm of $\mathbf{x}$ for $p = 2$, $q = 1$ is,

$$\|\mathbf{x}\|_{2,1} = \sqrt{x_1^2 + x_2^2} + \sqrt{x_2^2 + x_3^2} + \sqrt{x_3^2 + x_4^2} + \sqrt{x_4^2 + x_5^2}. \tag{9}$$

Suppose now that we use this norm in some recovery formulation. In such a situation, if, in the solution, we we want to keep (x_2, x_3), then in addition to $\mathbf{g}^2$, we should also keep $\mathbf{g}^1$ and $\mathbf{g}^3$. In a sense, $\mathbf{g}^2$ being non-zero gives some credit to its neighboring groups. Such a behavior may be questionable from a point of view that seeks sparsity, where it is desired to keep as few groups as possible. However, it might also be a desired behavior when x_i's are actually the analysis coefficients of some audio signal. Because, in our opinion, rather than a cluster of coefficients that start and end abruptly, a smoother appearing cluster is more likely in such a case.

Mixed Norms as Support Functions

Notice that for any vector $\mathbf{x}$ of length-N, we can write

$$\|\mathbf{x}\|_2 = \left(\sum_k |x_k|^2\right)^{1/2}, \tag{10}$$

as

$$\|\mathbf{x}\|_2 = \sup_{\mathbf{z}\in B}\langle \mathbf{x}, \mathbf{z}\rangle, \tag{11}$$

where $B \subset \mathbb{R}^N$ is the unit ball of the ℓ_2 norm, i.e.,

$$B = \{\mathbf{z} : \|\mathbf{z}\|_2 \leq 1\}. \tag{12}$$

Now, let us define $K_i \subset \mathbb{R}^N$ as

$$K_i = B \cap \{\mathbf{z} : z_j = 0, \text{ if } j \notin \{i_1, \dots i_M\}\}, \tag{13}$$

Remark 2 K_i may be regarded as the projection of B onto the coordinates active in the i^{th} group. □

We can now write $\|\mathbf{g}^i\|_2$ as,

$$\|\mathbf{g}^i\|_2 = \sup_{\mathbf{z}^i\in K_i} \langle \mathbf{x}, \mathbf{z}^i\rangle. \tag{14}$$

In this description, we refer to $\mathbf{z}^i$ as the dual variables for the i^{th} group.

Remark 3 Notice that for $\mathbf{z}^i$, i.e., the dual variables for $\mathbf{g}^i = \big(x_{i_1}, x_{i_2}, \dots, x_{i_M}\big)$, it suffices to store only the variables $\mathbf{z}^i_{i_1}, \mathbf{z}^i_{i_2}, \dots, \mathbf{z}^i_{i_M}$, since the rest of $\mathbf{z}^i_j$'s are equal to zero. □

We can now write $\|\mathbf{x}\|_{2,1}$ as,

$$\|\mathbf{x}\|_{2,1} = \sum_i \sup_{\mathbf{z}^i\in K_i} \langle \mathbf{x}, \mathbf{z}^i\rangle. \tag{15}$$

Remark 4 When the groups cover all of x_j's without overlap, the number of dual variables (that need to be stored) is equal to the dimension of $\mathbf{x}$. If the groups overlap, the number of dual variables exceed the dimension of $\mathbf{x}$. □

Remark 5 We can also write $\|\mathbf{x}\|_{2,1} = \sup_{\mathbf{z}\in K}\langle \mathbf{x}, \mathbf{z}\rangle$, where $K = \sum_i K_i$. However, it is more convenient to consider K_i's separately for the primal-dual algorithm. Nevertheless, when $K_i \cap K_j = \{0\}$, it might be preferrable to work with the sums of the sets in the implementation stage. □

Remark 6 Provided that K is a closed and bounded set, we can replace the expressions that involve '$\sup_{\mathbf{z}\in K}$' with '$\max_{\mathbf{x}\in K}$'. We note that this is the case for the sets of interest in the following, therefore we shall use 'max' from this point on. □

Example

For the example in Fig.1, the sets K_i are,

$$\begin{aligned}
K_1 &= \{\mathbf{z} : \sqrt{z_1^2 + z_2^2} \leq 1 \text{ and } z_3 = z_4 = z_5 = 0\}\\
K_2 &= \{\mathbf{z} : \sqrt{z_2^2 + z_3^2} \leq 1 \text{ and } z_1 = z_4 = z_5 = 0\}\\
K_3 &= \{\mathbf{z} : \sqrt{z_3^2 + z_4^2} \leq 1 \text{ and } z_1 = z_2 = z_5 = 0\}\\
K_4 &= \{\mathbf{z} : \sqrt{z_4^2 + z_5^2} \leq 1 \text{ and } z_1 = z_2 = z_3 = 0\}
\end{aligned} \tag{16}$$

Also, notice that $K_1 \cap K_3 = \{0\}$ and $K_1 + K_3$ is easy to describe :

$$K_1 + K_3 = \left\{\mathbf{z} : \sqrt{z_1^2 + z_2^2} \leq 1, \sqrt{z_3^2 + z_4^2} \leq 1 \text{ and } z_5 = 0\right\}. \tag{17}$$

Projections onto K_i

In the following, we will need to project vectors onto K_i. We note that given $\mathbf{x}$, its projection onto K_i can be obtained by projecting $\big(x_{i_1}, x_{i_2}, \dots, x_{i_M}\big)$ onto the unit ℓ_2 ball in $\mathbb{R}^M$ and setting the rest of the entries to zero. Analytically, the projection $\mathbf{p}$ is described as,

$$\begin{aligned}
\big(p_{i_1}, \dots, p_{i_M}\big) &= \big(x_{i_1}, \dots, x_{i_M}\big)\ \frac{1}{\max\big(\|x_{i_1}, \dots, x_{i_M}\|_2, 1\big)},\\
p_j &= 0, \text{ if } j \notin (i_1, i_2, \dots, i_M).
\end{aligned} \tag{18}$$

An Alternative Description

Yet another alternative description of mixed norms will be useful when we discuss convergence of the algorithms. For this, let π_{K_i} denote a linear projection operator defined as follows. For $\mathbf{t} = \pi_{K_i}\mathbf{z}$, we have,

$$t_j = \begin{cases} z_j, & \text{if } j \in (i_1, i_2, \dots, i_M),\\ 0, & \text{if } j \notin (i_1, i_2, \dots, i_M).\end{cases} \tag{19}$$

Notice that $\pi^T_{K_i} = \pi_{K_i}$. In this case, for $\|\mathbf{g}^i\|_2$ given as in (14), we can write

$$\|\mathbf{g}^i\|_2 = \max_{\mathbf{z}^i\in B}\langle \pi_{K_i}\mathbf{x}, \mathbf{z}^i\rangle \tag{20}$$

$$= \max_{\mathbf{z}^i\in B}\langle \mathbf{x}, \pi_{K_i}\mathbf{z}^i\rangle. \tag{21}$$

Therefore,

$$\|\mathbf{x}\|_{2,1} = \sum_i \max_{\mathbf{z}^i\in B} \langle \pi_{K_i}\mathbf{x}, \mathbf{z}^i\rangle. \tag{22}$$

Remark 7 If $K_i \cap K_j = \{0\}$, for all (i, j) pairs with $i \neq j$, then the spectral norm of $\sum_i \pi_{K_i}$ is unity. □

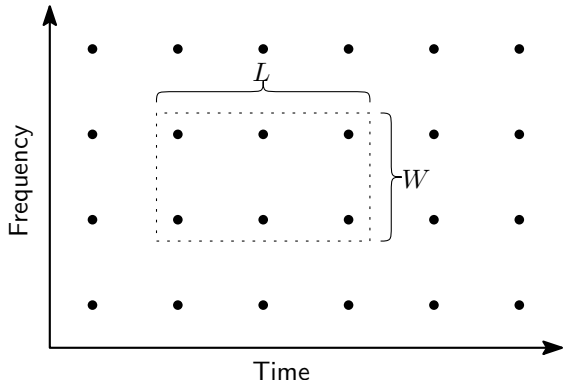

Fig. 2 Definition of a group in a time-frequency map.

2.2 Time and Frequency Persistence Using Mixed Norms

We now discuss how the mixed norms can be put to use for modeling the time/frequency persistence of the components.

Consider Fig. 2, which depicts the coefficients of a time-frequency frame. Here, a prototypical group is shown (dashed rectangle). The group covers a rectangle of L coefficients along the time-axis and W coefficients along the frequency axis. Imagine that we define the other groups by shifting this group along the time axis by multiples of s_L coefficients and along the frequency axis by multiples of s_W coefficients. When $L > W$, the mixed norm, defined by this collection of groups, is expected to assume lower values for 'tonal components' compared to 'transient components'. Similarly, if $W > L$, the resulting mixed norm assumes lower values for 'transient components' compared to 'tonal components'.

In the following, let $\|\cdot\|_{\text{ton}}$ and $\|\cdot\|_{\text{tr}}$ denote mixed norms which assume low values for the tonal and the transient components respectively. Using the description provided in Section 2.1, we write

$$\|\cdot\|_{\text{ton}} = \sum_{i=1}^{n} \max_{\mathbf{z}^i \in C_i} \langle \cdot, \mathbf{z}^i \rangle, \tag{23a}$$

$$\|\cdot\|_{\text{tr}} = \sum_{i=1}^{m} \max_{\mathbf{t}^i \in D_i} \langle \cdot, \mathbf{t}^i \rangle. \tag{23b}$$

Here, C_i and D_i denote projected ℓ_2-norm balls associated with different groups that are used in the definitions of the mixed norms, as discussed in Section 2.1.

2.3 Analysis/Synthesis Formulations

In the following, let A_1 and A_2 denote synthesis operators for time-frequency frames which are suitable for representing the tonal and the transient components respectively. Specifically, the columns of A_1 contain atoms with a high Q-factor (or, well localized in frequency), and the columns of A_2 contain atoms that are well localized in time. Using the the high-Q atoms from A_1, we hope to obtain a parsimonious representation of the tonal component. On the other hand, the time localized atoms from A_2 are more suitable for representing the transient component. We note that these two frames can be STFT frames or constant-Q frames [3, 40]. We also assume that the frames are tight with unit frame bound, i.e. $A_1\, A_1^T = A_2\, A_2^T = I$. However, we do not require that the frames be orthonormal bases, i.e., $A_i^T\, A_i$ may not be the identity operator but only a projection operator.

2.3.1 Analysis Prior Formulation

Given the mixture signal $\mathbf{y}$, the analysis prior formulation for the 'tonal+transient+residual' decomposition task is,

$$\min_{\mathbf{x}_1,\mathbf{x}_2} \max_{\substack{\mathbf{z}^i \in C_i \\ \mathbf{t}^i \in D_i}} \frac{1}{2}\|\mathbf{y} - \mathbf{x}_1 - \mathbf{x}_2\|_2^2 + \lambda_1 \sum_{i=1}^{n} \langle A_1^T\, \mathbf{x}_1, \mathbf{z}^i \rangle + \lambda_2 \sum_{i=1}^{m} \langle A_2^T\, \mathbf{x}_2, \mathbf{t}^i \rangle \tag{A}$$

Here, $\mathbf{x}_1$, and $\mathbf{x}_2$ denote the time-domain tonal and transient components respectively.

This formulation is referred to as an analysis prior formulation because regularization via mixed norms is imposed on the analysis coefficients of the components.

2.3.2 Synthesis Prior Formulation

In the setting described above, the synthesis prior formulation is,

$$\min_{\mathbf{x}_1,\mathbf{x}_2} \max_{\substack{\mathbf{z}^i \in C_i \\ \mathbf{t}^i \in D_i}} \frac{1}{2}\|\mathbf{y} - A_1\, \mathbf{x}_1 - A_2\, \mathbf{x}_2\|_2^2 + \lambda_1 \sum_{i=1}^{n} \langle \mathbf{x}_1, \mathbf{z}^i \rangle + \lambda_2 \sum_{i=1}^{m} \langle \mathbf{x}_2, \mathbf{t}^i \rangle. \tag{S}$$

Here, $\mathbf{x}_1$ and $\mathbf{x}_2$ denote the frame synthesis coefficients for the two components. In the time domain, the tonal component is given by $A_1\, \mathbf{x}_1$ and the transient component is given by $A_2\, \mathbf{x}_2$.

3 Description of the Algorithms

3.1 A General Algorithm

Note that both the analysis and synthesis prior formulations are of the form

$$\min_{\mathbf{x}_1,\mathbf{x}_2} \max_{\substack{\mathbf{z}^i\in C_i\\ \mathbf{t}_i\in D_i}} J(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z}^1,\ldots,\mathbf{z}^n,\mathbf{t}^1,\ldots,\mathbf{t}^m) \tag{24}$$

where $J(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z},\mathbf{t})$ is convex with respect to $\mathbf{x}_i$'s and concave with respect to $\mathbf{z}^i$'s and $\mathbf{t}^i$'s. In this setting, a general primal-dual algorithm is provided in Algorithm 2.

Algorithm 2 A Primal Dual Algorithm

Initialize $\mathbf{x}_i \leftarrow 0$, $\hat{\mathbf{x}}_i \leftarrow 0$, $\bar{\mathbf{x}}_i \leftarrow 0$, $\mathbf{z}^j \leftarrow 0$, $\mathbf{t}^k \leftarrow 0$, for $i = 1,2$, $j = 1,\ldots,n$, $k = 1,\ldots,m$.
repeat

$$(\mathbf{z},\mathbf{t}) \leftarrow \underset{\substack{\mathbf{v}^i\in C_i\\ \mathbf{r}^i\in D_i}}{\operatorname{argmax}} -\frac{1}{2\,\gamma}\left(\sum_{i=1}^{n}\|\mathbf{z}^i-\mathbf{v}^i\|_2^2+\sum_{i=1}^{m}\|\mathbf{t}^i-\mathbf{r}^i\|_2^2\right) + J(\bar{\mathbf{x}}_1,\bar{\mathbf{x}}_2,\mathbf{v},\mathbf{r}) \tag{D}$$

$$(\mathbf{x}_1,\mathbf{x}_2) \leftarrow \underset{\mathbf{u}_1,\mathbf{u}_2}{\operatorname{argmin}} \frac{1}{2\,\gamma}\sum_{i=1}^{2}\|\mathbf{x}_i-\mathbf{u}_i\|_2^2 + J(\mathbf{u}_1,\mathbf{u}_2,\mathbf{z},\mathbf{t}) \tag{P}$$

$\bar{\mathbf{x}}_i \leftarrow 2\,\mathbf{x}_i - \hat{\mathbf{x}}_i$, for $i = 1,2$.
$\hat{\mathbf{x}}_i \leftarrow \mathbf{x}_i$, for $i = 1,2$.
until convergence

Algorithm 2 is an adaptation of the primal-dual algorithm proposed in [7, 15]. For related earlier work, with simpler, yet computationally heavier methods, we also refer to [24, 32].

The main steps of the algorithm consist of proximal steps [12] on the primal and dual variables – these are the steps labeled (P) and (D) respectively. In practice, these steps alone also lead to a convergent scheme (see [42]). However, adding a low cost correction step, that is reminiscent of an 'extragradient' step [24], leads to a formally convergent algorithm [7, 15].

Choice of γ

Algorithm 2 contains a step parameter γ. In order to ensure the convergence of the algorithm, we can take,

$$\gamma \le \left[\max\left(\lambda_1\left\lceil\frac{L_1\,W_1}{s_{L_1}\,s_{W_1}}\right\rceil,\lambda_2\left\lceil\frac{L_2\,W_2}{s_{L_2}\,s_{W_2}}\right\rceil\right)\right]^{-1}. \tag{25}$$

Here, (L_1,W_1) and (s_{L_1},s_{W_1}) determine the group size and shift parameters for the tonal component (see Section 2.2). Similarly, (L_2,W_2) and (s_{L_2},s_{W_2}) are the associated parameters for the transient component.

The convergence of Algorithm 2 under this choice of γ is discussed in Appendix A. Below, we derive the steps (D) and (P) for the analysis and synthesis prior formulations.

3.2 Steps for the Analysis Prior Formulation

Recall the analysis formulation in (A). In the following, we denote the convexoconcave function in (A) as $J_A(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z},\mathbf{t})$, i.e.,

$$J_A(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z},\mathbf{t}) = \frac{1}{2}\|\mathbf{y}-\mathbf{x}_1-\mathbf{x}_2\|_2^2 + \lambda_1\sum_{i=1}^{n}\langle A_1^T\,\mathbf{x}_1,\mathbf{z}^i\rangle + \lambda_2\sum_{i=1}^{m}\langle A_2^T\,\mathbf{x}_2,\mathbf{t}^i\rangle \tag{26}$$

Let us now derive the steps of the algorithm.

Realizing (D) *for the Analysis Prior Formulation*

In this case, we need to solve decoupled constrained minimization problems. Specifically, let

$$\mathbf{v}^{i*} = \underset{\mathbf{v}^i\in C_i}{\operatorname{argmin}} \frac{1}{2\,\gamma}\,\|\mathbf{v}^i\|_2^2 - \frac{1}{\gamma}\langle\mathbf{v}^i,\mathbf{z}^i\rangle - \lambda_1\,\langle A_1^T\,\bar{\mathbf{x}}_1,\mathbf{v}^i\rangle. \tag{27}$$

Note that the right hand side may be written as a quadratic plus a constant term. Discarding the constant (since it does not change the minimizer) we obtain,

$$\mathbf{v}^{i*} = \underset{\mathbf{v}^i\in C_i}{\operatorname{argmin}} \left\|\mathbf{v}^i-\left(\mathbf{z}^i+\gamma\,\lambda_1\,A_1^T\,\bar{\mathbf{x}}_1\right)\right\|_2^2. \tag{28}$$

Finally, since C_i is a closed and convex set, we can express the minimizer as the projection of the term enclosed in parentheses to C_i. That is, if P_{C_i} denotes the projection operator onto C_i, we have

$$\mathbf{v}^{i*} = P_{C_i}\left(\mathbf{z}^i+\gamma\,\lambda_1\,A_1^T\,\bar{\mathbf{x}}_1\right). \tag{29}$$

Similarly, it follows that,

$$\mathbf{r}^{i*} = P_{D_i}\left(\mathbf{t}^i+\gamma\,\lambda_2\,A_2^T\,\bar{\mathbf{x}}_2\right), \tag{30}$$

where P_{D_i} denotes the projection operator onto D_i.

Realizing (P) *for the Analysis Prior Formulation*

Notice that the function to be minimized in this case is convex and differentiable :

$$f_A(\mathbf{u}_1,\mathbf{u}_2)=\frac{1}{2\,\gamma}\sum_{i=1}^{2}\|\mathbf{x}_i-\mathbf{u}_i\|_2^2+J_A(\mathbf{u}_1,\mathbf{u}_2,\mathbf{z},\mathbf{t}) \quad (31)$$

The derivatives with respect to $\mathbf{u}_1$ and $\mathbf{u}_2$ should evaluate to zero at the minimizers $\mathbf{u}_1^*$, $\mathbf{u}_2^*$:

$$0=\gamma^{-1}\big(\mathbf{u}_1^*-\mathbf{x}_1\big)+\mathbf{u}_1^*+\mathbf{u}_2^*-\mathbf{y}+\lambda_1\sum_{i=1}^{n}A_1\,\mathbf{z}^i \quad (32)$$

$$0=\gamma^{-1}\big(\mathbf{u}_2^*-\mathbf{x}_2\big)+\mathbf{u}_1^*+\mathbf{u}_2^*-\mathbf{y}+\lambda_2\sum_{i=1}^{m}A_2\,\mathbf{t}^i. \quad (33)$$

Rearranging, we obtain,

$$\begin{bmatrix}\mathbf{u}_1^*\\ \mathbf{u}_2^*\end{bmatrix}=\begin{bmatrix}\gamma+1 & \gamma\\ \gamma & \gamma+1\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{x}_1+\gamma\big(\mathbf{y}-\lambda_1\sum_{i=1}^{n}A_1\,\mathbf{z}^i\big)\\ \mathbf{x}_2+\gamma\big(\mathbf{y}-\lambda_2\sum_{i=1}^{m}A_2\,\mathbf{t}^i\big)\end{bmatrix} \quad (34)$$

3.3 Steps for the Synthesis Prior Formulation

Recall the synthesis prior formulation in (S). In the following, we shall denote the convexoconcave function in (S) by $J_S(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z},\mathbf{t})$, i.e.,

$$J_S(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z},\mathbf{t})=\frac{1}{2}\|\mathbf{y}-A_1\,\mathbf{x}_1-A_2\,\mathbf{x}_2\|_2^2+\lambda_1\sum_{i=1}^{n}\langle\mathbf{x}_1,\mathbf{z}^i\rangle+\lambda_2\sum_{i=1}^{m}\langle\mathbf{x}_2,\mathbf{t}^i\rangle. \quad (35)$$

Realizing (D) *for the Synthesis Prior Formulation*

Once again this case leads to decoupled constrained minimization problems. Specifically, we have,

$$\mathbf{v}^{i*}=\operatorname*{argmin}_{\mathbf{v}^i\in C_i}\frac{1}{2\,\gamma}\,\|\mathbf{v}^i\|_2^2-\frac{1}{\gamma}\langle\mathbf{v}^i,\mathbf{z}^i\rangle-\lambda_1\,\langle\bar{\mathbf{x}}_1,\mathbf{v}^i\rangle \quad (36)$$

$$=\operatorname*{argmin}_{\mathbf{v}^i\in C_i}\big\|\mathbf{v}^i-\big(\mathbf{z}^i+\gamma\,\lambda_1\,\bar{\mathbf{x}}_1\big)\big\|_2^2 \quad (37)$$

$$=P_{C_i}\big(\mathbf{z}^i+\gamma\,\lambda_1\,\bar{\mathbf{x}}_1\big) \quad (38)$$

Similarly, it follows that,

$$\mathbf{r}^{i*}=P_{D_i}\big(\mathbf{t}^i+\gamma\,\lambda_2\,\bar{\mathbf{x}}_2\big). \quad (39)$$

Realizing (P) *for the Synthesis Prior Formulation*

As in the analysis prior formulation, the function to be minimized in this case is convex and differentiable :

$$f_S(\mathbf{u}_1,\mathbf{u}_2)=\frac{1}{2\,\gamma}\sum_{i=1}^{2}\|\mathbf{x}_i-\mathbf{u}_i\|_2^2+J_S(\mathbf{u}_1,\mathbf{u}_2,\mathbf{z},\mathbf{t}) \quad (40)$$

The derivatives with respect to $\mathbf{u}_1$ and $\mathbf{u}_2$ should evaluate to zero at the minimizers $\mathbf{u}_1^*$, $\mathbf{u}_2^*$:

$$0=\gamma^{-1}\big(\mathbf{u}_1^*-\mathbf{x}_1\big)+A_1^T\,A_1\,\mathbf{u}_1^*+A_1^T\,A_2\,\mathbf{u}_2^*-A_1^T\,\mathbf{y}+\lambda_1\sum_{i=1}^{n}\mathbf{z}^i \quad (41)$$

$$0=\gamma^{-1}\big(\mathbf{u}_2^*-\mathbf{x}_2\big)+A_2^T\,A_2\,\mathbf{u}_2^*+A_2^T\,A_1\,\mathbf{u}_1^*-A_2^T\,\mathbf{y}+\lambda_2\sum_{i=1}^{m}\mathbf{t}^i. \quad (42)$$

For the sake of notation, let us define, $P_i=A_i^T\,A_i$. We remark that, thanks to the Parseval frame assumption, P_i is a projection operator that projects the coefficients to the range space of A_i^T. Consequently, it is idempotent, i.e. $P_i^2=P_i$.

Rearranging the equalities, we obtain,

$$\begin{bmatrix}I+\gamma\,P_1 & \gamma\,A_1^T\,A_2\\ \gamma\,A_2^T\,A_1 & I+\gamma\,P_2\end{bmatrix}\begin{bmatrix}\mathbf{u}_1^*\\ \mathbf{u}_2^*\end{bmatrix}=\begin{bmatrix}\mathbf{x}_1+\gamma\,\big(A_1^T\,\mathbf{y}-\lambda_1\,\sum_i\mathbf{z}^i\big)\\ \mathbf{x}_2+\gamma\,\big(A_2^T\,\mathbf{y}-\lambda_2\,\sum_i\mathbf{t}^i\big)\end{bmatrix} \quad (43)$$

To solve this system, let us multiply both sides of this equality by the matrix

$$\begin{bmatrix}I+\gamma\,P_1 & -\gamma\,A_1^T\,A_2\\ -\gamma\,A_2^T\,A_1 & I+\gamma\,P_2\end{bmatrix}. \quad (44)$$

This gives,

$$\begin{bmatrix}I+2\gamma\,P_1 & 0\\ 0 & I+2\gamma\,P_2\end{bmatrix}\begin{bmatrix}\mathbf{u}_1^*\\ \mathbf{u}_2^*\end{bmatrix}=\begin{bmatrix}\mathbf{x}_1-\gamma\,\lambda_1\,Z\\ \mathbf{x}_2-\gamma\,\lambda_2\,T\end{bmatrix}+\gamma\begin{bmatrix}A_1^T\big\{\mathbf{y}+A_1\,(\mathbf{x}_1-\gamma\,\lambda_1\,Z)-A_2\,(\mathbf{x}_2-\gamma\,\lambda_2\,T)\big\}\\ A_2^T\big\{\mathbf{y}+A_2\,(\mathbf{x}_2-\gamma\,\lambda_2\,T)-A_1\,(\mathbf{x}_1-\gamma\,\lambda_1\,Z)\big\}\end{bmatrix}=\begin{bmatrix}\mathbf{c}_1\\ \mathbf{c}_2\end{bmatrix} \quad (45)$$

where $Z=\sum_i\mathbf{z}^i$, $T=\sum_i\mathbf{t}^i$. To solve this system, we note that

$$P_i\big(I+2\gamma\,P_i\big)\,\mathbf{u}_i^*=\big(1+2\gamma\big)\,P_i\,\mathbf{u}_i^*=P_i\,\mathbf{c}_i. \quad (46)$$

Therefore,

$$\mathbf{u}_i^*=\mathbf{c}_i-\frac{2\,\gamma}{1+2\,\gamma}P_i\,\mathbf{c}_i \quad (47)$$

Remark 8 Note that the equality $P_i\, A_i^T = A_i^T$ can be used to arrange the expressions. Specifically, if we define

$$\mathbf{e}_1 = \mathbf{x}_1 - \gamma\, \lambda_1\, Z \tag{48}$$

$$\mathbf{e}_2 = \mathbf{x}_2 - \gamma\, \lambda_2\, T, \tag{49}$$

then,

$$\mathbf{c}_1 = \mathbf{e}_1 + \gamma\, \overbrace{A_1^T \Big(\mathbf{y} + A_1\, \mathbf{e}_1 - A_2\, \mathbf{e}_2 \Big)}^{\mathbf{f}_1} \tag{50}$$

$$\mathbf{c}_2 = \mathbf{e}_2 + \gamma\, \underbrace{A_2^T \Big(\mathbf{y} + A_2\, \mathbf{e}_2 - A_1\, \mathbf{e}_1 \Big)}_{\mathbf{f}_2}. \tag{51}$$

Finally, we have that $P_i\, \mathbf{c}_i = P_i\, \mathbf{e}_i + \gamma\, \mathbf{f}_i$. □

4 Heuristic Primal-Dual Schemes

We previously noted that (see Remark 4), when the groups, used to define the mixed norm, overlap, the number of dual variables exceed the number of primal variables. In practice, this may lead to memory problems, especially if the overlap between the groups is high. In order to avoid this, we present below a heuristic scheme based on the primal-dual algorithms discussed in Section 3. In the following, we restrict ourselves to the analysis prior formulation. However, similar schemes can be devised for the synthesis prior formulation as well.

Now let $\mathbf{z}$ denote the dual variables associated with $A_1^T\, \mathbf{x}_1$ or $A_2^T\, \mathbf{x}_2$. Also, let $\mathcal{N}(z_j)$ denote a neighborhood-vector associated with z_j, like $\mathcal{N}(z_j) = \big(z_j, z_{j_1}, \ldots, z_{j_m}\big)$, where m possibly depends on j. Suppose that there is an asssociated neighborhood for each coefficient. In this case, if the neighborhoods contain multiple elements, there will be overlaps between the neighboorhoods of different coefficients. We note that this is not necessarily the case for the groups used to define mixed norms in Section 2.1. For mixed norms, it is possible to have a grouping system where the groups contain multiple elements, but no two groups contain the same element.

For the setting described above, let $\mathcal{P}_j$ be an operator that maps $\mathbf{z}$ to a number, defined solely in terms of $\mathcal{N}_j$. Specifically, we experimented with $\mathcal{P}_j$ defined as,

$$\mathcal{P}_j(\mathbf{z}) = z_j\, \frac{1}{\max\big(\|\mathcal{N}(z_j)\|, 1\big)}, \tag{52}$$

where $\|\mathcal{N}(z_j)\|$ denotes some norm of the vector $\mathcal{N}(z_j)$. Specific choices are discused in Experiment 4.

Given $\mathcal{P}_j$, we propose Algorithm 3, as a modification of Algorithm 2. Here, the idea is to replace the projections in the (D) step for the analysis prior formulation (see Section 3.2) with $\mathcal{P}_j$. One important difference is that, the projections in Section 3.2 map vectors to vectors. Here, the operators $\mathcal{P}_j$ map vectors to single coefficients. Another difference is that, Algorithm 3 requires significantly fewer dual variables than Algorithm 2. This makes Algorithm 3 more efficient in terms of memory usage.

Algorithm 3 A Primal-Dual Scheme

Initialize $\mathbf{x}_1 \leftarrow 0$, $\mathbf{x}_2 \leftarrow 0$, $\mathbf{z} \leftarrow 0$, $\mathbf{t} \leftarrow 0$. Set γ.
repeat

$$(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2) \leftarrow \underset{\mathbf{u}_1, \mathbf{u}_2}{\operatorname{argmin}} \frac{1}{2\,\gamma} \sum_{i=1}^{2} \|\mathbf{x}_i - \mathbf{u}_i\|_2^2 + \frac{1}{2}\, \|\mathbf{y} - \mathbf{u}_1 - \mathbf{u}_2\|_2^2 + \lambda_1\, \langle \mathbf{u}_1, A_1^T\, \mathbf{z} \rangle + \lambda_2\, \langle \mathbf{u}_2, A_2^T\, \mathbf{t} \rangle \quad \text{(P2)}$$

$\bar{\mathbf{z}} \leftarrow \mathbf{z} + \lambda_1\, \gamma\, A_1^T\, (2\bar{\mathbf{x}}_1 - \mathbf{x}_1)$
$\bar{\mathbf{t}} \leftarrow \mathbf{t} + \lambda_2\, \gamma\, A_2^T\, (2\,\bar{\mathbf{x}}_2 - \mathbf{x}_2)$
for all j **do**
  $z_j = \mathcal{P}_j(\bar{\mathbf{z}})$.
end for
for all k **do**
  $t_k = \mathcal{P}_k(\bar{\mathbf{t}})$.
end for
$\mathbf{x}_1 = \bar{\mathbf{x}}_1$
$\mathbf{x}_2 = \bar{\mathbf{x}}_2$
until convergence

We do not have a proof of convergence for Algorithm 3. However, we have observed convergence in practice.

Although Algorithm 3 is not formally convergent, it has a very flexible form. In particular, one could modify the operators $\mathcal{P}_j$ in order to adapt the algorithm to different structures or trends that are observed/expected in the coefficients.

Algorithm 3 may be regarded as the primal-dual counterpart of the algorithms proposed in [26] for iterative shrinkage type algorithms. In [26], the authors modify ISTA by replacing the shrinkage step with operators that have more desirable properties. Here, the primal-dual algorithms do not contain shrinkage operators but instead contain projection operators. In line with this difference, we propose to replace the projection operators with operators that mimick projections.

5 Experiments

Matlab code for the experiments is available at ‘ `http://web.itu.edu.tr/ibayram/CoSep/` ’.

Experiment 1 (Restoring an Old Record)

The observed signal is an old recording of a tune played by kaval. The signal has 22×10^4 samples at a

sampling frequency of 32 KHz (i.e. it is 7 sec in duration). The record has periodically occuring artifacts, due to the particular method of recording. The artifacts appear as vertical bars in the spectrogram (see Fig. 3a), which can also be observed in the time-domain signal (see the signal plotted with a thin line in Fig 4).

We model the kaval tune as a tonal component and the artifact as a transient component. For the tonal component, we use an STFT with a Hamming window[3] of length 2048 samples (64 msec), and a Hop-size of 512 samples (16 msec). For the transient component, we use an STFT with a Bartlett window of length 512 samples (16 msec), and a Hop-Size of 256 samples (8 msec). In terms of the terminology introduced in Section 2.2, we choose the group sizes that define the mixed norms as follows. For the tonal component we set $L = 16$, $W = 2$, $s_L = 4$, $s_W = 2$. For the transient component, we set $L = 2$, $W = 16$, $s_L = 2$, $s_W = 4$. Also, we set $\lambda_1 = \lambda_2 = 0.005$ – these values are chosen manually to ensure that the residual is very low and a decent separation is achieved.

We employ Algorithm 2 for the analysis prior formulation. The spectra of the resulting tonal and transient components are shown in Fig. 3(b,c). Perceptually, the tonal component is almost free of the artifacts, wheras in the transient component, the tune is heard faintly. We also show in Fig. 4 a time-domain excerpt, where an artifact is visible around the middle of the original signal (thin line). Observe that the tonal component (thick line) has grasped the harmonic component from the given distorted mixture.

We note that the restoration performance achieved with the transient+tonal model is not easy to achieve with a denoising formulation that involves a quadratic data discrepancy term. Specifically, we could also set the problem as a denoising problem of the form

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \lambda \|A^T\mathbf{x}\|_{\text{ton}}, \tag{53}$$

where $\|A^T\mathbf{x}\|_{\text{ton}}$ is a mixed norm prior placed on the analysis coefficients of $\mathbf{x}$. However, in such a formulation, the quadratic term would fail to provide a fair description of the artifacts. On the other hand, the transient+tonal decomposition models both the signal and the artifacts better. In order to test this claim, we used the same STFT frame and the same mixed norm as the tonal component above. We selected λ manually in order to achieve a fair amount of denoising while fairly preserving the harmonics (which are weak to start with). The spectrogram of the denoised signal is shown in Fig 5. Although the artifacts are somewhat suppressed, the performance is poorer than that of the transient+tonal decomposition formulation. Also, perceptually, the denoising result is less pleasing compared to the decomposition formulation[4]. □

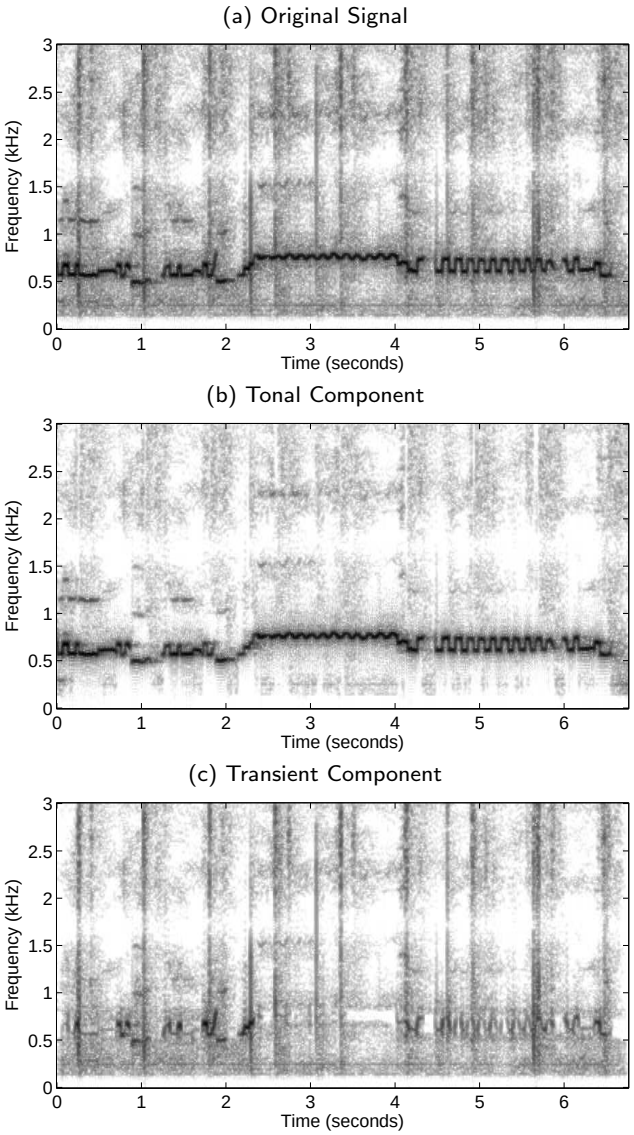

Fig. 3 Restoration of an old recording. (a) The spectrogram of the original recording – a tune played by kaval. The recording contains artifacts which appear as vertical lines in the spectrogram. Our goal is to eliminate these artifacts. (b) The tonal component obtained by employing Algorithm 2 with the analysis prior. The artifact is significantly suppressed. (c) The resulting transient component. This component contains mostly the artifacts. The tune is faintly heard.

Experiment 2 (Speed Comparison with Existing Algorithms)

[3] We actually normalize the windows in order to make the STFT a tight frame. For the examples in this paper, the normalized windows do not significantly differ from the reported ones (i.e. Bartlett, Hamming, etc.), because of the redundancy of the frames.

[4] The results can be listened to at
‘ `http://web.itu.edu.tr/ibayram/CoSep/` ’.

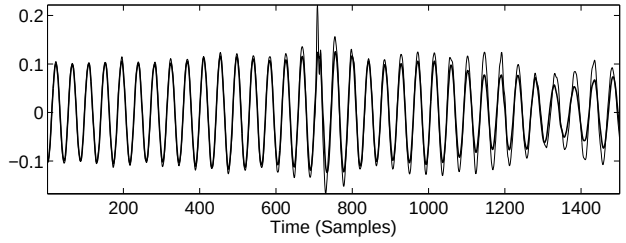

Fig. 4 Time-domain excerpts from the signals in Experiment 1. Thin : Original Signal, Thick : Tonal Component. The inharmonic component which occurs in the middle of the excerpt due to the undesired artifacts of recording is significantly suppressed in the tonal component.

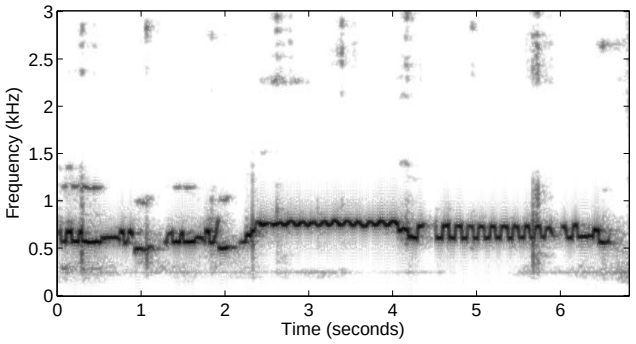

Fig. 5 The denoised signal obtained by modeling the desired signal in Experiment 1 with a tonal component (see the text). Observe that the denoised signal contains more of the artifacts than the tonal component shown in Fig. 3b.

In order to evaluate the convergence speed of the primal-dual algorithm, we compared its performance with an accelerated ISTA algorithm, for the synthesis prior formulation. Specifically, we use 'FISTA' [5] described in Algorithm 4.

Algorithm 4 FISTA

Initialize the variables. Set $t = 1$.

repeat

$$w_i \leftarrow z_i + \alpha\, A_i^T\,(y - A_1\, z_1 - A_2\, z_2),\ i = 1, 2, \tag{54}$$

$$\tilde{w}_i \leftarrow w_i^k,\ i = 1, 2, \tag{55}$$

$$w_1^k \leftarrow \underset{\theta_1}{\operatorname{argmin}}\, \frac{1}{2}\|w_1 - \theta_1\|_2^2 + \lambda_1\, \|\theta_1\|_a, \tag{56}$$

$$w_2^k \leftarrow \underset{\theta_2}{\operatorname{argmin}}\, \frac{1}{2}\|w_2 - \theta_2\|_2^2 + \lambda_2\, \|\theta_2\|_b. \tag{57}$$

$$\tilde{t} \leftarrow \frac{1 + \sqrt{1 + 4\,t^2}}{2}, \tag{58}$$

$$z_i \leftarrow w_i^k + \left(\frac{t-1}{\tilde{t}}\right)\,\left(w_i^k - w_i^{k-1}\right)\ i = 1, 2, \tag{59}$$

$$t \leftarrow \tilde{t} \tag{60}$$

until convergence

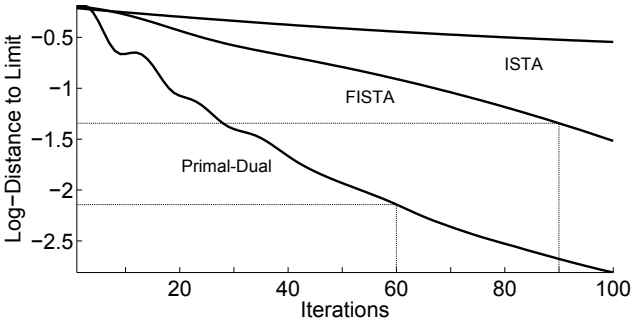

Fig. 6 Log distance to the limit with respect to iterations for ISTA, FISTA and the proposed Primal-dual algorithm. The iterations of (F)ISTA are not directly comparable to the iterations of the primal-dual algororithm. In the text, we argue that N iterations of ISTA correspond roughly to $2N/3$ iterations of the primal-dual algorithm in terms of computational cost. Therefore 90 iterations of (F)ISTA cost (computationally) as much as 60 iterations of the primal-dual algorithm.

We use the signal described in Experiment 1. For a specific mixed norm definition, we employ three different algorithms for the synthesis prior. These are, ISTA (Algorithm 1), FISTA (Algorithm 4) and the primal-dual algorithm (Algorithm 2). First, we ran each algorithm for 5000 iterations, in order to reach approximately their limit. Then, we monitored the distance to these estimated limits as iterations progress. The natural logarithm of these distances (normalized so that the norm of the associated limit is unity) are depicted in Fig. 6.

We note that the computational cost of the iterations are different for (F)ISTA and the primal-dual algorithm, therefore this graph should be interpreted with care. Below we provide an approximate comparison between the computational costs of the iterations of the algorithms.

For FISTA, the proximal maps in (56) and (57) can be realized using an iterative scheme [2]. Although this leads to iterations within iterations, and therefore is not desirable, we observed that nevertheless, few inner iterations (as few as 5) are sufficient to approximately realize (56) and (57) in practice. Each of these inner iterations involve additions and some simple projections. Compared to a forward or inverse STFT, these are relatively low cost operations. Neglecting the cost of these inner iterations, it can be argued that the main cost of an iteration is proportional to the number of forward and inverse STFTs. We observe that each iteration of FISTA requires two forward and two inverse transforms, adding to a total of four 'transforms'. A similar argument can be made for ISTA.

For the primal-dual algorithm, we can similarly argue that the main cost is again due to the number of STFTs. We observe that each iteration can be realized

with two forward and four inverse STFTs, adding to a total of six transforms.

Comparing the number of transforms for the two types of algorithms, we can say that an iteration of the primal-dual algorithm requires 3/2 times more computational power than an iteration of (F)ISTA. Therefore, we should really be comparing N iterations of ISTA/FISTA with $2N/3$ iterations of the primal-dual algorithm. The grid lines in Fig. 6 facilitate this comparison. We observe that 60 iterations of the primal-dual algorithm takes us closer to the limit than 90 iterations of (F)ISTA.

Although the primal-dual algorithm converges faster (per computation) compared to FISTA, we think that FISTA performs acceptably for the synthesis-prior formulation. However, this is not the case for the analysis prior formulation. For FISTA applied on the analysis prior formulation, the main issue is the realization of the proximal maps (56), (57). As in the synthesis prior formulation, these could be achieved with an iterative scheme (see [1]) – leading to inner iterations within outer iterations. However, in this case, each inner iteration itself involves a forward as well as an inverse transform, and the cost of an ISTA type algorithm grows rapidly. We note that applying FISTA to these inner iterations as well (as done in the TV 'denoising' step in [4]) would not totally solve this problem since even a few inner iterations require multiple STFTs. In contrast, the primal-dual algorithm avoids these problems. Hence, we think that the primal dual approach is especially suitable for analysis prior formulations.

As a final remark, we should note that the analysis of cost above for both the primal dual and (F)ISTA is a rather crude one. The neglected parts do also have costs in practice. Therefore the convergence speed comparison should be interpreted with caution.

Experiment 3 (Separation of Ney and Darbuka)

We have at hand a mixture signal that consists of a 'ney' tune, accompanied by 'darbuka' (a percussion instrument). Our goal is to separate the two components. The signal has 33×10^4 samples at a sampling frequency of 22 KHz (i.e. it is 15 sec in duration).

In order to separate the two components, we model the ney tune as a tonal component, the darbuka part as a transient component and employ Algorithm 2. We consider both the analysis and the synthesis prior formulations in order to see their differences/similarities. We use the same parameters for the analysis and synthesis prior formulations. For the tonal component, we use an STFT with a Hamming window of length 1024 samples (47 msec) and a Hop-size of 256 samples (12 msec). We use the same mixed norms as in Example 1. We also set $\lambda_1 = 0.005$, $\lambda_2 = 0.006$.

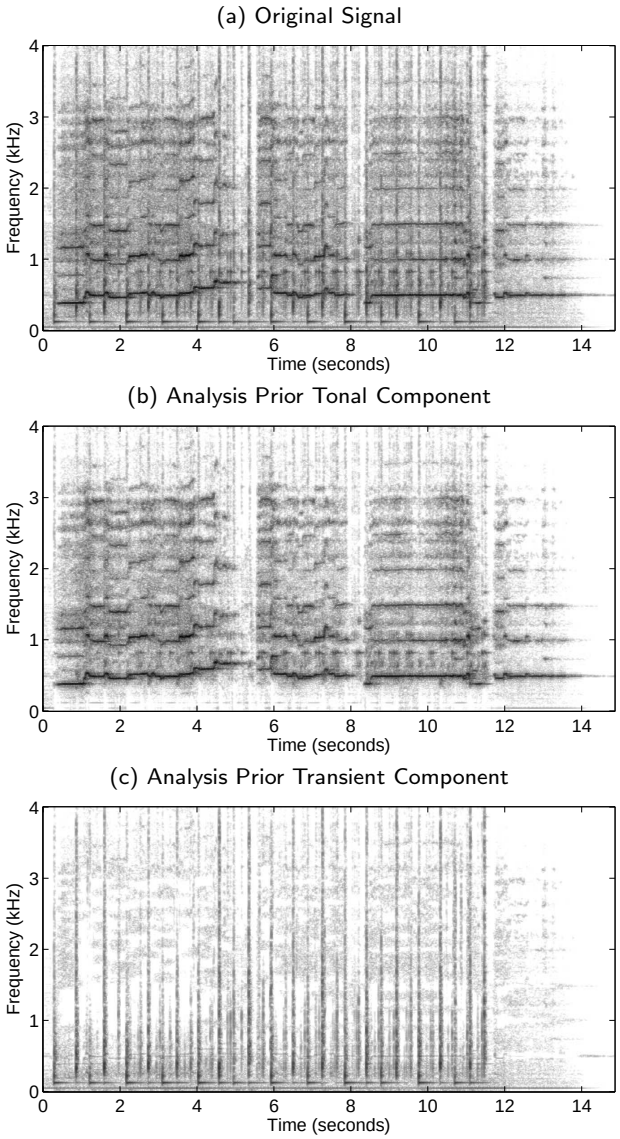

Fig. 7 Separation of the components of a musical signal. (a) The spectrogram of the mixture signal. The signal is composed of a tune played by ney, accompanied by darbuka. The input SNR for the ney ('tonal') and darbuka ('transient') components are 10.52 dB and -10.52 dB respectively. Notice that the two SNRs are related by a factor of -1. (b,c) The tonal and the transient component obtained by the analysis prior formulation (SNRs are 16.41 and 5.72 dB respectively).

For the analysis prior formulation, Algorithm 2 produces the two components whose spectrograms are shown in Fig. 7(b,c). For the synthesis prior, the solution gives us the STFT coefficients of the components. These coefficients are shown in Fig. 9(a,b). Using these coefficients, we reconstruct the time-domain components. The spectrograms of the resulting time-domain components are shown in Fig. 9(c,d).

We note that although the analysis prior frame coefficients (Fig. 7(b,c)) and the synthesis prior coefficients

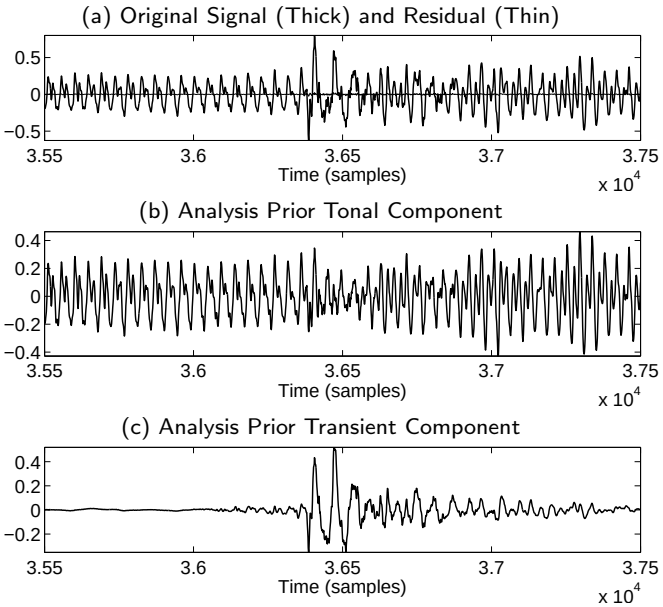

Fig. 8 Time-domain excerpts from the signals in Experiment 3. (a) Original signal (thick) and the residual (thin), (b) the tonal component for the analysis prior, (c) the transient component for the analysis prior.

(Fig. 9(a,b)) look quite different, the time-domain components do not show a marked difference. This is also evident from a comparison of Fig. 7(b,c) and Fig. 9(c,d). This similarity is somewhat contrary to our expectations, since the analysis and synthesis prior formulations are not equivalent when the frames are redundant [14]. Increasing the redundancy of the frames increases the dimension of the nullspaces of the synthesis operators and therefore, the steps labeled (P) for the analysis and synthesis prior formulations can lead to different outcomes. In this regard, we expected the difference between the analysis and the synthesis prior solutions to increase with increasing redundancy. However, even though we increased the redundancy until memory problems start to occur, we have not been able to see a significant difference between the solutions (see [11] for a recent interesting discussion – also [6] includes a comparison of the analysis and synthesis prior formulations for a reconstruction problem where data is incoherently sampled).

A similar observation could be made for the time domain signals in Fig. 8,10 as well. Both the analysis and synthesis formulations have achieved a fair decomposition. Specifically, the locally periodic components are captured by the tonal component and the rapid changes due to the addition of percussive elements are captured by the transient component in the results of both formulations. However, it is not easy to point to unique characteristics or to distinguish the two results. □

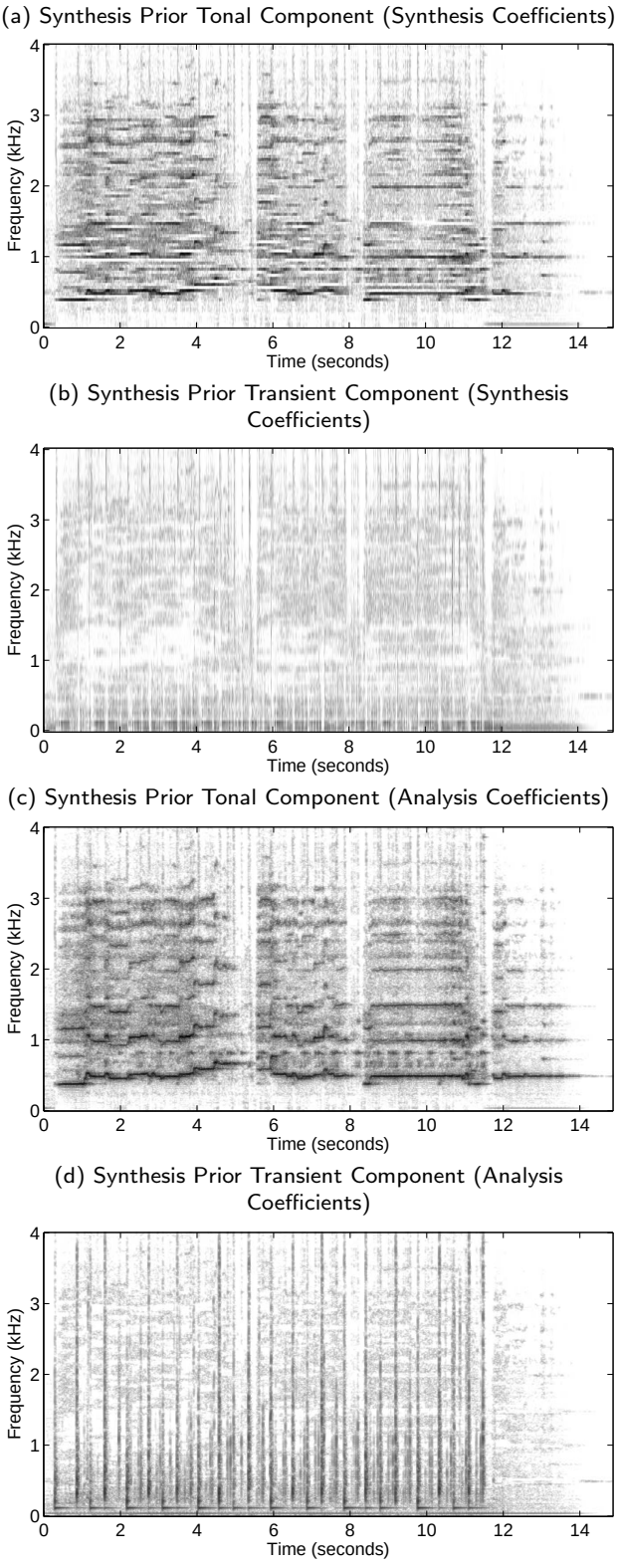

Fig. 9 The spectrograms of the components in Experiment 3, for the synthesis prior formulation. The resulting SNRs are, 16.67 dB for the tonal component and 6.07 dB for the transient component. (a,b) The synthesis coefficients of the tonal and the transient component. (c,d) The analysis coefficients of the tonal and the transient components. Observe that the spectrograms in (c,d) are similar to the ones in Fig. 7(b,c).

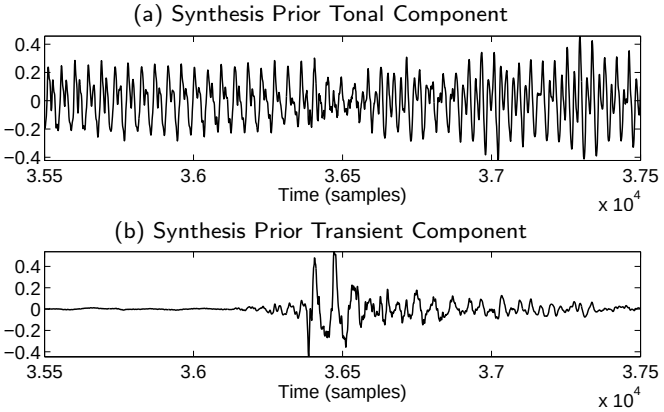

Fig. 10 Time-domain excerpts from the signals in Experiment 3. (a) the tonal component for the synthesis prior, (b) the transient component for the synthesis prior.

Experiment 4 (Performance of Heuristic Algorithms) In order to test the utility of the scheme in Section 4, we employed Algorithm 3 for the problem in Experiment 3.

Recall that Algorithm 3 requires a definition of $\mathcal{P}_j$. In view of (52), this is achieved by selecting $\mathcal{N}_j$ and defining $\|\mathcal{N}_j\|$. We select $\mathcal{N}_j$ as a rectangle in the time-frequency plane, centered at z_j, of length L and width W, as in Fig. 2.

We took $\|\mathcal{N}_j\|$ to be a weighted ℓ_2 norm, where the weighting is achieved by a 2D Bartlett window, centered around z_j (or t_j). The neighborhoods for the tonal component are defined by $L = 15$, $W = 3$. The neighborhoods for the transient component are defined by $L = 3$, $W = 15$. In this setting, Algorithm 3 produces the components shown in Fig. 11 (also see the time domain signals in Fig. 12).

Perceptually, the decomposition is less successful compared to those obtained in Experiment 3. However, the spectrograms of the resulting signals follow the desired pattern.

□

Remark 9 Experiments 3 and 4 aim to achieve a musically meaningful decomposition. In order to achieve this, a simple model is imposed on the components and the problem is formulated as a convex minimization problem. This in turn leads to a simple algorithm. The trade-off between the simplicity of the model/algorithm can be likened to rate distortion optimized coding [33], where a trade-off between distortion and bit-rate exists – see also [23] in this line.

More audio examples can be found at
‘ `http://web.itu.edu.tr/ibayram/CoSep/` ’ .

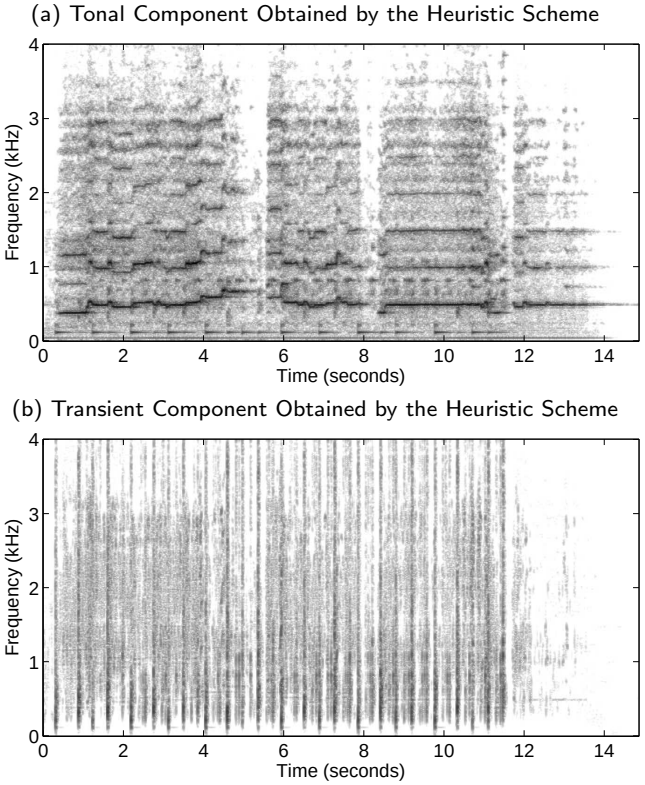

Fig. 11 Components obtained by the Heuristic PD algorithm, using a weighted ℓ_2 norm. (a) The tonal component (SNR = 14.89 dB), (b) the transient component (SNR = 4.36 dB).

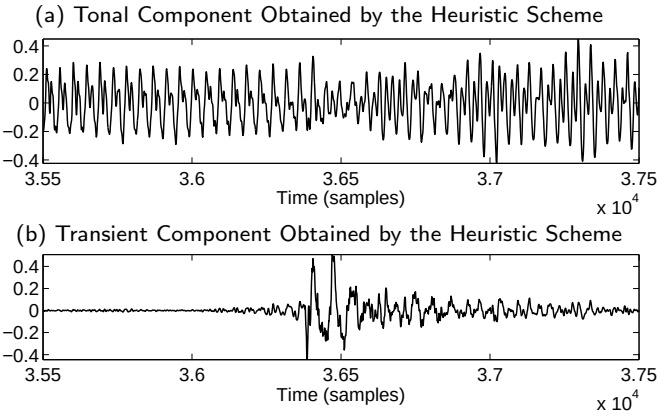

Fig. 12 Time-domain excerpts from the signals in Experiment 4. The same excerpt as in Fig. 8 and Fig. 10 is used. The residual is practically zero and is not shown. (a) The tonal component obtained by the heuristic scheme, (b) the transient component obtained by the heuristic scheme.

6 Conclusion

We considered the audio decomposition problem, where the goal is to separate components with different characteristics. Specifically, we focused on mixed-norm formulations. By writing the mixed norms as support functions, we derived equivalent saddle point problems for

both analysis and synthesis prior formulations. Following this, we proposed a primal-dual algorithm for the resulting saddle point problems. We demonstrated the utility of the algorithms on real audio signals. In particular, we showed that the algorithms can also be useful for restoration tasks which involve additive distortions that characteristically differ from the signal of interest.

The primal-dual algorithms, although formally convergent, can require a lot of memory if the groups are large and the overlap between the groups approach group sizes. To address this issue, we also a derived flexible heuristic scheme that can achieve similar decompositions with a better use of memory. The heuristic algorithm is also of interest because of its flexible form. This flexibility allows to quickly test different models for the components, through a change of the neighborhood shape and the utilized 'norm'.

A Conditions of Convergence

The analysis/synthesis prior algorithms described in Section 3 have a single parameter γ that needs to be selected. Below we provide an upper bound on γ that ensures convergence.

A.1 A General Problem and Algorithm

Consider the problem,

$$\min_{\mathbf{x}} \max_{\mathbf{z}\in K} \frac{1}{2}\,\|\mathbf{y} - A\,\mathbf{x}\|_2^2 + \langle F\,\mathbf{x}, \mathbf{z}\rangle \tag{61}$$

where K is a closed convex set and F is a matrix. For this problem, consider Algorithm 5.

Algorithm 5 A Convergent Primal Dual Algorithm

Initialize $\mathbf{z}^0 \leftarrow 0$, $\mathbf{x}^0 \leftarrow 0$, $\bar{\mathbf{x}}^0 \leftarrow 0$, $k \leftarrow 0$.
repeat

$$\mathbf{z}^{k+1} \leftarrow \operatorname*{argmax}_{\mathbf{z}\in K} -\frac{1}{2\,\gamma}\,\|\mathbf{z}^k - \mathbf{z}\|_2^2 + \langle F\,\bar{\mathbf{x}}^k, \mathbf{z}\rangle. \tag{62}$$

$$\mathbf{x}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{x}} \frac{1}{2\,\gamma}\,\|\mathbf{y}-\mathbf{x}^k\|_2^2 + \frac{1}{2}\,\|\mathbf{y}-A\,\mathbf{x}\|_2^2 + \langle F\,\mathbf{x}, \mathbf{z}^{k+1}\rangle \tag{63}$$

$\bar{\mathbf{x}}^{k+1} \leftarrow 2\,\mathbf{x}^{k+1} - \mathbf{x}^k$.
$k \leftarrow k+1$.
until convergence

Proposition 1 *Let ρ be the spectral radius of F. If*

$$\gamma \le 1/\rho, \tag{64}$$

then the sequence $(\mathbf{x}^k, \mathbf{z}^k)$ produced by Algorithm 5 converges to a saddle point of the functional in (61).

Algorithm 5 along with its proof of convergence follows by adapting the algorithm/discussion in [7,15]. Below we discuss how it relates to the synthesis and analysis prior formulations/algorithms discussed in Section 3.

A.2 Relation Between the Problems

Let us now relate (A) with (61). For this, we define the super vectors $\mathbf{x}$ and $\mathbf{z}$ as,

$$\mathbf{x} = \begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix}\mathbf{z}^1\\ \vdots\\ \mathbf{z}^n\\ \mathbf{t}^1\\ \vdots\\ \mathbf{t}^m\end{bmatrix}. \tag{65}$$

Also, define the matrices F, E as (recall the definition of π_{K_i}'s in Section 2.1),

$$F = \begin{bmatrix}\lambda_1\,\pi_{C_1} & 0\\ \vdots & \vdots\\ \lambda_1\,\pi_{C_n} & 0\\ 0 & \lambda_2\,\pi_{D_1}\\ \vdots & \vdots\\ 0 & \lambda_2\,\pi_{D_m}\end{bmatrix}, \quad E = \begin{bmatrix}A_1^T & 0\\ 0 & A_2^T\end{bmatrix}. \tag{66}$$

In this setting, the problem in (A) is equivalent to the problem,

$$\min_{\mathbf{x}} \max_{\mathbf{z}\in C} \frac{1}{2}\,\left\|\mathbf{y} - \begin{bmatrix}I & I\end{bmatrix}\mathbf{x}\right\|_2^2 + \langle F\,E\,\mathbf{x}, \mathbf{z}\rangle. \tag{67}$$

where C is defined as the Cartesian products of the sets C_i $(i = 1, \ldots, n)$ and D_j $(j = 1, \ldots, m)$ – see (23). Also, 'I' denotes the identity matrix that applies to $\mathbf{x}_1$ or $\mathbf{x}_2$. Therefore, in view of Prop. 1, we can say that the analysis prior algorithm converges if $\gamma < 1/\rho$ where ρ is the spectral radius of $F\,E$. We shall consider specific cases in Section 5. Provided that A_1^T and A_2^T are associated with tight frames, we have that $\rho(F\,E) \le \rho(F)$. Thus, $\rho(F)$ is of interest.

Similarly, the synthesis prior problem is equivalent to the problem,

$$\min_{\mathbf{x}} \max_{\mathbf{z}\in C} \frac{1}{2}\,\left\|\mathbf{y} - \begin{bmatrix}A_1 & A_2\end{bmatrix}\mathbf{x}\right\|_2^2 + \langle F\,\mathbf{x}, \mathbf{z}\rangle. \tag{68}$$

Thus the synthesis prior algorithm converges if $\gamma < 1/\rho$ where ρ is the spectral radius of F. Therefore in both problems, we see that estimating $\rho(F)$ is of interest.

For $\lambda_1 = \lambda_2 = 1$, $\rho(F)$ may be regarded as a measure of how much the mixed norm groups overlap. Recall that when mixed norms have overlapping groups, a specific coefficient may appear in more than one group. To that end, the number of elements in all of the groups exceed the number of total time-frequency coefficients. The ratio gives an upper bound (actually a tight one) on $\rho(F)$ (again, when $\lambda_i = 1$). This ratio is easy to compute under the setting discussed in Section 2.1.

For a general choice of λ_1, λ_2, suppose that the group sizes for the tonal component are L_1, W_1 and the shift parameters are s_{L_1}, s_{W_1}. Also, let the associated parameters for

the transient component be L_1, W_1, s_{L_2}, s_{W_2}. In this case, an upper bound on $\rho(F)$ is,

$$\rho(F) \leq \max\left(\lambda_1 \left\lceil \frac{L_1\, W_1}{s_{L_1}\, s_{W_1}} \right\rceil, \lambda_2 \left\lceil \frac{L_2\, W_2}{s_{L_2}\, s_{W_2}} \right\rceil\right) \tag{69}$$

In view of the discussion above, which guarantees convergence if $\gamma\,\rho(F) \leq 1$, we note that the rhs of (69) may be taken as γ^{-1}.

Acknowledgements We thank Prof. Barış Bozkurt, Bahcesehir University, Istanbul, Turkey for comments and providing the signals used in the experiments.

We also thank the reviewers for their constructive remarks.

References

1. Bayram, I.: Denoising formulations based on support functions (2011). Unpublished Manuscript, available at ‘ http://web.itu.edu.tr/ibayram/DenSup.pdf ’
2. Bayram, I.: Mixed-norms with overlapping groups as signal priors. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP) (2011)
3. Bayram, I.: An analytic wavelet transform with a flexible time-frequency covering (2012). Manuscript, available from ‘ http://web.itu.edu.tr/ibayram/AnDWT/ ’
4. Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems **18**(11), 2419–2434 (2009)
5. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences **1**, 183–202 (2009)
6. Candès, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted l1 minimization. Journal of Fourier Analysis and Applications **14**(5), 877–905 (2008)
7. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision **40**(1), 120–145 (2011)
8. Chen, P.Y., Selesnick, I.W.: Translation invariant shrinkage of group sparse signals (2012). Manuscript, available from ‘ http://eeweb.poly.edu/iselesni/ogs/ ’
9. Christensen, M.G., Jakobsson, A., Andersen, S.V., Jensen, S.H.: Linear amplitude decomposition for sinusoidal audio coding. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP) (2005)
10. Christensen, O.: An Introduction to Frames and Riesz Bases. Birkhäuser (2003)
11. Cleju, N., Jafari, M.G., Plumbley, M.D.: Choosing analysis or synthesis recovery for sparse reconstruction. In: Proc. Eur. Sig. Proc. Conf (EUSIPCO) (2012)
12. Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In: H.H. Bauschke, R.S. Burachik, P.L. Combettes, V. Elser, D.R. Luke, H. Wolkowicz (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer, New York (2011)
13. Daudet, L., Torrésani, B.: Hybrid representations for audiophonic signal encoding. Signal Processing **82**(11), 1595–1617 (2002)
14. Elad, M., Milanfar, P., Rubinstein, R.: Analysis versus synthesis in signal priors. Inverse Problems **23**(3), 947–968 (2007)
15. Esser, E., Zhang, X., Chan, T.F.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sciences **3**(4), 1015–1046 (2010)
16. Fadilli, M.J., Starck, J.L., Bobin, J., Moudden, Y.: Image decomposition and separation using sparse representations: An overview. Proceedings of the IEEE **98**(6), 983–994 (2010)
17. Figueiredo, M.A.T., Bioucas-Dias, J.M., Nowak, R.D.: Majorization-minimization algorithms for wavelet-based image restoration. IEEE Trans. Image Processing **16**(12), 2980–2991 (2007)
18. Hamdy, K., Ali, M., Tewfik, H.: Low bit rate high quality audio coding with combined harmonic and wavelet representations. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP) (1996)
19. Hiriart-Urruty, J.B., Lemaréchal, C.: Fundamentals of Convex Analysis. Springer (2004)
20. Hunter, D.R., Lange, K.: A tutorial on MM algorithms. Amer. Statist. **58**(1), 30–37 (2004)
21. Jacob, L., Obozinski, G., Vert, J.P.: Group lasso with overlap and graph lasso. In: Proc. 26^{th} Int. Conf. on Machine Learning (2009)
22. Jaillet, F., Torrésani, B.: Time-frequency jigsaw puzzle: adaptive multiwindow and multilayered Gabor representations. Int. J. for Wavelets and Multiresolution Information Processing **5**(2), 293–316 (2007)
23. Jensen, T., Østergaard, J., Dahl, J., Jensen, S.H.: Multiple-description l_1-compression. IEEE Trans. Signal Processing **59**(8), 3699–3711 (2011)
24. Korpelevich, G.: The extragradient method for finding saddle points and other problems. Ekonomika i Matematcheskie Metody **12**, 747–756 (1976)
25. Kowalski, M.: Sparse regression using mixed norms. J. of Appl. and Comp. Harm. Analysis **27**(3), 303–324 (2009)
26. Kowalski, M., Siedenburg, K., Dörfler, M.: Social sparsity! Neighborhood systems enrich structured shrinkage operators (2012). Manuscript, available from ‘ http://hal.archives-ouvertes.fr/hal-00691774 ’
27. Kowalski, M., Torrésani, B.: Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients. Signal, Image and Video Processing **3**(3), 251–264 (2009)
28. Levine, S., Smith, J.O.: A sines+transients+noise audio representation for data compression and time/pitch-scale modifications. In: Proc. 105th Convention of the AES (1998)
29. McAulay, R., Quatieri, T.: Speech analysis synthesis based on a sinusoidal representation. IEEE Trans. Acoust., Speech, and Signal Proc. **34**(4), 744–754 (1986)
30. Molla, S., Torrésani, B.: A hybrid scheme for encoding audio signal using hidden markov models of waveforms. J. of Appl. and Comp. Harm. Analysis **18**(2), 137–166 (2005)
31. Nam, S., Davies, M.E., Elad, M., Gribonval, R.: Cosparse analysis modelling – Uniqueness and algorithms. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP) (2011)
32. Popov, L.D.: A modification of the Arrow-Hurwicz method for search of saddle points. Matematicheskie Zametki **28**(5), 777–784 (1980)
33. van Schijndel, N.H., et al.: Adaptive RD optimization hybrid sound coding. J. Audio Eng. Soc. **56**(10), 787–809 (2008)
34. Selesnick, I.W.: Resonance-based signal decomposition: A new sparsity-enabled signal analysis method. Signal Processing **91**(12)

35. Selesnick, I.W., Bayram, I.: Oscillatory + transient signal decomposition using overcomplete rational-dilation wavelet transforms. In: Proceedings of SPIE (Wavelets XIII) (2009)
36. Selesnick, I.W., Figueiredo, M.A.T.: Signal restoration with overcomplete wavelet transforms : Comparison of analysis and synthesis priors. In: Proceedings of SPIE (Wavelets XIII) (2009)
37. Siedenburg, K., Dörfler, M.: Structured sparsity for audio signals. In: Proc. Int. Conf. on Digital Audio Effects (DAFx) (2011)
38. Smith, J.O., Serra, X.: "PARSHL: An analysis/synthesis program for nonharmonic sounds based on a sinusoidal representation. In: Proc. International Computer Music Conference (1987)
39. Starck, J.L., Elad, M., Donoho, D.: Redundant multiscale transforms and their application for morphological component analysis. Advances in Imaging and Electron Physics **132** (2004)
40. Velasco, G.A., Holighaus, N., Dörfler, M., Grill, T.: Constructing an invertible constant-q transform with nonstationary Gabor frames. In: Proc. Int. Conf. on Digital Audio Effects (DAFx) (2011)
41. Verma, T.S., Levine, S., Meng, T.H.Y.: Transient modeling synthesis : A flexible transient analysis/synthesis tool for transient signals. In: Proc. ICMC (1997)
42. Zhu, M., Chan, T.F.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration (2008). UCLA CAM Report [08-34]