Exploiting the Rules of the TF-MUSIC and Spatial Smoothing to Enhance the DOA Estimation for Coherent and Non-stationary Sources

Ruslan Zhagypar, Kalamkas Zhagyparova, and Muhammad Tahir Akhtar
School of Engineering and Digital Sciences,
Nazarbayev University, Nur-Sultan, Kazakhstan.
E-mails: {ruslan.zhagypar, kalamkas.zhagyparova, muhammad.akhtar}@nu.edu.kz, akhtar@ieee.org

***Abstract*—The paper introduces a combination of several techniques and methods to tackle the problem of Direction-of-Arrival (DOA) estimation for coherent and non-stationary signals under severe noise conditions. Until now, these properties of the signals were studied by many researchers, yet the solutions were developed separately. At the first stage of the algorithm, the Spatial Time-Frequency Distribution (STFD) matrix is derived, which accounts for non-stationarity of the signals and allows denoising. The forward-backward spatial smoothing is applied to the STFD matrix to solve the problem of signals coherency. The main principle of the Root MUltiple SIgnal Classification (Root-MUSIC), namely, solving for the roots of the polynomial to obtain the signals DOA is exploited. The algorithm was simulated with coherent chirp signals with different Signal-to-Noise Ratio (SNR) values to observe its efficiency compared to the conventional Root-MUSIC and Time-Frequency MUSIC (TF-MUSIC) methods.**

I. Introduction

The aim of sound source localization (SSL) is to automatically determine the positions of sound sources. There are two components of a source position that can be estimated as part of SSL: Direction-of-arrival (DOA) estimation and Distance estimation [1]. DOA is important in various fields, starting from human-robot interaction in robotics, progressing through rescue scenarios without visual contact and ending up with target tracking. The DOA estimation methods in real-life scenarios need to take into account that more than one sound source might be active in the environment. Therefore it is also necessary to estimate the position of multiple simultaneous sound sources.

Many approaches exist in the literature to address the DOA estimation problem. According to the classical literature review in this topic, these approaches can be broadly divided in four categories: time delay-based, beamforming-based, subspace-based and learning-based approaches [1]. This paper will discuss advancements in subspace-based DOA estimation techniques.

Major efforts have been devoted to solving the DOA estimation problem by array signal processing methods. Particularly, the high-resolution MUltiple SIgnal Classification (MUSIC) algorithm, which represents the most classical approach to spatial spectrum estimation, was initially developed in [2]. The method is based on the orthogonality property of signal and noise subspaces derived from performing the eigenvalue decomposition (EVD) on an input covariance matrix. The second most studied counterpart of the original algorithm is known as Root-MUSIC, which estimates the DOA by determining the roots of a polynomial formed from the noise eigenvectors [3]. The above-mentioned approaches imply using narrowband, non-coherent and stationary signal sources and low-level noise for adequate performance, however, these requirements are rarely met in real-life conditions. It is known that *coherent* signals, which tend to have similar frequency components, deteriorate the performances of the frequency estimating algorithms. However, [4][5] suggest applying the spatial smoothing technique to weaken the restriction of non-coherence. Besides, most of the signals used in practice, including human speech, are *non-stationary*, that is, they have variable statistical characteristics with time. The dependence of signal frequency on time significantly reduces the effectiveness of conventional methods. Therefore, the authors of [6][7] proposed a revolutionary approach that leverages the properties of spatial time-frequency (TF) distributions. Furthermore, the TF analysis represents the energy distribution of the signal across all TF points. It allows *denoising* by separating signal components from white noise.

The main focus of this paper is to advance the conventional Root-MUSIC for DOA estimation irrespective of the level of noise and independent of coherency and stationarity of the applied signals. In comparison with existing literature [4][5], the spatial time-frequency distribution (STFD) matrix is formed instead of the regular covariance matrix. According to [8], the former one provides better signal selectivity due to different TF signatures of source signals. Another advantage of STFD matrix is noise suppression as the power of noise is spread over the entire TF plane. In contrast to [6][7], the spatial smoothing technique was applied to coherent signals to restore the full-rank property of the STFD matrix. At last, DOA for multiple sources were estimated from the roots of a polynomial as opposed to spatial spectrum search in [4]-[7]. Hence, the method introduced in this paper expands the scope of above-mentioned methods by simultaneously handling both non-stationary and coherent signals.

The remaining of this paper is organized as follows. In Section II, preliminary information about TF concepts and the signal model are presented. The key features of the proposed method are discussed in details in Section III and the simulation results are reported in Section IV. Finally, Section V concludes the paper and defines the future study direction.

II. Preliminaries

A. TF Concepts

There exist different approaches for deriving the TF distribution (TFD) of a signal. Common methods include Short Time Fourier Transform (STFT), Spectrogram and Gabor Transform, Rihaczek Distribution and Wavelet Transform [9]. However, the core distribution that is connected to all other TF Distributions (TFDs) via a certain time-lag kernel is known as Wigner-Ville Distribution (WVD). Considering the original signal $s(t)$ received by a microphone, its analytic associate $z(t)$ is used in the TFD calculations. Thus, the WVD of the analytic signal, $W_{zz}(t,f)$, is formulated as [9]

$$W_{zz}(t,f) = \mathcal{F}_{\tau\to f}\{z(t+\frac{\tau}{2})\ z^*(t-\frac{\tau}{2})\}, \tag{1}$$

where t is the discrete-time given in time-samples, f is the discrete frequency defined with frequency bins, τ is the time lag, $\mathcal{F}_{\tau\to f}$ stands for the Discrete Fourier Transform (DFT) operator, and $*$ represents the conjugation. The usage of the analytic associate is motivated by an attempt to eliminate impractical artifacts on the TF plane caused by the interaction between positive and negative frequencies. Analytic signal contains energy of the signal only at positive frequencies in the frequency domain. One way of obtaining such a signal is by using the Hilbert transform. The product of inside of DFT in (1) is termed as instantaneous autocorrelation function (IAF) and denoted as $K_{zz}(t,\tau)$. Correspondingly, the distribution of $z_1(t)$ and $z_2(t)$ is described by the cross-WVD function:

$$W_{z_1z_2}(t,f) = \mathcal{F}_{\tau\to f}\{z_1(t+\frac{\tau}{2})\ z_2^*(t-\frac{\tau}{2})\}. \tag{2}$$

In the case when WVD is performed on a signal which contains more than one frequency components, there might be additional cross-terms on the TF plane. These terms reduce the usefulness of the TFD [9]. Therefore, a special time-lag kernel, $G(t,\tau)$, is applied to the IAF, making a smoothed version of WVD of $z_1(t)$ and $z_2(t)$ [10]:

$$D_{z_1z_2}(t,f) = \mathcal{F}\{G(t,\tau) \circledast K_{z_1z_2}\}, \tag{3}$$

where $K_{z_1z_2}$ is the IAF of the signals $z_1(t)$ and $z_2(t)$, and $\circledast$ represents the convolution. The time-lag kernel acts as a filter implemented by different window functions. The particular kernel function used in the paper is the Blackman-Harris window.

B. Signal Model

The model consists of an M-element Uniform Linear Array (ULA) that receives the signals from a P number of sources. To a certain extent, the assumption of a far-field environment may not be applied due to spatial smoothing, which is discussed later in the paper. The wavefront delay between adjacent sensors in ULA is calculated as [11]

$$\tau_p = \frac{d\sin(\theta_p)}{c}, \tag{4}$$

where c is the speed of light, θ_p is the DOA for the pth source, and d represents the spacing between array elements. It is recommended that the spacing is equal to half of the signal wavelength [12]. Assuming that there are N time samples, an instantaneous mixing model is given by

$$\boldsymbol{x}(t) = \boldsymbol{A}(\theta)\,\boldsymbol{s}(t) + \boldsymbol{n}(t), \quad t = 1, \ldots, N, \tag{5}$$

where $\boldsymbol{s}(t)$ is $P\times 1$ vector for the source signals, $\boldsymbol{n}(t)$ is $M\times 1$ vector for additive white Gaussian noise, and $\boldsymbol{A}(\theta)$ is $M\times P$ propagation matrix which contains the delay information of each signal source at every array element, and where M denotes the number of sensors.

III. Proposed Method

A. Array STFD matrix

STFD matrix represents auto- and cross-TFDs of signals from all array elements. Since the matrix considers the number of sensors, the third space dimension is introduced. The covariance matrix of the signals in (5) becomes *time-dependent* when the stationarity assumption is removed [10]:

$$\boldsymbol{R}_{xx}(t,\tau) = \boldsymbol{A}(\theta)\boldsymbol{R}_{ss}(t,\tau)\,\boldsymbol{A}(\theta)^H + \sigma_n^2\delta(\tau)\,\boldsymbol{I}, \tag{6}$$

where $\boldsymbol{R}_{xx}(t,\tau)$ is the $M\times M$ covariance matrix of the recorded signals at specific time instant and time-lag, $\boldsymbol{R}_{ss}(t,\tau)$ is the $P\times P$ covariance matrix of the source signals, σ_n^2 is the additive noise variance, $\delta(\tau)$ is the Dirac delta function, and $\boldsymbol{I}$ is the $M\times M$ identity matrix.

One may notice the similarity between the IAF and the covariance function, which is described as

$$\begin{aligned}\boldsymbol{R}_{xx}(t,\tau) &= \mathbb{E}\{\boldsymbol{K}_{xx}(t,\tau)\} \\ &= \mathbb{E}\left\{\boldsymbol{x}\left(t+\frac{\tau}{2}\right)\boldsymbol{x}^H\left(t+\frac{\tau}{2}\right)\right\},\end{aligned} \tag{7}$$

where $\mathbb{E}$ defines the expectation operator, and $\boldsymbol{K}_{xx}$ is the matrix where each entry represents the IAF of the corresponding signals.

Using (3) and (7), we rewrite the expression in (6) to derive the STFD matrix of the recorded signals:

$$\boldsymbol{D}_{xx}(t,f) = \boldsymbol{A}\boldsymbol{D}_{ss}(t,f)\,\boldsymbol{A}^H + \sigma_n^2\boldsymbol{I}, \tag{8}$$

where $\boldsymbol{D}_{ss}(t,f)$ is the STFD matrix of the source signals. It is worth noting that the diagonal and off-diagonal entries of the STFD matrix correspond to *auto-TFDs* and *cross-TFDs*, respectively. The structure of the matrix allows applying the same subspace methods that are usually done on covariance matrices.

B. Post-processing of STFD matrix

The first step in reducing the effects of cross terms to a notable degree is spatial averaging, which is formulated as

$$D_{\text{avg}}(t,f) = \frac{1}{M}\sum_{m=1}^{M} D_{mm}(t,f), \tag{9}$$

where $D_{\text{avg}}(t,f)$ represents the averaged TFD, and $D_{mm}(t,f)$ is the TFD of the mth sensor signals. This effect is explained by averaging the auto-terms located on the same spots on the TF planes while the cross-terms are allocated in different spots.

The usage of STFD matrix permits to denoise the TF representation of the signals. It is implemented by setting a threshold to the TF points' energies of $D_{\text{avg}}(t, f)$. The idea is to select those points with energies higher than user-defined threshold, ε, and reject the ones with low energies that may come from noise [10]. In this study, the threshold is empirically chosen to be 4% of the energy of the maximum TF point:

$$\varepsilon = 0.04 \times \max(D_{avg}(t, f)). \tag{10}$$

The final step for obtaining the averaged STFD matrix, which would be suitable for the DOA estimation, is determining the averages of the chosen TF points after setting the threshold:

$$\boldsymbol{D}_{xx} = \frac{1}{n_{\text{points}}} \sum_{i=1}^{n_{\text{points}}} \boldsymbol{D}_{xx}(t_i, f_i), \tag{11}$$

n_{points} in the expression refers to the total number of TF points. It is during this stage the matrix entries take single values instead of TF matrices.

C. Spatial Smoothing for Coherent Signals

The spatial smoothing technique has been developed for stationary signals under uniform white noise [4]. Applied to a covariance matrix, its performance degrades when at least one of the conditions is not met. However, in this study, the case with non-stationary signals is already handled by the developed STFD matrix. The basic idea of the MUSIC algorithm requires the covariance matrix, or the STFD matrix in this paper, to be full rank to keep the noise subspace orthogonal to the signal subspace. The full-rank assumption becomes no longer valid provided that two or more signals are coherent [5]. Thus, the aim of the spatial smoothing is to diagonalize a matrix and decorrelate the signal sources [13].

The algorithm divides the ULA into L overlapping subarrays of size K, which lies in the range $[P+1; M]$. There are two ways of creating the subarrays depending on whether the sensors are grouped in a forward or a backward direction. In the forward smoothing, the received signal vector similar to (5) is created at l-th subarray as

$$\boldsymbol{x}_l(t) = [x_l(t), x_{l+1}(t), \ldots, x_{l+K-1}(t)]^T = \boldsymbol{U}_l \boldsymbol{x}(t), \tag{12}$$

where $\boldsymbol{U}_l = [\boldsymbol{0}_{K\times(l-1)} \; \boldsymbol{I}_K \; \boldsymbol{0}_{K\times(M-K-l+1)}]$ is the selection matrix which is used to select the sensors from l to $l+K-1$ when multiplied to complete array model [4][14]. Correspondingly, separate STFD matrices can be found for each subarray. Hence, the total STFD matrix of the forward smoothing, $\boldsymbol{D}^f$, is derived by averaging all STFD matrices of L subarrays:

$$\boldsymbol{D}^f = \frac{1}{L} \sum_{l=1}^{L} \boldsymbol{D}_l^f = \frac{1}{L} \sum_{l=1}^{L} \boldsymbol{U}_l \boldsymbol{D}_{xx} \boldsymbol{U}_l^H, \tag{13}$$

where $\boldsymbol{D}_l^f$ is the STFD matrix of lth subarray in forward smoothing.

In backward smoothing, since the sensors are selected in reverse order, the l-th subarray includes sensors numbered from $(l+K-1)$ to l. Therefore, the output vector of the subarray, $\boldsymbol{y}_l(t)$, is expressed as

$$\boldsymbol{y}_l(t) = \boldsymbol{U}_l \boldsymbol{J} \boldsymbol{x}^*(t), \tag{14}$$

where $\boldsymbol{J}$ is an $M \times M$ exchange matrix with ones on the anti-diagonal entries and zeros on others. Thus, the counterpart of (13) for backward smoothing is defined as

$$\boldsymbol{D}^b = \frac{1}{L} \sum_{l=1}^{L} \boldsymbol{D}_l^b = \frac{1}{L} \sum_{l=1}^{L} \boldsymbol{U}_l \boldsymbol{J} \boldsymbol{D}_{xx}^* \boldsymbol{J} \boldsymbol{U}_l^H, \tag{15}$$

where $\boldsymbol{D}_l^b$ is the STFD matrix of lth subarray in backward smoothing, and $\boldsymbol{D}_{xx}^*$ is the conjugate of the non-smoothed STFD matrix in (8).

Finally, the spatially smoothed STFD matrix is the average of (13) and (15):

$$\boldsymbol{D}^{fb} = \frac{\boldsymbol{D}^f + \boldsymbol{D}^b}{2}. \tag{16}$$

The smoothed STFD matrix would reach full rank when the time of smoothing L is equal to the half of the number of coherent sources. At this point, all the eigenvalues become non-zero, which confirms the decorrelation ability of the spatial smoothing technique [5].

D. Root-MUSIC algorithm for DOA estimation

The working principle of the Root-MUSIC algorithm is based on orthogonality of source and noise subspaces [3]. After performing the EVD on the spatially smoothed STFD matrix, the eigenvalues are sorted in descending manner. Hence, considering $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$ are the eigenvalues for the spatially smoothed STFD matrix $\boldsymbol{D}^{fb}$, and $v_1 \geq v_2 \geq \ldots \geq v_P$ are eigenvalues for the signal part only, the following expression is derived as

$$\lambda_i = \begin{cases} v_i + \sigma_n^2, & i = 1, 2, \ldots, P \\ \sigma_n^2, & i = P+1, \ldots, M \end{cases}. \tag{17}$$

As (17) suggests, the first P eigenvectors of all $\boldsymbol{q}_1$, $\boldsymbol{q}_2$, $\ldots, \boldsymbol{q}_M$ span the signal as well as noise subspace and are denoted as $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\ldots, \boldsymbol{e}_P$, while the remaining $M-P$ eigenvectors span the noise subspace only. The subspace-based methods such as Root-MUSIC assume a Hermitian matrix, which has *orthogonal* eigenvectors for distinct eigenvalues

$${\boldsymbol{e}_k}^H \boldsymbol{q}_i = 0, \quad i = P+1, \; \ldots, \; M \text{ and } k = 1, \; \ldots, \; P. \tag{18}$$

In the spectral MUSIC algorithm, the DOA are estimated by the power spectrum function, which results in peaks when the source angles are found. For that, the function utilizes the expression (18) in the denominator and is given as

$$\begin{aligned} P_{\text{MUSIC}}(\theta) &= \frac{1}{\sum_{i=P+1}^{M} |\boldsymbol{e}^H \boldsymbol{q}_i|^2} \\ &= \frac{1}{\sum_{i=P+1}^{M} \boldsymbol{e}^H \boldsymbol{q}_i {\boldsymbol{q}_i}^H \boldsymbol{e}} = \frac{1}{\boldsymbol{e}^H \boldsymbol{Q} \boldsymbol{Q}^H \boldsymbol{e}}, \end{aligned} \tag{19}$$

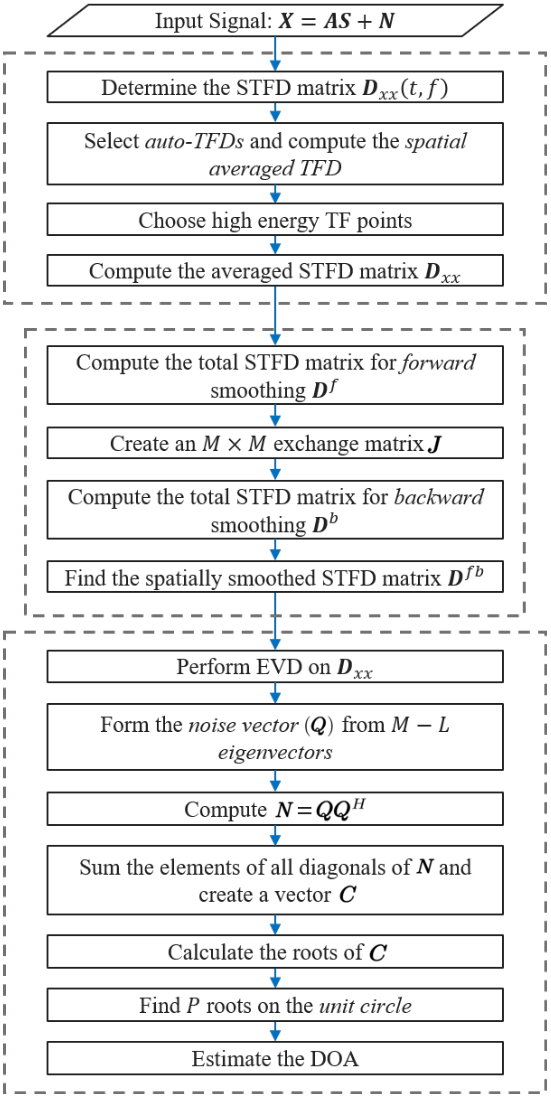

Fig. 1. The flowchart of the proposed method.

where $\boldsymbol{Q}$ denotes the $M \times 1$ noise vector. In fact, the expression (18) can be viewed as the $\mathcal{Z}$-transform of $\boldsymbol{q}_i$

$$Q_i(z)|_{z=e^{jw_1}} = \sum_{n=0}^{M-1} q_i\,(n)\,z^{-n} = \boldsymbol{e}_1{}^H \boldsymbol{q}_i = 0. \qquad (20)$$

Therefore, the Root-MUSIC considers the denominator of (19) as a polynomial in $\mathcal{Z}$-domain

$$F(z) = \boldsymbol{e}^H(z)\boldsymbol{Q}\boldsymbol{Q}^H\boldsymbol{e}(z) = \boldsymbol{e}^H(z)\boldsymbol{N}\boldsymbol{e}(z), \qquad (21)$$

where $\boldsymbol{e}^H = [1 \quad z^{-1} \quad z^{-2} \quad \ldots \quad z^{-(M-1)}]$, and $\boldsymbol{N}$ is the $(M-1) \times (M-1)$ matrix containing the information about the coefficients of the polynomial. As a matter of fact, a special vector $\boldsymbol{C}$ containing the coefficients is created by summing the elements in each diagonal of the matrix.

Ideally, there will be P number of roots of the polynomial, lying on the unit circle, that represent the DOA for the incoming signals [15]. However, the roots might deviate from the unit circle due to impact of noise. The actual angles of the sources are determined for each root as

$$\theta_k = \arcsin[\frac{\lambda}{2\pi d}\arg(z_k)], \quad k = 1, \ \ldots, \ P, \qquad (22)$$

where $\arg(z_k)$ is the argument of the kth root on the unit circle, and θ_k is the corresponding DOA for kth source.

The proposed method consists of three main components: the STFD matrix, the spatial smoothing technique, and the Root-MUSIC algorithm. The flowchart of the proposed method is represented in Fig. 1.

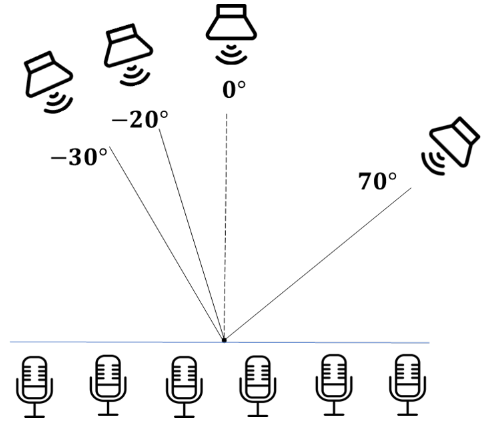

Fig. 2. A pictorial representation of the simulation set-up.

IV. Simulations

The simulation set-up diagram is illustrated in Fig. 2. As shown, we consider a ULA with $M = 6$ to evaluate and compare the performance of the proposed method with the Root-MUSIC [3] and the TF-MUSIC [7] algorithms. The spacing between the array elements are taken as half of the signal wavelength. There are $P = 4$ source signals that impinge on the sensor array from different directions $-30°$, $-20°$, $0°$, $70°$, where the minus sign represents the source lying to the left of the reference point.

A. Experiment 1: Coherent sources

In many applications, there is the effect of signal reverberation when the signals get reflected from various surfaces, which results in the duplicates of the original signal. Hence, to simulate the coherent sources, the model assumes two identical signals arriving at the array from $-30°$ and $0°$ directions. As only two sources are considered coherent, the number of spatial smoothing is equal to half of it, i.e. $L = 1$. The source are set to be stationary with constant angular frequencies $[\frac{2\pi}{5}, \frac{4\pi}{5}, \frac{2\pi}{5}, \frac{3\pi}{5}]$, respectively. The SNR value is chosen to be 20 dB. The results of the TF-MUSIC are originally given in the form of spectral plots, e.g. Fig. 3 shows the spectral plot for the first experiment. Thus, the peak values in TF-MUSIC in following experiments are also read from the similar plots. The experiment results are illustrated in Table I.

TABLE I
RESULTS OF EXPERIMENT 1

	DOA Angle			
	$\theta_1 = -30°$	$\theta_2 = -20°$	$\theta_3 = 0°$	$\theta_4 = 70°$
Root-MUSIC	-6.34	-19.98	23.55	69.98
TF-MUSIC	-	-20	-2	70
Proposed TF-Root-MUSIC	-30.01	-19.83	-0.01	69.96

TABLE II
RESULTS OF EXPERIMENT 2

	DOA Angle			
	$\theta_1 = -30°$	$\theta_2 = -20°$	$\theta_3 = 0°$	$\theta_4 = 70°$
Root-MUSIC	-49.9	-8.85	17.66	38.57
TF MUSIC	-	-20	0.5	70
Proposed TF-Root-MUSIC	-30.01	-19.83	-0.01	69.96

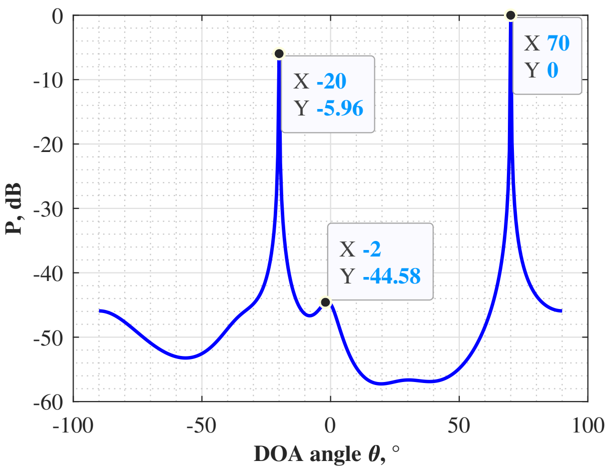

Fig. 3. A Spectral plot of TF-MUSIC for the Experiment 1.

According to Table I, the Root-MUSIC and TF-MUSIC algorithms demonstrate valid results only for non-coherent sources located at $-20°$ and $70°$. However, neither methods can correctly determine the DOA for the coherent sources. In fact, the Fig. 3 depicts how the TF-MUSIC spectral plot cannot spot the peak of the signal arriving from $-30°$. While the proposed TF-Root-MUSIC algorithm outputted the DOA for all sources accurately because of the applied spatial smoothing technique.

B. Experiment 2: Non-stationary sources

The experiment considers non-coherent and non-stationary signals. The Linear Frequency Modulated (LFM) chirp signals are used to represent the sources since they are the most common examples of non-stationary signals. The angular frequencies of the signals are modulated starting from $[\frac{2\pi}{5}, \frac{4\pi}{5}, \frac{\pi}{5}, \frac{3\pi}{5}]$ to $[\frac{4\pi}{5}, \frac{2\pi}{5}, \frac{3\pi}{5}, \pi]$. The Blackman-Harris window is chosen as the time-lag kernel for smoothing the TFD, which has 512 time samples and 512 frequency bins. The SNR value is equal to 20 dB. The results are given in Table II.

As expected, the Table II shows that the TF-MUSIC and the proposed TF-Root-MUSIC methods accurately estimate the DOA for all sources. The results for the Root-MUSIC make it explicit that the algorithm is not capable of determining the DOA for non-stationary sources.

TABLE III
RESULTS OF EXPERIMENT 3 FOR THE ROOT-MUSIC

	DOA Angle			
SNR (dB)	$\theta_1 = -30°$	$\theta_2 = -20°$	$\theta_3 = 0°$	$\theta_4 = 70°$
10	-49.90	-8.87	17.60	38.61
0	-54.46	-0.10	21.26	46.75
-5	-49.64	-9.08	17.79	38.76
-10	-50.00	-9.24	18.14	38.56
-15	-50.16	-7.31	19.07	37.43
-20	-50.58	-8.16	12.91	30.75

C. Experiment 3: Coherent and Non-stationary sources at different SNR values

The experiment is organized to analyze the performance of each algorithm when the coherent and non-stationary sources have different SNR values. The angular frequencies are linearly varied from $[\frac{2\pi}{5}, \frac{\pi}{5}, \frac{2\pi}{5}, \frac{3\pi}{5}]$ to $[\frac{4\pi}{5}, \frac{3\pi}{5}, \frac{4\pi}{5}, \pi]$. It should be noted that the signals arriving from $-30°$ and $0°$ are coherent. The performance of each algorithm is examined under different levels of noise (SNR). The results are given separately for each method in Tables III, IV, and V.

The Table III shows that although the TF-MUSIC can handle the non-stationary signals, it has poor performance at retrieving the DOA for the coherent signals due to the lack of spatial smoothing. This is the reason for having no values in the first column of the TF-MUSIC. Whereas the Table IV illustrates that the Root-MUSIC fails at determining the DOA mainly because of nonstationarity of the signals. It can be seen from the Table V that the proposed method demonstrates accurate performance for both non-stationary and coherent signals even at low SNR values.

V. CONCLUSION

An efficient method for DOA estimation of non-stationary and coherent sources under adverse noise conditions has been developed in the paper. The proposed approach uses a combination of spatial TF analysis, spatial smoothing technique, and the direction estimating principal of the Root-MUSIC. In addition, once the STFD matrix is calculated, the TF points with higher energies than the threshold are selected for denoising. The forward-backward spatial smoothing is used to decorrelate the sources and restore the rank of the

TABLE IV
RESULTS OF EXPERIMENT 3 FOR THE TF-MUSIC

SNR (dB)	DOA Angle $\theta_1 = -30°$	$\theta_2 = -20°$	$\theta_3 = 0°$	$\theta_4 = 70°$
10	-	-25.5	0	70
0	-	25	0	70.5
-5	-	-25	0	70
-10	-	-24	0	69.5
-15	-	-24.5	0.5	66
-20	-	-25.5	-2.5	65.5

TABLE V
RESULTS OF EXPERIMENT 3 FOR THE PROPOSED TF-ROOT-MUSIC

SNR (dB)	DOA Angle $\theta_1 = -30°$	$\theta_2 = -20°$	$\theta_3 = 0°$	$\theta_4 = 70°$
10	-30.25	-20.28	0.06	69.94
0	-29.40	-20.22	-0.04	70.19
-5	-29.23	-19.93	-0.34	72.19
-10	-29.64	-17.63	-0.16	73.17
-15	-29.28	-19.67	1.19	66.77
-20	-29.92	-19.92	-3.03	65.06

STFD matrix. Ultimately, the derived matrix is processed by the conventional time-domain Root-MUSIC to extract the DOA information. Simulation results illustrate that the proposed method outperforms the classical Root-MUSIC and TF-MUSIC methods under various conditions.

One of the main limitations is that the algorithm cannot be applied when the number of sources is greater than the sensors $(P \geq M)$. It can be improved by using the Blind Source Separation (BSS) as a preprocessing step as it was suggested in [16]. It allows reducing the problem down to separate single sources, the DOA for which later can be estimated individually. Thus, due to the outstanding capabilities of the BSS, it is expected to expand the application scenarios of the proposed approach

ACKNOWLEDGMENT

This work was partially supported by the Faculty Development Competitive Research Grants Program of Nazarbayev University under Grant Number 110119FD4525.

REFERENCES

[1] C. Rascon and I. Meza, “Localization of sound sources in robotics: A review,” *Robotics and Autonomous Systems*, vol. 96, pp. 184-210, 2017.

[2] R. O. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Trans. Antennas Propagation*, vol. 36, no. 4, pp. 532-544, 1988.

[3] B. D. Rao and K. V. S. Hari, “Performance analysis of Root-Music,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1939-1949, Dec. 1989.

[4] J. Wen, B. Liao and C. Guo, “Spatial smoothing based methods for direction-of-arrival estimation of coherent signals in nonuniform noise,” *Digital Signal Process.*, vol. 67, pp. 116-122, 2017.

[5] Q. Chen and L. Ruolun, “On the explanation of spatial smoothing in MUSIC algorithm for coherent sources,” in *Proc. IEEE Int. Conf. Information Sci. Tech.*, 2011, pp. 699-702.

[6] M. G. Amin and Y. Zhang, “Direction Finding Based on Spatial Time-Frequency Distribution Matrices,” *Digital Signal Process.*, vol. 10, pp. 325–339, 2000.

[7] A. Belouchrani and M. G. Amin, “Time-frequency MUSIC,” *IEEE Signal Process. Lett.*, vol. 6, no. 5, pp. 109-110, 1999.

[8] A. Belouchrani, M. G. Amin, N. Thirion-Moreau, and Y. D. Zhang, “Source Separation and Localization Using Time-Frequency Distributions: An Overview,” *IEEE Signal Process. Magazine*, vol. 30, no. 6, pp. 97–107, 2013.

[9] B. Boashash, *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, 2nd ed. Massachusetts: Academic Press, 2016, pp. 331-518.

[10] B. Boashash and A. Aïssa-El-Bey, “Multisensor Time-Frequency Signal Processing: A tutorial review with illustrations of performance enhancement in selected application areas,” *Digital Signal Process*, vol. 77, pp. 153-186, June 2018.

[11] W. Coventry, C. Clemente, and J. Soraghan, “Polynomial Root-MUSIC Algorithm for Efficient Broadband Direction of Arrival Estimation,” in *Proc. 2017 Sensor Signal Process. for Defence Conference (SSPD)*, London, pp. 1-5, 2017.

[12] H. Tang, “DOA estimation based on MUSIC algorithm,” M.S. thesis, Kalmar, May 2015.

[13] J. Pan, M. Sun, Y. Wang and X. Zhang, “An Enhanced Spatial Smoothing Technique With ESPRIT Algorithm for Direction of Arrival Estimation in Coherent Scenarios,” *IEEE Trans. Signal Process.*, vol. 68, pp. 3635-3643, 2020.

[14] K. V. S. Hari and U. Gummadavelli, “Effect of spatial smoothing on the performance of subspace methods in the presence of array model errors,” *Automatica*, vol. 30, no. 1, pp. 11–26, 1994.

[15] H. K. Hwang, Z. Aliyazicioglu, M. Grice, A. Yakovlev, and P. Lu. “Direction of arrival estimation using polynomial root intersection for multidimensional estimation (prime),” in *Proc. IMECS Conference*, 2008.

[16] N.A. Khan, S. Ali, M. Mohammadi, and M. Haneef, “Direction of arrival estimation of sources with intersecting signature in time–frequency domain using a combination of IF estimation and MUSIC algorithm,” *Multidimensional Systems and Signal Process.*, vol. 31, pp. 549–567, 2020.