EMPLOYING PHASE INFORMATION FOR AUDIO DENOISING

İlker Bayram

Istanbul Technical University, Istanbul, Turkey

ABSTRACT

Spectral audio denoising methods usually make use of the magnitudes of a time-frequency representation of the signal. However, if the time-frequency frame consists of quadrature pairs of atoms (as in the short-time Fourier transform), then the phases of the coefficients also follow a predictable pattern, for which simple models are viable. In this paper, we propose a scheme that takes into account the phase information of the signals for the audio denoising problem. The scheme requires to minimize a cost function composed of a diagonally weighted quadrature data term and a fused-lasso type penalty. We formulate the problem as a saddle point search problem and propose an algorithm that numerically finds the solution. Based on the optimality conditions of the problem, we present a guideline on how to select the parameters of the problem. We discuss the performance and the influence of the parameters through experiments.

Index Terms— Audio denoising, non-negative garrote, total variation, fused lasso, audio phase.

1. INTRODUCTION

Spectral audio denoising methods usually modify the coefficient *magnitudes* of a time-frequency frame, even if the frame employs quadrature pairs of atoms, as in the short-time Fourier transform (STFT) – see e.g. [11]. This can be attributed to the fact that, although the magnitudes of successive time-frequency samples (for fixed frequency) exhibit high correlation, the correlation between the phases is low [11]. Nevertheless, the phases of the coefficients also follow a predictable pattern, for which simple models are viable. In this paper, we propose a scheme for audio denoising, that takes into account both the phase and the magnitude of the time-frequency coefficients. The proposed formulation requires the minimization of a convex cost function that contains a diagonally weighted quadratic data term and a fused-lasso type penalty [24], where the latter is composed of the sum of a total variation term and an ℓ_1 penalty.

Effective noise removal can be achieved by shrinking / thresholding the spectrogram magnitudes [11]. Following this modification on the magnitudes, one adds back the phase of the noisy signal to reconstruct the denoised signal. Using the noisy signal's phase is not without motivation – it is shown in [14] that the expected value of the noisy phase is equal to the phase of the clean signal (unlike the case for the magnitude). In methods that modify the spectrogram, if the time-frequency coefficients are treated independently (e.g. as in the spectral subtraction method [5]), the denoised signal suffers from a phenomenon called musical noise, due to the existence of isolated time-frequency atoms after denoising [8]. Appearance of isolated time-frequency components can be prevented or reduced by taking into account the behavior of the neighboring time-frequency atoms [14, 8] or working in time-frequency blocks of atoms, rather than individual atoms [26] – see also [18, 21, 3, 19, 22, 10] for recent work in this context. The reason for this is that the time-frequency distribution of audio follows predictable patterns with well-formed groups. This is especially easy to see if one observes just the magnitudes – at a particular band, whenever there is some activity, the magnitude varies smoothly over time and this behavior does not depend on the exact center frequency of the component under consideration. On the other hand, the phases of the coefficients may change rapidly – the coefficients gain a certain amount of phase, depending on the center frequency of the component, as one traverses the subband (see [4] for a discussion on how to construct a prior for audio signals, based on this observation). Although this gain in phase depends on the (unknown) center frequency of the component, it is roughly constant, which forms a key observation for the formulation in this paper.

We argue that the observation about the predictability of the phase can be used along with the magnitudes of the time-frequency coefficients to devise a denoising scheme that does not suffer from musical noise. Specifically, based on a simple model of audio, we propose a convex minimization problem, the solution of which determines the denoised estimate in Section 2. We also provide a guideline on the selection of the parameters of the formulation in this section. We present an algorithm that solves the formulation based on a saddle-point interpretation of the minimization problem, in Section 3. We demonstrate the utility of the method on real audio signals and discuss the how the parameters of the formulation affect the denoised estimate in Section 4.

2. PROBLEM FORMULATION

In this section, we describe and discuss the proposed formulation that uses the phase information of the STFT coefficients. We start with an observation on the phases, based on a simple model for audio signals. Following this, we present a formulation that requires the solution of a convex minization problem. The selection of the parameters used in the formulation and the relation of the formulation with existing work are also discussed briefly.

Notation for the STFT

We use the following definition for the STFT. Let $g(k)$ denote a smooth finite-support discrete-time window function. Let S denote the number of subbands used in the STFT and set $\theta_l = 2\pi l/S$ for l an integer in $[0, S-1]$. Using $g(k)$, we define the bandpass functions $g_l(k) = g(k)\, \exp(i\, \theta_l\, k)$. Finally, let T denote the hop-size for the STFT. Under these definitions, $X(l,j)$, the STFT of a discrete-time function $x(k)$, is defined as

$$X(l,j) = \langle x(\cdot), g_l(\cdot - j\,T)\rangle \tag{1a}$$

$$= \sum_k x(k)\, g_l^*(k - j\,T). \tag{1b}$$

E-mail : ibayram@itu.edu.tr

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

2.1. A Simple Model for Audio Signals

We model the audio signal as a linear combination of sinusoids, each of which is weighted by a piecewise-constant weight function. That is, we assume that the signal of interest is of the form

$$x(k) = \sum_m r_m(k)\, \exp(i\, \omega_m k). \tag{2}$$

Here, $r_m(k)$ denotes the piecewise-constant weight function for the m^{th} component. We refer to [20, 23] for more detailed discussions of sinusoidal models.

Now, let X denote the STFT of $x(k)$ and $\omega_m \approx \theta_l$ for some (m, l) pair. Moreover, suppose that $r_m(k)$ is non-zero and constant in the support of $g_l\big(k - (j+t)\,T\big)$ for $t = 0, 1, 2, \ldots K$. In that case, we will have

$$X(l, j) \approx \exp\big(-i\, \omega_m\, T\, t\big)\, X(l, j + t) \tag{3}$$

for $t = 0, 1, 2, \ldots K$. Observe that if we *knew* the value of ω_m, we could fairly predict $X(j + t, l)$ from $X(j, l)$, for $t = 0, 1, 2, \ldots K$. Unfortunately, in general, we do not expect to have knowledge of ω_m's. Nevertheless, in [4], we argued that if N is sufficiently large, then $|\omega_m - \theta_l|$ can be made arbitrarily small for a certain l. In this case, we can predict the value of $X(j + t, l)$ using $X(j, l)$ and a phase update term that depends only on θ_l. This, in turn, leads to a simple prior that can be used as a regularizer in audio restoration formulations. However, we also demonstrated in [4] that even if $|\omega_m - \theta_l|$ is small, the same sinusoidal component leaks in neighboring channels (e.g. the channels indexed by '$l \pm 1$') and one needs the value of ω_m in order to predict the value of X in those leakage channels (see the discussion in Section 2 of [4]). Therefore, in practice, we either need to estimate ω_m's or we need to find a prediction scheme that does not rely on the knowledge of ω_m's. We investigate the feasibility of the latter choice in this paper.

Notice now that, although the ratio

$$\frac{X(l, j)}{X(l, j + 1)} \approx \exp(-i\, \omega_m\, T) \tag{4}$$

depends on the unknown value ω_m, it remains *constant* if we replace 'j' with '$j + t$' on the left hand side, for $t = 1, 2, \ldots, K$. Observe also that this ratio has unit magnitude. It follows that if we define

$$\hat{X}(l, j) = |X(l, j)|\, \frac{X(l, j + 1)}{|X(l, j + 1)|}, \tag{5}$$

then

$$X(l, j) \approx \hat{X}(l, j)\, \alpha_l(j) \tag{6}$$

for some piecewise constant function $\alpha_l(j)$. Note also that further errors in this approximate equality may arise because the constant pieces of $r_m(k)$ may have partial overlaps with the supports of the window function $g(n - jT)$. Notice also that $\alpha_l(j)$'s can be taken as zero for a significant portion of the spectrogram since the original spectrogram is expected to be sparse. In summary, $\alpha_l(j)$ is expected to be

(i) piecewise constant with magnitude less than unity,

(ii) sparse.

2.2. Problem Formulation

Suppose now that we observe the noisy signal $y = x + n$. Also, let Y denote the STFT coefficients. Similarly as above, we define

$$\hat{Y}(l, j) = |Y(l, j)|\, \frac{Y(l, j + 1)}{|Y(l, j + 1)|}. \tag{7}$$

Under this definition, we propose to estimate X from Y as,

$$\hat{X} = \hat{Y}\, \hat{\alpha}, \tag{8}$$

where

$$\hat{\alpha} = \underset{|\alpha(l,j)| \leq 1}{\operatorname{argmin}}\ \frac{1}{2} \|Y - \hat{Y}\, \alpha\|_2^2 + \lambda_1\ \mathrm{TV}(\alpha) + \lambda_2\, \|\alpha\|_1. \tag{9}$$

Here, we define the total variation functional as $\mathrm{TV}(\alpha) = \|D\, \alpha\|_1$, where '$D$' is a difference operator defined so that for $\beta = D\, \alpha$, we have

$$\beta(l, j) = \alpha(l, j) - \alpha(l, j + 1). \tag{10}$$

Referring to the two points brought forward at the end of Section 2.1,

(i) the TV term ensures that the phase of the reconstruction at a particular frame are predictable from neighboring frames,

(ii) the ℓ_1 term ensures that the overall spectrogram of the reconstruction is sparse.

2.3. Selection of the Parameters

A vector $\hat{\alpha}$ solves (9) only if [17],

$$\hat{Y}^* \left(Y - \hat{Y}\, \hat{\alpha}\right) = \lambda_1\, D^T\, u + \lambda_2\, v, \tag{11}$$

for some u, v with $|u(l, j)| \leq 1$, $|v(l, j)| \leq 1$. Suppose now that the reconstruction is fairly successful so that $\hat{Y}\, \hat{\alpha} \approx X$. In that case, we have $Y - \hat{Y}\, \hat{\alpha} \approx N$, where N is the STFT of the noise term. Noting also $|\hat{Y}^*| = |Y| = |X + N|$, from (11) we obtain,

$$|X\, N + N^2| \approx |\lambda_1\, D^T\, u + \lambda_2\, v|. \tag{12}$$

If we further consider x as a white Gaussian signal, independent of n and assume $|XN + N^2| \approx |XN| + |N|^2$, then we obtain, after some rearranging,

$$\mathbb{E}(|X\, N + N^2|) \approx \|g\|_2^2\, \sigma^2\, \left(\frac{2}{\pi}\, \mathrm{SNR} + 1\right), \tag{13}$$

where $\mathrm{SNR} = \|x\|_2 / \|n\|_2$, σ^2 is the noise variance and $\|g\|_2^2$ is the energy of the time-frequency atom used to obtain the STFT coefficients (see (1)).

Noting now that both $|u|$ and $|v|$ are bounded by unity (and $\sigma(D) = 2$), we can argue that $(2\lambda_1 + \lambda_2)$ should be on the order of

$$\|g\|_2^2\, \sigma^2\, \left(\frac{2}{\pi}\, \mathrm{SNR} + 1\right). \tag{14}$$

However, the relative weight of λ_1 and λ_2 eventually depends on the amount of 'tonal' vs. 'transient' components in x (see e.g. [25, 13] for discussions). We refer to Experiment 2 for a brief discussion of the effects of changing the relative weights of the parameters.

2.4. Relation with the Non-Negative Garrote

The formulation outlined above may be compared with the 'non-negative garrote' formulation by Breiman [7]. In [7], (for the 'orthonormal design' case) the idea is to estimate a sparse vector x from noisy observations y as, $\hat{x} = y\,\hat{\alpha}$, where $\hat{\alpha}$ is obtained by solving the minimization problem

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \frac{1}{2}\|y - y\,\alpha\|_2^2 + \lambda\,\|\alpha\|_1. \tag{15}$$

Notice that $\hat{\alpha}$ does not directly give the estimate but is used to 'shape' the noisy observations. The ℓ_1 penalty term leads to a sparse $\hat{\alpha}$, which in turn leads to a sparse estimate.

In contrast to the non-negative garrote, in order to make use of the phases, we obtain the denoised signal by shaping $\hat{Y}$, defined in (7), instead of the observed signal's STFT, namely Y. Another difference is the inclusion of a TV term so as to ensure that the amount of phase gained in consecutive time-frequency samples is roughly the same. If we discard the TV term and use Y instead of $\hat{Y}$ in (9), as a direct application of [7] would suggest, we obtain a separable problem where the STFT coefficients are treated independently of their time-frequency neighbors. This in turn leads to reconstructions with 'islands' in the time-frequency representation – especially for low SNR reconstructions, isolated time-frequency atoms start to appear in the reconstructions. These are typically perceived as musical noise [26]. The inclusion of the TV term in (9) allows us to make use of the observation in Section 2.1 about the phases of the neighboring time-frequency coefficients. This helps reduce musical noise, because isolated time-frequency atoms are further penalized. This effect is also demonstrated in Experiment 2 in Section 4.

3. A MINIMIZATION ALGORITHM

3.1. A Saddle Point Problem

The ℓ_1 norm of a complex vector $x = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}$ can be written as,

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| = \sup_{z \in B_\infty} \Re\big\{\langle x, z\rangle\big\}, \tag{16}$$

where B_∞ is the unit ball of the ℓ_∞ norm in $\mathbb{C}^n$ (this constrains $|z_i| \le 1$) and $\Re$ denotes the real part. We can thus express the minimization problem in (9) as

$$\min_{|\alpha(l,j)|\le 1} \max_{\substack{|u(l,j)|\le 1,\\ |v(l,j)|\le 1}} \frac{1}{2}\,\|Y - \hat{Y}\,\alpha\|_2^2 + \lambda_1\,\Re\big\{\langle D\,\alpha, u\rangle\big\} + \lambda_2\,\Re\big\{\langle \alpha, v\rangle\big\}. \tag{17}$$

Note here that the spectral norm of D is 2.

Remark 1. In order to avoid working with complex numbers, we can also regard $\alpha(l,j)$ as a two-component vector. According to that view, the time-frequency map may be regarded as a vector field where each time-frequency coefficient actually represents a two-dimensional vector. This interpretation allows us to drop 'real parts' in the minimization expressions – see [4] for the details of such a development.

3.2. The Algorithm

In order to numerically solve the saddle point problem in (17), we employ the algorithm discussed in [9, 15]. The algorithm requires computing certain projection operators $P_{c\,B}$. Here, $c\,B$ denotes the unit ball of the ℓ_∞ norm in $\mathbb{C}^n$, weighted by the constant c. For $q = P_{c\,B}(s)$, this projection can easily be realized by setting,

$$q(l,j) = s(l,j)\,\left[\max\left(\frac{|s(l,j)|}{c}, 1\right)\right]^{-1}. \tag{18}$$

Input SNR (dB)	5	10	15	20
Yu et al.[26]	16.79	20.13	22.71	24.82
Proposed	15.54	17.92	19.82	21.76
Prop. + Wiener	17.30	20.42	23.03	25.55

Table 1: Denoising Results. 'Proposed' refers to the reconstruction in (8). 'Prop. + Wiener' refers to the proposed method followed by empirical Wiener filtering [16, 26].

Algorithm 1 A Saddle Point Algorithm

1: Initialize α, p, $\bar{\alpha}$, u, v to zero.
2: **repeat**
3: $\quad u \leftarrow P_{\lambda_1\,B}\big(u + D^T\,\bar{\alpha}/2\big)$
4: $\quad v \leftarrow P_{\lambda_2\,B}\big(v + D^T\,\bar{\alpha}/2\big)$
5: $\quad \alpha \leftarrow P_B\big(\alpha - (D^T\,u + v)/2\big)$
6: $\quad \bar{\alpha} \leftarrow 2\alpha - p$
7: $\quad p \leftarrow \alpha$
8: **until** Some convergence criterion is met

This algorithm is fairly easy to implement but takes some time to convergence (1-2 minutes on a regular PC, for the experiments in the following section). One of course may utilize alternative algorithms (such as ADMM – see e.g. [6]) for this problem. It would also be of interest to adapt the finite-converging algorithms proposed for 1-dimensional TV denoising/fused lasso problem for real valued data (see e.g. [12, 2] and the references therein). Here, our main goal is to evaluate the formulation and demonstrate the utility of including the phase information. We hope to investigate faster algorithms that (approximately) solve the problem in the near future.

4. EXPERIMENTS

Experiment 1. Our first experiment demonstrates the denoising performance of the proposed formulation. We use a tune played by a stringed instrument as the clean signal. We add Gaussian noise to this signal and obtain various noisy observations with different SNRs. For each observation, we set $(\lambda_1, \lambda_2) = (5\,\beta, 0.75\,\beta)$, where $\beta = \|g\|_2^2\,\sigma^2\,\big(2\,\pi^{-1}\,\mathrm{SNR} + 1\big)$ (see (14)). Following [26], we also apply an empirical Wiener filter [16] to further enhance the SNR. The resulting SNR values are given in Table 1. In order to give an idea about the performance with respect to the state of the art, we also include in Table 1, the SNR values achieved by the block-thresholding method of Yu et al. [26]. Also, Fig. 1 shows the spectrograms of the signals from the experiment with input SNR = 5 dB. Observe that the proposed method preserves the harmonics that are also preserved by the block thresholding method. Further, some of the harmonics, parts of which are cleared by block thresholding, are retained by the proposed method, thanks to the TV term. Note also that empirical filtering boosts the SNR significantly, by removing some of the bias due to (over) suppression introduced by the ℓ_1 term (see also [10] for a discussion of this point).

Experiment 2. In this experiment, we demonstrate how the TV term allows to control the amount of musical noise. Consider the noisy signal spectrogram shown in Fig. 2a. The observation is a

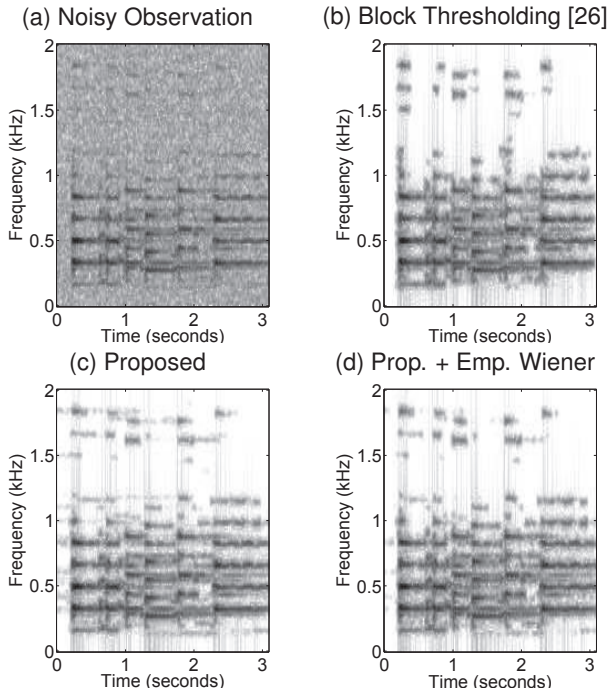

Fig. 1: Spectrograms of the signals in Experiment 1. (a) The noisy observation signal, SNR = 5 dB. (b) The denoised signal obtained by the block thresholding method of [26], SNR = 16.79 dB. (c) The denoised estimate obtained by the proposed formulation, SNR = 15.54 dB. (d) Result of applying an empirical Wiener filter [16] to the signal in (c), SNR = 17.30 dB.

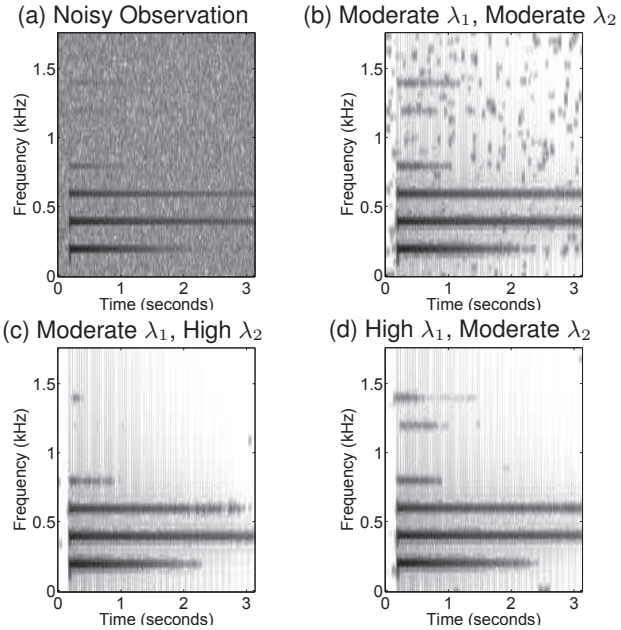

Fig. 2: Spectrograms of the signals in Experiment 2. (a) Spectrogram of a single note played by guitar contaminated with additive white noise, SNR = 5 dB. (b,c,d) Reconstruction obtained by the proposed scheme with different regularization parameters. For a fixed value of β (see the text), the weight parameters are (b) $(\lambda_1, \lambda_2) = (\beta, \beta)$, (c) $(\lambda_1, \lambda_2) = (\beta, 4\beta)$, (d) $(\lambda_1, \lambda_2) = (4\beta, \beta)$.

single note played by guitar, sampled at 11 KHz contaminated with white Gaussian noise (SNR= 5 dB). For this experiment, let $\beta = \|g\|_2^2 \, \sigma^2 \left(2\,\pi^{-1}\,\mathrm{SNR} + 1\right)$ as in Experiment 1.

If we choose the regularization parameters as $(\lambda_1, \lambda_2) = (\beta, \beta)$, we obtain a reconstruction as shown in Fig. 2b (SNR = 18.45 dB). We can interpret this case as an example for moderate TV regularization and moderate ℓ_1 regularization. We observe that such a choice leads to a reconstruction with a fair preservation of the harmonics but musical noise is visible in the spectrogram.

If we increase the weight of the ℓ_1 term while keeping the TV weight constant and set $(\lambda_1, \lambda_2) = (\beta, 4\,\beta)$, we obtain a reconstruction as shown in Fig. 2c (SNR = 16.44 dB). Musical noise is indeed suppressed but at the expense of suppressing the higher order harmonics. This is also evident from the relatively low SNR of the reconstruction.

If we increase the weight of the TV term and set $(\lambda_1, \lambda_2) = (4\beta, \beta)$, the reconstruction is as shown in Fig. 2d (SNR = 19.01 dB). Observe that increasing the weight of the TV term, instead of the weight of the ℓ_1 term helps reduce musical noise with a modest suppression of the higher order harmonics.

The results above indicate that, while the ℓ_1 term helps suppress the noise in general, the TV term is instrumental in the elimination of isolated time-frequency components.

5. CONCLUSION AND OUTLOOK

We introduced a denoising scheme that makes use of the phases and magnitudes of the time-frequency coefficients of the observed signal. The proposed formulation is based on the predictable behavior of the coefficient phases along a subband. Although the experiments in this paper are performed with Gaussian noise, the formulation is applicable in more general settings as long as the phase of the noise STFT coefficients do not fit the pattern discussed in Section 2. We also note that the approach may be modified for use in more general problems, such as audio clipping, inpainting or restoration applications (see [1] and the references therein).

Based on optimality conditions, we noted that a linear combination of the parameters λ_1, λ_2 should be on the order of a constant that depends on the SNR. However, the two parameters affect the reconstructed signal in different ways, as discussed in Experiment 2. The ratio, or the relative weight of the parameters therefore is an important choice. We think that this ratio should be chosen based on a function that takes as input the relative energies of the tonal and the transient components of the signal. Providing such a function would be of interest since as it would help select the parameters automatically.

Another direction might be to make use of the phase information in adaptive block-based estimation schemes as in [26]. In that case, estimation of α can be performed based on certain (adaptive) groups. We hope to pursue this idea in future work.

6. REFERENCES

[1] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley. Audio inpainting. *IEEE Trans. Audio, Speech and Language Proc.*, 20(3):922–932, March 2012.

[2] İ. Bayram. A divide-and-conquer algorithm for 1D total variation denoising. `http://web.itu.edu.tr/ibayram/TVDenoise/TVDenoise.pdf`. Manuscript, 2013.

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

[4] İ. Bayram and M. E. Kamasak. A simple prior for audio signals. *IEEE Trans. Audio, Speech and Language Proc.*, 21(6):1190–1200, June 2013.

[5] S. F. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 27(2):113–120, April 1979.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[7] L. Breiman. Better subset regression using the nonnegative garrote. *Technometrics*, 37(4):373–384, November 1995.

[8] O. Cappé. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor. *IEEE Trans. Speech and Audio Proc.*, 2(2):345–349, April 1994.

[9] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, May 2011.

[10] P. Y. Chen and I. W. Selesnick. Translation-invariant shrinkage/thresholding of group sparse signals. *Signal Processing*, 94:476–489, January 2014.

[11] I. Cohen and S. Gannot. Spectral enhancement methods. In J. Benesty, M. M. Sondhi, and Y. Huang, editors, *Springer Handbook of Speech Processing*. Springer, 2008.

[12] L. Condat. A direct algorithm for 1D total variation denoising. *IEEE Signal Processing Letters*, 20(11):1054–1057, November 2013.

[13] L. Daudet and B. Torrésani. Hybrid representations for audiophonic signal encoding. *Signal Processing*, 82(11):1595–1617, November 2002.

[14] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 32(6):1109–1121, December 1984.

[15] 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*, 3(4):1015–1046, November 2010.

[16] S. P. Ghael, A. M. Sayeed, and R. G. Baraniuk. Improved wavelet denoising via empirical Wiener filtering. In *Proc. SPIE Wavelet Applications in Signal and Image Proc.*, 1997.

[17] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis*. Springer, 2004.

[18] M. Kowalski. Sparse regression using mixed norms. *J. of Appl. and Comp. Harm. Analysis*, 27(3):303–324, November 2009.

[19] M. Kowalski, K. Siedenburg, and M. Dörfler. Social sparsity! neighborhood systems enrich structured shrinkage operators. *IEEE Trans. Signal Processing*, 61(10):2498–2511, May 2013.

[20] R. J. McAulay and T. F. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 34(4):744–754, August 1986.

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

[22] K. Siedenburg and M. Dörfler. Persistent time-frequency shrinkage for audio denoising. *Journal of the Audio Eng. Soc.*, 61(1/2):29–38, January 2014.

[23] 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.

[24] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B*, 67(1):91–108, 2005.

[25] T. S. Verma, S. Levine, and T. H. Y. Meng. Transient modeling synthesis : A flexible transient analysis/synthesis tool for transient signals. In *Proc. ICMC*, 1997.

[26] G. Yu, S. Mallat, and E. Bacry. Audio denoising by time-frequency block thresholding. *IEEE Trans. Signal Processing*, 56(5):1830–1839, May 2008.