TIME-FREQUENCY KERNEL DESIGN FOR SPARSE JOINT-VARIABLE SIGNAL REPRESENTATIONS

Branka Jokanović, Moeness G. Amin, Yimin D. Zhang, and Fauzia Ahmad

Center for Advanced Communications, Villanova University, Villanova, PA 19085, USA

ABSTRACT

Highly localized quadratic time-frequency distributions cast nonstationary signals as sparse in the joint-variable representations. The linear model relating the ambiguity domain and time-frequency domain permits the application of sparse signal reconstruction techniques to yield high-resolution time-frequency representations. In this paper, we design signal-dependent kernels that enable the resulting time-frequency distribution to meet the two objectives of reduced cross-term interference and increased sparsity. It is shown that, for random undersampling schemes, the new adaptive kernel is superior to traditional reduced interference distribution kernels.

Index Terms— Kernel design, reduced interference distribution, sparse representation, time-frequency analysis

1. INTRODUCTION

The primary goal of reduced interference quadratic time-frequency distributions (QTFDs), applied to nonstationary signals, is to remove, or at least significantly attenuate, the cross-terms [1, 2]. These terms, which result from the bilinear data products underlying the quadratic distributions, act as interference and clutter the time-frequency signal representations. This could, in turn, lead to misinterpretations of the signal local power concentrations and misreading of the signal time-frequency signature, including the instantaneous frequencies. To solve this problem, many signal-independent and signal-dependent reduced interference distributions (RIDs) and their fast implementations have been devised [3-8]. The former involve applying a fixed two-dimensional (2-D) low-pass kernel to the ambiguity function, which amounts to smoothing the Wigner-Ville distribution. The employed kernel attempts to capture the auto-terms that pass through and cluster around the origin in the ambiguity domain, while giving low responses to cross-terms that are distant from the time-lag and Doppler frequency axes. In the signal-dependent RID approach, the kernel shape changes according to the signal component structures and can be irregular.

QTFDs and most existing methods for time-frequency analysis are defined for uniformly sampled data. Uniform under-sampling below the Nyquist rate causes aliasing and must, therefore, be avoided when involving Fourier transform or Fourier basis. One way to overcome aliasing is by using random sampling schemes [9-14]. In the presence of randomly undersampled data, the RIDs are faced with the combined task of reducing cross-terms and acting as an interpolator or estimator of the missing samples. The data interpolation and signal reconstruction capabilities of either the signal-dependent or signal-independent approaches have never been investigated or become part of the overall kernel design.

In this paper, we consider the problem of auto-term preservations and cross-term suppressions under the auspices of sparse signal representation. In particular, we design a new signal-dependent time-frequency kernel that meets the two objectives of reduced cross-term interference and increased sparsity under a large number of missing data samples. Nonstationary signals, completely or partially characterized by their instantaneous frequencies, are sparse in the time-frequency domain [9-13]. In this respect, they can be locally reconstructed from few random observations. To achieve the above two objectives, the low-pass filtering and the sparse data properties should both play a role and become an integral part of the overall design paradigm. It is shown that the added sparsity criterion of the resulting time-frequency representation renders the kernel robust to missing data. Using an optimization procedure, we obtain the kernel which satisfies the required constraints. Once the kernel is designed, the corresponding time-frequency signal representation can be obtained through the use of Cohen's class, i.e., using 2-D Fourier transform of the modified ambiguity function. In this case, the new distribution becomes a member of QTFDs.

The remainder of the paper is organized as follows. Section 2 provides a brief review of the RIDs with signal-independent and signal-dependent kernels. In Section 3, we introduce the new kernel design approach, followed by the mathematical formulation and discrete-time implementations. Section 4 contains examples, which compare the proposed approach with commonly used kernels when processing signals with missing samples and

noise. Conclusion is provided in Section 5.

2. REDUCED INTERFERENCE DISTRIBUTIONS

The reduced interference time-frequency distribution in polar coordinates $RID(\rho,\varphi)$ is formulated as the 2-D Fourier transform of the product of the ambiguity function $A(r,\psi)$ and the kernel function $\Phi(r,\psi)$, expressed as [15]:

$$RID(\rho,\varphi) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(r,\psi)\Phi(r,\psi)re^{j2\pi r\rho\sin(\psi-\varphi)}drd\psi , \quad (1)$$

where the ambiguity function and the kernel function are defined in terms of the radius r and the aspect angle ψ. The ambiguity function in the polar coordinates can be obtained through direct calculation of polar samples or by interpolation of the rectangular form $A(\theta,\tau)$, which for signal $x(t)$ is defined as:

$$A(\theta,\tau) = \int_{-\infty}^{\infty} x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2})e^{-j\theta t}dt , \quad (2)$$

where $(\cdot)^*$ denotes complex conjugation. The kernel function acts as a low-pass filter in the ambiguity domain and places different weights on the ambiguity function samples. Majority of signals have auto-terms located near the origin and around the two axes in the ambiguity domain, whereas the cross-terms appear distant from the origin. This property motivated the introduction of various kernels with low-pass filtering characteristics in order to suppress cross-terms and preserve the auto-term shape. It is noted that the nonstationary signals considered are deterministic and not nonstationary random processes [16].

Existing time-frequency kernels can be divided into two forms: signal-independent and signal-dependent. The former includes Choi-Williams kernel [3], Margenau-Hill [1], and Born-Jordan kernel [1]. Table 1 shows some most commonly used signal-independent kernels. One of the drawbacks of these kernels is that they have fixed shapes in the ambiguity domain and, therefore, are inflexible in accommodating a large class of signals to achieve optimum tradeoff between auto-term preservation and cross-term suppression, even under the inclusion of the adjustable parameter σ. Additionally, and more importantly, these kernels are not designed for data with missing samples. The primary goal of these kernels has been invariably the suppression of the cross-terms. On the other hand, missing samples in the time domain may cause significant artifacts in the ambiguity function [12]. These artifacts resemble noise in the sense that they spread over the entire ambiguity domain, but they exhibit strong presence along the $\tau = 0$ axis because of the modeled impulsive nature of the missing data samples. Therefore, while signal-independent kernels, which act as low pass filters in the ambiguity domain, reduce noise and artifacts to some extent, the contribution of the artifacts remains strong at low frequencies and along the $\tau = 0$ axis. Because most signal-independent kernels, including those listed in Table 1, are designed to satisfy the marginal properties, i.e.,

$$\Phi(0,\tau) = 1,$$
$$\Phi(\theta,0) = 1, \quad (3)$$

the resulting time-frequency distribution generally exhibits vertical strips (lines) in the presence of missing samples (e.g., Fig. 4(b)).

On the other hand, signal-dependent kernel design is formulated as an optimization problem [4] under two separate constraints. The first constraint forces the kernel to have low-pass filtering characteristics, which lends itself to cross-term suppression. With the second constraint, we can specify the volume under the optimal kernel. Such optimization can be performed either globally or locally. The adaptive optimal kernel [5] provides better representation of the local signal characteristics than its non-adaptive counterparts. With no missing samples, the adaptive kernel generally outperforms the signal-independent kernels. However, under random data observations, existing signal-dependent kernels may be misguided when finding the optimal solution. The presence of the artifacts in the ambiguity domain, particularly along the entire $\tau = 0$ axis, can be wrongly interpreted as signal regions of interest, which undesirably causes the kernel shape to favor the artifacts, resulting in highly cluttered time-frequency distributions. This shortcoming must be overcome and a new kernel design approach is required to provide robustness to missing samples.

Table 1. Examples of Doppler-lag domain kernels

Distribution	Kernel $\Phi(\theta,\tau)$
Choi-Williams	$e^{-\frac{\theta^2\tau^2}{2\sigma^2}}$
Margenau-Hill	$\cos(\theta\tau/2)$
Born-Jordan	$\mathrm{sinc}(2\sigma\theta\tau)$

3. PROPOSED ADAPTIVE KERNEL DESIGN

3.1 Kernel Function Requirement

In order to properly devise RIDs under missing data samples, we note that the artifacts due to missing data assume random non-clustered spreading in the time-frequency domain. This is in contrast with the localization property of the desired signal components. Since localization and sparsity are indicative to the same signal behavior, significant improvement in artifact reduction, while preserving the signal time-frequency signature, can be achieved by enforcing the sparsity of the time-frequency distribution.

Toward this end, the proposed time-frequency kernel is multitasking and achieves the following objectives:

1) Reduce or significantly attenuate the cross-terms;
2) Preserve the signal auto-terms;
3) Robust to missing time-domain samples.

The first two requirements are well known in quadratic time-frequency signal representations and, for several decades, various approaches have been proposed aiming at cross-term suppression while preserving the signal auto-terms [1-8]. To the best of our knowledge, the third task, which is implemented to enforce the sparsity of the resulting time-frequency representation, has not been explored in time-frequency kernel design since it was generally assumed that full data volume is available with no missing observations.

The recent development of compressed sensing and sparse signal reconstructions motivates sparsity-based time-frequency kernel design. Since a large class of nonstationary signals has sparse time-frequency representations, various procedures have recently been proposed for the analysis of incomplete data in the time-frequency domain [9-13]. However, these methods would observe cross-terms removal, auto-terms preservation, and sparsity property in time-frequency domain separately. The block diagram of Fig. 1 presents the existing approach for calculating the RID in the presence of missing time samples. It is clear that both cross-term removal and sparsity constraint are included in the final result; however, their applications are decoupled, necessitating a sequential cascading process. This observation has motivated a search for a new approach that would combine the aforementioned three tasks and perform them simultaneously.

3.2. Mathematical Formulation

We follow an approach that searches for the optimum radial attenuation $\sigma(\psi)$ as a function of the aspect angle, which would fulfill the three desired tasks. The optimization problem is formulated as follows:

$$\underset{\sigma}{\text{minimize}} \quad L_2(\sigma) + \lambda L_1(\sigma)$$

$$\text{subject to} \quad \Phi(r,\psi) = e^{-\frac{r^2}{2\sigma^2(\psi)}}, \quad (4)$$

$$\frac{1}{2\pi}\int \sigma^2(\psi)d\psi \leq \alpha,$$

where

$$L_2(\sigma) = \iint_{r\ \psi} |A(r,\psi) - A(r,\psi)\Phi(r,\psi)|^2\ rdrd\psi \quad (5)$$

and

$$L_1(\sigma) = \iint_{\rho\ \varphi} \left| \iint_{r\ \psi} A(r,\psi)\Phi(r,\psi)re^{j2\pi r\rho\sin(\psi-\varphi)}drd\psi \right| d\rho d\varphi \quad (6)$$

are, respectively, the l_2 and l_1-norm terms, and λ represents the regularization parameter.

In the proposed objective function, two terms are minimized. The first term is an averaged l_2-norm error between the original and the kernelled ambiguity functions.
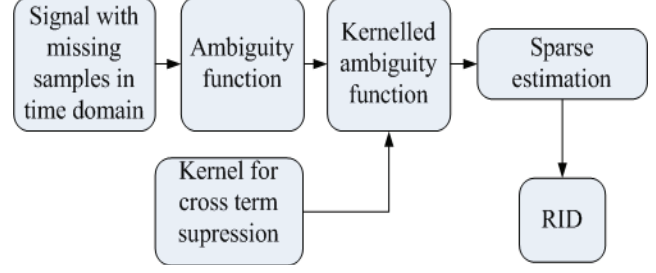

Fig. 1. Existing approach of computing RID in the case of incomplete data.

Therefore, minimization of this term amounts to energy preservation in the ambiguity domain. The minimization of the second term imposes sparsity in the time-frequency domain and its role is to provide robustness against missing data samples. It is known that, in general, the solution of an objective function which contains an l_1-norm regularized term provides sparsity of that term. In our case, the cross-terms adversely affect time-frequency sparsity. Therefore, the designed kernel would avoid the inclusion of the artifacts in the ambiguity domain, In contrast to the traditional approach which uses l_2 norm, the l_1 estimate is not a linear function of $A(r,\psi)$ warranting the use of a more complex procedure when considering the objective function in (4). The new role of the kernel in imposing sparsity of the corresponding time-frequency distribution marks a significant change in the way we look at kernels and advocates adaptive kernel design.

Since we assume that auto-terms are around the origin in the ambiguity domain, we specify the kernel class as radial Gaussian. This formulation places a condition that the desired kernel acts as a low-pass filter. The choice of α in (4) influences the tradeoff between cross-term suppression and auto-term preservation.

3.3. Discrete-Time Implementation

RIDs are often computed for discrete-time signals. The discretized form of our optimization problem is as follows:

$$\underset{\sigma}{\text{minimize}} \quad \|\mathbf{A} - \mathbf{A} \circ \mathbf{\Phi}\|_2^2 + \lambda \left\| \mathbf{W}(\mathbf{A} \circ \mathbf{\Phi})\mathbf{W}^T \right\|_1$$

$$\text{subject to} \quad \Phi(p,q) = e^{-\frac{p^2}{2\sigma^2(q)}}, \quad (7)$$

$$\frac{1}{2\pi}\sum \sigma^2(q) \leq \alpha,$$

where $\mathbf{W}$ represents the discrete Fourier transform matrix, p and q are discretized versions of the radius and angle, respectively, and the symbol ' $\circ$ ' denotes the Hadamard product. We also follow the same approach as in the adaptive optimal kernel and observe the signal charac-teristics locally [4]. In (7), $\mathbf{A}$ corresponds to the short-time ambiguity function, i.e., the ambiguity function of the

windowed signal centered at time instant t_c (refer to Fig. 2). After computing σ and the corresponding kernel function, the 2-D Fourier transform is performed which corresponds to time t_c. In (7), the vectorized versions of the matrices are used and are subject to the l_1 and l_2 vector definitions.

In order to determine the value of σ, the gradient method is used. Starting from an initial value σ_0 for $i = 0$, the new value in the $(i+1)$th iteration is obtained as:

$$\sigma_{i+1}(q) = \sigma_i(q) - \varepsilon \nabla_i \qquad (8)$$

where ε is the step size and ∇_i is the gradient of the objective function, described in (7), at the ith iteration. Note that finding the gradient involves computing the derivative of the l_1 penalty term. This term is not differentiable, thus we use the sub-gradient strategy to define a gradient and update σ at each iteration [17]. We define the sub-gradient of our objective function $f(\sigma)$ as:

$$\nabla f(\sigma) = \begin{cases} \nabla L_2(\sigma) + \lambda \operatorname{sign}(L_1(\sigma)), & |L_1(\sigma)| > 0, \\ \nabla L_2(\sigma) + \lambda, & L_1(\sigma) = 0, \ \nabla L_2(\sigma) < -\lambda, \\ \nabla L_2(\sigma) - \lambda, & L_1(\sigma) = 0, \ \nabla L_2(\sigma) > -\lambda, \\ 0, & L_1(\sigma) = 0, -\lambda \le \nabla L_2(\sigma) \le \lambda, \end{cases}$$
$$(9)$$

where $\operatorname{sign}(\cdot)$ denotes the sign function.

4. SIMULATION RESULTS

In this section, we demonstrate the effectiveness of the proposed kernel through simulations.

In the first example, we observe the sum of two signals with polynomial phase:

$$x(t) = e^{j2\pi(0.05t + 0.05t^2/64 + 0.1t^3/64^2)}$$
$$+ e^{j2\pi(0.15t + 0.05t^2/64 + 0.1t^3/64^2)},$$

where $1 \le t \le 64$. 40% of the noise-free data is randomly missing. The ambiguity functions of data with and without missing samples are shown in Fig. 3. It is evident that the missing samples create a similar effect as that of additive noise, as discussed in [11, 12]. It can be observed that the ambiguity function of data with missing samples exhibit artifacts over the entire ambiguity domain, which is much more cluttered compared to the ambiguity domain of the noise-free full data. Also, we can observe that part of the artifacts reside along the $\tau = 0$ axis, which corresponds to impulse type behavior in the time domain. Thus, missing samples can be seen as a special type of impulsive noise.

We compare the QTFD obtained from the proposed kernel with the Choi-Williams distribution, adaptive optimal kernel distribution, and Spectrogram. The results which are depicted in Fig. 4, clearly demonstrate the effectiveness of the proposed approach. With missing samples, the Spectrogram and Choi-Williams distribution in Figs. 4(b) and 4(c) fail to capture the signal signature. Additionally, it
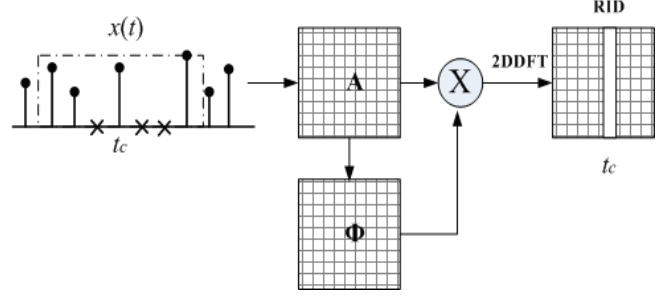

Fig. 2. Illustration of computing the time-frequency representation of one time slice. Missing samples are denoted by symbol '×'.
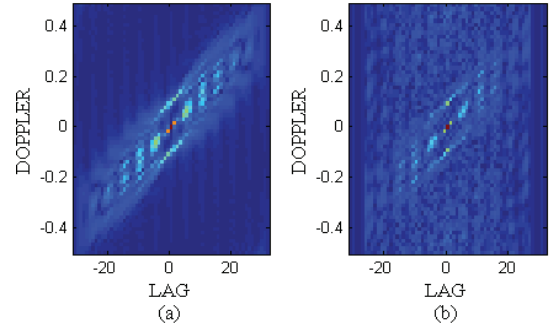

Fig. 3. Ambiguity function for data (a) with no missing samples; (b) for data with 40% missing samples.

is evident that Choi-Williams distribution shows vertical lines in time-frequency domain. These lines are impulses which the Choi-Williams distribution attempts to capture since it assumes them to be a part of the signal. On the other hand, the adaptive optimal kernel distribution with a high kernel volume in Fig. 4(a) signifies the two signal components, but also includes cross-terms. Our approach, as shown in Fig. 4(d), successfully suppresses the cross-terms.

In the next example, we consider data consisting of two crossing chirps when 50% of the samples is missing (Fig. 5). Similar observations, as in previous example, hold true for this data. Besides enhancing the auto-term concentration and providing a cross-term-free distribution, our approach yields less noisy representation compared to the signal-dependent kernel.

5. CONCLUSION

In this paper, we proposed an adaptive kernel design, which lends itself to satisfying the combined objective of bilinear cross-term suppression and sparse reconstruction in the time-frequency domain. It was shown that the new signal-dependent time-frequency kernel outperforms traditional signal-independent kernels, which only impose low-pass filtering in the ambiguity domain. It is also superior to signal-dependent kernels which are designed without sparsity considerations.
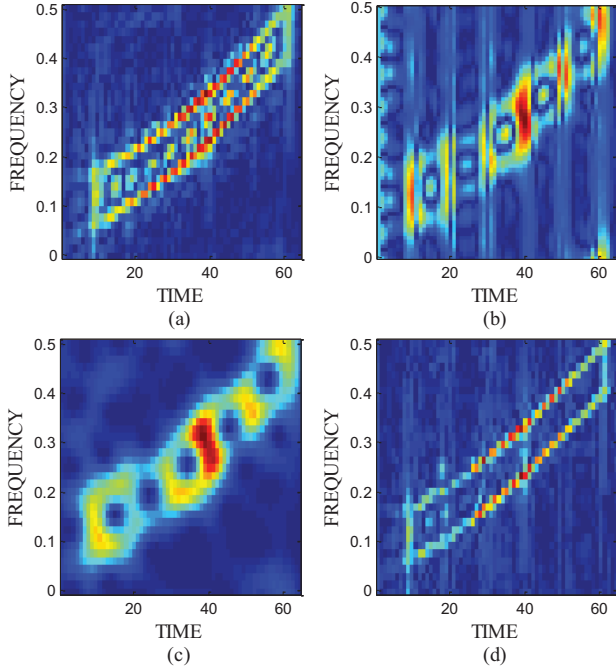
Fig. 4. RIDs of a signal consisting of two polynomial phase signals when 40% of data is missing. (a) adaptive optimal kernel; (b) Choi-Williams distribution; (c) Spectrogram; (d) proposed approach.
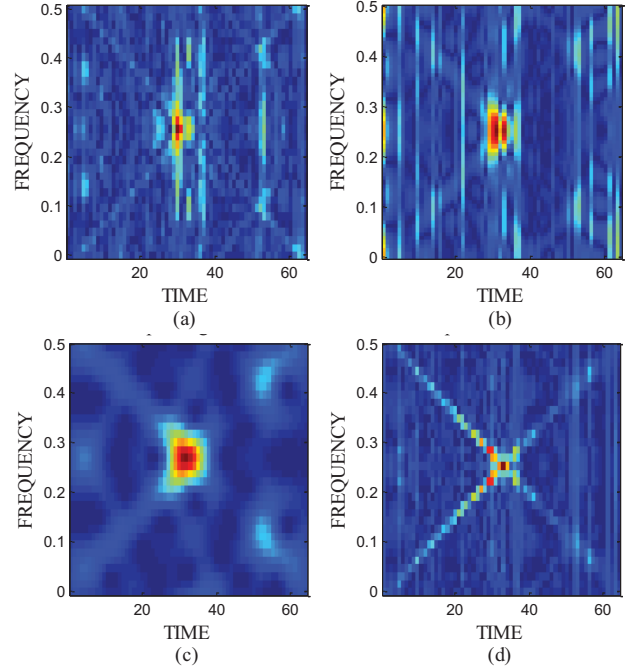

Fig. 5. RIDs of a signal consisting of two chirp components when 50% of data is missing. (a) adaptive optimal kernel; (b) Choi-Williams distribution; (c) Spectrogram; (d) proposed approach.

REFERENCES

[1] B. Boashash, *Time Frequency Analysis*, Oxford, UK: Elsevier, 2003.

[2] A. Belouchrani, M. G. Amin, N. Thirion-Moreau, and Y. D. Zhang, "Source separation and localization using time-frequency distributions," *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 97-107, Nov. 2013.

[3] H. I. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. Acoust. Speech Signal Process.*, vol.37, no.6, pp.862-871, Jun. 1989.

[4] R. G. Baraniuk and D. L. Jones, "A signal-dependent time-frequency representation: optimal kernel design," *IEEE Trans. Signal Process.*, vol.41, no.4, pp.1589-1602, Apr. 1993.

[5] D. L. Jones and R. G. Baraniuk, "An adaptive optimal-kernel time-frequency representation," *IEEE Trans. Signal Process.*, vol.43, no.10, pp.2361-2371, Oct. 1995.

[6] M. G. Amin and W. J. William, "High spectral resolution time–frequency distribution kernels," *IEEE Trans. Signal Process.*, vol. 46, no. 10, Oct. 1998.

[7] S. B. Hearon and M. G. Amin, "Minimum-variance time-frequency distribution kernels," *IEEE Trans. Signal Process.,* vol. 43, no. 5, pp. 1258-1262, May 1995.

[8] M. G. Amin, "Spectral decomposition of the time-frequency distribution kernels," *IEEE Trans. Signal Process.*, vol. 42, pp. 1156-1165, May 1994.

[9] P. Flandrin and P. Borgnat, "Time-frequency energy distributions meet compressed sensing," *IEEE Trans. Signal Process.*, vol.58, no.6, pp. 2974-2982, Jun. 2010.

[10] Y. D. Zhang and M. G. Amin, "Compressive sensing in nonstationary array processing using bilinear transforms," in *Proc. IEEE SAM*, Hoboken, NJ, June 2012.

[11] LJ. Stankovic, I. Orovic, S. Stankovic, and M. G. Amin, "Compressive sensing based separation of non-stationary and stationary signals overlapping in time-frequency," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4562-4572, Sep. 2013.

[12] Y. D. Zhang, M. G. Amin, and B. Himed, "Reduced interference time-frequency representations and sparse reconstruction of undersampled data," in *Proc. European Signal Proc. Conf.*, Marrakech, Morocco, Sep. 2013.

[13] B. Jokanović, M. G. Amin, and S. Stanković, "Instantaneous frequency and time-frequency signature estimation using compressive sensing," in *Proc. SPIE*, vol. 8714, Baltimore, MD, May 2013.

[14] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

[15] T. Alieva, and M. J. Bastiaans, "Phase-space distributions in quasi-polar coordinates and the fractional Fourier transform," *JOSA A*, vol. 17, no. 12, pp. 2324-2329, 2000.

[16] M. G. Amin, "Time-frequency spectrum analysis and estimation for nonstationary random processes," in B. Boashash (ed.), *Time-Frequency Signal Analysis: Methods and Applications*, Longman Cheshire, 1992.

[17] M. Schmidt, G. Fung, and R. Rosales, "Fast optimization methods for l_1 regularization: A comparative study and two new approaches," in *Proc. European Conf. Machine Learning,* Warsaw, Poland, Sep. 2007.