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 $X[n]$, $n = 0, ..., N-1$ be a nonstationary, zero-mean, circularly symmetric complex, finite-length or, equivalently, periodic[1] random process with autocorrelation $C_X[n_1, n_2] \triangleq \mathrm{E}\{X[n_1]X^*[n_2]\}$ (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]

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

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., $\bar{R}_X[n,k] = \mathrm{E}\{R_X[n,k]\}$, with the RD defined as $R_X[n,k] \triangleq \sum_m X[n]X^*[n-m]e^{-j\frac{2\pi}{N}km} = X[n]\hat{X}^*[k]e^{-j\frac{2\pi}{N}nk}$ [1, 2] (here, $\hat{X}[k] \triangleq \sum_n X[n]e^{-j\frac{2\pi}{N}kn}$ 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 $\bar{R}_X[n,k]$ 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.

In [8], we proposed a "compressive" estimator of the Wigner-Ville spectrum that exploits TF sparsity to reduce the amount of measurement data required (similar in spirit to a sufficient statistic). This is useful if the measurements are transmitted over a low-rate channel or if dedicated measurement devices are available: one obtains a bit-rate reduction in the first case and a reduction of the number of measurement devices in the second. The method of [8] achieves a compression in the measurement space by means of a compressed sensing (CS) technique. The measurements are inner products of the observed signal with noise-like signals constructed as random linear combinations of Gabor functions.

In this paper, we present an alternative compressive spectrum estimator that also uses a CS technique for reducing the amount of measurement data. The difference from [8] (besides the fact that here we consider the RS and use a finite-blocklength, discrete setting) is the nature of the measurements, which are samples of the ambiguity function (AF) of $x[n]$ at randomly chosen TF lag positions. This is especially interesting if dedicated devices for computing AF samples are available [9, 10]. Algorithmically, our compressive spectrum estimator is similar to the TF analysis technique presented in [11]. However, the motivation and background are quite different: whereas [11] considers TF analysis in a deterministic setting (with the goal of improving the TF localization of the Wigner distribution), we consider nonstationary spectral estimation. We furthermore provide a bound on the mean-square error of our RS estimator.

This paper is organized as follows. In Section 2, we review smoothed-RD estimators of the RS, which provide a basis for the proposed compressive estimator. In Section 3, we develop the compressive estimator and present a bound on its mean-square error. Finally, numerical results are provided in Section 4.

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.,

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

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).

[1] It will be convenient to consider length-N functions as periodic functions with period N.

The EAF is the expectation of the AF $A_X[m,l] \triangleq \sum_n X[n]X^*[n-m]\, e^{-j\frac{2\pi}{N}ln}$, i.e., $\bar{A}_X[m,l] = \mathrm{E}\{A_X[m,l]\}$. The 2-D DFT relation (1) holds also for the RD and AF, i.e.,

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

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

$$m_X^{(\phi)} \,\triangleq\, \frac{\sum_m\sum_l \phi[m,l]\,|\bar{A}_X[m,l]|^2}{\sum_m\sum_l |\bar{A}_X[m,l]|^2}\,, \tag{3}$$

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

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 $e^{j\frac{2\pi}{N}ln}x[n-m]$. 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

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

Here, $\psi[n,k]$ is a convolution kernel that remains to be chosen. The 2-D DFT of $\hat{R}_X[n,k]$,

$$\hat{A}_X[m,l] \,\triangleq\, \frac{1}{N}\sum_n\sum_k \hat{R}_X[n,k]\, e^{-j\frac{2\pi}{N}(ln-mk)}, \tag{5}$$

is an estimator of the EAF $\bar{A}_X[m,l]$. From (5), (4), and (2),

$$\hat{A}_X[m,l] \,=\, \Psi[m,l]\, A_x[m,l]\,, \tag{6}$$

where the weighting function $\Psi[m,l]$ is the 2-D DFT of $\psi[n,k]$, i.e., $\Psi[m,l] \triangleq \frac{1}{N}\sum_n\sum_k \psi[n,k]\, e^{-j\frac{2\pi}{N}(ln-mk)}$.

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) $\varepsilon \triangleq \mathrm{E}\big\{\big\|\hat{R}_X - \bar{R}_X\big\|_2^2\big\}$ can be written as $\varepsilon = B^2 + V$ with the squared bias $B^2 \triangleq \big\|\mathrm{E}\{\hat{R}_X\} - \bar{R}_X\big\|_2^2$ and the variance $V \triangleq \mathrm{E}\big\{\big\|\hat{R}_X - \mathrm{E}\{\hat{R}_X\}\big\|_2^2\big\}$. The MVU design minimizes V under the constraint $B^2=0$. For a circularly symmetric Gaussian process $X[n]$ with a small EAF support $\mathcal{A}$, it can be shown (cf. [5]) that the MVU design yields

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

and that the resulting MSE is approximately given by

$$\varepsilon_{\mathrm{MVU}} \,\approx\, \frac{1}{N}\|\bar{R}_X\|_2^2\,|\mathcal{A}|\,, \tag{8}$$

where $|\mathcal{A}| = \sum_{(m,l)\in\mathcal{A}} 1$ is the "area" of $\mathcal{A}$. According to (8), $\varepsilon_{\mathrm{MVU}}$ is smaller if the process is more underspread (i.e., if $|\mathcal{A}|$ 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 $\mathcal{A}$ around the origin, and thus we can define $\mathcal{A}$ to be the *effective* EAF support. Since $|\mathcal{A}|$ is small, $\Psi_{\mathrm{MVU}}[m,l]$ in (7) is well localized around the origin. Due to the 2-D DFT relating $\psi_{\mathrm{MVU}}[n,k]$ and $\Psi_{\mathrm{MVU}}[m,l]$, the convolution kernel $\psi_{\mathrm{MVU}}[n,k]$ is then a smooth function, and thus (4) corresponds to a *smoothing* of the RD. Therefore, the RS estimator (4) using $\psi_{\mathrm{MVU}}[n,k]$,

$$\hat{R}_{X,\mathrm{MVU}}[n,k] \,\triangleq\, \frac{1}{N}\sum_{n'}\sum_{k'} \psi_{\mathrm{MVU}}[n-n',k-k']\, R_x[n',k']\,, \tag{9}$$

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

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

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

3. COMPRESSIVE RS ESTIMATOR

In this section, we propose a compressive extension of the MVU smoothed-RD estimator $\hat{R}_{X,\mathrm{MVU}}[n,k]$ 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 $\bar{A}_X[m,l]$ is *effectively* supported in a centered rectangular region $\mathcal{A} \triangleq [-\Delta m/2+1, \Delta m/2] \times [-\Delta l/2+1, \Delta l/2]$ of area $|\mathcal{A}| = \Delta m \Delta l$, where Δm and Δl are even integers dividing N. Because $\hat{A}_{X,\mathrm{MVU}}[m,l]=0$ outside $\mathcal{A}$ (see (10)) and $\hat{A}_{X,\mathrm{MVU}}[m,l]$ is the 2-D DFT of $\hat{R}_{X,\mathrm{MVU}}[n,k]$, we can subsample $\hat{R}_{X,\mathrm{MVU}}[n,k]$ by the factors $\Delta n = N/\Delta l$ with respect to n and $\Delta k = N/\Delta m$ with respect to k. This gives the *subsampled RS estimator*

$$\hat{R}_{X,\mathrm{sub}}[p,q] \,\triangleq\, \hat{R}_{X,\mathrm{MVU}}[p\Delta n, q\Delta k]\,,\qquad p=0,\ldots,\Delta l-1\,,\; q=0,\ldots,\Delta m-1\,.$$

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

$$\hat{R}_{X,\mathrm{MVU}}[n,k] \,=\, \sum_{p=0}^{\Delta l-1}\sum_{q=0}^{\Delta m-1} \theta[n-p\Delta n, k-q\Delta k]\, \hat{R}_{X,\mathrm{sub}}[p,q]\,, \tag{11}$$

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

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

$$\mathbf{a} = \mathbf{U}\mathbf{r}\,,\quad \text{with } \mathbf{U} \,\triangleq\, \frac{\Delta n\Delta k}{N}\,\mathbf{F}_{\Delta l}^{*}\otimes\mathbf{F}_{\Delta m}\,, \tag{12}$$

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

RS $\bar{R}_{X,\text{sub}}[p,q] \triangleq \bar{R}_X[p\Delta n, q\Delta k]$ are significant, the others being approximately zero. Equivalently, $\mathbf{r}$ is an *effectively S-sparse* vector. The positions of the significant samples are unknown but a rough estimate of their number, S, is assumed known.

The proposed compressive estimator is based on (12) and the following considerations. Let us randomly select $M \ll |\mathcal{A}|$ samples of the AF $A_x[m,l]$ within the effective EAF support $\mathcal{A}$. These AF samples correspond to $M \ll |\mathcal{A}|$ entries of the length-$|\mathcal{A}|$ vector $\mathbf{a}$. Let $\mathbf{m}$ denote the length-M vector consisting of the selected samples, arranged in arbitrary order. Because of (12), the vectors $\mathbf{m}$ and $\mathbf{r}$ are related via the "measurement equation"

$$\mathbf{m} = \boldsymbol{\Phi}\mathbf{r}\,,$$

where the $M \times |\mathcal{A}|$ "measurement matrix" $\boldsymbol{\Phi}$ comprises those M rows of $\mathbf{U}$ in (12) that correspond to the selected entries of $\mathbf{a}$. Hence, $\boldsymbol{\Phi}$ is obtained by randomly selecting $M \ll |\mathcal{A}|$ rows from the $|\mathcal{A}| \times |\mathcal{A}|$ matrix $\mathbf{U}$.

From the "compressed measurements" $\mathbf{m}$, we then recover the S-sparse vector $\mathbf{r}$ by means of the following convex optimization problem (*basis pursuit*) [13]:

$$\hat{\mathbf{r}} \,=\, \arg\min_{\mathbf{r}'} \|\mathbf{r}'\|_1 \quad \text{subject to } \ \boldsymbol{\Phi}\mathbf{r}' = \mathbf{m}\,. \tag{13}$$

This is justified (for M chosen large enough) by the following result of CS theory [13, 14].

Consider a vector $\mathbf{r}$ of length Q (not necessarily sparse) and the vector of length $M \le Q$ defined by $\mathbf{m} = \boldsymbol{\Phi}\mathbf{r}$, where the $M \times Q$ matrix $\boldsymbol{\Phi}$ is formed by randomly selecting M rows from a unitary (possibly up to a factor) $Q \times Q$ matrix $\mathbf{U}$. Let $\mu \triangleq \sqrt{Q}\max_{i,j}|(\mathbf{U})_{i,j}|$ (known as the coherence *of $\mathbf{U}$), and let $\mathbf{r}_S$ be the S-sparse approximation to $\mathbf{r}$ that is obtained by zeroing all entries of $\mathbf{r}$ except the S entries with largest magnitudes. Then if*

$$M \ge C(\log Q)^4 \mu^2 S\,, \tag{14}$$

the vector $\hat{\mathbf{r}}$ in (13) satisfies with overwhelming probability[2]

$$\|\hat{\mathbf{r}} - \mathbf{r}\|_2 \,\le\, D\,\frac{\|\mathbf{r} - \mathbf{r}_S\|_1}{\sqrt{S}}\,. \tag{15}$$

Here, C and D are two positive constants.

Thus, if $\mathbf{r}$ is effectively S-sparse (i.e., $\|\mathbf{r} - \mathbf{r}_S\|_1$ is small) and if M is sufficiently large, then the result $\hat{\mathbf{r}}$ of basis pursuit in (13) is a good approximation to $\mathbf{r}$ (i.e., $\|\hat{\mathbf{r}} - \mathbf{r}\|_2$ 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 EAF support $\mathcal{A}$, where M satisfies (14) with an empirically chosen constant C. Arrange these samples in the length-M vector $\mathbf{m}$.
2. Construct the $M \times |\mathcal{A}|$ matrix $\boldsymbol{\Phi}$ by randomly selecting M rows from the matrix $\mathbf{U}$ in (12).
3. Using $\boldsymbol{\Phi}$, calculate an estimate $\hat{\mathbf{r}}$ of $\mathbf{r}$ from $\mathbf{m}$ by means of basis pursuit (13). The vector $\hat{\mathbf{r}}$ contains the reconstructed samples of the subsampled RS estimator $\hat{R}_{X,\text{sub}}[p,q]$, which will be denoted as $\tilde{R}_{X,\text{sub}}[p,q]$.
4. Apply the interpolation (11) to $\tilde{R}_{X,\text{sub}}[p,q]$, i.e., calculate

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

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 $\varepsilon_{\text{CS}} \triangleq \mathrm{E}\big\{\big\|\hat{R}_{X,\text{CS}} - \bar{R}_X\big\|_2^2\big\}$ 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 $\varepsilon_{\text{MVU}} = \mathrm{E}\big\{\big\|\hat{R}_{X,\text{MVU}} - \bar{R}_X\big\|_2^2\big\}$ of the basic (noncompressive) estimator $\hat{R}_{X,\text{MVU}}[n,k]$ in (9) is bounded as

$$\varepsilon_{\text{MVU}} \,\le\, \|\bar{R}_X\|_2^2 \left(m_X^{(\mathcal{I}_{\bar{\mathcal{A}}})} + \frac{|\mathcal{A}|}{N}\right). \tag{16}$$

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

An additional error component is due to the S compression-recovery stage and corresponds to the MSE component $\Delta\varepsilon \triangleq \mathrm{E}\big\{\big\|\hat{R}_{X,\text{CS}} - \hat{R}_{X,\text{MVU}}\big\|_2^2\big\}$. As before, let S be the effective sparsity of the subsampled RS $\bar{R}_{X,\text{sub}}[p,q]$, i.e., the number of significant samples of $\bar{R}_{X,\text{sub}}[p,q]$ (we note, however, that the bound presented below holds for arbitrary $S \in \{1,\dots,\Delta m\Delta l\}$). Let us index the subsampled TF locations (p,q) in the order of decreasing RS magnitudes $\big|\bar{R}_{X,\text{sub}}[p,q]\big|$, such that (p_r,q_r) corresponds to the rth largest RS magnitude, $\big|\bar{R}_{X,\text{sub}}[p_r,q_r]\big|$. 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

$$\Delta\varepsilon \,\le\, D^2 \|\bar{R}_X\|_2^2 \,\frac{\Delta m\Delta l - S}{S}\,\sigma_X(S)\,, \tag{17}$$

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

$$\sigma_X(S) \,\triangleq\, \frac{\Delta n \Delta k}{\|\bar{R}_X\|_2^2} \sum_{r=S+1}^{\Delta m\Delta l} P_r\,, \tag{18}$$

where $P_r \triangleq \mathrm{E}\big\{\big|\hat{R}_{X,\text{sub}}[p_r,q_r]\big|^2\big\} = \mathrm{E}\big\{\big|\hat{R}_{X,\text{MVU}}[p_r\Delta n, q_r\Delta k]\big|^2\big\}$. 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

$$P_r \,\approx\, \frac{1}{N^2}\bigg|\sum_n\sum_k \psi_{\text{MVU}}[p_r\Delta n - n, q_r\Delta k - k]\,\bar{R}_X[n,k]\bigg|^2 + \frac{1}{N}\sum_n\sum_k \big|\psi_{\text{MVU}}[p_r\Delta n - n, q_r\Delta k - k]\big|^2 \big|\bar{R}_X[n,k]\big|^2.$$

This expression involves the RS $\bar{R}_X[n,k]$, and it shows that P_r is small if $\bar{R}_X[n,k]$ is small within a neighborhood of $(p_r\Delta n, q_r\Delta k)$ or, said differently, if $(p_r\Delta n, q_r\Delta k)$ is located outside a broadened version of the effective support of $\bar{R}_X[n,k]$ (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 $S\Delta n\Delta k$ is approximately equal to the area of the broadened effective support of $\bar{R}_X[n,k]$.

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 ε_{CS} of the compressive estimator $\hat{R}_{X,\text{CS}}[n,k]$:

$$\varepsilon_{\text{CS}} \,\le\, \big(\sqrt{\varepsilon_{\text{MVU}}} + \sqrt{\Delta\varepsilon}\big)^2.$$