A Simple Prior For Audio Signals

İlker Bayram and Mustafa Kamasak

Abstract—We propose a simple prior for restoration problems involving oscillatory signals. The prior makes use of an underlying analytic frame decomposition with narrow subbands. Other than this, the prior does not have any other parameters, which makes it simple to use and apply. We demonstrate the utility of the proposed prior through some real audio restoration experiments.

Index Terms—Inverse problem, audio prior, audio inpainting, saddle point.

I. INTRODUCTION

An inverse problem requires to reconstruct an object x, given distorted and noisy observations, y. A variational approach formulates this task as a minimization problem of the form,

$$\hat{x} = \underset{u}{\operatorname{argmin}} f(y, u) + g(u). \qquad (1)$$

Here, the functional $f(y, \cdot)$, called the 'data term', penalizes deviations from the observations and, $g(\cdot)$, called the 'prior term', penalizes deviations from our model/prior expectations regarding x. The functional for the data term is determined by the distortion and the noise properties affecting the observation. Consequently, the data term can be considered 'independent' of the properties of x. In contrast, the prior term is chosen by taking into account the particular family of signals that the desired object belongs to. However, from a practical point of view, it is not sufficient that the prior be a good descriptor of the family – it is also desired that the prior be easy to work with. In this paper, our goal is to propose an effective prior for audio signals, that is also easy to work with.

Bayesian formulations, that make use of the probability distribution of the expected object, provide alternatives to the variational formulation in (1). For audio, various prior distributions have been proposed [1–5]. Among these, [1–3] study priors that also take into account the time/frequency persistence of audio signals by employing hyperparameters for modelling the distributions of the time-frequency frame[1] coefficients. Typically, by sampling the posterior distribution

İ. Bayram is with the Dept. of Electronics and Communication Engineering, Istanbul Technical University, Istanbul, Turkey (e-mail : ibayram@itu.edu.tr).

M. Kamasak is with the Dept. of Computer Engineering, Istanbul Technical University, Istanbul, Turkey (e-mail : kamasak itu.edu.tr).

[1]A *frame* may be regarded as a collection of atoms that span a space of interest in a stable manner – see [6] for a detailed discussion.
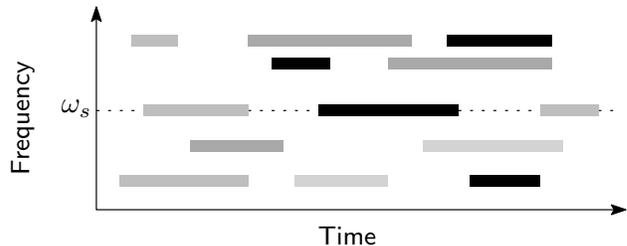
Fig. 1. We think of an audio signal as a linear combination of constant-frequency components with time-varying weights. We also assume that the weights can be well-approximated by piecewise constant functions. Here, notice that the weight for the middle component (with frequency ω_s) can be obtained by demodulating the component by ω_s.

(given the observations) through Markov Chain Monte Carlo techniques, statistics of the posterior distribution are obtained, which are in turn useful for estimation – in particular, minimum mean square estimates are obtained simply by estimating the mean of the posterior distributions (also see [7, 8] in this context). Alternative works, such as [4, 5] work with individual time-frequency domain coefficients. These latter approaches do not involve hyperparameters that link neighboring coefficients, but in turn they do not make use of the time/frequency persistence properties of audio signals.

Recently, mixed norms have been proposed for modelling the time-frequency coefficients of audio [9]. In contrast to the ℓ_1 norm, which promotes individual sparsity, mixed norms allow a small number of clusters/groups of coefficients, where each group is non-sparse. Thus mixed norms are more suitable, compared to the ℓ_1 norm, for capturing clusters observed in time-frequency coefficients of typical audio signals. Also, by adjusting the size, shape and weighting within the groups, components with different characteristics (such as tonal or transient) can be captured [10, 11] – also see [12–16] for related discussion/applications. Unlike the priors used in Bayesian formulations mentioned above, mixed norms do not involve hyperparameters to be estimated. However, one still needs to decide the group structure (the size, shape, weights etc.).

In this paper, we propose a prior that does not require to select parameters/structures. The idea rests on a simplified version of the sinusoidal models in [17, 18]. Specifically, we think of an audio signal as a linear combination of constant-frequency components, whose weights can be well approximated by piecewise constant functions (see Fig. 1).

Under this model, if $w_f(k)$ denotes the (discrete-time) weight function of a component (with center frequency 'f'), the total variation of this function, defined as [19, 20],

$$\sum_k |w_f(k) - w_f(k+1)| \tag{2}$$

is expected to be small. In addition, since the number of such constant-frequency components is also expected to be small, a prior as described in three stages below, becomes feasible.

(i) Partition the audio signal into narrow subbands,
(ii) Demodulate each subband to obtain its 'baseband' counterpart (or the weight function),
(iii) Compute the total variation of each 'baseband' and sum these resulting values.

We remark that for a signal as in Fig. 1, such a prior is expected to take smaller values than a random signal of the same length. Although the three-stage description above is valid, we actually realize the proposed prior in a simpler way. In particular, the 'partitioning' in (i) is realized by applying a time-frequency frame with narrow subbands. For this, an STFT or an analytic wavelet frame[2] with a high Q-factor[3] (see e.g. [21, 22]) may be utilized. After that, we work with the frame coefficients and steps (ii), (iii) are realized by an equivalent 'phase corrected difference operator' applied on the coefficients of each subband. The details are provided in Section II.

We also noted that the simplicity of use is an important criterion for the utility of a prior. To that end, we shall also discuss a 'dual' expression of the proposed prior. The dual expression allows to rewrite (1) as a saddle point problem. This in turn allows to employ several effective schemes for saddle point problems which have recently been proposed [23–27] (also see [28, 29]).

A. Some Remarks About the Proposed Prior

- The proposed prior is based on the simple signal model, described in Fig. 1. Originally, the model does not include frequency-varying components like chirps or effects like vibrato/tremolo. However, such components can be handled by local approximations, provided the parameters of the underlying time-frequency frame are adequate. We provide some discussion on the limitations of the prior in Section II-B.

[2]Here, by 'analytic wavelet frame', we refer to a discrete wavelet transform where the wavelets' Fourier transforms are one-sided (i.e. the wavelets contain either negative or positive frequencies).

[3]The Q-factor of a function is defined to be the ratio of its center frequency to its bandwidth.

- The prior is not adaptive – it does not depend on the underlying signal. One could propose an adaptive scheme by employing a prior segmentation step as in [30] so as to obtain stationary segments, followed by a frequency decomposition of these segments. Alternatively, based on a parametric modulated stationary signal model as in [31], one could first estimate the parameters of the underlying components and make use of this additional knowledge. Although these approaches/models are viable, they also complicate the model/prior. In this paper, our goal is to develop a prior which can be useful, although resting on a model that might be rather simplistic for real signals.
- For inverse problems, an alternative to the variational or Bayesian formulations in (1) is to employ matching pursuit [32] schemes. In those approaches, the prior knowledge is embedded directly into the procedure for obtaining a solution. For instance, under a sparsity assumption in a certain dictionary, one tries to fulfill a given objective using as few atoms as possible from the dictionary. This is done by selecting the atoms one by one, starting with the most important ones. For audio, other structures than sole sparsity can be incorporated by using fixed combinations of atoms rather than single atoms – for a discussion, we refer to [33].
- The proposed prior operates directly on the signal, not on the coefficients of a particular representation of the signal. In this regard, the proposed prior is an instance of an analysis prior [34]. Although a synthesis counterpart of the proposed prior could also be feasible, such a discussion is beyond the scope of the paper.

Notation

Throughout the paper, vector fields defined over the time-frequency parameter space are denoted with a bar as in $\bar{z}$. The value of this vector field at a time-frequency point (s, k) is denoted by $\bar{z}_{s,k}$ and is a vector. Also, throughout the text, we use the same inner product symbol $\langle ., . \rangle$ for functions, for vector fields and for individual vectors. For instance, if $\bar{z}$ and $\bar{u}$ are two vector fields defined over the parameter space (s, k) we have the equivalence $\langle \bar{z}, \bar{u} \rangle = \sum_{s,k} \langle \bar{z}_{s,k}, \bar{u}_{s,k} \rangle$, where the two inner products on the left and right hand sides of this equality have different domains.

Outline

In Section II, we present some observations regarding time-frequency coefficients of frames. Based on these observations, we give the formal description of the proposed prior. We also provide in this Section, a discussion on

the limitations of the prior. A 'dual' description of the prior, which facilitates its use in variational formulations, is derived in Section III. Two different applications are discussed and algorithms are provided in Section IV, along with experiments that employ the algorithms. Section V is the conclusion.

II. DESCRIPTION OF THE PRIOR

A. Motivation and Computation Details

Consider a single component signal with fixed frequency ω_c and variable amplitude $a(t)$ given as,

$$x(t) = a(t) e^{-j(\omega_c t + \phi_0)}. \tag{3}$$

Suppose we sample this signal with period T, to produce the discrete-time signal

$$x_d(n) = a(Tn) e^{-j(\omega_c T n + \phi_0)}. \tag{4}$$

Provided that $a(t)$ varies slowly, or T is small enough, we will have

$$|x_d(n)| = a(Tn) \approx a\big(T(n+1)\big) = |x_d(n+1)| \tag{5}$$

In words, the magnitudes of the neighboring samples are approximately the same. For phases, we observe that

$$\angle x_d(n+1) = T \omega_c (n+1) + \phi_0 \tag{6}$$
$$= (T \omega_c n + \phi_0) + T \omega_c \tag{7}$$
$$= \angle x_d(n) + T \omega_c. \tag{8}$$

In words, the phases are the same up to a correction factor. Combining these two observations, we can write

$$x_d(n+1) - e^{-j \omega_c T} x_d(n) \approx 0. \tag{9}$$

In this simple model, ω_c can also be replaced by a time-dependent frequency like $\omega_c + \epsilon(t)$, so that $x(t) = a(t) \exp\big(-j(\omega_c + \epsilon(t)) t + \phi_0\big)$. Provided that $\epsilon(t)$ is small, we would still expect (9) to hold. This is the main observation that our prior depends on.

More generally, consider a multicomponent continuous-time signal given as,

$$x(t) = \sum_i a_i(t) e^{-j(\omega_i t + \phi_i)}. \tag{10}$$

Also let $f(t) = g(t) e^{-j(\omega_c t + \phi)}$ be a function, well localized in the time-frequency plane. We will use the time-shifts of $f(t)$ to sample $x(t)$. Note that,

$$\langle x, f \rangle = \int x(t) f^*(t) dt$$
$$= \sum_i \int a_i(t) g(t) e^{-j\big((\omega_i - \omega_c) t + (\phi_i - \phi)\big)} dt. \tag{11}$$

Now suppose that $\omega_m \approx \omega_c$ for some m and $a_m(t) g(t)$ is well concentrated (say, around $t = t_0$) [4]. In this case, (11) will approximately evaluate to

$$e^{-j\big((\omega_m - \omega_c) t_0 + (\phi_m - \phi)\big)} \int a_m(t) g(t) dt. \tag{12}$$

Now if $a_m(t)$ changes slowly enough so that

$$a_m(t) g(t - T) \approx a_m(t - T) g(t - T), \tag{13}$$

and

$$(\omega_m - \omega_c) T \ll 1, \tag{14}$$

we will have,

$$\langle x(t), f(t - T) \rangle \approx e^{-j\omega_m T} \langle x(t), f(t) \rangle. \tag{15}$$

On the other hand, if (13) or (14) is violated, (15) is no longer expected to hold.

Based on these observations, we propose a prior for signals with slowly varying harmonic components, as follows. Suppose we have a collection of discrete-time, analytic time-frequency atoms $f_{s,k}(n)$, where s denotes the subband parameter and k denotes the time-shift parameter (these may be the atoms of an STFT transform or an analytic wavelet frame with a high-Q factor as in [21, 22]). For subband s, we assume that $f_{s,k+1}(n) = f_{s,k}(n - N_s)$, i.e. the time-shift between the consecutive atoms in the sth band is given by N_s. Also, let the center frequency of the sth band be denoted by ω_s. Now let

$$c_{s,k} = \langle x, f_{s,k} \rangle = \sum_n x(n) f_{s,k}^*(n). \tag{16}$$

In this setting, we define our prior $P(x)$ as,

$$P(x) = \sum_{s,k} \big| c_{s,k} - c_{s,k+1} e^{j \omega_s N_s} \big|. \tag{17}$$

Observe that for $\omega_s = 0$,

$$\sum_k \big| c_{s,k} - c_{s,k+1} e^{j \omega_s N_s} \big| = \sum_k \big| c_{s,k} - c_{s,k+1} \big| \tag{18}$$

is nothing but the total variation of $c_{s,k}$, viewed as a function of k. For $\omega_s \neq 0$, $c_{s,k}$, is no longer expected to be a lowpass or a piecewise constant sequence. In this case,

$$\sum_k \big| c_{s,k} - c_{s,k+1} e^{j \omega_s N_s} \big| \tag{19}$$

is equivalent to first demodulating $c_{s,k}$ (by ω_s) and then computing the total variation.

---

[4]This is to ensure that $(\omega_m - \omega_c) G \ll 1$, where G denotes the approximate length of the time support of $g(t)$. In this case, we will have $e^{-j\big((\omega_m - \omega_c) t + (\phi_m - \phi)\big)} \approx e^{-j\big((\omega_m - \omega_c) t_0 + (\phi_m - \phi)\big)}$ on the time support of $a_k(t) g(t)$.

B. Limitations of the Model

The computation of the prior can be summarized as follows. Given the time-frequency coefficients of the signal, $c_{s,k}$,

(i) first apply a phase corrected difference operator along each subband (this is equivalent to computing '$c_{s,k} - c_{s,k+1} e^{j \omega_s N_s}$' – see (17)),

(ii) then compute the ℓ_1 norm of the phase corrected difference coefficients.

If we interpret the ℓ_1 norm as a rough measure of sparsity, we can argue that, in order for step (ii) to yield a small number, it is necessary that the difference spectrogram produced by step (i) be sparse. In order to understand when this is indeed the case and test whether this is true for audio signals, we conducted a few experiments using different signals.

1) Synthetic Signal: Our first experiment is performed on a synthetic signal. The spectrogram of the signal is shown in Fig. 2a. The sampling frequency is taken as 32 kHz. For the STFT, we used a window length of 2048 samples (64 msec) and a Hop-size of 128 samples (4 msec). The signal consists of two components. The first component is a constant frequency sinusoid, multiplied by a Hamming window. The frequency of the component is chosen so as to match the center frequency of a particular subband of the STFT. The second component is a linear chirp whose frequency varies slowly – again this is obtained by windowing an infinitely long chirp signal with a Hamming window. For both components, we observe some leakage to the neighboring subbands.

We applied the phase corrected difference operator to the STFT coefficients of the synthetic signal to obtain the difference spectrogram shown in Fig. 2b. For the first component, there appears a hollow canal in between two subbands with relatively higher magnitude. The hollow canal is produced in the subband whose center frequency matches the frequency of the signal component. In the neighboring leakage subbands, the suppression is not as strong, due to a violation of (14). For the second component, as the frequency of the chirp approaches the center frequency of a particular subband, (14) starts to hold and cancellation occurs (see the short white segments pointed to by the arrows). Although the prior is based on a model that does not include frequency varying components, we see that, such components can still be handled to some extent since they can be approximated by a sequence of shorter constant frequency components. This being said, we also observe that if the chirp frequency is not close to the center frequency of any subband, the suppression provided by the phase corrected difference operator is weak. This also means that,
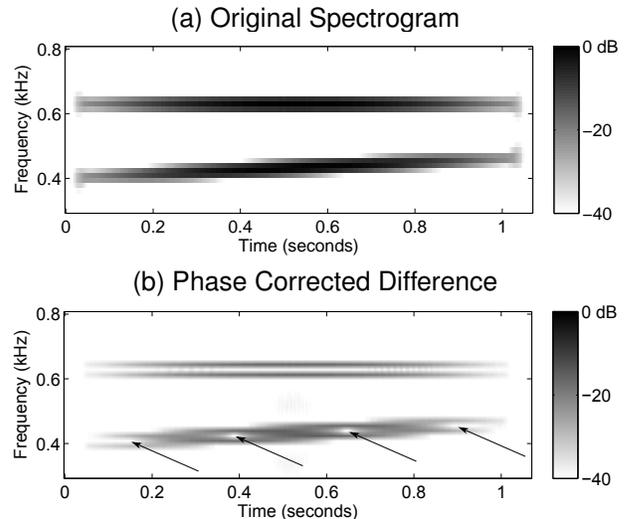
Fig. 2. (a) Spectrogram of a synthetic signal with two components. (b) The spectrogram after the application of the phase corrected difference operator. The difference operator creates a long hollow canal for the constant frequency component. For the chirp component, these canals, pointed to by the arrows, appear whenever the frequency of the chirp approaches the center frequency of a subband.

for a constant frequency component, if the frequency of the component does not match the center frequency of any subband, then the prior does not necessarily yield a small value for the signal. Nevertheless, in practice, we can try to avoid this problem by increasing the number of subbands so that there will be a subband that roughly matches the frequency of a given component. Below, we test whether this is indeed possible, using a real audio signal.

One last observation is regarding the robustness of the prior against changes in the amplitude of the components. Although the model in Fig. 1 assumes that the weight functions are piecewise constant, the weight functions for the components above are not constant but are slowly varying (actually Hamming functions). Since the variation is slow (note that the time-support of the Hamming windows are 1 sec) compared to the time-support of the atoms (128 msec), the prior is effective in this case.

2) Stringed Instrument: Our second experiment is performed on an excerpt played by tanbur (a stringed instrument). The signal is 3-sec long with a sampling frequency of 32 kHz (a total of 96000 samples). The (zoomed-in) spectrogram of the signal is shown in Fig. 3a. For the STFT, we used a window length of 4096 samples (128 msec) and a Hop-size of 256 samples (8 msec). The spectrogram, after the application of the phase corrected difference operator is shown in Fig. 3b. Applying the phase corrected difference operator reduces the energy of the spectrogram significantly.

But more importantly[5], we observe that the difference spectrogram now contains hollow canals, similar to the ones in the experiment above with the synthetic signal. Similarly, we think that such canals are produced whenever the frequency of a certain component matches the center frequency of a subband. In that case, the condition (14) is in effect and the phase corrected difference locally yields small outputs. For the subbands neighboring these hollow canals (leakage subbands), the condition (14) no longer holds and therefore the phase corrected difference fails to produce small outputs.

3) Vibrato: In a third experiment, we consider a violin vibrato. The parameters of interest are as follows. Sampling frequency is 44.1 kHz, the window length and Hop-size for the STFT are 4096 samples (93 msec) and 256 samples (6 msec) respectively. The original spectrogram and the phase corrected difference spectrogram are shown in Fig. 3c,d. For this signal, the frequency content varies more rapidly than the tanbur signal in the previous experiment. In turn, the phase corrected difference operator is less successful in suppressing the spectrogram. In particular, we do not observe long hollow canals as in the previous two experiments. However, we still observe that the difference operator has been able to cancel some portions of the spectrogram.

More experiments using the tanbur signal can be found in Section IV. We now consider an alternative description of the prior, that is useful for rearranging variational formulations as in (1).

III. A DUAL DESCRIPTION OF THE PRIOR

In this section, we present another description of $P(x)$ in the form,

$$P(x) = \max_{\bar{z} \in B} \langle K x, \bar{z} \rangle \tag{20}$$

where K is a linear map and B is a closed convex set. This description is useful for deriving saddle point or dual problems associated with variational formulations as in (1) that employ $P(x)$. This issue is further discussed in Section IV.

In the following, we assume that our input signal x is real-valued. Recall that the prior makes use of a collection of analytic time-frequency atoms $\left\{f_{s,k}\right\}_{(s,k) \in I}$, where I is some index set. Becasue of the analyticity property, the atoms are complex-valued. In order to work with real variables, we split the real and imaginary parts of the atoms as $f_{s,k}^r$ and $f_{s,k}^i$ (so that $f_{s,k} = f_{s,k}^r + j f_{s,k}^i$). Based on

[5]We have observed that such a reduction in energy occurs even if the underlying signal was uncorrelated noise. Indeed one could justify the occurence of such a reduction under some stationarity conditions.
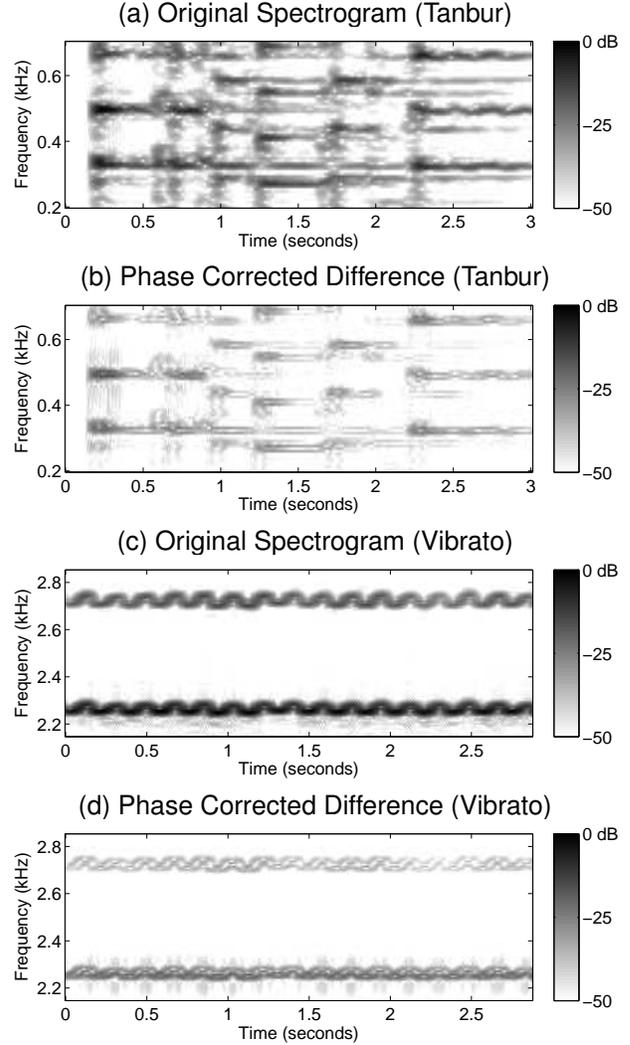
Fig. 3. (a) A zoom into the spectrogram magnitude of a piece by a stringed instrument. (b) Phase corrected difference operator applied to the STFT of the signal in (a). Observe that the phase corrected difference operator creates hollow canals similar to the ones in Fig. 2b. (c) A zoom into the spectrogram of a violin vibrato. (d) Phase corrected difference operator applied to the STFT of the signal in (c).

the time-frequency coefficients $c_{s,k}$ defined in (16), let us define a vector field $\bar{c}_{s,k}$, taking values in $\mathbb{R}^2$ as,

$$\bar{c}_{s,k} = \begin{bmatrix} \text{real}(c_{s,k}) \\ -\text{imag}(c_{s,k}) \end{bmatrix} = \begin{bmatrix} \langle x, f_{s,k}^r \rangle \\ \langle x, f_{s,k}^i \rangle \end{bmatrix}. \tag{21}$$

Notice that the mapping that takes x to $\bar{c}$ is linear. Let A denote this linear operator that maps x to $\bar{c}$. In this case, A^T is described as,

$$A^T \bar{c} = \sum_{s,k} \bar{c}_{s,k}^1 f_{s,k}^r + \bar{c}_{s,k}^2 f_{s,k}^i. \tag{22}$$

Remark 1. If the collection of atoms $\{f_{s,k}\}_{(s,k)\in I}$ forms a Parseval frame [6], then it can be shown that $A^T A = I$.

In this setting, $P(x)$ in (17) can be written as,

$$P(x) = \sum_{s,k} \|\bar{c}_{s,k} - R_s \bar{c}_{s,k+1}\|_2, \qquad (23)$$

where R_s is the rotation matrix associated with the sth subband, defined as

$$R_s = \begin{bmatrix} \cos(\omega_s N_s) & \sin(\omega_s N_s) \\ -\sin(\omega_s N_s) & \cos(\omega_s N_s) \end{bmatrix}. \qquad (24)$$

Now, let us denote $\bar{d}_{s,k} = \bar{c}_{s,k} - R_s \bar{c}_{s,k+1}$. Also, let Δ denote the linear mapping from $\bar{c}$ to $\bar{d}$, i.e. $\bar{d} = \Delta \bar{c}$. Note that if $\bar{e} = \Delta^T \bar{d}$, then

$$\bar{e}_{s,k} = \bar{d}_{s,k} - R_s^T \bar{d}_{s,k-1}. \qquad (25)$$

Finally, observe that if $v \in \mathbb{R}^2$, then $\|v\|_2 = \max_{u \in B_2} \langle v, u \rangle$, where B_2 is the unit ball of the ℓ_2 norm in $\mathbb{R}^2$. Therefore, if $\bar{z}$ is a vector field over the parameter space (s,k), then

$$\sum_{s,k} \|\bar{d}_{s,k}\|_2 = \sum_{s,k} \max_{\bar{z}_{s,k} \in B_2} \langle \bar{d}_{s,k}, \bar{z}_{s,k} \rangle = \max_{\bar{z}_{s,k} \in B_2} \langle \bar{d}, \bar{z} \rangle. \qquad (26)$$

Using our notation so far, we can now write,

$$P(x) = \max_{\bar{z}_{s,k} \in B_2} \langle \Delta A x, \bar{z} \rangle \qquad (27)$$

$$= \max_{\bar{z}_{s,k} \in B_2} \langle x, A^T \Delta^T \bar{z} \rangle. \qquad (28)$$

In the following, these expressions will be invoked to obtain saddle point problems.

IV. SUITABLE FORMULATIONS, EXPERIMENTS

A. Linear Inverse Problems

Consider a problem of the form,

$$\min_x \frac{1}{2} \|y - H x\|_2^2 + \lambda P(x), \qquad (29)$$

where H is a matrix. Using the alternative description in Section III, we can write this as a saddle point problem [35] as,

$$\min_x \max_{\bar{z}_{s,k} \in B_2} \frac{1}{2} \|y - H x\|_2^2 + \lambda \langle \Delta A x, \bar{z} \rangle. \qquad (30)$$

In general, the solution can be obtained by adapting primal-dual algorithms (see e.g. [23, 25, 27]). Specifically, if H is invertible, we can say more about the solution :

Proposition 1. Suppose that H is invertible. Let $C = (H^T H)^{-1}$. Also, let $\bar{z}^*$ be a vector field that satisfies

$$\bar{z}^* \in \operatorname*{argmin}_{\bar{z}_{s,k} \in B_2} \|y - \lambda H^{-T} A^T \Delta^T \bar{z}\|_2^2. \qquad (31)$$

Then, $x^* = H^{-1} y - \lambda C A^T \Delta^T \bar{z}^*$ minimizes (29).

Proof: See the appendix. ∎

Denoising: One special case of interest, in (29), is $H = I$. In this case, (29) reads as,

$$\min_x \frac{1}{2} \|y - x\|_2^2 + \lambda P(x), \qquad (32)$$

This could also be regarded as a denoising formulation. The denoising problem is of interest in more general settings as well since its solution is equivalent to realizing the 'proximal map' associated with the proposed norm, which appears in many algorithms in convex optimization [36, 37]. The following is a corollary of Prop. 1.

Corollary 1. Let $\bar{z}^*$ be a vector field that satisfies

$$\bar{z}^* \in \operatorname*{argmin}_{\bar{z}_{s,k} \in B_2} \|y - \lambda A^T \Delta^T \bar{z}\|_2^2. \qquad (33)$$

Then, $x^* = y - \lambda A^T \Delta^T \bar{z}^*$ is a solution of (32).

In this case, the following algorithm produces the denoised signal.

Algorithm 1 Denoising Algorithm

Initialize $\bar{z} \leftarrow 0$. Also, let ρ be a constant such that $I - \rho \lambda^2 \Delta A A^T \Delta^T$ is positive semi-definite.
repeat
  $\tilde{z} \leftarrow \bar{z} + \rho \lambda \Delta A (y - \lambda A^T \Delta^T \bar{z})$
  for all (s,k) **do**
    $\bar{z}_{s,k} \leftarrow \tilde{z}_{s,k} / \max(\|\tilde{z}_{s,k}\|_2, 1)$
  end for
until convergence
$x^* \leftarrow y - \lambda A^T \Delta^T \bar{z}$

Remark 2. For the constant ρ used in the algorithm, if the underlying frame is a Parseval frame [6], then $\rho < 1/(2\lambda^2)$ works.

The derivation of the algorithm is provided in the appendix.

Experiment 1. To test the effectiveness of the prior, we perform a denoising experiment[6]. The clean signal is a 3-second tune played by tanbur (a stringed instrument), sampled at 32 kHz (a total of 10^5 samples). The (zoomed

[6]Matlab codes for the experiments can be found at 'http://web.itu.edu.tr/ibayram/AudioPr/'

(a) Clean Signal   (b) Noisy Signal

(c) Proposed Prior   (d) Method of Yu et al.
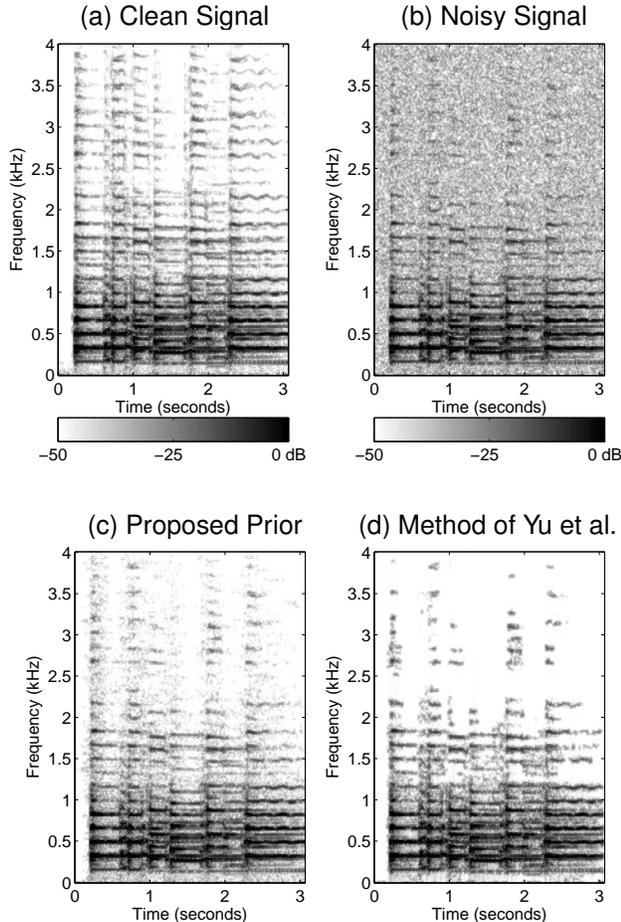
Fig. 4. Denoising experiment. (a) Original signal, (b) Noisy signal, SNR = 17.55 dB, (c) Denoised signal with the proposed prior, SNR = 24.91 dB, (c) Denoised with the method of Yu et al. [38], SNR = 23.95 dB.

in) spectrogram of the original signal is shown in Fig. 4a. We produce the 'observation signal' by adding iid Gaussian noise to the clean signal (SNR = 17.55 dB). The spectrogram of the noisy signal is shown in Fig. 4b.

To define the prior, we employ an STFT with a Hamming window[7] of length 4096 (corresponding to 128 ms) and a Hop-Size of 256 (corresponding to 8 ms) – therefore the redundancy is approximately 16. We input the noisy signal to Algorithm 1 with $\lambda = 0.008$. The resulting spectrogram is depicted in Fig. 4c. The output SNR is 24.91 dB.

In order to provide a comparison, we also applied the block-thresholding method of Yu et al. [38] [8]. The resulting

spectrogram[9] is depicted in Fig. 4d. In this case, the output SNR is 23.95 dB.

SNR values aside[10] (which is not a very reliable measure in terms of perceptual quality), there are differences between the spectrograms of the denoised signals. We think that the reason is the following. Observe that the clean signal is very rich in harmonics. However, high order harmonics are relatively faint and they seem to be lost in noise if we look at the noisy signal's spectrogram. Because of this, the block-thresholding method, which takes into account only the magnitudes of coefficients, eliminates these components completely (observe especially the high-frequency regions, wiped-out by the method). On the other hand, the proposed prior works with the phases of the coefficients as well. This could help the denoising algorithm to extract more (albeit slightly) of the harmonics. Nevertheless, if we increase λ further, we observed that the high-frequency components are supressed for our algorithm as well. □

Experiment 2. In this experiment, we apply the denoising algorithm to the clean 'Glockenspiel' signal shown in Fig. 5(a,c), in order to better understand the denoising behavior of the prior. Due to the nature of the instrument, the signal consists of attacks (at the onsets) followed by a tonal component. The attacks are clearly visible, both in the time domain signal and the spectrogram. As the time-frequency frame, we use an STFT with the same parameters as in Experiment 1. Also, λ is set to 0.05.

The spectrogram of the denoised signal is depicted in Fig. 5b. Because our prior penalizes deviations from a tonal behavior, in the denoised signal, the attacks are significantly suppressed. This is also evident from the time-domain residual signal where the energy is mostly concentrated at the onsets. In a sense, denoising lead to a decomposition of the signal into its 'transient' and 'tonal' components, although that was not intended. Such decompositions are of interest for coding as well as processing purposes – we refer the reader to [8, 10, 39, 40] for detailed discussions.

We note that the conclusion in the previous paragraph also points to a shortcoming of the proposed prior. The prior is not very suitable for representing 'transients' like attacks as in this example, or clicks, very rapid changes etc. Nevertheless, we note that the same objection could be made for the sinusoidal models as in [17, 18], which have proved useful.

□

---

[7]To be precise, we employ a 'normalized' version of a Hamming window. Normalization is performed to make the STFT a Parseval frame. In this setting, the normalized window is close to the original window.

[8]We used the code that was made available on the webpage of the authors.

[9]This is the spectrogram obtained by applying the STFT to the time-domain signal, i.e. the magnitudes of the 'analysis coefficients' of the STFT.

[10]We also note that especially for higher levels of noise, block-thresholding yields a higher SNR than our algorithm.

(a) Original Glockenspiel Signal

(b) 'Denoised' Glockenspiel Signal

(c) Original Glockenspiel Signal

(d) Residual After Denoising
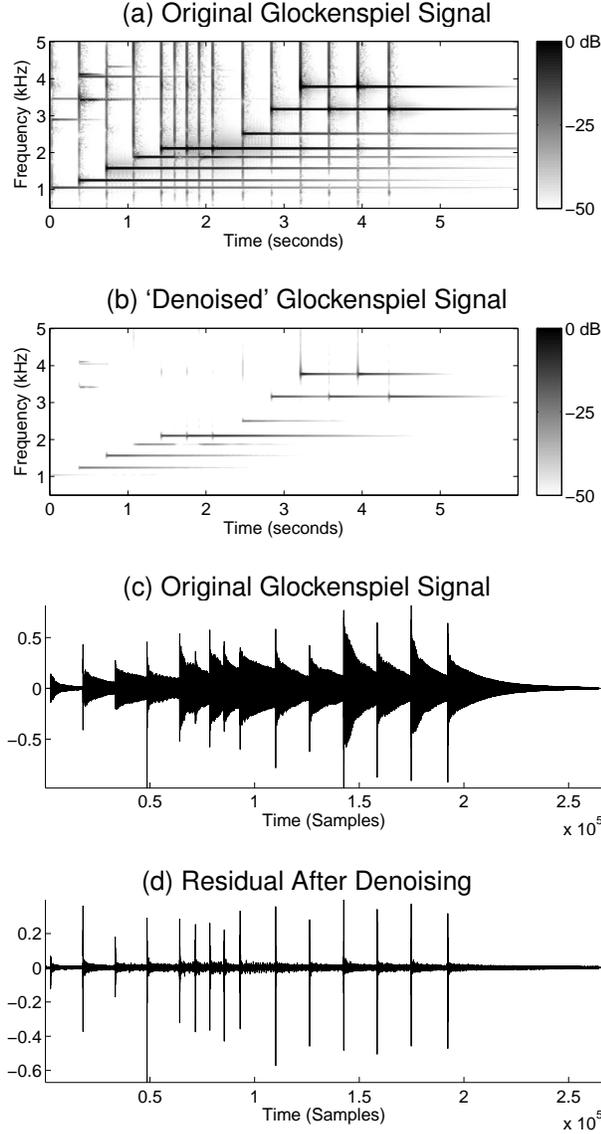
Fig. 5. Decomposition of Harmonic and Transient Components. (a) The original Glockenspiel signal. The onsets cause the thin vertical bars in the spectrogram. (b) In the denoised signals, the onsets have been significantly suppressed. (c) Time domain representation Glockenspiel signal. (d) The residual, i.e. the difference between the original and the denoised signals, consist mainly of the onsets.

B. Recovering Missing Segments

In a second application, we try to recover missing segments of an audio signal. The scenario is as follows. There is an underlying original signal $x(n)$, but somehow we only observe a subset of the samples. Our task is to construct a signal that agrees with the given data.

The problem discussed above appears in different settings. In particular, packet loss concealment schemes in streaming audio can be formulated as above [41, 42]. In this context, see [41] for an algorithm based sinusoidal models, [42] for a related coding scheme that makes use of the time persistence property of the packets. The problem has also been called an instance of an 'audio inpainting' problem in [43] and is handled by a matching pursuit type algorithm – also see [44] for a time-frequency regression formulation.

Here, we will make use of the proposed prior. For this, let N denote the indices of the observed samples. We formulate the problem as,

$$\min_u P(u) \text{ subject to } u(n) = x(n) \text{ for } n \in N. \quad (34)$$

Now let S denote the set of signals which agree with the data, i.e.,

$$S = \{u : u(n) = x(n) \text{ for } n \in N\}. \quad (35)$$

Notice that S is a convex set. Using S, we can rewrite our problem (34) as,

$$\min_{u \in S} \max_{\bar{z}_{s,k} \in B_2} \langle \bar{z}, \Delta A u \rangle. \quad (36)$$

This is a saddle point problem. A simple algorithm that converges in practice [23] is to take small steps in z and u successively (see Algorithm 2). Formally convergent schemes can be found in [24–29].

Algorithm 2 A Primal-Dual Algorithm

Initialize $\bar{z} \leftarrow 0$. Set γ to a small value.

$$\text{Initialize } u(n) \leftarrow \begin{cases} x(n), & \text{for } n \in N, \\ 0, & \text{for } n \notin N. \end{cases} \quad (37)$$

repeat

$$u \leftarrow \underset{v \in S}{\operatorname{argmin}} \frac{1}{2\gamma} \|u - v\|_2^2 + \langle A^T \Delta^T \bar{z}, v \rangle \quad (P)$$

$$\bar{z} \leftarrow \underset{\bar{t}_{s,k} \in B_2}{\operatorname{argmax}} -\frac{1}{2\gamma} \|\bar{t} - \bar{z}\|_2^2 + \langle \bar{t}, \Delta A u \rangle \quad (D)$$

until convergence

We now discuss how to realize the steps of Algorithm 2.

Realizing (P)*:* Let

$$v^* = \underset{v \in S}{\operatorname{argmin}} \frac{1}{2\gamma} \|u - v\|_2^2 + \langle A^T \Delta^T \bar{z}, v \rangle \quad (38)$$

This can be written as,

$$v^* = \underset{v \in S}{\operatorname{argmin}} \|(u - \gamma A^T \Delta^T \bar{z}) - v\|_2^2. \quad (39)$$

Now let $\tilde{u} = u - \gamma A^T \Delta^T \bar{z}$. Then we have,

$$v^*(n) = \begin{cases} \tilde{u}(n), & \text{for } n \notin N, \\ x(n), & \text{for } n \in N. \end{cases} \quad (40)$$

Realizing (56): Let

$$\bar{t}^* = \operatorname*{argmax}_{\bar{t}_{s,k} \in B_2} -\frac{1}{2\gamma} \|\bar{t} - \bar{z}\|_2^2 + \langle \bar{t}, \Delta A u \rangle. \tag{41}$$

Equivalently, we have,

$$\bar{t}^* = \operatorname*{argmax}_{\bar{t}_{s,k} \in B_2} \left\| (\bar{z} + \gamma \Delta A u) - \bar{t} \right\|_2^2 \tag{42}$$

Now let $\tilde{z} = \bar{z} + \gamma \Delta A u$. Then,

$$\bar{t}_{s,k}^* = \tilde{z}_{s,k} \frac{1}{\max(\|\tilde{z}_{s,k}\|_2, 1)} \tag{43}$$

Experiment 3. We apply the proposed formulation (34) on an audio signal. We use the signal in Experiment 1 (reproduced in Fig. 6a for ease of reference). To obtain the 'observation signal', we set to zero a segment of length 500 samples, out of every 3000 samples. The spectrogram of the observation signal is shown in Fig. 6b (SNR = 9.26 dB). The new vertical bars in the spectrogram are due to the missing segments.

To define $P(\cdot)$, we used an STFT frame as in Experiment 1 but with half the redundancy (i.e. the hop size is 512 samples). The spectrogram of the reconstructed signal is shown in Fig. 6c (SNR = 19.10 dB). Time-domain details are shown in Fig. 6d. Perceptually, the distortion due to the missing segments is significantly suppressed.

To compare the performance of the algorithm with the method in [43], we ran another experiment using the same algorithm. In this case, we set the first 10ms (320 samples) segment out of every 100ms (3200 samples) frame to zero. This results in an input SNR of 9.7 dB. After applying the algorithm, the SNR increases to 20.5 dB. Although the utilized signal is different from that in [43], we note that the SNR value is comparable (see the right-end values in Fig.2c of [43]). □

V. CONCLUSION

In this paper, we introduced a simple prior for audio signals, that is also easy to work with, in different settings. The prior could be interpreted as an adaptation of total variation to modulated signals. We have shown how to use the prior in two different applications and evaluated its performance using real audio signals. We hope to utilize the prior in other applications such as restoration, audio sensor fusion, etc. in the near future.

The proposed prior concentrates more on the 'tonal' component [39] of the audio. To that end, it could be complemented with an additional term for handling 'transient' components. However, this would also complicate the prior, which we tried to avoid in the first place. We think that a modification
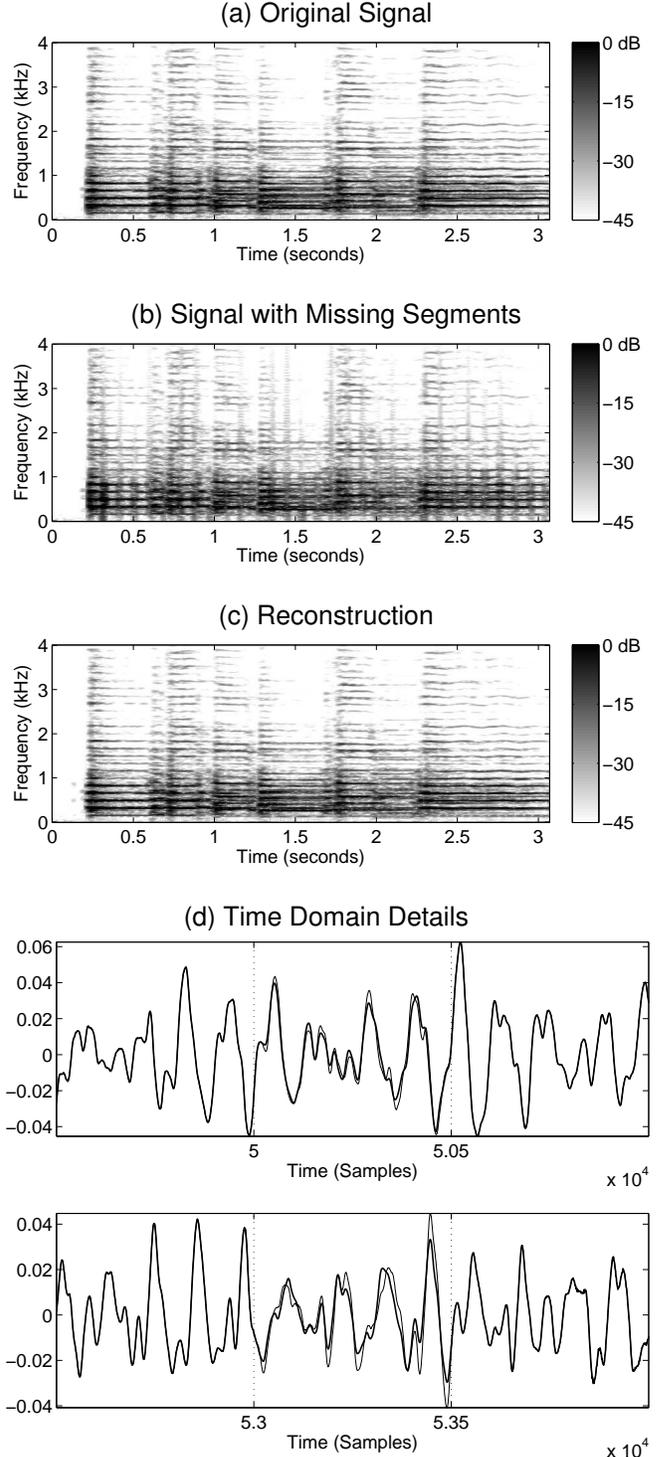
Fig. 6. Audio Inpainting Experiment. (a) Spectrogram of the original signal (b) Spectrogram of the observation signal with missing segments (c) Spectrogram of the reconstruction (d) Time domain details of the original signal (thin) and the reconstruction (thick). The missing segments lie between the dashed lines.

9

that is sensitive to the complexity of the prior would be an interesting topic to pursue.

Another possible extension might be to take into account the existence of harmonics in the stationary parts of audio signals. Note that the current definition of the prior, which adds the 'demodulated total variation' of each subband is based on the assumption that there are few subbands with activity at each point in time. However, for audio, the frequencies of these subbands are usually related. In order to take this into account, the simple sum (over the subbands) can be modified or replaced with a more suitable function.

ACKNOWLEDGEMENTS

APPENDIX
DERIVATION OF PROP. 1

The derivation below is essentially an adaptation of a related derivation regarding 2D total variation in [45] – also see [46] for an alternative.

Recalling (30), our goal is to find x^* and $\bar{z}^*$, with $\bar{z}^*_{s,k} \in B_2$, such that,

$$x^* = \operatorname*{argmin}_{x} \frac{1}{2}\|y - H\,x\|_2^2 + \lambda\,\langle \Delta\,A\,x, \bar{z}^*\rangle, \tag{44}$$

$$\bar{z}^* \in \operatorname*{argmax}_{\bar{z}_{s,k}\in B_2}\langle \Delta\,A\,x^*, \bar{z}\rangle. \tag{45}$$

By differentiating the rhs of (44) with respect to x, we see that x^* satisfies,

$$H^T\left(H\,x^* - y\right) + \lambda\,A^T\,\Delta^T\,\bar{z}^* = 0. \tag{46}$$

But if H has full column-rank, then we can solve for x^* as

$$x^* = (H^T\,H)^{-1}\big(H^T\,y - \lambda\,A^T\,\Delta^T\,\bar{z}^*\big). \tag{47}$$

Note that if H is square, then $(H^T\,H)^{-1} = H^{-1}\,H^{-T}$. In this case, if we plug (47) into (45), we obtain,

$$\bar{z}^* \in \operatorname*{argmax}_{\bar{z}_{s,k}\in B_2}\big\langle y - \lambda\,H^{-T}\,A^T\,\Delta^T\,\bar{z}^*, H^{-T}\,A^T\,\Delta^T\,\bar{z}\big\rangle. \tag{48}$$

Let us write $B = H^{-T}\,A^T\,\Delta^T$ for simplicity. The last equation implies that

$$\langle y - \lambda\,B\,\bar{z}^*, B\,(\bar{z}^* - \bar{z})\rangle \geq 0. \tag{49}$$

for all $\bar{z}$ such that $\bar{z}_{s,k} \in B_2$. But observe that,

$$\|y - \lambda\,B\,\bar{z}\|_2^2 = \|y - \lambda\,B\,\bar{z}^*\|_2^2 + \|\lambda\,B\,(\bar{z}^* - \bar{z})\|_2^2 \\ + 2\lambda\,\langle y - \lambda\,B\,\bar{z}^*, B\,(\bar{z}^* - \bar{z})\rangle. \tag{50}$$

In view of (49) and the nonnegativity of the second term, we deduce that $\|y - \lambda\,B\,\bar{z}\|_2^2$ assumes its minimum at $\bar{z}^*$. This is (31). The final statement of the proposition follows from (47).

APPENDIX
DERIVATION OF ALGORITHM 1

The algorithm obtains the minimizer by solving (33). We will present a derivation of the algorithm as an instance of a gradient projection algorithm [47, 48]. Alternatively, it can be viewed as an instance of a forward-backward algorithm [36, 45] or derived through a majorization-minimization scheme [49].

Let us recall the gradient projection algorithm. Consider a problem of the form

$$\min_{t\in D} f(t) \tag{51}$$

where D is a convex set in $\mathbb{R}^n$ and $f(t)$ is a differentiable function whose gradient is bounded as,

$$\|\nabla f(t) - \nabla f(u)\|_2 \leq \alpha\|t - u\|_2 \tag{52}$$

for some α. Also, let $P_D(\cdot)$ denote the projection operator onto D. In this setting, the gradient projection algorithm is as given below.

Algorithm 3 Gradient Projection

Set $\rho < 2/\alpha$.
repeat
$\quad \tilde{t} \leftarrow t - \rho\,\nabla f(t)$ (S1)
$\quad t \leftarrow P_D(\tilde{t})$ (S2)
until convergence

We remark that (33) is essentially a problem of the form (51) where $f(\bar{z}) = \frac{1}{2}\left\|y - \lambda\,A^T\,\Delta^T\,\bar{z}\right\|_2^2$ and D is the set of vector fields $\bar{z}$ such that $\bar{z}_{s,k} \in B_2$. Notice that D is convex in this case. Also notice that,

$$\nabla f(\bar{z}) = \lambda\,\Delta\,A\left(\lambda\,A^T\,\Delta^T\,\bar{z} - y\right). \tag{53}$$

In view of this, if we define $M = \lambda^2\,\Delta\,A\,A^T\,\Delta$, we can write,

$$\nabla f(\bar{z}) - \nabla f(\bar{u}) = M(\bar{z} - \bar{u}), \tag{54}$$

Let us now derive an upper bound on the spectral norm of M. If the underlying frame is Parseval, it can be shown that $A^T\,A = I$. Consequently, $A\,A^T$ is a projection operator, therefore,

$$\|A\,A^T\,v\|_2 \leq \|v\|_2. \tag{55}$$

10

Recalling the definition of Δ (see the text following (24)), we can write it as $\Delta = I - \mathbf{R}$. Here, $\mathbf{R}$ is the linear operator that delays $\bar{c}_{s,k}$ along k and then applies the unitary rotation matrix to each entry of the vector field. Therefore the spectral norm of $\mathbf{R}$ is unity. In turn, we find that,

$$\|\Delta u\|_2 \le 2\|u\|_2. \tag{56}$$

Combining (55) and (56), we find that

$$\|\nabla f(\bar{z}) - \nabla f(\bar{u})\|_2 \le 4\lambda^2 \|\bar{z} - \bar{u}\|_2. \tag{57}$$

In view of (53) and (57), for (33), the step labeled (S1) can now be written as,

$$\tilde{z} \leftarrow \bar{z} + \rho\,\lambda\,\Delta\,A(y - \lambda\,A^T\,\Delta^T\,\bar{z}) \tag{58}$$

where $\rho < 1/(2\lambda^2)$.

The step labelled (S2) should project $\tilde{z}_{s,k}$ onto B_2 for all (s, k) pairs. Notice that for a specific (s, k), the projection is given by,

$$\tilde{z}_{s,k}/\max(\|\tilde{z}_{s,k}\|_2, 1). \tag{59}$$

References

[1] O. Dikmen and A. T. Cemgil, "Gamma Markov random fields for audio source modelling," *IEEE Trans. Audio, Speech and Language Proc.*, vol. 10, no. 3, pp. 589 – 601, Mar. 2010.

[2] P. J. Wolfe, S. J. Godsill, and W.-J. Ng, "Bayesian variable selection and regularization for time-frequency surface estimation," *J. R. Statist. Soc. B*, vol. 66, no. 3, pp. 575–589, 2004.

[3] C. Févotte, B. Torrésani, L. Daudet, and S. J. Godsill, "Sparse linear regression with structured priors and application to denoising of musical audio," *IEEE Trans. Audio, Speech and Language Proc.*, vol. 16, no. 1, pp. 174–185, Jan. 2008.

[4] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Trans. Audio, Speech and Language Proc.*, vol. 13, no. 5, pp. 845–856, Sep. 2005.

[5] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," *EURASIP Journal on Applied Signal Processing*, vol. 7, pp. 1110–1126, 2005.

[6] O. Christensen, *An Introduction to Frames and Riesz Bases*. Birkhäuser, 2003.

[7] M. Kowalski and T. Rodet, "An unsupervised algorithm for hybrid/morphological signal decomposition," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, 2011.

[8] S. Molla and B. Torrésani, "A hybrid scheme for encoding audio signal using hidden markov models of

waveforms," *J. of Appl. and Comp. Harm. Analysis*, vol. 18, no. 2, pp. 137–166, Mar. 2005.

[9] M. Kowalski, "Sparse regression using mixed norms," *J. of Appl. and Comp. Harm. Analysis*, vol. 27, no. 3, pp. 303–324, Nov. 2009.

[10] M. Kowalski and B. Torrésani, "Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients," *Signal, Image and Video Processing*, vol. 3, no. 3, pp. 251–264, 2009.

[11] K. Siedenburg and M. Dörfler, "Structured sparsity for audio signals," in *Proc. Int. Conf. on Digital Audio Effects (DAFx)*, 2011.

[12] I. Bayram, "Mixed-norms with overlapping groups as signal priors," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, 2011.

[13] J. J. Fuchs, "Extension of the global matched filter to structured groups of atoms: Application to harmonic signals," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, 2011.

[14] G. Peyré and J. Fadili, "Group sparsity with overlapping partition functions," in *Proc. Eur. Sig. Proc. Conf (EUSIPCO)*, 2011.

[15] L. Jacob, G. Obozinsky, and J. P. Vert, "Group lasso with overlap and graph lasso," in *Proc. Int. Conf. Machine Learning (ICML)*, 2009.

[16] P.-Y. Chen and I. W. Selesnick, "Overlapping group shrinkage/thresholding and denoising," manuscript, 2012.

[17] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 34, no. 4, pp. 744–754, Aug. 1986.

[18] J. O. Smith and X. Serra, ""PARSHL: An analysis/synthesis program for nonharmonic sounds based on a sinusoidal representation," in *Proc. International Computer Music Conference*, 1987.

[19] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, pp. 227–238, 1996.

[20] F. I. Karahanoğlu, I. Bayram, and D. Van De Ville, "A signal processing approach to generalized 1-D total variation," *IEEE Trans. Signal Processing*, vol. 59, no. 11, pp. 5265–5274, Nov. 2011.

[21] G. A. Velasco, N. Holighaus, M. Dörfler, and T. Grill, "Constructing an invertible constant-q transform with non-stationary Gabor frames," in *Proc. Int. Conf. on Digital Audio Effects (DAFx)*, 2011.

[22] I. Bayram, "An analytic wavelet transform with a flexible time-frequency covering," manuscript, 2012, 'http://web.itu.edu.tr/ibayram/AnDwt.pdf'.

[23] M. Zhu and T. F. Chan, "An efficient primal-dual

hybrid gradient algorithm for total variation image restoration," May 2008, UCLA CAM Report [08-34], May, 2008.

[24] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, "An algorithm for minimizing the Mumford-Shah functional," in *Proc. IEEE Int. Conf. on Computer Vision*, 2009.

[25] E. Esser, X. Zhang, and T. F. Chan, "A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science," *SIAM J. Imaging Sciences*, vol. 3, no. 4, pp. 1015–1046, Nov. 2010.

[26] X. Zhang, M. Burger, and S. Osher, "A unified primal-dual algorithm framework based on bregman iteration," *J. Sci. Comput.*, vol. 46, no. 1, pp. 20–46, Jan. 2011.

[27] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, May 2011.

[28] L. D. Popov, "A modification of the Arrow-Hurwicz method for search of saddle points," *Mathematical Notes*, vol. 28, no. 5, pp. 845–848, 1980.

[29] G. Korpelevic, "An extragradient method for finding saddle points and for other problems," *Ekon. Mat. Metody*, vol. 12, no. 4, pp. 747–756, 1976.

[30] P. Prandoni, M. Goodwin, and M. Vetterli, "Optimal time segmentation for signal modeling and compression," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, 1997.

[31] M. G. Christensen, A. Jakobsson, and S. V. A. adn S. H. Jensen, "Amplitude modulated sinusoidal signal decomposition for audio coding," *IEEE Signal Processing Letters*, vol. 13, no. 7, pp. 389–392, Jul. 2006.

[32] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[33] L. Daudet, "Sparse and structured decompositions of signals with the molecular matching pursuit," *IEEE Trans. Audio, Speech and Language Proc.*, vol. 14, no. 5, pp. 1808–1816, Sep. 2006.

[34] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, Jun. 2007.

[35] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1996.

[36] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM J. Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, Nov. 2005.

[37] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. Springer, New York, 2011.

[38] G. Yu, S. Mallat, and E. Bacry, "Audio denoising by time-frequency block thresholding," *IEEE Trans. Signal Processing*, vol. 56, no. 5, pp. 1830–1839, May 2008.

[39] L. Daudet and B. Torrésani, "Hybrid representations for audiophonic signal encoding," *Signal Processing*, vol. 82, no. 11, pp. 1595–1617, Nov. 2002.

[40] F. Jaillet and B. Torrésani, "Time-frequency jigsaw puzzle: adaptive multiwindow and multilayered Gabor representations," *Int. J. for Wavelets and Multiresolution Information Processing*, vol. 5, no. 2, pp. 293–316, 2007.

[41] M. Lagrange and S. Marchand, "Long interpolation of audio signals using linear prediction in sinusoidal modelling," *J. Audio Eng. Soc.*, vol. 53, no. 10, pp. 891–905, Oct. 2005.

[42] D. Giacobello, M. N. Murthi, M. G. Christensen, S. H. Jensen, and M. Moonen, "Estimation of frame independent and enhancement components for speech communication over packet networks," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, 2010.

[43] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *IEEE Trans. Audio, Speech and Language Proc.*, vol. 20, no. 3, pp. 922–932, Mar. 2012.

[44] P. J. Wolfe and S. J. Godsill, "Interpolation of missing data values for audio signal restoration using a Gabor regression model," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, 2005.

[45] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Proc.*, vol. 18, no. 11, pp. 2419–2434, 2009.

[46] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1-2, pp. 89–97, January-March 2004.

[47] A. A. Goldstein, "Convex programming in Hilbert space," *Bull. Amer. Math. Soc.*, vol. 70, no. 5, pp. 709–710, 1964.

[48] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1995.

[49] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Proc.*, vol. 16, no. 12, pp. 2980–2991, Dec. 2007.