A FAST GRIFFIN-LIM ALGORITHM

Nathanaël Perraudin[1], *Peter Balazs*[2], *Peter L. Søndergaard*[2]

[1] EPFL, Switzerland, nathanael.perraudin@epfl.ch
[2] Acoustics Research Institute, Vienna, Austria
peter.soendergaard@oeaw.ac.at, peter.balazs@oeaw.ac.at

ABSTRACT

In this paper, we present a new algorithm to estimate a signal from its short-time Fourier transform modulus (STFTM). This algorithm is computationally simple and is obtained by an acceleration of the well-known Griffin-Lim algorithm (GLA). Before deriving the algorithm, we will give a new interpretation of the GLA and formulate the phase recovery problem in an optimization form. We then present some experimental results where the new algorithm is tested on various signals. It shows not only significant improvement in speed of convergence but it does as well recover the signals with a smaller error than the traditional GLA.

Index Terms— Magnitude-only reconstruction, Short-time Fourier transform, Phase reconstruction, time-scale modification (TSM), signal estimation, spectrogram inversion

I. INTRODUCTION

Time-frequency representations, in particular *Gabor transforms* [1], i.e. the sampled Short-Time Fourier transforms (STFT), are ubiquitous in signal processing. Gabor transforms describe a signal in time and frequency simultaneously. This transformation is fast (thanks to the Fast Fourier transform (FFT)) and provides a good tool for signal modification. If the magnitude squared of the STFT is understood to be the "localized time-frequency power spectrum", the phase remains a complicated object which is difficult to modify appropriately. As a consequence, most of the transformations on the STFT work with the magnitude or the magnitude squared (the spectrogram), leaving the phase unchanged or sometime completely dropped. Since the STFT is a redundant construction, the obtained coefficients usually does not form a valid spectrogram (ie: there exists no signal having exactly this spectrogram.)

For instance, in the case of adaptive filtering like denoising, the magnitude of the STFT is often modified without any modification of the phase [2].

Furthermore, accurate reconstruction of a signal from its spectrogram is also important. This is known as the *phase recovery* problem: recovering a signal from the amplitude of some measurements, only. In the influential paper [3], it was proven that for frames with sufficiently high redundancy, a signal can be reconstructed from the magnitude of its frame coefficients only (up to a global phase factor). Recent results [4] put the necessary redundancy at $\frac{4L-4}{4}$ for a frame for $\mathbb{C}^L$.

The notion of valid spectrogram plays a very important role in the problem: the STFT has to verify a so-called "consistency criterion" [5], [6]. In fact, the set of complex STFT coefficients is a proper subset of the coefficient space i.e. taking an array of complex coefficients usually does not correspond to the STFT of a signal. As a result, modifying the magnitude of the STFT does not in general lead to a valid spectrogram.

From this difference, we consider two different problems respectively called:

- Phase recovery: constructing a signal from a valid spectrogram and no phase information.
- STFT magnitude approximation: constructing a signal from a non-valid STFT magnitude and eventually some starting phase.

Both of those problems can be solved using our algorithm which is highly inspired by the Griffin-Lim algorithm (GLA) [7]. This last method (see section IV) performs iteratively two projections. We propose to consider the difference between two iterations. Doing so, we lose all theoretical guarantees of convergence. However, this new structure allows a more accurate and faster convergence. We also expect it to be compatible with GLA modification presented in [5], [8], [9].

After presenting briefly the Gabor transform, we will give a new interpretation of the problem and the GLA. This will lead to a new proposition of algorithm which we call the fast Griffin-Lim algorithm (FGLA). We will then present simulation results.

It should be noted that both the GLA and the algorithm presented in this paper can be applied to *any* frame, and not just to Gabor frames. However, for clarity and simplicity, we shall consider only the Gabor case in this paper.

II. GABOR THEORY

In this contribution, we consider Gabor systems $\mathcal{G}(g, a, M)$ in $\mathbb{C}^L$. All signals and windows on $\mathbb{C}^L$ are considered to have periodic boundary conditions. For $g \in \mathbb{C}^L$, and integer $a, M > 0$, we define the Gabor system

$$\mathcal{G}(g, a, M) := \left(g_{m,n} = g[\cdot - na]e^{2\pi i m \cdot /M}\right)_{n,\ m}, \qquad (1)$$

where $m = 0, \ldots, M-1$ is the index of the frequency-channel and $n = 0, \ldots, N-1$ is the index of the time-position. If $\mathcal{G}$ is also a *frame* [10], we refer to the system as a *Gabor frame*. For $x \in \mathbb{C}^L$, the corresponding Gabor transform is given by

$$(\mathbf{G}x)[m + nM] = \langle x, g_{m,n}\rangle = \sum_{l=0}^{L-1} x[l]\overline{g_{m,n}[l]}, \qquad (2)$$

with the analysis operator $\mathbf{G}$ that is given by the matrix

$$\mathbf{G}[m + nM, l] := \mathbf{G}_{g,a,M}[m + nM, l] := \overline{g_{m,n}[l]}.$$

Gabor synthesis is performed by applying the conjugate transpose of $\mathbf{G}$ to a coefficient sequence $c \in \mathbb{C}^{MN}$. The action of the synthesis operator can be equivalently described as

$$x_{syn}[l] = (\mathbf{G}^* c)[l] = \sum_{m,n} c[m + nM]g[l - na]e^{2\pi i m l/M}. \qquad (3)$$

The concatenation $\mathbf{S} = \mathbf{G}^*\mathbf{G}$ of analysis and synthesis operators is called the *frame operator*.

Reconstruction can be realized using the so-called *canonical dual* system, obtained by inverting $\mathbf{S}$ and defined as

$$\gamma_{m,n} = \mathbf{S}^{-1} g_{m,n}. \qquad (4)$$

In the particular case of Gabor frames, the canonical dual system is again a Gabor frame, i.e. it equals $\mathcal{G}(\gamma_{0,0}, a, M)$. Thus we refer to $\gamma_{0,0} = \mathbf{S}^{-1} g$ as the canonical dual window.

In this case the synthesis operator of $\gamma_{0,0}$ coincides with the pseudo-inverse of the original analysis operator, i.e. $\mathbf{G}^*_{\gamma,a,M} = \mathbf{G}^\dagger$. So the inversion formula reads

$$x[l] = \sum_{m,n} \langle x, g_{m,n} \rangle \gamma_{m,n}[l] = \mathbf{G}^\dagger \mathbf{G} x[l]. \tag{5}$$

A particular way to modify the coefficients is by multiplication by a fixed *symbol* s.

$$Mf = \sum_{m,n} s_{m,n} \langle x, g_{m,n} \rangle \gamma_{m,n}.$$

Such operators, called *multipliers* can be defined for all kind of frames [11], and find applications in acoustics, see e.g. [12].

III. THE PROBLEM

The problem can be expressed as finding a signal $x^* \in \mathbb{R}^L$ (or more generally in $\mathbb{C}^L$) from given a given set of non-negative coefficients s , such that the magnitude of the STFT of x^*: $|\mathbf{G}x|$, is as close as possible to s. The ℓ_2-norm will be used as measure of closeness. Mathematically, we formulate the problem in the following form:

Given a frame $\mathbf{G}$ and real positive coefficients $s = |s|$, x^* is the solution of

$$\text{minimize}_{x \in \mathbb{R}^L} \| |\mathbf{G}x| - s \|_2 \tag{6}$$

We will call s a valid STFT magnitude if there exists an x such that $|\mathbf{G}x| = s$.

For convenience, we define an equivalent problem with the optimization variable on the coefficient side.

$$\text{minimize}_{c \in \mathbb{C}^{MN}} \| |c| - s \|_2 \text{ s. t. } \exists x \in \mathbb{R}^L \mid c = \mathbf{G}x \tag{7}$$

Those definitions lead naturally to a measure of error:

$$E(x) = \frac{\| |\mathbf{G}x| - s \|_2}{\|s\|_2} \tag{8}$$

For convenience, instead of (8), we use the signal to noise ratio of the STFT magnitude. This can be expressed

$$SSNR(x) = -10 \log_{10}(E(x)) \tag{9}$$

IV. THE GRIFFIN-LIM ALGORITHM (GLA)

The GLA (named after their authors) was presented in 1984 in [7]. It aims at estimating a signal from its modified short time Fourier transform. The GLA is a version of the double-projection algorithm originally suggested by Gerchberg and Saxton [13] for solving the phase recovery problem in terms of the Fourier transform. The Gerchberg-Saxton works for a non-redundant system (the Fourier transform) by considering additional side-constraints to make the solution unique. The GLA algorithm on the other hand works for redundant systems without any side constraints, where the uniqueness of the solution comes via the redundancy.

The GLA proceeds by projecting a signal iteratively onto two different sets in $\mathbb{C}^{\frac{L}{a} \times M}$ denoted by $\mathcal{C}_1$ and $\mathcal{C}_2$.

$\mathcal{C}_1$ is the set of admissible points for problem (7). It is also the set of coefficients c that can be reached from $x \in \mathbb{R}^L$ through the frame $\mathbf{G}$, i.e. the range of G:

$$\mathcal{C}_1 = \{c \mid \exists x \in \mathbb{R}^L \text{ s. t. } c = \mathbf{G}x\} \tag{10}$$

This meets the hard constraint of problem (7). Note that $\mathcal{C}_1$ is the set that satisfies the consistency criterion [14]. By [10] we can express the projection in the following way:

$$P_{\mathcal{C}_1}(c) = \mathbf{G}\mathbf{G}^\dagger c \tag{11}$$

Let $\mathcal{C}_2$ to be the set of coefficients minimizing (7) without necessary satisfying the hard constraint. It is simply given by

$$\mathcal{C}_2 = \left\{ c \in \mathbb{C}^{MN} \middle| |c| = s \right\}.$$

The projection onto $\mathcal{C}_2$ is simply equivalent to forcing the magnitude of s to be c elementwise:

$$P_{\mathcal{C}_2}(c) = s \cdot e^{\cdot i \angle c}. \tag{12}$$

The GLA can now be formulated (cf algo 1).

Algorithm 1 Griffin-Lim algorithm (GLA)

Fix the initial phase $\angle c_0$
Initialize $c_0 = s \cdot e^{\cdot i \angle c_0}$
Iterate for $n = 1, 2, ...$
$c_n = P_{\mathcal{C}_1}\left(P_{\mathcal{C}_2}(c_{n-1})\right)$
Until convergence
$x^* = \mathbf{G}^\dagger c_n$

Improvements of the GLA can be found in the literature. In [5], an approximate way to perform the projection $P_{\mathcal{C}_1}$ is proposed. As the projection operator is highly structured, it is normally applied using a fast algorithm, and this structure cannot be exploited in the approximation. We have therefore chosen not to use this approximation in this paper.

In [15], [8] the *Real-Time Spectrogram Inversion* RTISI algorithm, which is an extension of the GLA was proposed. Reconstruction is performed piece by piece by using again GLA and a clever starting point. A further improvement is presented in [9]. In the next section, we propose a different modification for the GLA. It should be possible to combine both modifications into one algorithm, however the detailed analysis of this is beyond the scope of this contribution.

V. THE FAST GRIFFIN LIM ALGORITHM (FGLA)

Equations (6) and (7) define the problem in an optimization form. However, classic optimization algorithms cannot easily reach a solution since both (7) and (6) are not convex. Phase recovery was recently expressed as a convex optimization problem in [16], [17]. However, nowadays, the heavy computation cost of the method makes it unsuitable for long signal (i.e. $L > 128$). In this contribution, we rather propose to search for the solution of the non convex problem (7). In fact, we need to find the intersection of the two sets $\mathcal{C}_1$ and $\mathcal{C}_2$. Iterative projections would converge to an optimal solution if both sets would be convex. Our idea is to make larger steps at each iteration. To do so, we will use the information available in the previous iterations.

More precisely, we will replace the update rule of the Griffin-Lim

$$c_n = P_{\mathcal{C}_1}\left(P_{\mathcal{C}_2}(c_{n-1})\right) \tag{13}$$

by

$$c_n = P_{\mathcal{C}_1}\left(P_{\mathcal{C}_2}(c_{n-1} + \alpha_n \Delta c_{n-1})\right) \tag{14}$$

where $\Delta c_n = c_n - c_{n-1}$. At convergence, (14) and (13) are equivalent. However, (14) is a faster way to converge to the solution. Indeed the parameter $\alpha_n \Delta c_{n-1}$ increases the steps depending on the current iterations values.

The similar trick is used in the algorithm called "FISTA" (fast iterative shrinkage thresholding algorithm) [18] that speeds up the algorithm "ISTA". In this method, they provide the optimal sequence of α_n that optimizes the convergence. In our case, the computation of such sequence remains still an open question, due to the non convexity of our problem. Thus, in the following, we have considered the simple case: α being a constant.

Using (14), we define the algorithm 2 called the Fast Griffin-Lim algorithm (FGLA). We observe that the heavy part of the computation takes place into the projection $P_{\mathcal{C}_1}$ which happens only once per iteration in both algorithms. Hence we assume the computation cost per iteration to be equivalent in both algorithms. Since the projection only involves pure Gabor analysis and synthesis, efficient algorithms [19] for these operators can be used.

Note that changing the update rule suppresses all theoretical guarantees of convergence. This is another open issue of this contribution.

Algorithm 2 Fast Griffin-Lim algorithm (FGLA)

Fix the initial phase $\angle c_0$
Initialize $c_0 = s \cdot e^{i \angle c_0}$, $t_0 = P_{\mathcal{C}_2}\left(P_{\mathcal{C}_1}(c_0)\right)$
Iterate for $n = 1, 2, \ldots$
  $t_n = P_{\mathcal{C}_1}\left(P_{\mathcal{C}_2}(c_{n-1})\right)$
  $c_n = t_n + \alpha_n(t_n - t_{n-1})$
  Update α_n
Until convergence
$x^* = \mathbf{G}^\dagger c_n$

VI. NUMERICAL RESULTS

In this section we present three different experiments of phase reconstruction. We remind the reader that as the problem is not convex, the algorithms will converge most likely to a local minimum depending on the starting point.

Experiments were done using classical windows. Those presented in this paper use a Nuttall (figure 1), a Gaussian (figures 2) and a Hann window (figure 3). We choose as frame parameters $a = 32$, $M = 256$. This makes a redundancy of 8 that assures all the information to lie in the spectrogram [3].

Using different parameters or windows lead to similar results. A reproducible research addendum can be downloaded at http://unlocbox.sourceforge.net/rr/fgla/. From this archive, parameters can be easily changed and other configuration tested.

In the first example, we aim at finding a signal from its spectrogram (phase reconstruction). In this specific case we do know that such a signal exists. The initial phase is simply set to zero and the number of iterations for both algorithm is fixed to $10'000$. In figure 1, we observe that the FGLA does not only converge faster (better average slopes), but also to points with smaller error. Note that for the signal 'bat', the new algorithm was able to perform perfect reconstruction. This signal is very short, only 400 samples. We also observe that, using the FGLA, the $SSNR$ is not strictly increasing from one iteration to another. However in average, the $SSNR$ is increasing.

In the second example, we start from a signal, compute the Gabor coefficients, apply a spectrogram multiplier and reconstruct a new signal as good as we can. In that case, signals fitting exactly the modified STFTM usually do not exist. As a consequence, we are looking for the signal with the best spectrogram approximation. The applied multiplier is random. This multiplier is chosen because it modifies the spectrogram in a significant way and, in that case, algorithm usually need more iterations to converge. Other multipliers gives similar results. The initial phase, this time, is not set to zero like in the previous experiment, but we keep the original phase of the STFT. We fixed the maximum number of iterations to $1'000$ as well. Generally, the new algorithm converges faster. The $SSNR$ is sometime improved, but not in a very significant manner.

In the third and last experiment, we analyze the effect of α onto the FGLA. Figure 3 displays tests for various constants α. $\alpha = 1$ seems to be the limit of stability of the algorithm. $\alpha = 0$ correspond to the Griffin-Lim algorithm. Increasing α leads to better results with some optimal value near 1 but not bigger. As a consequence, 0.99 has been chosen for the other experiments.

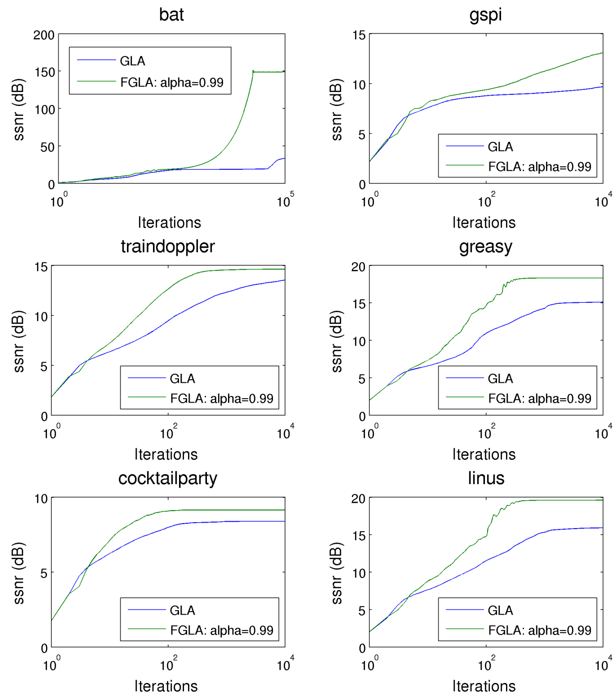

Figure 1. Phase recovery problem: SSNR through iterations for the GLA and the FGLA.

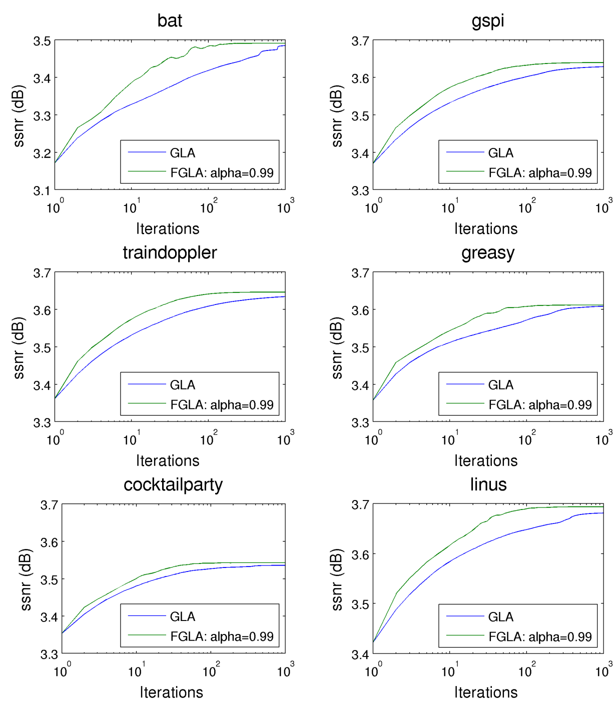

Figure 2. STFT magnitude optimization problem: SSNR through iterations for the GLA and the FGLA.

The algorithm presented in this paper has been incorporated as an option for the `frsynabs` function in the the LTFAT toolbox, [20].

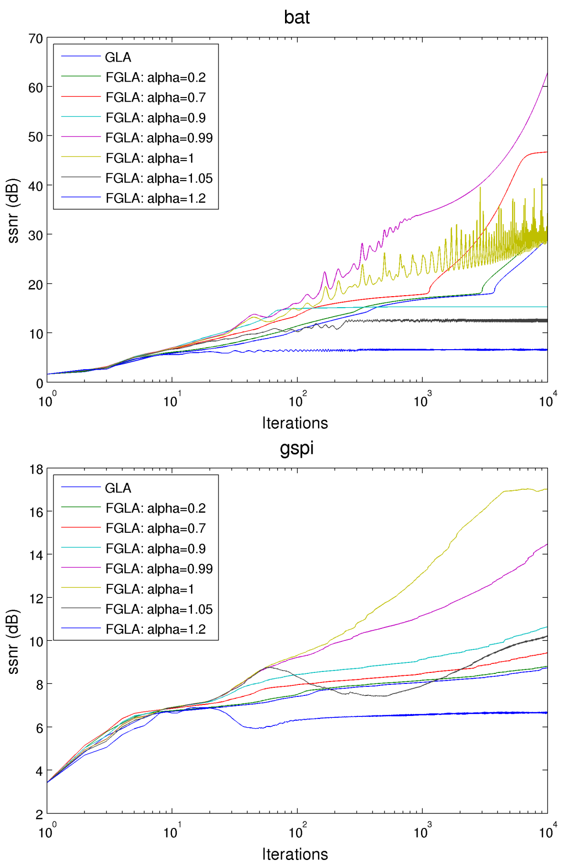

Figure 3. Influence of the parameter alpha onto the FGLA.

VII. CONCLUSION

In this paper, we have presented the phase recovery problem in the form of an optimization problem. This approach allows us to give a new interpretation of the GLA in order to be able to speed it up. We proposed an algorithm (FGLA) that was indeed faster but seems also to converge to better points. However, any theoretical guarantee of convergence has been lost in the process. Practically, our algorithm can replace the GLA at a very low cost of implementation and computation. In our further research ,we will look for a convergence proof and an optimal sequence of α_n and possible merge our algorithm with the RTISI real-time GLA algorithm.

Acknowledgment

This work was supported by the Austrian Science Fund (FWF) START-project FLAME ("Frames and Linear Operators for Acoustical Modeling and Parameter Estimation"; Y 551-N13).

VIII. REFERENCES

[1] H. G. Feichtinger and T. Strohmer, Eds., *Gabor Analysis and Algorithms*, Boston, 1998.

[2] P. Majdak, P. Balazs, W. Kreuzer, and M. Dörfler, "A time-frequency method for increasing the signal-to-noise ratio in system identification with exponential sweeps," in *Proceedings of the 36th International Conference on Acoustics, Speech and Signal Processing, ICASSP 2011, Prag*, 2011.

[3] R. Balan, P. Casazza, and D. Edidin, "On signal reconstruction without phase," *Applied and Computational Harmonic Analysis*, vol. 20, no. 3, pp. 345–356, 2006.

[4] B. G. Bodmann and N. Hammen, "Stable phase retrieval with low-redundancy frames," *arXiv preprint arXiv:1302.5487*, 2013.

[5] J. Le Roux, H. Kameoka, N. Ono, and S. Sagayama, "Fast signal reconstruction from magnitude stft spectrogram based on spectrogram consistency," in *Proc. 13th International Conference on Digital Audio Effects (DAFx-10)*, 2010, pp. 397–403.

[6] J. Le Roux and E. Vincent, "Consistent Wiener filtering for audio source separation," *Signal Processing Letters, IEEE*, vol. 20, no. 3, pp. 217–220, 2013.

[7] D. Griffin and J. Lim, "Signal estimation from modified short-time fourier transform," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 32, no. 2, pp. 236–243, 1984.

[8] X. Zhu, G. Beauregard, and L. Wyse, "Real-time signal estimation from modified short-time fourier transform magnitude spectra," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 5, pp. 1645–1653, 2007.

[9] V. Gnann and M. Spiertz, "Improving RTISI Phase Estimation With Energy Order and Phase Unwrapping," in *Proc. of International Conference on Digital Audio Effects DAFx*, vol. 10, 2010.

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

[11] P. Balazs, "Basic definition and properties of Bessel multipliers," *Journal of Mathematical Analysis and Applications*, vol. 325, no. 1, pp. 571–585, January 2007. [Online]. Available: http://dx.doi.org/10.1016/j.jmaa.2006.02.012

[12] P. Balazs, B. Laback, G. Eckel, and W. A. Deutsch, "Time-frequency sparsity by removing perceptually irrelevant components using a simple model of simultaneous masking," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 1, pp. 34–49, 2010. [Online]. Available: http://www.kfs.oeaw.ac.at/xxl/mask/mask.pdf

[13] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optik*, vol. 35, no. 2, pp. 237–250, 1972.

[14] J. Le Roux, N. Ono, and S. Sagayama, "Explicit consistency constraints for stft spectrograms and their application to phase reconstruction," *Proc. SAPA*, pp. 23–28, 2008.

[15] G. T. Beauregard, X. Zhu, and L. Wyse, "An efficient algorithm for real-time spectrogram inversion," in *Proceedings of the 8th International Conference on Digital Audio Effects*, 2005, pp. 116–118.

[16] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, 2012.

[17] D. L. Sun and J. O. Smith III, "Estimating a signal from a magnitude spectrogram via convex optimization," *arXiv preprint arXiv:1209.2076*, 2012.

[18] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[19] P. L. Søndergaard, "Efficient Algorithms for the Discrete Gabor Transform with a long FIR window," *J. Fourier Anal. Appl.*, vol. 18, no. 3, pp. 456–470, 2012.

[20] P. L. Søndergaard, B. Torrésani, and P. Balazs, "The Linear Time Frequency Analysis Toolbox," *International Journal of Wavelets, Multiresolution Analysis and Information Processing*, vol. 10, no. 4, 2012.