COMPRESSIVE NONSTATIONARY SPECTRAL ESTIMATION USING PARSIMONIOUS RANDOM SAMPLING OF THE AMBIGUITY FUNCTION

Alexander Jung, Georg Tauböck, and Franz Hlawatsch

Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Gusshausstrasse 25/389, A-1040 Vienna, Austria; e-mail: ajung@nt.tuwien.ac.at

ABSTRACT

We propose a compressive estimator for the discrete Rihaczek spectrum (RS) of a time-frequency sparse, underspread, nonstationary random process. The new estimator uses a compressed sensing technique to achieve a reduction of the number of measurements. The measurements are randomly located samples of the ambiguity function of the observed signal. We provide a bound on the mean-square estimation error and demonstrate the performance of the estimator by means of simulation results. The proposed RS estimator can also be used for estimating the Wigner-Ville spectrum (WVS) since for an underspread process the RS and WVS are almost equal.

Index Terms—Nonstationary spectral estimation, Rihaczek spectrum, Wigner-Ville spectrum, compressed sensing, basis pursuit

1. INTRODUCTION

Let $X[n]$, $n = 0, \ldots, N - 1$, be a nonstationary, zero-mean, circularly symmetric complex, finite-length, or equivalently, periodic random process with autocorrelation $C_x[n_1, n_2] \equiv \mathbb{E}\{X[n_1]X^*[n_2]\}$ (here, $n_1, n_2 \in \mathbb{Z}$ due to periodic continuation). An alternative characterization of the second-order statistics of $X[n]$ is given by the Rihaczek spectrum (RS) [1–3]

$$\hat{R}_X[n, k] \triangleq \sum_m C_X[n, n-m] e^{-j \frac{2\pi}{N} km},$$

where $k$ is a discrete frequency variable (sums are from 0 to $N - 1$ unless noted otherwise). The RS is the expectation of the Rihaczek distribution (RD), i.e., $\hat{R}_X[n, k] = \mathbb{E}\{R_X[n, k]\}$, with the RD defined as $R_X[n, k] = \mathbb{E}\{X[n]X^*[n]e^{-j \frac{2\pi}{N} km}\} = X[n]X^*[k]e^{-j \frac{2\pi}{N} km}$ [1, 2] (here, $X[k] = \sum_n X[n]e^{-j \frac{2\pi}{N} nk}$ is the discrete Fourier transform (DFT) of $X[n]$). If $X[n]$ is an underspread process [3], the RS can be viewed as a “time-dependent power spectrum” or “mean-time-frequency (TF) energy distribution,” and it is approximately equal to the Wigner-Ville spectrum [2–4]. The underspread property means that process components that are not very close in the TF plane are approximately uncorrelated [3]; this property is satisfied by many practical processes.

Here, we consider the problem of estimating the RS $\hat{R}_X[n, k]$ of an underspread process $X[n]$ from a single observed process realization $x[n]$. Existing spectrum estimators for underspread processes include smoothed versions of the Wigner distribution of $x[n]$ [2, 4–6] and estimators using a local cosine basis expansion [7] or a Gabor expansion [8] of $x[n]$. In this paper, we assume that the process $X[n]$ is effectively TF sparse in addition to being underspread. Effective TF sparsity means that most RS values within the total TF region considered are almost zero. For example, in many applications the TF spectrum is effectively supported in a few relatively small regions of the TF plane (corresponding to TF localized signal components) and thus almost zero in the rest of the TF plane.

2. SMOOTHED-RD ESTIMATORS OF THE RS

The proposed RS estimator is a compressive extension of a smoothed-RD estimator of the RS. In this section, we consider smoothed-RD estimators satisfying an approximate minimum-variance-unbiased property (see [5] for the Wigner-Ville case). We first need to establish some background.

2.1. Expected Ambiguity Function and Underspread Processes

The expected ambiguity function (EAF) of a nonstationary random process $X[n]$ is defined as [3]

$$\bar{A}_X[m, l] \triangleq \sum_n C_X[n, n-m] e^{-j \frac{2\pi}{N} ln},$$

where $m$ and $l$ denote discrete time lag and discrete frequency lag, respectively. The RS and EAF are related by a 2-D DFT, i.e.,

$$\hat{A}_X[m, l] = \frac{1}{N} \sum_n \sum_k \hat{R}_X[n, k] e^{-j \frac{2\pi}{N} (ln-mk)}. \hspace{1cm} (1)$$

This work was supported by FWF project “Statistical Inference” (S10603-N13) within the National Research Network SISE and by WWTF project SPORTS (MA 07-004).

It will be convenient to consider length-$N$ functions as periodic functions with period $N$. ©2009 IEEE 642
The EAF is the expectation of the AF $A_X[m, l] = \sum_n X[n] X^*[n-m] e^{-j \frac{2\pi}{N}(n-m)}$, i.e., $A_X[m, l] = E[A_X[m, l]]$. The 2-D DFT relation (1) holds also for the RD and AF, i.e.,

$$A_X[m, l] = \frac{1}{N} \sum_n \sum_k R_X[n, k] e^{-j \frac{2\pi}{N}(n-m-k)}.$$  \(2\)

A process $X[n]$ is said to be underspread if its EAF $A_X[m, l]$ is well concentrated around the origin $(m, l) = (0, 0)$; its RS $R_X[n, k]$ will then be a smooth function. The degree of underspreadness of $X[n]$ can be measured by a TF correlation moment \[3, 12\]

$$m_X^{(\omega)} = \frac{\sum_m \sum_l |\hat{A}_X[m, l]|^2}{\sum_m \sum_l |A_X[m, l]|^2},$$  \(3\)

where $\phi[m, l]$ is a suitable weighting function that is generally increasing for radially increasing $m$ and $l$. For an underspread process, $m_X^{(\omega)}$ will be small ($< 1$).

2.2. Definition and Design of Smoothed-RD Estimators

It is natural to consider an RS estimator $\hat{R}_X[n, k]$ that is quadratic in $x[n]$ and covariant to TF shifts of the type $e^{j \frac{2\pi}{N}n} x[n-m]$. Such estimators belong to the discrete-time-frequency version of Cohen’s class \[2\], and can thus be written in terms of the RD of $x[n]$ as

$$\hat{R}_X[n, k] \triangleq \frac{1}{N} \sum_{n'} \sum_{k'} \psi[n-n', k-k'] R_X[n', k'].$$  \(4\)

Here, $\psi[n, k]$ is a convolution kernel that remains to be chosen. The 2-D DFT of $\hat{R}_X[n, k]$, $\hat{A}_X[m, l]$, is an estimator of the EAF $A_X[m, l]$. From (5), (4), and (2),

$$\hat{A}_X[m, l] = \Psi[m, l] A_X[m, l],$$  \(6\)

where the weighting function $\Psi[m, l]$ is the 2-D DFT of $\psi[n, k]$, i.e., $\Psi[m, l] = \sum_{n'} \sum_{k'} \psi[n-n', k-k'] e^{-j \frac{2\pi}{N}(n-m-k)}$.

Following \[5\], we design $\psi[n, k]$ or, equivalently, $\Psi[m, l]$ such that $\hat{R}_X[n, k]$ is approximately a minimum variance unbiased (MVU) estimator. The mean-square error (MSE) $\varepsilon \triangleq E[\|\hat{R}_X - R_X\|^2]$ can be written as $\varepsilon = B^2 + V$ with the squared bias $B^2 \triangleq E[\|\hat{R}_X\|^2]$ and the variance $V \triangleq E[\|\hat{R}_X - E[\hat{R}_X]\|^2]$.

The MVU design minimizes $V$ under the constraint $B^2 = 0$. For a circularly symmetric Gaussian process $X[n]$ with a small EAF support $A$, it can be shown (cf. \[5\]) that the MVU design yields

$$\Psi_{MVU}[m, l] = \begin{cases} 1, & (m, l) \in A \\ 0, & \text{elsewhere,} \end{cases}$$  \(7\)

and that the resulting MSE is approximately given by

$$\varepsilon_{MVU} \approx \frac{1}{N} \| \hat{R}_X \|^2 |A|,$$  \(8\)

where $|A| = \sum_{(m, l) \in A} 1$ is the “area” of $A$. According to (8), $\varepsilon_{MVU}$ is smaller if the process is more underspread (i.e., if $|A|$ is smaller). In practice, the EAF is rarely exactly limited to a small support. However, in the underspread case, it decays quickly outside some (small) region $A$ around the origin, and thus we can define $A$ to be the effective EAF support. Since $|A|$ is small, $\varepsilon_{MVU}[m, l]$ in (7) is well localized around the origin. Due to the 2-D DFT relating $\psi_{MVU}[n, k]$ and $\Psi_{MVU}[m, l]$, the convolution kernel $\psi_{MVU}[n, k]$ is then a smooth function, and thus (4) corresponds to a smoothing of the RD. Therefore, the RS estimator (4) using $\psi_{MVU}[n, k]$, $\hat{R}_{X,MVU}[n, k] \triangleq \frac{1}{N} \sum_{n'} \sum_{k'} \psi_{MVU}[n-n', k-k'] R_X[n', k']$, \(9\)

will be called a smoothed-RD estimator. According to (6) and (7), the corresponding EAF estimator is obtained as

$$\hat{A}_{X,MVU}[m, l] = \begin{cases} A_X[m, l], & (m, l) \in A \\ 0, & \text{elsewhere.} \end{cases}$$  \(10\)

That is, $\hat{A}_{X,MVU}[m, l]$ equals the AF $A_X[m, l]$ within $A$ and it is zero outside $A$.

3. COMPRESSIVE RS ESTIMATOR

In this section, we propose a compressive extension of the MVU smoothed-RD estimator $\hat{R}_{X,MVU}[n, k]$ in (9).

3.1. Formulation of the Compressive RS Estimator

We assume that the process $X[n]$ is underspread in the sense that its EAF $A_X[m, l]$ is effectively supported in a centered rectangular region $A \triangleq [-\Delta m/2, +\Delta m/2] \times [-\Delta l/2, +\Delta l/2]$ of area $|A| = \Delta m \Delta l$, where $\Delta m$ and $\Delta l$ are even integers dividing $N$. Because $\hat{A}_{X,MVU}[m, l] = 0$ outside $A$ (see (10)) and $\hat{A}_{X,MVU}[m, l]$ is the 2-D DFT of $\hat{R}_{X,MVU}[n, k]$, we can subsample $\hat{R}_{X,MVU}[n, k]$ by the factors $\Delta n = N/\Delta m$ with respect to $n$ and $\Delta k = N/\Delta m$ with respect to $k$. This gives the subsampled RS estimator

$$\hat{R}_{X,sub}[p, q] \triangleq \hat{R}_{X,MVU}[p \Delta n, q \Delta k],$$

with $p = 0, \ldots, \Delta l - 1, q = 0, \ldots, \Delta m - 1$.

We can recover $\hat{R}_{X,MVU}[n, k]$ from $\hat{R}_{X,sub}[p, q]$ by means of the interpolation

$$\hat{R}_{X,MVU}[n, k] = \sum_{p=0}^{\Delta l-1} \sum_{q=0}^{\Delta m-1} \theta[n-p \Delta n, k-q \Delta k] \hat{R}_{X,sub}[p, q],$$

with $\theta[n, k] \triangleq \frac{\sin(n \Delta n/|A|) \sin(k \Delta k/|A|)}{\sin(\Delta n/|A|)} e^{j \frac{2\pi}{N}n}$.

Let us represent $\hat{R}_{X,sub}[p, q]$ by the $\Delta l \times \Delta m$ matrix $R$ with elements $(R)_{p,q} = \hat{R}_{X,sub}[p-1, q-1]$, for $p = 1, \ldots, \Delta l$ and $q = 1, \ldots, \Delta m$, and let $r \triangleq \text{vec}(R)$ denote the vector of length $\Delta l \Delta m = |A|$ that results from stacking all columns of $R$. Thus, $r$ represents $\hat{R}_{X,sub}[p, q]$. Similarly, let $A$ be the $\Delta l \times \Delta m$ matrix comprising those samples of $\hat{A}_{X,MVU}[m, l]$ in (10) that are located within the effective support region $A$, i.e., $(A)_{l,m} = \hat{A}_{X,MVU}[m - \Delta m/2, 1-l \Delta l/2] = A_{X,MVU}[m - \Delta m/2, 1-l \Delta l/2]$, for $l = 1, \ldots, \Delta l$ and $m = 1, \ldots, \Delta m$. Furthermore, let us stack all columns of $A$ into the vector $a \triangleq \text{vec}(A)$ of length $\Delta l \Delta m = |A|$. Thus, $a$ contains all AF values located in $A$. From the 2-D DFT relation (5), it then follows that

$$a = U r,$$

with $U \triangleq \frac{\Delta n \Delta k}{N} F_{2\Delta l} \otimes F_{2\Delta m}$, \(12\)

where $\otimes$ denotes the Kronecker product and, e.g., $F_{2\Delta l}$ denotes the length-$2\Delta l$ DFT matrix, i.e., $(F_{2\Delta l})_{k,l} \triangleq e^{-j \frac{2\pi}{N}k l}$. We note that $U$ is a $|A| \times |A|$ matrix (with $|A| = \Delta m \Delta l$) that is unitary up to a factor.

In what follows, we assume that the process $X[n]$ is effectively TF sparse in the sense that at most $S$ samples of the subsampled
Consider a vector \( \hat{m} = \Phi r \),
where the \( M \times |A| \) “measurement matrix” \( \Phi \) comprises those \( M \) rows of \( U \) in (12) that correspond to the selected entries of \( a \). Hence, \( \Phi \) is obtained by randomly selecting \( M \ll |A| \) rows from the \( |A| \times |A| \) matrix \( U \).
From the “compressed measurements” \( \hat{m} \), we then recover the \( S \)-sparse vector \( r \) by means of the following convex optimization problem (basis pursuit) [13]:

\[
\hat{r} = \arg \min_{r} \| r \|_1 \quad \text{subject to} \quad \Phi r = \hat{m}.
\]

(13)
This is justified (for \( M \) chosen large enough) by the following result of CS theory [13, 14].
Consider a vector \( r \) of length \( Q \) (not necessarily sparse) and the vector of length \( M \leq Q \) defined by \( \hat{m} = \Phi r \), where the \( M \times Q \) matrix \( \Phi \) is formed by randomly selecting \( M \ll |A| \) rows from a unitary (possibly up to a factor) \( Q \times Q \) matrix \( U \). Let \( \mu = \sqrt{Q \max(|U|)} \) (known as the coherence of \( U \)), and let \( r_S \) be the \( S \)-sparse approximation to \( r \) that is obtained by zeroing all entries of \( r \) except the \( S \) entries with largest magnitudes. Then if

\[
M \geq C \log^4 Q \mu^2 S,
\]
the vector \( \hat{r} \) in (13) satisfies with overwhelming probability\(^2\)

\[
\| \hat{r} - r \|_2 \leq \frac{D \| r - r_S \|_1}{\sqrt{S}}.
\]

(15)
Here, \( C \) and \( D \) are two positive constants.
Thus, if \( r \) is effectively \( S \)-sparse (i.e., \( \| r - r_S \|_1 \) is small) and if \( M \) is sufficiently large, then the result \( r \) of basis pursuit in (13) is a good approximation to \( r \) (i.e., \( \| \hat{r} - r \|_2 \) is small).

The proposed compressive RS estimation algorithm can now be stated as follows.

1. Calculate \( M \) samples of the AF \( A_X[m, l] \) with locations \( (m, l) \) randomly selected within the effective support \( A \), where \( M \) satisfies (14) with an empirically chosen constant \( C \). Arrange these samples in the length-\( M \) vector \( m \).
2. Construct the \( M \times |A| \) matrix \( \Phi \) by randomly selecting \( M \) rows from the matrix \( U \) in (12).
3. Using \( \Phi \), calculate an estimate \( \hat{r} \) of \( r \) from \( \hat{m} \) by means of basis pursuit (13). The vector \( \hat{r} \) contains the reconstructed samples of the subsampled RS estimator \( \hat{R}_{X,\text{sub}}[p, q] \), which will be denoted as \( \hat{R}_{X,\text{sub}}[p, q] \).
4. Apply the interpolation (11) to \( \hat{R}_{X,\text{sub}}[p, q] \), i.e., calculate

\[
\hat{R}_{X,\text{CS}}[n, k] = \sum_{p=0}^{\Delta l-1} \sum_{q=0}^{\Delta m-1} \theta[n-p\Delta n, k-q\Delta k] \hat{R}_{X,\text{sub}}[p, q].
\]

This defines the compressive RS estimator.

\(^2\)That is, the probability of (15) not being true decreases exponentially for growing \( M \).

### 3.2. Performance Bound

We now present (without proof, because of space limitations) an upper bound on the MSE \( \varepsilon_{\text{CS}} \equiv \varepsilon \left( \| \hat{R}_{X,\text{CS}} - R_X \|_2^2 \right) \) of the proposed compressive estimator \( \hat{R}_{X,\text{CS}}[n, k] \). This bound holds for a circularly symmetric complex Gaussian process \( X[n] \). It consists of two parts, which will be considered separately and then combined.

First, using Isserlis’ formula [5], it can be shown that the MSE \( \varepsilon_{\text{MVU}} \equiv \varepsilon \left( \| \hat{R}_{X,\text{MVU}} - R_X \|_2^2 \right) \) of the basic (noncompressive) estimator \( \hat{R}_{X,\text{MVU}}[n, k] \) in (9) is bounded as

\[
\varepsilon_{\text{MVU}} \leq \| R_X \|_2^2 \left( \frac{m(\frac{Q}{A})}{N} + \frac{|A|}{N} \right).
\]

(16)
Here, \( m(\frac{Q}{A}) \) is the EAF moment (3) using the weighting function \( \phi[m, l] \equiv \mathcal{I}_A[m, l] \), which is defined in (1) if \((m, l) \notin A\) and 0 else. (Note that \( m(\frac{Q}{A}) = 0 \) if the EAF is exactly zero outside \( A \); also note that the approximation (8) formally follows from (16) by setting \( m(\frac{Q}{A}) = 0 \).) The bound (16) is valid for an arbitrary shape of \( A \). In our case, \( A = [-\Delta n/2 + 1, \Delta n/2] \times [-\Delta l/2 + 1, \Delta l/2] \) and thus \( |A| = \Delta m \Delta l \). The bound is small if \( Q/n \) is underspread (small \( m(\frac{Q}{A}) \) and \( |A| = \Delta m \Delta l \)).

An additional error component is due to the compression-recovery stage and corresponds to the MSE component \( \Delta \varepsilon \equiv \varepsilon \left( \| \hat{R}_{X,\text{CS}} - \hat{R}_{X,\text{MVU}} \|_2^2 \right) \). As before, let \( S \) be the effective sparsity of the subsampled RS \( \hat{R}_{X,\text{sub}}[p, q] \), i.e., the number of significant samples of \( \hat{R}_{X,\text{sub}}[p, q] \) (we note, however, that the bound presented below holds for arbitrary \( S \in \{1, \ldots, 2m, 2k\} \)). Let us index the subsampled TF locations \((p, q)\) in the order of decreasing RS magnitudes \( |\hat{R}_{X,\text{sub}}[p, q]| \); such that \((p_r, q_r)\) corresponds to the \( r \)-th largest RS magnitude, \( |\hat{R}_{X,\text{sub}}[p_r, q_r]| \). We assume that the number \( M \) of randomly selected AF samples is sufficiently large so that (15) is satisfied (a sufficient condition is, (14), with \( Q = \Delta m \Delta l \)). Then, using (15), it can be shown that \( \Delta \varepsilon \) is bounded as

\[
\Delta \varepsilon \leq D^2 \| R_X \|_2^2 \frac{\Delta m \Delta l - S}{S} \sigma_X(S),
\]

(17)
with the TF sparsity profile (here defined differently from [8])

\[
\sigma_X(S) \equiv \sum_{r=S+1}^{\infty} P_r,
\]

(18)
where \( P_r \equiv E \left[ |\hat{R}_{X,\text{sub}}[p_r, q_r]|^2 \right] = E \left[ |\hat{R}_{X,\text{MVU}}[\Delta n, \Delta l, q, k]|^2 \right] \). Good reconstruction accuracy (for a given \( S \)) requires that \( \sigma_X(S) \) is small. For an underspread process (i.e., with small \( \Delta m \Delta l \)), it can be shown that \( P_r \) can be approximated as

\[
P_r \approx \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{k=0}^{K-1} |\hat{R}_{X,\text{MVU}}[\Delta n, n, q, k]|^2 \| \hat{R}_{X,\text{sub}}[n, k] \|^2.
\]

This expression involves the RS \( \hat{R}_{X}[n, k] \), and it shows that \( P_r \) is small if \( \hat{R}_{X}[n, k] \) is small within a neighborhood of \((p, \Delta n, q, \Delta k)\) or, said differently, if \((p, \Delta n, q, \Delta k)\) is located outside a broadened version of the effective support of \( R_X[n, k] \) (the broadening depends on the TF spread of \( \psi_{\text{MVU}}[n, k] \)). Thus, \( \sigma_X(S) \) in (17), (18) will be small if \( S \) is chosen such that \( S \Delta m \Delta l \) is approximately equal to the area of the broadened effective support of \( R_X[n, k] \).

Using the triangle inequality for norms (see [8] for details), the MSE bounds (16) and (17) can now be combined into a bound on the MSE \( \varepsilon_{\text{CS}} \) of the compressive estimator \( \hat{R}_{X,\text{CS}}[n, k] \):

\[
\varepsilon_{\text{CS}} \leq \left( \sqrt{\varepsilon_{\text{MVU}}} + \sqrt{\Delta \varepsilon} \right)^2.
\]
We note that the bound (15) is known to be very loose [14], and thus our bound on $\Delta \varepsilon$ and, in turn, on $\varepsilon_{CS}$ will generally be quite pessimistic. These bounds are useful mostly in an asymptotic sense.

4. SIMULATION RESULTS

We generated 1000 realizations of a nonstationary Gaussian random process of length 256 whose RS is shown in Fig. 1(a). The TF sparsity profile $\sigma_X(S)$ and the EAF of the process are plotted in Fig. 2. From this figure, we can conclude that the process is only moderately TF sparse and underspread. In Fig. 1(b)–(d), we depict RS estimates (averaged over the 1000 realizations) for compression factors $N/M = 1$, 2, and 5 (note that $N/M = 1$ corresponds to the noncompressive smoothed-RD estimator $\hat{R}_{X,MV}[n,k]$ in (9)). In Fig. 1(b), we observe a certain smoothing in the averaged estimate $\hat{R}_{X,MV}[n,k]$; this is due to the weak underspreadness of the process. From Fig. 1(c), (d), we conclude that the compression does not result in drastic additional errors for $N/M$ up to about 5.

Fig. 3 shows the empirical normalized MSE (NMSE) of the compressive RS estimator $\hat{R}_{X,CS}[n,k]$ versus the compression factor $N/M$. One can observe a “graceful degradation” with increasing $N/M$.

5. CONCLUSION

We proposed a compressive spectrum estimator for TF sparse, underspread, nonstationary random processes. The measurements used by the estimator are samples of the ambiguity function of the observed signal that are randomly chosen within a small region of the TF lag plane. A compression (reduction of the number of measurements) is achieved by means of a compressed sensing technique. Simulation results demonstrated good average performance for compression factors up to about 5. Our method can also be used to estimate the Wigner-Ville spectrum and the autocorrelation.

6. ACKNOWLEDGMENT

The authors would like to thank Prof. Gerald Matz for helpful suggestions and comments.

![Fig. 1. Average performance of RS estimators: (a) RS of a nonstationary process $X[n]$ of length 256, (b)–(d) averaged RS estimates obtained with (b) the noncompressive estimator $\hat{R}_{X,MV}[n,k]$ (compression factor $N/M = 1$), (c) the compressive estimator $\hat{R}_{X,CS}[n,k]$ with $N/M = 2$, and (d) the compressive estimator with $N/M = 5$. The total TF region of size $256 \times 256$ is shown.](image)

![Fig. 2. (a) TF sparsity profile $\sigma_X(S)$ and (b) EAF magnitude $|\hat{A}_X[n,l]|$ of the simulated process. The normalized area of the effective EAF support $\mathcal{A}$ is $|\mathcal{A}|/N \approx 0.25$.](image)

![Fig. 3. Empirical normalized MSE of the compressive RS estimator versus the compression factor $N/M$.](image)

7. REFERENCES