Optimal Design of Orders of Discrete Fractional Fourier Transforms for Sparse Representations

Xiao-Zhi Zhang

Telephone: +86 20 3932 2246 Fax: +86 20 3932 2252 Email: zxz_usc@163.com
Engineering Research Center, University of South China, Hengyang, 421001, China.
School of Information Engineering, Guangdong University of Technology, Guangzhou, 510006, China;

*Bingo Wing-Kuen Ling

Telephone: +86 20 39322258 Fax: +86 20 39322232 Email: yongquanling@gdut.edu.cn
School of Information Engineering, Guangdong University of Technology, Guangzhou, 510006, China.

Ran Tao

Telephone: +86 10 68914257 Fax: +86 10 68914257 E-mail: rantao@bit.edu.cn
Department of Electronic Engineering, Beijing Institute of Technology, Beijing,100081, China

Zhi-Jing Yang

Telephone: + 86 13929558558 Fax: +86 20 3932 2252 Email: yzhj@gdut.edu.cn
School of Information Engineering, Guangdong University of Technology, Guangzhou, 510006, China.

Wai-Lok Woo

Telephone: +44 (0)19 1222 6682 Fax: +44 (0)19 1222 8180 Email: lok.woo@ncl.ac.uk
School of Electrical and Electronic Engineering, Newcastle University, Newcastle, NE1 7RU, United Kingdom.

Saeid Sanei

Telephone: +44 (0)11 5848 3453 Email: saeid.sanei@ntu.ac.uk
School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS, United Kingdom.

Kok-Lay Teo

Telephone: +618 9266 1115 Fax: +618 9266 3197 Email: K.L.Teo@curtin.edu.au
Department of Mathematics and Statistics, Curtin University, Perth, Western Australia, 6002, Australia.

Abstract

This paper proposes an optimal design of the orders of the discrete fractional Fourier transforms (DFrFTs) and construct an overcomplete transform using the DFrFTs with these orders for performing the sparse representations. The design problem is formulated as an optimization problem with an L_1 norm nonconvex objective function. To avoid all the orders of the DFrFTs to be the same, the exclusive OR of two constraints are imposed. The constrained optimization problem is further reformulated to an optimal frequency sampling problem. A method based on solving the roots of a set of harmonic functions is employed for finding the optimal sampling frequencies. As the designed overcomplete transform can exploit the physical meanings of the signals in terms of representing the signals as the sums of the components in the time frequency plane, the designed overcomplete transform can be applied to many applications.

Keywords — Discrete fractional Fourier transform, overcomplete transform, sparse representation, nonsmooth and nonconvex optimization, frequency sampling, harmonic functions.

1.   Introduction

Fractional Fourier transform (FrFT) maps the signals represented in the time domain to the signals represented in a domain corresponding to a line in the time frequency plane [1]. This line is obtained by rotating the x-axis of the time frequency plane by a certain angle. When the rotational angle is an integer multiple of π plus or minus half π, then this line is the frequency axis of the time frequency plane and the signals are represented in the frequency domain. When the rotational angle is an integer multiple of π, then this line is the time axis of the time frequency plane and the signals are represented in the time domain. In general, the

rotational angle could be any real number. Hence, this line could be inclined between the time axis and the frequency axis of the time frequency plane and the signals are represented in a domain which is partly a time domain and partly a frequency domain. Therefore, the FrFT can be understood as the generalization of both the Fourier transform (FT) and the time domain linear operations. As a result, the performances of all the existing applications based on either the FT or the time domain linear operations can be improved or maintained if the FT or the time domain linear operations is replaced by the appropriate FrFT. Because of this reason, many signal processing applications based on the representations of the signals in the DFrFT domains such as the filtering of signals [2]-[5], the sampling and the reconstructions of signals [6]-[8], the watermarking as well as the encryptions and the decryptions of images [9]-[10], the compressions of magnetic resonance images (MRIs) [11]-[13], etc, have been developed.

On the other hand, recent developed techniques on the sparse signal representations have been applied to many signal processing applications. It is well known that if the energies of the signals are only localized in few components, then the signals can be sparsely and efficiently represented [14]. For the compression applications, since only few coefficients are used for the representations of the signals, the coding gains are very high. For the denoising applications, as the signals are localized in few coefficients, the total numbers of coefficients corrupted by the noises are small. Hence, the signal to noise ratios are also very high. Therefore, the sparse representation plays an important role in the signal processing community. However, in general signals being sparse in the time domain are not sparse in the frequency domain and vice versa [14]. Also, many signals are neither sparse in the time domain nor sparse in the frequency domain, but they are sparse in the domains corresponding to some lines in the time frequency plane. In this case, the signals can be sparsely represented via performing the DFrFTs [15]-[17]. However, the optimal orders of the DFrFTs are unknown. That is, the optimal angles required to be rotated on the x-axis of the time frequency plane are unknown. If the DFrFTs are chosen with the inappropriate orders, then the transformed signals may not be sparse. In this case, the efficiencies of the representations of the signals could be very poor. For the compression applications, the coding gains will be very low. For the denoising applications, the signal to noise ratios will be also very low. Because of this reason, this motivates us to find the optimal orders of the DFrFTs so that the signals can be the sparsely and efficiently represented. In this case, the performances of the existing applications such as the compression applications and the denoising applications [11]-[13], [18] can be further improved.

In recent years, the sparse representations of signals based on the overcomplete transforms have attracted a lot of attentions. An overcomplete transform is to map the low dimensional signals in the time domain to the high dimensional signals in the transformed domain [19]. Among them, the overcomplete wavelets are the common overcomplete transforms used in the signal processing community. However, as the overcomplete wavelets are not memoryless, so the averaging effects will be introduced. Besides, there are some existing memoryless overcomplete transforms such as those based on the KSVD dictionary [20]. These overcomplete transforms are obtained via performing the training on the data sets. However, the vectors of these overcomplete transform matrices do not associate with any physical meaning. Therefore, it is hard to persuade the experts in other fields to apply these overcomplete transforms in their applications. On the other hand, the DFrFTs can represent the signals as the sums of the components in the time frequency plane. In particular, if the frequencies of the signals are linearly related to their times, then only few components in the time frequency plane can be used to represent the signals and the signals are sparsely represented. These components are in the lines of the time frequency plane. The direct proportional constants between the times and the frequencies of the signals are related to the slopes of the lines where these lines can be obtained by rotating the x-axis of the time frequency plane. Therefore, the sparse components can be interpreted as the signal components where their frequencies are linearly related to their times. It is worth noting that different orders of the DFrFTs correspond to different direct proportional constants relating

the frequencies and the times of the vectors of DFrFT matrices [1]. If different DFrFTs with different orders are employed for the representations, then a boarder class of signals can be represented. These physical interpretations of the signal representations help the experts in other fields to understand how the frequencies of the signals are related to their times. As a result, they can apply the DFrFTs to solve their problems in their fields. Because of these properties, this motivates us to design an overcomplete transform using different DFrFTs with different orders.

This paper proposes an optimal design of an overcomplete transform consisting of different DFrFTs with different orders for the sparse representations of the signals. The design of the orders of the DFrFTs was first proposed in [17]. Since the elements in the overcomplete transform matrix are expressed in term of the highly nonlinear polynomials of the trigonometric functions of the orders of the DFrFTs, the design problem is highly nonlinear and nonconvex. In general, it is very challenging to find the globally optimal solutions of the nonconvex optimization problems [21]. Moreover, the L_1 norm operator is usually used for formulating the sparse criteria. As the L_1 norm operator is not differentiable, the design problems are actually the nonsmooth optimization problems [22]. In this case, the conventional gradient descent approaches cannot be directly applied for finding the solutions of the nonsmooth optimization problems. Furthermore, since the transformed vectors can be broken down into the sub-vectors, the optimal transformed vectors can be obtained by finding the individual optimal sub-vectors. It is worth noting that each sub-vector is obtained by computing the matrix multiplication of the corresponding DFrFT matrix to the corresponding input vector. As the same input vector is multiplied to all the DFrFT matrices, all these individual optimal sub-vectors as well as all the optimal DFrFT matrices will be the same. To understand this phenomenon more, as the orders of the DFrFTs are in a continuous set, the feasible set of the optimization problem is continuous. Since the objective function is also continuous, the optimal orders of the DFrFTs will be localized within a small neighborhood even though they are different. However, this is not meaningful for practical applications. Therefore, this situation should be avoided. To address this issue, it is required to impose certain constraints to the optimization problem such that these optimal orders of DFrFTs are different and the absolute differences among them should be large enough to separate from each others. Therefore, this paper proposes to impose the parameter free constraints to the optimization problem to avoid all the orders of the DFrFTs to be the same and guarantee that they are separated from each others. In particular, the constraints that the objective function being either stationary (Here, the stationary points are the points where the gradients of the objective function evaluated at these points are equal to zero.) or nondifferentiable at the optimal orders of the DFrFTs are imposed to the optimization problem. In this case, the feasible set of the optimization problem is discrete. However, this results to the optimization problem subject to the XOR of two constraints. Nevertheless, this is not a conventional optimization problem. The existing optimization algorithms cannot be directly applied for finding the globally solution of this optimization problem. Therefore, it is required to propose a new method for finding the solution of this optimization problem.

To handle the above difficulties, the optimal design of the orders of the DFrFTs is reformulated as an optimal frequency sampling problem. The solution of this optimal frequency sampling problem is found via finding the roots of a set of harmonic functions. The outline of the rest of this paper is as follows. The DFrFT is reviewed in Section 2. The formulation of the design problem is presented in Section 3. A new method for finding the globally optimal solution of the optimization problem is shown in Section 4. An application example is illustrated in Section 5. Finally, a conclusion is drawn in Section 6.

2.   Review on DFrFT

Suppose that there are K discrete time finite length training signals and assume that all these training signals have the same lengths. Let N be the lengths of these training signals and $x_k(n)$ for $n = 0, \cdots, N-1$ and for $k = 0, \cdots, K-1$ be the impulse responses of these

training signals. In this paper, these training signals expressed in both the time domain and the DFrFT domains are represented in the vector form. Let

$$\mathbf{x}_k \equiv \left[x_k(0), \quad \cdots, \quad x_k(N-1)\right]^T \quad \text{for} \quad k = 0, \cdots, K-1 \tag{1}$$

be the vector form of these training signals. Here, the superscript "T" denotes the transposition operator. Denote $\mathfrak{R}^{a \times b}$ and $C^{a \times b}$ as the set of $a \times b$ real valued matrices and the set of $a \times b$ complex valued matrices, respectively. Let $\mathbf{F}_\alpha \in C^{N \times N}$ be a DFrFT matrix with the order $\alpha \in [-\pi, \pi)$.

There are many different definitions of DFrFTs [23], [24] and employing different definitions of DFrFTs will result to different transformed signals. However, it is found that the differences among the transformed signals based on different definitions of the DFrFTs are very small. It is worth noting that the DFrFT matrices can be expressed in the following form for all the definitions of the DFrFTs:

$$\mathbf{F}_\alpha = \mathbf{E} diag\left(e^{-j\alpha\boldsymbol{\beta}}\right)\mathbf{E}^T. \tag{2}$$

However, different definitions of the DFrFTs correspond to different $\mathbf{E}$. Here, $\mathbf{E}$ are the real valued unitary matrices which are independent of α. That is:

$$\mathbf{E}\mathbf{E}^T = \mathbf{E}^T\mathbf{E} = \mathbf{I}_N, \tag{3}$$

where $\mathbf{I}_N$ denotes the $N \times N$ identity matrix, $diag(\mathbf{z})$ denotes the diagonal matrix with its diagonal elements being equal to the elements of $\mathbf{z}$ and $e^{j\alpha\boldsymbol{\beta}}$ is denoted as

$$e^{j\alpha\boldsymbol{\beta}} \equiv \begin{bmatrix} e^{j\alpha\beta_0} \\ \vdots \\ e^{j\alpha\beta_{N-1}} \end{bmatrix}, \tag{4}$$

in which $\boldsymbol{\beta} = \left[\beta_0, \cdots, \beta_{N-1}\right]^T = \left[0, \cdots, N-1\right]^T$ and β_n for $n = 0, \cdots, N-1$ are also independent of α.

The definition of the DFrFT presented in this paper is based on that shown in [23]. Let $\mathbf{u}_n$ for $n = 1, \cdots, N$ be the column vectors of $\mathbf{E}$. Then, $\mathbf{E}$ can be obtained by performing the following procedures.

Step 1: Define $\mathbf{S}$ and $\mathbf{P}$ as two $N \times N$ matrices as follows:

$$\mathbf{S} = \begin{bmatrix} 2 & 1 & 0 & \cdots & 0 & 1 \\ 1 & 2\cos\left(\dfrac{2\pi}{N}\right) & 1 & \cdots & 0 & 0 \\ 0 & 1 & 2\cos\left(\dfrac{4\pi}{N}\right) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & 1 & 2\cos\left(\dfrac{2\pi(N-1)}{N}\right) \end{bmatrix} \tag{5}$$

and

$$\mathbf{P} = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 1 & \cdots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & \cdots & -1 & 0 \\ 0 & 1 & 0 & \cdots & 0 & -1 \end{bmatrix}. \tag{6}$$

Step 2: Since **S** and **P** are defined in the above, $\mathbf{PSP}^{-1}$ can be evaluated. It can be checked easily that $\mathbf{PSP}^{-1}$ is a block diagonal matrix. Define the upper diagonal block matrix and the lower diagonal block matrix as **Ev** and **Od**, respectively. That is:

$$\mathbf{PSP}^{-1} = \begin{bmatrix} \mathbf{Ev} & \mathbf{0} \\ \mathbf{0} & \mathbf{Od} \end{bmatrix}. \tag{7}$$

Here, both **Ev** and **Od** are two $\frac{N}{2} \times \frac{N}{2}$ matrices.

Step 3: Individually find the eigenvalues and the corresponding eigenvectors of **Ev** and **Od**. Here, the dimensions of the eigenvectors of both **Ev** and **Od** are $\frac{N}{2}$.

Step 4: Individually arrange the eigenvectors of **Ev** and **Od** according to the descending orders of their corresponding eigenvalues. Denote the arranged eigenvectors as $\mathbf{e}_k$ and $\mathbf{o}_k$ for $k = 0, \cdots, \frac{N}{2} - 1$, respectively.

Step 5: Set

$$\mathbf{u}_{2k} = \mathbf{P} \begin{bmatrix} \mathbf{e}_k^T & \mathbf{0}^T \end{bmatrix}^T \tag{8}$$

and

$$\mathbf{u}_{2k+1} = \mathbf{P} \begin{bmatrix} \mathbf{0}^T & \mathbf{o}_k^T \end{bmatrix}^T \tag{9}$$

for $k = 0, \cdots, \frac{N}{2} - 1$. Here, the dimensions of $\mathbf{u}_k$ for $k = 0, \cdots, N-1$ are N. Then, **E** is defined as:

$$\mathbf{E} = \begin{bmatrix} \mathbf{u}_0, & \cdots, & \mathbf{u}_{N-1} \end{bmatrix}. \tag{10}$$

For the further details, please refer to Table IV in [23].

3. Problem formulation

Suppose that the overcomplete transform consists of M DFrFTs with different orders. Denote the vector of the orders of the DFrFTs as

$$\boldsymbol{\alpha} \equiv \begin{bmatrix} \alpha_0, \cdots, \alpha_{M-1} \end{bmatrix}^T \in \begin{bmatrix} -\pi, \pi \end{bmatrix}^{M \times 1} \tag{11}$$

and the overcomplete transform matrix as

$$\mathbf{F} \equiv \begin{bmatrix} \mathbf{F}_{\alpha_0} \\ \vdots \\ \mathbf{F}_{\alpha_{M-1}} \end{bmatrix} \in C^{NM \times N}, \tag{12}$$

where $\mathbf{F}_{\alpha_m}$ are the DFrFT matrices with the orders of their transforms $\alpha_m \in \begin{bmatrix} -\pi, \pi \end{bmatrix}$ for $m = 0, \cdots, M-1$. Here, the overcomplete transform matrix is constructed by putting these DFrFT matrices together such that the total number of the rows of the overcomplete transform matrix is more than the total number of its columns. Let $\mathbf{c}_k \in C^{MN \times 1}$ for $k = 0, \cdots, K-1$ be the corresponding transformed vectors. Then, these transformed vectors can be represented as follows:

$$\mathbf{Fx}_k = \mathbf{c}_k \quad \text{for} \quad k = 0, \cdots, K-1. \tag{13}$$

The objective of the overcomplete transform matrix design problem is to find an optimal $\boldsymbol{\alpha}$ (which is equivalent to find the optimal DFrFT matrices) such that the transformed vectors $\mathbf{c}_k$ are sparse.

The common criteria for formulating the sparse objective functions are via the L_0 norm operator. For this case, the objective is to find the optimal orders of the DFrFTs such that the sum of all the L_0 norms of the transformed vectors is minimized. However, the globally optimal solutions of the L_0 norm optimization problems are in general not uniquely defined.

Hence, it is very difficult to select the globally optimal solutions. On the other hand, as discussed in [25], [26], the L_0 norm optimization problems can be approximated by the corresponding L_1 norm optimization problems and the solutions of the corresponding L_1 norm optimization problems will be also sparse if the restricted isometry conditions are satisfied. Therefore, this paper employs the L_1 norm operator to formulate the optimal design of the orders of DFrFTs.

Since $\mathbf{c}_k$ are complex valued, it is required to define an L_1 norm operator for the complex valued vectors. Let

$$\mathbf{c}_k \equiv \begin{bmatrix} c_{k,0}, & \cdots & , c_{k,MN-1} \end{bmatrix}^T \quad \text{for} \quad k = 0, \cdots, K-1. \tag{14}$$

The most common definition on the L_1 norm operator for the complex valued vectors is:

$$\left\| \mathbf{c}_k \right\|_{1,\odot} \equiv \sqrt{\sum_{n=0}^{MN-1} \left(\left| real\left(c_{k,n} \right) \right|^2 + \left| imag\left(c_{k,n} \right) \right|^2 \right)}. \tag{15}$$

However, as $\mathbf{F}\mathbf{x}_k = \mathbf{c}_k$ for $k = 0, \cdots, K-1$, $\left\| \mathbf{c}_k \right\|_{1,\odot}^2 = \left\| \mathbf{F}\mathbf{x}_k \right\|_{1,\odot}^2 = M \left\| \mathbf{x}_k \right\|^2$ for $k = 0, \cdots, K-1$. It can be seen that the two norms of the transformed vectors are independent of the orders of the DFrFTs. Therefore, the orders of the DFrFTs cannot be designed via the formulation with this definition of the L_1 norm operator. Also, as the phases of $\mathbf{c}_k$ for $k = 0, \cdots, K-1$ cannot be computed from their norm values, there is no control on the phase parts of the transformed vectors via this definition of the L_1 norm operator. Moreover, unlike the conventional FT, $\mathbf{c}_k$ for $k = 0, \cdots, K-1$ are not necessary to be the conjugate symmetric vectors for the real valued signals. For some applications such as the compression applications, some transformed coefficients are dropped. In this case, the reconstructed signals may be complex valued and severe distortions will be resulted. Similarly, for the communication applications, the noises are corrupted to the transformed coefficients. In this case, the transformed signals will be complex valued and the phase distortions of the decoded signals will also be very severe.

To address this problem, this paper proposes a new definition of the L_1 norm operator for complex valued vectors as follows:

$$\left\| \mathbf{c}_k \right\|_1 \equiv \sum_{n=0}^{MN-1} \left| real\left(c_{k,n} \right) \right| + \sum_{n=0}^{MN-1} \left| imag\left(c_{k,n} \right) \right|. \tag{16}$$

Here, both $real\left(\mathbf{c}_k \right)$ and $imag\left(\mathbf{c}_k \right)$ are the real valued vectors. As $\left\| \mathbf{c}_k \right\|_1 \equiv \left\| real\left(\mathbf{F}\mathbf{x}_k \right) \right\|_1 + \left\| imag\left(\mathbf{F}\mathbf{x}_k \right) \right\|_1$, they only involve the conventional L_1 norm operator instead of the square root of the conventional L_2 norm operator. Also, the objective function is convex with respect to $\mathbf{F}$. Because of these nice properties, this paper applies this new L_1 norm operator to $\mathbf{c}_k$ for $k = 0, \cdots, K-1$ for designing the orders of the DFrFTs. Now, the optimal design of the orders of the DFrFTs is formulated as the following optimization problem:

Problem ($\mathbf{Q}$)

$$\min_{\boldsymbol{\alpha}} \quad \sum_{k=0}^{K-1} \left(\sum_{n=0}^{MN-1} \left| real\left(c_{k,n} \right) \right| + \sum_{n=0}^{MN-1} \left| imag\left(c_{k,n} \right) \right| \right), \tag{17a}$$

$$\text{subject to} \quad \boldsymbol{\alpha} \equiv \begin{bmatrix} \alpha_0, \cdots, \alpha_M \end{bmatrix}^T \in \Re^{M \times 1}, \tag{17b}$$

$$\mathbf{F} \equiv \begin{bmatrix} \mathbf{F}_{\alpha_0} \\ \vdots \\ \mathbf{F}_{\alpha_{M-1}} \end{bmatrix} \in C^{NM \times N} \tag{17c}$$

$$\text{and} \quad \mathbf{F}\mathbf{x}_k = \mathbf{c}_k \quad \text{for} \quad k = 0, \cdots, K-1. \tag{17d}$$

It is worth noting that the feasible set of Problem ($\mathbf{Q}$) is continuous. That means, if an optimal order of the DFrFT is found, then all the points in a small neighbor of this optimal

6

order of the DFrFT could be the rest of the other optimal orders of the DFrFTs. However, this case should be avoided. Denote the optimal orders of the DFrFTs as α_m^* for $m = 0, \cdots, M-1$. It is required to find α_m^* for $m = 0, \cdots, M-1$ such that

$$\alpha_m^* \neq \alpha_n^* \quad \text{for} \quad m \neq n \quad \text{and for} \quad m, n \in \{0, \cdots, M-1\}. \tag{18}$$

To address this difficulty, it is worth noting that the objective function of the optimization problem evaluated at the locally optimal orders of the DFrFTs is either nondifferentiable or differentiable with their first order derivatives being equal to zero and their second order derivatives being positive. However, the required computational power for identifying those differentiable points with their first order derivatives being equal to zero and their second order derivatives being positive is very large. Therefore, this paper only identifies the stationary points and the nondifferentiable points of the objective function at the first stage of the algorithm. As the gradients of the objective function evaluated at those stationary points are equal to zero, we have:

$$\lim_{\alpha_m \to \alpha_m^*} \frac{d \left\| \tilde{\mathbf{c}}_{\alpha_m} \right\|_1}{d\alpha_m} = \lim_{\alpha_m \to \alpha_m^*} \frac{d}{d\alpha_m} \sum_{k=0}^{K-1} \left(\left\| real\left(\mathbf{F}_{\alpha_m} \mathbf{x}_k \right) \right\|_1 + \left\| imag\left(\mathbf{F}_{\alpha_m} \mathbf{x}_k \right) \right\|_1 \right) = 0. \tag{19}$$

On the other hand, the right hand limits of the gradients of the objective function are not equal to the left hand limits of the gradients of the objective function when the gradients of the objective function are evaluated at those nondifferentiable points. That is:

$$\lim_{\alpha_m \to \alpha_m^{*+}} \frac{d \left\| \tilde{\mathbf{c}}_{\alpha_m} \right\|_1}{d\alpha_m} = \neq \lim_{\alpha_m \to \alpha_m^{*-}} \frac{d \left\| \tilde{\mathbf{c}}_{\alpha_m} \right\|_1}{d\alpha_m}. \tag{20}$$

It is obvious to see that the set of the orders of the DFrFTs satisfying either the stationary condition or the nondifferentiable condition consists of a finite number of points. In other words, the feasible set of the optimization problem is discrete. Therefore, all the points in a small neighbor of the globally optimal solution of the optimization problem but not including this globally optimal solution are not in this feasible set. Because of this nice property, the above issue is solved. Therefore, the above two constraints are imposed to the optimization problem. Now, Problem ($\mathbf{Q}$) is reformulated as follows:

Problem ($\mathbf{P}$)

$$\min_{\boldsymbol{\alpha}} \quad \sum_{k=0}^{K-1} \left(\sum_{n=0}^{MN-1} \left| real\left(c_{k,n} \right) \right| + \sum_{n=0}^{MN-1} \left| imag\left(c_{k,n} \right) \right| \right), \tag{21a}$$

$$\text{subject to} \quad \boldsymbol{\alpha} \equiv \left[\alpha_0, \cdots, \alpha_M \right]^T \in \Re^{M \times 1}, \tag{21b}$$

$$\mathbf{F} \equiv \begin{bmatrix} \mathbf{F}_{\alpha_0} \\ \vdots \\ \mathbf{F}_{\alpha_{M-1}} \end{bmatrix} \in C^{NM \times N}, \tag{21c}$$

$$\mathbf{F}\mathbf{x}_k = \mathbf{c}_k, \tag{21d}$$

$$\text{and} \quad \lim_{\alpha_m \to \alpha_m^*} \frac{d \left\| \tilde{\mathbf{c}}_{\alpha_m} \right\|_1}{d\alpha_m} = 0 \quad \text{or} \quad \lim_{\alpha_m \to \alpha_m^{*+}} \frac{d \left\| \tilde{\mathbf{c}}_{\alpha_m} \right\|_1}{d\alpha_m} = \neq \lim_{\alpha_m \to \alpha_m^{*-}} \frac{d \left\| \tilde{\mathbf{c}}_{\alpha_m} \right\|_1}{d\alpha_m}. \tag{21e}$$

Here, Problem ($\mathbf{P}$) is subject to four conditions. The first three conditions defined in (21b), (21c) and (21d) are required to be satisfied simultaneously. However, the fourth condition defined in (21d) consists of two parts and these two parts are not satisfied simultaneously. Either one of these two parts is satisfied.

Now, define

$$\boldsymbol{\theta}_m \equiv \left[\cos\left(0 \times \alpha_m\right), \cdots, \cos\left((N-1) \times \alpha_m\right), \sin\left(0 \times \alpha_m\right), \cdots, \sin\left((N-1) \times \alpha_m\right) \right]^T \text{for} \quad m = 0, \cdots, M-1 \tag{22}$$

and

$$\boldsymbol{\theta} \equiv \left[\boldsymbol{\theta}_0, \cdots, \boldsymbol{\theta}_{M-1} \right]. \tag{23}$$

Then, the following theorem enables us to convert the highly nonlinear objective function of $\boldsymbol{\alpha}$ to a linear objective function of $\boldsymbol{\theta}$.

Theorem 1

$\exists \tilde{\mathbf{X}} \in \mathfrak{R}^{2NK \times 2N}$ such that Problem (**P**) is equivalent to the following optimization problem:

Problem ($\overline{\mathbf{P}}$)

$$\min_{\boldsymbol{\theta}} \quad \sum_{m=0}^{M-1} \left\| \tilde{\mathbf{X}} \boldsymbol{\theta}_m \right\|_1 , \tag{24a}$$

subject to $\quad \boldsymbol{\theta}_m \equiv \left[\cos(0 \times \alpha_m), \cdots, \cos\left((N-1) \times \alpha_m\right), \sin(0 \times \alpha_m), \cdots, \sin\left((N-1) \times \alpha_m\right) \right]^T ,$ (24b)

$$\boldsymbol{\theta} \equiv \left[\boldsymbol{\theta}_0, \cdots, \boldsymbol{\theta}_{M-1} \right], \tag{24c}$$

and $\quad \displaystyle\lim_{\alpha_m \to \alpha_m^*} \frac{d \left\| \tilde{\mathbf{c}}_{\alpha_m} \right\|_1}{d\alpha_m} = 0 \quad$ or $\quad \displaystyle\lim_{\alpha_m \to \alpha_m^{*+}} \frac{d \left\| \tilde{\mathbf{c}}_{\alpha_m} \right\|_1}{d\alpha_m} = \neq \lim_{\alpha_m \to \alpha_m^{*-}} \frac{d \left\| \tilde{\mathbf{c}}_{\alpha_m} \right\|_1}{d\alpha_m} .$ (24d)

Proof:

Let

$$\boldsymbol{\beta} \equiv \left[0, \cdots, N-1 \right]^T \tag{25}$$

and

$$\tilde{\mathbf{x}}_k \equiv \left[\tilde{x}_{k,0}, \quad \cdots \quad , \tilde{x}_{k,N-1} \right]^T \equiv \mathbf{E}^T \mathbf{x}_k \quad \text{for} \quad k = 0, \cdots, K-1 . \tag{26}$$

Then, we have:

$$\mathbf{F}_{\alpha_m} \mathbf{x}_k = \mathbf{E} diag\left(e^{-j\alpha_m \boldsymbol{\beta}} \right) \tilde{\mathbf{x}}_k = \mathbf{E} diag\left(\tilde{\mathbf{x}}_k \right) e^{-j\alpha_m \boldsymbol{\beta}} \tag{27}$$

for $\quad m = 0, \cdots, M-1 \quad$ and for $\quad k = 0, \cdots, K-1$. Denote

$$\mathbf{X}_k \equiv \mathbf{E} diag\left(\tilde{\mathbf{x}}_k \right) \equiv \begin{bmatrix} x_{k,0,0} & \cdots & x_{k,0,N-1} \\ \vdots & \ddots & \vdots \\ x_{k,N-1,0} & \cdots & x_{k,N-1,N-1} \end{bmatrix} \quad \text{for} \quad k = 0, \cdots, K-1 . \tag{28}$$

Then, we have:

$$\mathbf{F}_{\alpha_m} \mathbf{x}_k = \mathbf{X}_k e^{-j\alpha_m \boldsymbol{\beta}} \quad \text{for} \quad m = 0, \cdots, M-1 \quad \text{and for} \quad k = 0, \cdots, K-1 . \tag{29}$$

Denote

$$\tilde{\mathbf{X}}_a \equiv \begin{bmatrix} \mathbf{X}_0 \\ \vdots \\ \mathbf{X}_{K-1} \end{bmatrix} \tag{30}$$

and

$$\tilde{\mathbf{X}} \equiv \begin{bmatrix} \tilde{\mathbf{X}}_a & \mathbf{0}_{NK \times N} \\ \mathbf{0}_{NK \times N} & \tilde{\mathbf{X}}_a \end{bmatrix} . \tag{31}$$

Then, we have

$$\sum_{m=0}^{M-1} \left\| \tilde{\mathbf{X}} \boldsymbol{\theta}_m \right\|_1 = \sum_{m=0}^{M-1} \left\| \begin{bmatrix} \tilde{\mathbf{X}}_a & \mathbf{0}_{NK \times N} \\ \mathbf{0}_{NK \times N} & \tilde{\mathbf{X}}_a \end{bmatrix} \begin{bmatrix} \cos(0 \times \alpha_m) \\ \vdots \\ \cos\left((N-1) \times \alpha_m\right) \\ \sin(0 \times \alpha_m) \\ \vdots \\ \sin\left((N-1) \times \alpha_m\right) \end{bmatrix} \right\|_1$$

$$= \sum_{m=0}^{M-1} \left(\left\| \tilde{\mathbf{X}}_a \begin{bmatrix} \cos(0 \times \alpha_m) \\ \vdots \\ \cos\left((N-1) \times \alpha_m\right) \end{bmatrix} \right\|_1 + \left\| \tilde{\mathbf{X}}_a \begin{bmatrix} \sin(0 \times \alpha_m) \\ \vdots \\ \sin\left((N-1) \times \alpha_m\right) \end{bmatrix} \right\|_1 \right)$$

$$= \sum_{m=0}^{M-1} \left(\left\| real\left(\tilde{\mathbf{X}}_a e^{-j\alpha_m \boldsymbol{\beta}} \right) \right\|_1 + \left\| imag\left(\tilde{\mathbf{X}}_a e^{-j\alpha_m \boldsymbol{\beta}} \right) \right\|_1 \right)$$

$$= \sum_{m=0}^{M-1} \left(\left\| real \left(\begin{bmatrix} \mathbf{X}_0 \\ \vdots \\ \mathbf{X}_{K-1} \end{bmatrix} e^{-j\alpha_m \boldsymbol{\beta}} \right) \right\|_1 + \left\| imag \left(\begin{bmatrix} \mathbf{X}_0 \\ \vdots \\ \mathbf{X}_{K-1} \end{bmatrix} e^{-j\alpha_m \boldsymbol{\beta}} \right) \right\|_1 \right)$$

$$= \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} \left(\left\| real \left(\mathbf{X}_k e^{-j\alpha_m \boldsymbol{\beta}} \right) \right\|_1 + \left\| imag \left(\mathbf{X}_k e^{-j\alpha_m \boldsymbol{\beta}} \right) \right\|_1 \right) \qquad (32)$$

$$= \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} \left(\left\| real \left(\mathbf{F}_{\alpha_m} \mathbf{x}_k \right) \right\|_1 + \left\| imag \left(\mathbf{F}_{\alpha_m} \mathbf{x}_k \right) \right\|_1 \right)$$

The result follows directly. This completes the proof. ∎

It can be seen in (28) that there are N elements in each row of $\mathbf{X}_k$ for $k = 0, \cdots, K-1$. By defining a finite impulse response filter with the values of its impulse response equal to the values of these N elements of a row of $\mathbf{X}_k$ for $k = 0, \cdots, K-1$, and denoting the frequency response of the filter corresponding to the pth row of $\mathbf{X}_k$ as $\tilde{X}_{k,p}(\omega)$ for $k = 0, \cdots, K-1$ and for $p = 0, \cdots, N-1$, then we have

$$\tilde{X}_{k,p}(\omega) = \sum_{n=0}^{N-1} x_{k,p,n} e^{-j\omega n} . \qquad (33)$$

Hence,

$$\sum_{m=0}^{M-1} \left\| \tilde{\mathbf{X}} \boldsymbol{\theta}_m \right\|_1 = \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} \left(\left\| real \left(\mathbf{X}_k e^{-j\alpha_m \boldsymbol{\beta}} \right) \right\|_1 + \left\| imag \left(\mathbf{X}_k e^{-j\alpha_m \boldsymbol{\beta}} \right) \right\|_1 \right)$$

$$= \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} \left(\left\| real \left(\begin{bmatrix} x_{k,0,0} & \cdots & x_{k,0,N-1} \\ \vdots & \ddots & \vdots \\ x_{k,N-1,0} & \cdots & x_{k,N-1,N-1} \end{bmatrix} \begin{bmatrix} e^{-j\alpha_m 0} \\ \vdots \\ e^{-j\alpha_m (N-1)} \end{bmatrix} \right) \right\|_1 + \left\| imag \left(\begin{bmatrix} x_{k,0,0} & \cdots & x_{k,0,N-1} \\ \vdots & \ddots & \vdots \\ x_{k,N-1,0} & \cdots & x_{k,N-1,N-1} \end{bmatrix} \begin{bmatrix} e^{-j\alpha_m 0} \\ \vdots \\ e^{-j\alpha_m (N-1)} \end{bmatrix} \right) \right\|_1 \right) . \quad (34)$$

$$= \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} \sum_{p=0}^{N-1} \left(\left| real \left(\sum_{n=0}^{N-1} x_{k,p,n} e^{-j\alpha_m n} \right) \right| + \left| imag \left(\sum_{n=0}^{N-1} x_{k,p,n} e^{-j\alpha_m n} \right) \right| \right)$$

$$= \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} \sum_{p=0}^{N-1} \left(\left| real \left(\tilde{X}_{k,p}(\alpha_m) \right) \right| + \left| imag \left(\tilde{X}_{k,p}(\alpha_m) \right) \right| \right)$$

It is worth noting that the signals represented in the frequency domain are 2π periodic, so the frequency domain can be characterized by $[-\pi, \pi)$. On the other hand, the set of the orders of the DFrFTs is also characterized by $[-\pi, \pi)$. From (34), we can see that the objective function of Problem ($\bar{\mathbf{P}}$) is $\sum_{m=0}^{M-1} \sum_{k=0}^{K-1} \sum_{p=0}^{N-1} \left(\left| real \left(\tilde{X}_{k,p}(\alpha_m) \right) \right| + \left| imag \left(\tilde{X}_{k,p}(\alpha_m) \right) \right| \right)$. Therefore, finding the globally optimal solution of Problem ($\bar{\mathbf{P}}$) is equivalent to find M optimal sampling frequencies $\alpha_m \in [-\pi, \pi)$ for $m = 0, \cdots, M-1$ such that $\sum_{m=0}^{M-1} \sum_{k=0}^{K-1} \sum_{p=0}^{N-1} \left(\left| real \left(\tilde{X}_{k,p}(\alpha_m) \right) \right| + \left| imag \left(\tilde{X}_{k,p}(\alpha_m) \right) \right| \right)$ is minimized. In other words, it is equivalent to find $\alpha_m \in [-\pi, \pi)$ for $m = 0, \cdots, M-1$ such that the absolute sum of both the real parts and the imaginary parts of the frequency responses of all these NK filters evaluated at these sampling frequencies is minimized.

However, Problem ($\bar{\mathbf{P}}$) is different from the conventional filter design problems [21], [22]. This is because the filter coefficients in the conventional filter design problems are unknown and they are required to be determined. Here, the filter coefficients in Problem ($\bar{\mathbf{P}}$) are known. However, the sampling frequencies are unknown and they are required to be determined. Also, for the conventional filter design problems, the domains of the frequency responses of the filters are $[-\pi, \pi)$. It is a continuous set. On the other hand, as the constraints on the orders of the DFrFTs are imposed to Problem ($\bar{\mathbf{P}}$), the orders of the DFrFTs are in a discrete set. Moreover, for the conventional filter design problems, we have:

$$\tilde{X}_{k,p}(\omega) = \sum_{n=0}^{N-1} x_{k,p,n} \cos(n\omega) - j \sum_{n=0}^{N-1} x_{k,p,n} \sin(n\omega)$$

$$= \begin{bmatrix} \cos(0\omega), & \cdots, & \cos(N-1)\omega \end{bmatrix} \begin{bmatrix} x_{k,p,0} \\ \vdots \\ x_{k,p,N-1} \end{bmatrix} - j \begin{bmatrix} \sin(0\omega), & \cdots, & \sin(N-1)\omega \end{bmatrix} \begin{bmatrix} x_{k,p,0} \\ \vdots \\ x_{k,p,N-1} \end{bmatrix}. \quad (35)$$

Since the optimization variables of the conventional filter design problems are the filter coefficients, the objective functions of these optimization problems are the linear functions of the optimization variables. However, the objective function of Problem ($\bar{\mathbf{P}}$) is:

$$\tilde{X}_{k,p}(\omega_m) = \sum_{n=0}^{N-1} x_{k,p,n} \cos(n\omega_m) - j \sum_{n=0}^{N-1} x_{k,p,n} \sin(n\omega_m)$$

$$= \begin{bmatrix} x_{k,p,0}, & \cdots, & x_{k,p,N-1} \end{bmatrix} \begin{bmatrix} \cos(0\omega_m) \\ \vdots \\ \cos(N-1)\omega_m \end{bmatrix} - j \begin{bmatrix} x_{k,p,0}, & \cdots, & x_{k,p,N-1} \end{bmatrix} \begin{bmatrix} \sin(0\omega_m) \\ \vdots \\ \sin(N-1)\omega_m \end{bmatrix}. \quad (36)$$

Since the optimization variables of the optimal sampling problem are the sampling frequencies and the objective function of the optimal sampling problem involves the sine and cosine functions of the optimization variables, the objective function of Problem ($\bar{\mathbf{P}}$) is a nonlinear function of the optimization variables.

In fact, Problem ($\bar{\mathbf{P}}$) is a nonsmooth optimization problem. As discussed in Section 1, the conventional gradient descent approaches [22] cannot be directly applied for finding its locally optimal solutions. Moreover, Problem ($\bar{\mathbf{P}}$) is a nonconvex optimization problem. In general, it is very difficult to find its globally optimal solution [21]. This is because the total number of the locally optimal solutions are unknown and it is different to guarantee to reach all the locally optimal solutions. Furthermore, one of the constraints of Problem ($\bar{\mathbf{P}}$) is neither the conventional equality nor the conventional inequality constraints. In fact, it involves the XOR of two constraints. Existing algorithms cannot be directly applied for finding the solution of this optimization problem due to the XOR of two constraints. To tackle these difficulties, a new solution method is proposed below.

4. Solution method

Define

$$f_{k,p}^c(\omega) \equiv \sum_{q=0}^{N-1} x_{k,p,q} \cos(q\omega) \quad (37)$$

and

$$f_{k,p}^s(\omega) \equiv \sum_{q=0}^{N-1} x_{k,p,q} \sin(q\omega) \quad (38)$$

for $k = 0, \cdots, K-1$, for $p = 0, \cdots, N-1$ and $\forall \omega \in [-\pi, \pi)$, as well as

$$f(\omega) \equiv \sum_{k=0}^{K-1} \sum_{p=0}^{N-1} \left(\left| f_{k,p}^c(\omega) \right| + \left| f_{k,p}^s(\omega) \right| \right). \quad (39)$$

From (33), we have

$$f_{k,p}^c(\omega) = real\left(\tilde{X}_{k,p}(\omega)\right) \quad (40)$$

and

$$f_{k,p}^s(\omega) = -imag\left(\tilde{X}_{k,p}(\omega)\right). \quad (41)$$

It is obvious to see that the objective function of Problem ($\bar{\mathbf{P}}$) is $\sum_{m=0}^{M-1} \sum_{k=0}^{K-1} \sum_{p=0}^{N-1} \left(\left| real\left(\tilde{X}_{k,p}(\alpha_m)\right) \right| + \left| imag\left(\tilde{X}_{k,p}(\alpha_m)\right) \right| \right) = \sum_{m=0}^{M-1} f(\alpha_m)$. Hence, in order to minimize the objective function of Problem ($\bar{\mathbf{P}}$), it is equivalent to find M different sampling frequencies $\alpha_m \in [-\pi, \pi)$ for $m = 0, \cdots, M-1$ in the discrete feasible set defined by either (19) or (20)

10

such that $\sum_{m=0}^{M-1} f(\alpha_m)$ is minimized. Recall the above discussion, the objective function of the optimization problem is either stationary or nondifferentiable when it is evaluated at the locally optimal sampling frequencies. For the nondifferentiable points, since the harmonic functions are always differentiable, the objective function is not differentiable only at the points where the absolute operators are not differentiable at these points. This implies that the operands inside the absolute operators in the objective function are equal to zero when they are evaluated at these sampling frequencies. Hence, in order to find these nondifferentiable points, it only requires to find the sampling frequencies such that the operands inside the absolute operators in the objective function are equal to zero. For the stationary points, the operands inside the absolute operators in the objective function are not equal to zero when they are evaluated at these sampling frequencies, but the gradient of the objective function is equal to zero when it is evaluated at these sampling frequencies. However, as the objective function involves the sum of the absolute values of the polynomials of the harmonic functions and the signs of the polynomials of the harmonic functions are not known before the sampling frequencies are determined, computing the gradient of the objective function is challenging. To address this difficulty, it is worth noting that the objective function is piecewise differentiable and the points at the boundaries of the pieces are the nondifferentiable points. As these nondifferentiable points are already found and the signs of the polynomials of the harmonics functions within these pieces remain unchanged, the signs of the polynomials of the harmonics functions within these pieces can be found by evaluating the signs of the polynomials of the harmonics functions just around these nondifferentiable points. As a result, the objective function can be expressed as the sum of the signs of the polynomials of the harmonics functions just around these nondifferentiable points multiplying the polynomials of the harmonic functions. Since the derivatives of the polynomials of the harmonic functions are still the polynomials of the harmonic functions, finding the stationary points of the objective function becomes finding the roots of a new set of the polynomials of the harmonic functions. The details are as follows:

Let $L_{k,p}^c$ for $k = 0, \cdots, K-1$ and for $p = 0, \cdots, N-1$ be the total number of the roots of $f_{k,p}^c(\omega)$ in $[-\pi, \pi)$. Denote these roots as $\omega_{k,p,l_{k,p}^c}^c$ for $l_{k,p}^c = 0, \cdots, L_{k,p}^c - 1$. Also, denote the set of these roots as

$$\Omega_{k,p}^c \equiv \left\{ \omega_{k,p,0}^c, \cdots, \omega_{k,p,L_{k,p}^c-1}^c \right\} . \tag{42}$$

Similarly, let $L_{k,p}^s$ for $k = 0, \cdots, K-1$ and for $p = 0, \cdots, N-1$ be the total number of the roots of $f_{k,p}^s(\omega)$ in $[-\pi, \pi)$. Denote these roots as $\omega_{k,p,l_{k,p}^s}^s$ for $l_{k,p}^s = 0, \cdots, L_{k,p}^s - 1$. Also, denote the set of these roots as

$$\Omega_{k,p}^s \equiv \left\{ \omega_{k,p,0}^s, \cdots, \omega_{k,p,L_{k,p}^s-1}^s \right\} . \tag{43}$$

Define $\Omega \equiv \{-\pi, \pi\} \cup \bigcup_{k=0}^{K-1} \bigcup_{p=0}^{N-1} \left(\Omega_{k,p}^c \cup \Omega_{k,p}^s \right)$. Suppose that the elements in Ω are arranged in the ascending order. Let $\tilde{L} \equiv 2 + \sum_{k=0}^{K-1} \sum_{p=0}^{N-1} L_{k,p}^s + L_{k,p}^c$ be the total number of elements in Ω. Define these sorted elements as ω_i for $i = 0, \cdots, \tilde{L}-1$. That is, $\omega_i < \omega_j$ for $i < j$ and for $i, j \in \{0, \cdots, \tilde{L}-1\}$. Then, we have

$$\left| f_{k,p}^c(\omega) \right| = \text{sign}\left(f_{k,p}^c(\omega_i^+) \right) f_{k,p}^c(\omega) \tag{44}$$

$\forall \omega \in (\omega_i, \omega_{i+1})$, for $i = 0, \cdots, \tilde{L}-2$, for $k = 0, \cdots, K-1$ and for $p = 0, \cdots, N-1$, and

$$\left| f_{k,p}^s(\omega) \right| = \text{sign}\left(f_{k,p}^s(\omega_i^+) \right) f_{k,p}^s(\omega) \tag{45}$$

$\forall \omega \in (\omega_i, \omega_{i+1})$, for $i = 0, \cdots, \tilde{L}-2$, for $k = 0, \cdots, K-1$ and for $p = 0, \cdots, N-1$,

as well as

$$f(\omega) = \sum_{k=0}^{K-1} \sum_{p=0}^{N-1} \left(\text{sign}\left(f_{k,p}^c \left(\omega_i^+ \right) \right) f_{k,p}^c (\omega) + \text{sign}\left(f_{k,p}^s \left(\omega_i^+ \right) \right) f_{k,p}^s (\omega) \right) \tag{46}$$

$\forall \omega \in (\omega_i, \omega_{i+1})$ and for $i = 0, \cdots, \tilde{L} - 2$. Define

$$\tilde{f}_{k,p}^c (\omega) \equiv - \sum_{q=0}^{N-1} x_{k,p,q} q \sin(q\omega) \tag{47}$$

for $k = 0, \cdots, K - 1$ and for $p = 0, \cdots, N - 1$, and

$$\tilde{f}_{k,p}^s (\omega) \equiv \sum_{q=0}^{N-1} x_{k,p,q} q \cos(q\omega) \tag{48}$$

for $k = 0, \cdots, K - 1$ and for $p = 0, \cdots, N - 1$.

Then,

$$\frac{d}{d\omega} f(\omega) = \sum_{k=0}^{K-1} \sum_{p=0}^{N-1} \left(\text{sign}\left(f_{k,p}^c \left(\omega_i^+ \right) \right) \frac{d}{d\omega} f_{k,p}^c (\omega) + \text{sign}\left(f_{k,p}^s \left(\omega_i^+ \right) \right) \frac{d}{d\omega} f_{k,p}^s (\omega) \right)$$
$$= \sum_{k=0}^{K-1} \sum_{p=0}^{N-1} \left(\text{sign}\left(f_{k,p}^c \left(\omega_i^+ \right) \right) \tilde{f}_{k,p}^c (\omega) + \text{sign}\left(f_{k,p}^s \left(\omega_i^+ \right) \right) \tilde{f}_{k,p}^s (\omega) \right) \tag{49}$$

$\forall \omega \in (\omega_i, \omega_{i+1})$ and for $i = 0, \cdots, \tilde{L} - 2$. Denote the total number of the roots of $\frac{d}{d\omega} f(\omega)$ for $\omega \in (\omega_i, \omega_{i+1})$ as $\hat{L}_i$ for $i = 0, \cdots, \tilde{L} - 2$. Denote these elements as ω_{i,j_i} for $j_i = 0, \cdots, \hat{L}_i - 1$ and for $i = 0, \cdots, \tilde{L} - 2$. Denote the set of the sampling frequencies corresponding to these nondifferential points and these stationary points as

$$\hat{\Omega} \equiv \left\{ \omega_0, \omega_{0,0}, \cdots \omega_{0,\hat{L}_0 - 1}, \omega_1, \omega_{1,0}, \cdots \omega_{1,\hat{L}_1 - 1}, \cdots, \omega_{\tilde{L}-2}, \omega_{\tilde{L}-2,0}, \cdots \omega_{\tilde{L}-2,\hat{L}_{\tilde{L}-2}-1}, \omega_{\tilde{L}-1} \right\}. \tag{50}$$

To characterize the set $\hat{\Omega}$, it only requires to find the roots of (37), (38) and $\frac{d}{d\omega} f(\omega)$ $\forall \omega \in (\omega_i, \omega_{i+1})$ and for $i = 0, \cdots, \tilde{L} - 2$. As all these functions are the polynomials of the harmonic functions, the elements in $\hat{\Omega}$ can be found efficiently via applying the existing methods for finding the roots of the polynomials of the harmonic functions. In particular, the polynomials of the harmonic functions are expressed as the conventional polynomials via the half angle formula [27], [28]. Then, the roots of these conventional polynomials are found via the Newton's method [29]. More precisely, the guesses of the solutions of the polynomials are initialized. Then, these guesses of the solutions are updated by subtracting the current guesses of the solutions to the ratios of the functional values to the gradients of the functions evaluated at these current guesses. By iterating this procedure until the algorithm converges, the roots of the polynomials are found. By generating the other initial guesses, different roots may be obtained. By repeating these procedures, eventually all the roots of the polynomials are found.

Once $\hat{\Omega}$ is determined, the functional values of these nondifferentiable points and these stationary points are computed. Next, these nondifferentiable points and these stationary points are sorted in the ascending order. Finally, the first M nondifferentiable points or the stationary points with their second order derivatives being positive are selected. Denote the set of these M selected points as $\tilde{\Omega}$. Although there are $1 + \sum_{p=0}^{\tilde{L}-2} \left(\hat{L}_p + 1 \right)$ points in $\hat{\Omega}$ in which $\sum_{p=0}^{\tilde{L}-2} \hat{L}_p$ points are the stationary points and $\tilde{L}$ points are the nondifferential points, it is not required to compute the second order derivatives of all these $\sum_{p=0}^{\tilde{L}-2} \hat{L}_p$ stationary points.

Only some points are required to compute their second order derivatives. Therefore, the required computational power of the proposed algorithm is very low. By constructing the optimal overcomplete transform using the DFrFTs with different orders, the transformed

vectors are sparse. Compared to the existing penalized techniques, our proposed method is can find the optimal orders of the DFrFTs in a more effective and efficient manner.

The result is summarized in the following theorem:

Theorem 2

The M elements in $\tilde{\Omega}$ are the M globally optimal orders of the DFrFTs.

Proof:

Since the M elements in $\tilde{\Omega}$ are the nondifferentiable points and the stationary points with their second order derivatives being positive and they have been considered in the whole frequency range $[-\pi, \pi]$, these M elements in $\tilde{\Omega}$ are the M globally optimal orders of the DFrFTs. This completes the proof. ∎

The algorithm can be summarized as follows:

Algorithm 1

Step 1:

Based on a predefined definition of the DFrFT, compute the matrix $\mathbf{E}$.

Step 2:

Compute $\tilde{\mathbf{x}}_k = \mathbf{E}^T \mathbf{x}_k$ for $k = 0, \cdots, K-1$ and

$$\mathbf{X}_k = \mathbf{E} diag\left(\tilde{\mathbf{x}}_k\right) = \begin{bmatrix} x_{k,0,0} & \cdots & x_{k,0,N-1} \\ \vdots & \ddots & \vdots \\ x_{k,N-1,0} & \cdots & x_{k,N-1,N-1} \end{bmatrix} \quad \text{for} \ \ k = 0, \cdots, K-1.$$

Step 3:

Find the roots of the following equations using the method summarized in Algorithm 2:

$$f_{k,p}^c\left(\omega\right) \equiv \sum_{q=0}^{N-1} x_{k,p,q} \cos\left(q\omega\right) \quad \text{for} \ \ k = 0, \cdots, K-1 \ \ \text{and for} \ \ p = 0, \cdots, N-1,$$

and

$$f_{k,p}^s\left(\omega\right) \equiv \sum_{q=0}^{N-1} x_{k,p,q} \sin\left(q\omega\right) \quad \text{for} \ \ k = 0, \cdots, K-1 \ \ \text{and for} \ \ p = 0, \cdots, N-1.$$

Composite the discrete sets $\Omega_{k,p}^c$ and $\Omega_{k,p}^s$ for $k = 0, \cdots, K-1$ and for $p = 0, \cdots, N-1$ using the roots of $f_{k,p}^c\left(\omega\right)$ and $f_{k,p}^s\left(\omega\right)$, respectively.

Step 4:

Composite the discrete set Ω by union $\Omega_{k,p}^c$ and $\Omega_{k,p}^s$ for $k = 0, \cdots, K-1$ and for $p = 0, \cdots, N-1$ as well as the set of the boundary points $\{-\pi, \pi\}$. Sort the elements in Ω in the ascending order and obtain the elements ω_i for $i = 0, \cdots, \tilde{L}-1$.

Step 5:

Find the roots of the following equation:

$$\frac{d}{d\omega} f\left(\omega\right) = \sum_{k=0}^{K-1} \sum_{p=0}^{N-1} \left(\text{sign}\left(f_{k,p}^c\left(\omega_i^+\right)\right) \frac{d}{d\omega} f_{k,p}^c\left(\omega\right) + \text{sign}\left(f_{k,p}^s\left(\omega_i^+\right)\right) \frac{d}{d\omega} f_{k,p}^s\left(\omega\right) \right)$$

$$= \sum_{k=0}^{K-1} \sum_{p=0}^{N-1} \left(\text{sign}\left(f_{k,p}^c\left(\omega_i^+\right)\right) \tilde{f}_{k,p}^c\left(\omega\right) + \text{sign}\left(f_{k,p}^s\left(\omega_i^+\right)\right) \tilde{f}_{k,p}^s\left(\omega\right) \right)$$

$\forall \omega \in \left(\omega_i, \omega_{i+1}\right)$ and for $i = 0, \cdots, \tilde{L}-2$ using the method summarized in Algorithm 2. Composite the set $\hat{\Omega}$ by union Ω and the set of the roots of $\frac{d}{d\omega} f\left(\omega\right)$.

Step 6:

Find the first M nondifferentiable points or the stationary points with their second order derivatives being positive. Composite the set $\tilde{\Omega}$ using these M points.

Step 7:

Construct the optimal overcomplete transform using the DFrFTs with these M orders.

Algorithm 2

Step 1:

13

The polynomials of the harmonic functions are expressed as the conventional polynomials via the half angle formula. Denote an obtained polynomial as $g(x)$ and the order of this polynomial as $\tilde{N}$. Define an acceptable value on the absolute differences of the obtained solutions between two consecutive iterations as ε.

Step 2:

Initialize $m = 1$.

Step 3:

Initialize $k = 1$ and a guess of a solution of the polynomial as $x_{k,m}^*$.

Step 4:

Compute $x_{k+1,m}^* = x_{k,m}^* - \dfrac{g\left(x_{k,m}^*\right)}{\left.\dfrac{dg(x)}{dx}\right|_{x=x_{k,m}^*}}$ and iterate this step until $\left|x_{k+1,m}^* - x_{k,m}^*\right| \le \varepsilon$.

Step 5:

Compute the order of the DFrFT based on the obtained solution in Step 4 using the inverse half angle formula.

Step 6: Increment the value of m and go back to Step 3 until $m = \tilde{N}$.

Since Algorithm 2 is based on the iterative based Newton method, the convergence of the algorithm depends on the convergence condition of the Newton method. In particular, the algorithm is guaranteed to be converged if $\left|\left.\dfrac{dg(x)}{dx}\right|_{x=x_{k,m}^*}\right| \le 1$ for all k and for $m = 1, \cdots, \tilde{N}$.

Besides, as there are $1 + \sum_{p=0}^{\tilde{L}-2}\left(\hat{L}_p + 1\right)$ elements in $\hat{\Omega}$ in which $-\pi$ and π are the two boundary frequencies in $\hat{\Omega}$, only $-1 + \sum_{p=0}^{\tilde{L}-2}\left(\hat{L}_p + 1\right)$ sampling frequencies are required to be computed. Therefore, the required computational power for finding all the stationary points and the nondifferential points is bounded by $-1 + \sum_{p=0}^{\tilde{L}-2}\left(\hat{L}_p + 1\right)$ multiplied to the required computational power for finding a root of a polynomial with its order being bounded by the maximum order of all the polynomials. As discussed in the above, only some points are required to compute their second order derivatives, so the total required computational power of our proposed method is low.

5. Application example

In this Section, an example on a frequency modulated transmitted signal is illustrated to demonstrate the application value of the proposed method.

Let $m(t)$, $r(t)$, $\eta(t)$, A and c be the message, the transmitted signal, the complex valued zero mean additive white Gaussian noise, the gain of the transmitted signal and the frequency gain of the transmitted signal, respectively. That is:

$$r(t) = Ae^{j\left(c\int_{-\infty}^{t} m(\tau)d\tau\right)} + \eta(t). \tag{51}$$

We have

$$r(n) = Ae^{j\left(c\sum_{k=-\infty}^{n} m(k)\right)} + \eta(n) \tag{52}$$

for the corresponding discrete time signal. Here, $A = 1$ is chosen for the illustration because of the normalization reason. c is set at 0.3. This is because it is small enough to prevent the occurrence of the overflow of the phase out of the 2π range. A complex valued zero mean additive white Gaussian noise is added to the signal. It is used to model the noisy environments in the practical situations. The variances of both $real(\eta(n))$ and

$imag(\eta(n))$ are set at 0.1. It is the common noise level used in the practical situations. To demonstrate the application value of the proposed method, the image "Lena" with the size 256×256 is transmitted. The first 255 columns of the image are regarded as the training signals and the last column of the image is regarded as the test signal. That is, $N = 256$ and $K = 255$. Now, the last column of the image is represented in the form of (52). Denote the transmitted test signal as $x_{test}(n)$ for $n = 0, \cdots, N-1$. To construct the overcomplete transform, it is required to determine the value of M. It is worth noting that M is the overcomplete ratio. That is, it is the ratio of the total number of the rows of the overcomplete transform matrix to that of its columns. As the transformed signal based on a positive order of the DFrFT is related to that based on the corresponding negative order of the DFrFT, M should be chosen as twice of the overcomplete ratio. Besides, the larger the value of M results to a more sparse representation of the signal, but it requires a more computational effect. Therefore, M is used to tradeoff between the sparsity of the representation and the computational cost. In many practical applications, the overcomplete ratios are either chosen as two or three. Therefore, both $M = 4$ and $M = 6$ are employed for the illustration in this paper.

According to (34), we define the corresponding objective function as follows:

$$J(\omega) \equiv \sum_{k=0}^{K-1} \sum_{p=0}^{N-1} \left(\left| real\left(\tilde{X}_{k,p}(\omega) \right) \right| + \left| imag\left(\tilde{X}_{k,p}(\omega) \right) \right| \right) \quad \forall \omega \in [-\pi, \pi). \tag{53}$$

From (17a), (32) and (34), we can see that the functional value of $J(\omega)$ is equal to the sum of the L_1 norms of the real parts and the imaginary parts of all the transformed vectors. This is also equal to the objective functional value of Problem ($\mathbf{Q}$). Figure 1 plots $J(\omega)$ $\forall \omega \in [-\pi, \pi)$ and for $M = 4$. Here, the definition of the DFrFT is based on that discussed in [23]. It can be seen that the two positive values of ω corresponding to the lowest two objective functional values of $J(\omega)$ in which $J(\omega)$ are either nondifferentiable or stationary with their second order derivatives being positive are $\alpha_0 = 0.2060\pi$ and $\alpha_1 = 0.2780\pi$. By using these two orders of DFrFTs, $\mathbf{F} = \begin{bmatrix} \mathbf{F}_{\alpha_0} \\ \mathbf{F}_{\alpha_1} \end{bmatrix}$ can be constructed. As $\mathbf{F}\mathbf{x}_k = \mathbf{c}_k$ for $k = 0, \cdots, K-1$, $\mathbf{c}_k$ can be represented as the linear combination of the columns of $\mathbf{F}$ with the linear combinational coefficients being defined by $\mathbf{x}_k$. Here, the columns of $\mathbf{F}_{\alpha_0}$ and $\mathbf{F}_{\alpha_1}$ are the basis vectors of the DFrFT matrices using the DFrFTs with the orders α_0 and α_1, respectively. These basis vectors tell the corresponding relationships between the times and the frequencies of the signal. As a result, this representation exploits this important time frequency information of the signal. On the other hand, since these two values of ω are not in the set $\left\{ -\pi, -\dfrac{\pi}{2}, 0, \dfrac{\pi}{2}, \pi \right\}$, this implies that the signal with the highest sparsity is neither represented in the time domain nor in the frequency domain. Instead, the signal has the highest sparsity represented in the DFrFT domains with the orders of the DFrFTs being equal to 0.2060π and 0.2780π. Figure 2a and Figure 2b show the real parts and the imaginary parts of the coefficients of the signal represented in the time domain, respectively. This is equivalent to represent the signal in the DFrFT domain using the DFrFT with $\alpha = 0$. Figure 3a and Figure 3b show the real parts and the imaginary parts of the coefficients of the signal represented in the frequency domain, respectively. This is equivalent to represent the signal in the DFrFT domain using the DFrFT with $\alpha = \dfrac{\pi}{2}$. Figure 4a and Figure 4b show the real parts and the imaginary parts of the coefficients of the signal represented in the DFrFT domain using the DFrFT with $\alpha = 0.2060\pi$, respectively, and Figure 4c and Figure 4d show the real parts and the imaginary parts of the coefficients of the

signal represented in the DFrFT domain using the DFrFT with $\alpha = 0.2780\pi$, respectively. The total numbers of the coefficients with their magnitudes being larger than 10% of their maximum magnitudes as well as the percentages of these large value coefficients are shown in Table 1. It can be seen clearly that there are 186 real valued coefficients and 147 imaginary valued coefficients with their magnitudes being larger than 10% of their maximum magnitudes for the DFrFT with $\alpha = 0.2060\pi$, which contributes 65.0% of the total number of the coefficients. On the other hand, there are 158 real valued coefficients and 167 imaginary valued coefficients with their magnitudes being larger than 10% of their maximum magnitudes for the DFrFT with $\alpha = 0.2780\pi$, which contributes 63.5% of the total number of the coefficients. However, there are 92.2%, 78.9% and 78.9% of the coefficients with their magnitudes being larger than 10% of their maximum magnitudes for the DFrFT with $\alpha = 0$,

$\alpha = \dfrac{\pi}{2}$ and $\alpha = \pi$, respectively. From here, it can be concluded that the signals represented in the DFrFT domains with the orders being equal to 0.2060π and 0.2780π are more sparse than that represented in the time domain, the swapped time domain and in the frequency domain.
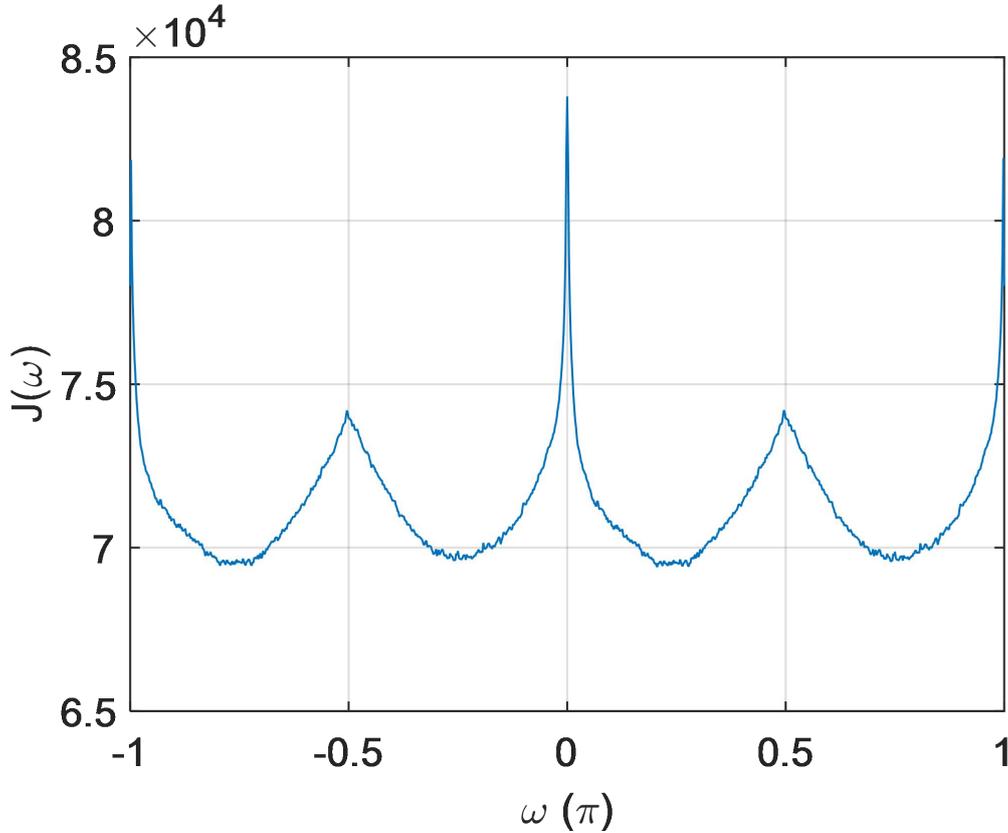
Figure 1. $J(\omega)$ with the definition of the DFrFT discussed in [23].

(a)   (b)
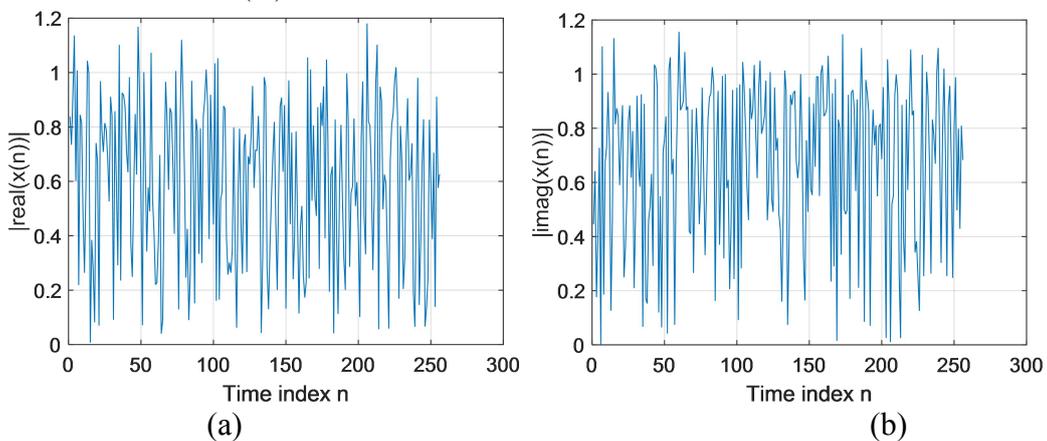
Figure 2. (a) The real part coefficients and (b) the imaginary part coefficients of the signal represented in the DFrFT domain using the DFrFT with $\alpha = 0$.
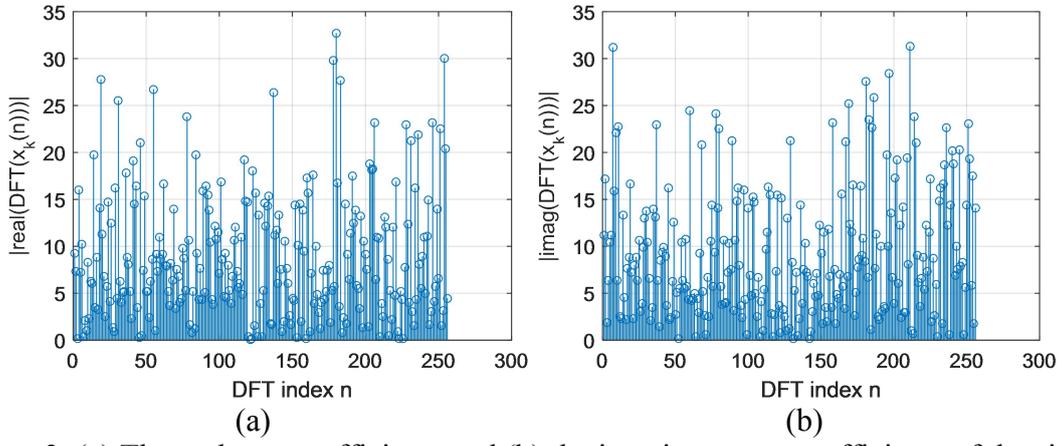
Figure 3. (a) The real part coefficients and (b) the imaginary part coefficients of the signal represented in the DFrFT domain using the DFrFT with $\alpha = \dfrac{\pi}{2}$.
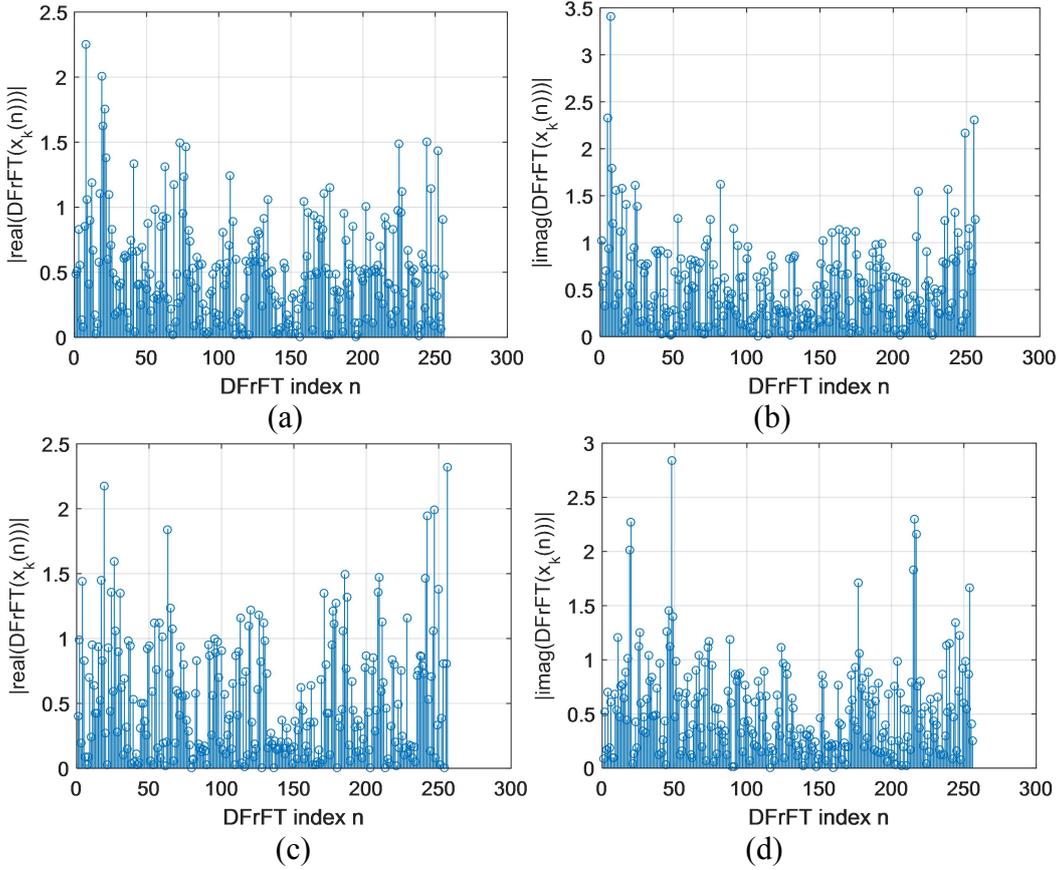
Figure 4. (a) The real part coefficients and (b) the imaginary part coefficients of the signal represented in the DFrFT domain using the DFrFT with $\alpha = 0.2060\pi$. (c) The real part coefficients and (d) the imaginary part coefficients of the signal represented in the DFrFT domain using the DFrFT with $\alpha = 0.2780\pi$.

Table 1. The total numbers of the coefficients with their magnitudes being larger than 10% of their maximum magnitudes as well as the percentages of these large value coefficients with $M = 4$.

Orders of the DFrFTs	Total numbers of large real valued coefficients	Total numbers of large imaginary valued coefficients	Total numbers of large value coefficients	Percentages of large value coefficients
$\alpha = 0$	229	243	472	92.2%
$\alpha = 0.5\pi$	203	201	404	78.9%
$\alpha = \pi$	199	205	404	78.9%
$\alpha = 0.2060\pi$	186	147	333	65.0%
$\alpha = 0.2780\pi$	158	167	325	63.5%

On the other hand, as the total number of the local minima of $J(\omega)$ is much more than 4, there are more than 4 nondifferentiable points and stationary points with their second order derivatives being positive in $\hat{\Omega}$. In fact, one may choose another even value of M which is larger than 4 and employ more than 2 DFrFT matrices with different orders to construct the overcomplete transform. Here, if $M = 6$ is chosen, then the obtained three positive values of the orders of the DFrFTs are $\alpha = 0.2300\pi$, $\alpha = 0.2060\pi$ and $\alpha = 0.2180\pi$. In this case, the total numbers of the coefficients with their magnitudes being larger than 10% of their maximum magnitudes as well as the percentages of these large value coefficients are shown in Table 2.

Table 2. The total numbers of the coefficients with their magnitudes being larger than 10% of their maximum magnitudes as well as the percentages of these large value coefficients with $M = 6$.

Orders of the DFrFTs	Total numbers of large real valued coefficients	Total numbers of large imaginary valued coefficients	Total numbers of large value coefficients	Percentages of large value coefficients
$\alpha = 0$	225	239	464	90.6%
$\alpha = 0.5\pi$	198	202	400	78.1%
$\alpha = \pi$	188	197	385	75.2%
$\alpha = 0.2300\pi$	178	169	347	67.8%
$\alpha = 0.2060\pi$	188	144	332	64.8%
$\alpha = 0.2180\pi$	147	205	352	68.8%

6. Conclusions

The objective of this paper is to design the optimal orders of the DFrFTs and construct an overcomplete transform using the DFrFTs with these orders for performing the sparse representations. The design of the orders of the DFrFTs is formulated as an optimization problem. In order to obtain a practical and a meaningful representation, the DFrFTs should have different orders. To address this difficulty, the constraint that the objective function evaluated at these locally optimal solutions is either stationary or nondifferentiable is imposed. In this case, the continuous feasible set becomes a discrete feasible set. Then, the optimal order design problem is further reformulated as the optimal frequency sampling problem. In particular, the absolute sum of both the real parts and the imaginary parts of the frequency responses of a set of filters evaluated at the sampling frequencies is minimized. By finding the roots of a set of harmonic functions and sorting the roots of these harmonic functions in the ascending order, the first M optimal orders of the DFrFTs which are either the nondifferentiable points or the stationary points with their second order derivatives being positive are taken as the M globally optimal orders of the DFrFTs. Since there is a finite number of nondifferentiable points and the stationary points, it only requires to find the roots of a finite number of harmonic functions. Also, only some orders of the DFrFTs are required to compute their second order derivatives. Therefore, the total required computational power of our proposed method is low. From the application viewpoints, as the transformed signals are represented in the DFrFT domains corresponding to the lines in the time frequency plane, the designed overcomplete transform can exploit the physical natures of the signals.

References

[1] L. B. Almeida, "The fractional Fourier transform and time-frequency representations," *IEEE Transactions on Signal Processing*, vol. 42, no. 11, pp. 3084-3091, 1994.

[2] A. Kutay, H. M. Ozaktas, O. Ankan, and L. Onural, "Optimal filtering in fractional Fourier domains," *IEEE Transactions on Signal Processing*, vol. 45, no. 5, pp. 1129-1143, 1997.

[3] Suba R. Subramaniam, Bingo Wing-Kuen Ling and Apostolos Georgakis, "Filtering in rotated time-frequency domains with unknown noise statistics," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 489-493, 2012.

[4] Xiao-Zhi Zhang, Bingo Wing-Kuen Ling, Suba R. Subramaniam and Apostolos Georgakis, "Globally optimal joint design of two sets of mask coefficients in two different predefined rotated axes of time frequency plane for signal restoration applications," *Digital Signal Processing*, vol. 40, pp. 31-39, 2015.

[5] C. Vijaya and J. S. Bhat, "Signal compression using discrete fractional Fourier transform and set partitioning in hierarchical tree," *Signal Processing*, vol. 86, pp. 1976-1983, 2006.

[6] Ayush Bhandari and Pina Marziliano, "Sampling and reconstruction of sparse signals in fractional Fourier domain," *IEEE Transactions on Signal Processing*, vol. 17, no. 3, pp. 221-224, 2010.

[7] Deyun Wei and Yuan-Min Li, "Generalized sampling expansions with multiple sampling rates for lowpass and bandpass signals in the fractional Fourier transform domain," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4861-4874, 2016.

[8] Ran Tao, Bing Deng, Wei-Qiang Zhang and Yue Wang, "Sampling and sampling rate conversion of band limited signals in the fractional Fourier transform domain," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 158-171, 2008.

[9] Igor Djurovic, Srdjan Stankovic and Ioannis Pitas, "Digital watermarking in the fractional Fourier transformation domain," *Journal of Network and Computer Applications*, vol. 24, pp. 167-173, 2001.

[10] Bryan Hennelly and John T. Sheridan, "Optical image encryption by random shifting in fractional Fourier domains," *Optics Letters*, vol. 28, no. 4, pp. 269-71, 2003.

[11] Vicente Parot, Carlos Sing-Long, Carlos Lizama, Cristian Tejos, Sergio Uribe and Pablo Irarrazaval, "Application of the fractional Fourier transform to image reconstruction in MRI," *Magnetic Resonance in Medicine*, vol. 68, pp. 17-29, 2012.

[12] I. Samil Yetik, M. Alper Kutay and Haldun M. Ozaktas, "Image representation and compression with the fractional Fourier transform," *Optics communications*, vol. 197, pp. 275-278, 2001.

[13] C. K. Sarseena and R. B. Yadhu, "Fractional Fourier domain MRI reconstruction using compressive sensing under different random sampling scheme," *International Journal of Emerging Science and Engineering*, vol. 2, no. 6, 2004.

[14] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization", *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197-2202, 2003.

[15] Shengheng Liu, Tao Shan, Ran Tao, Yimin D. Zhang, Guo Zhang, Feng Zhang and Yue Wang, "Sparse discrete fractional Fourier transform and its applications," *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6582-6595, 2014.

[16] Ljubisa Stankovic, Irena Orovic, Srdjan Stankovic, Moeness Amin, "Compressive sensing based separation of nonstationary and stationary signals overlapping in time-frequency," *IEEE Transactions on Signal Processing*, vol. 61, no. 18, pp. 4562-4572, 2013.

[17] Zhijing Yang, Chunmei Qing, Bingo Wing-Kuen Ling, Wai Lok Woo and Saeid Sanei, "Optimal overcomplete kernel design for sparse representations via discrete fractional Fourier transforms," *International Symposium on Communication Systems, Networks & Digital Signal Processing*, pp. 1-4, 2012.

[18] HongXia Bu, Xia Bai and Ran Tao, "Compressed sensing SAR imaging based on sparse representation in fractional Fourier domain," *Science China, Information Sciences*, vol. 55, no. 8, pp. 1789-1800, 2012.

[19] Mostafa Sadeghi, Massoud Babaie-Zadeh and Christian Jutten, "Dictionary learning for sparse representation: a novel approach," *IEEE Signal Processing Letters*, vol. 20, no. 12, pp. 1195-1198, 2013.

[20] Michal Aharon, Michael Elad and Alfred Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311-4322, 2006.

[21] C. Y. F. Ho, B. W. K. Ling, L. Benmesbah, T. C. W. Kok, W. C. Siu and K. L. Teo, "Two-channel linear phase FIR QMF bank minimax design via global nonconvex optimization programming," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4436-4441, 2010.

[22] C. Y. F. Ho, B. W. K. Ling, Z. W. Chi, M. Shikh-Bahaei, Y. Q. Liu and K. L. Teo, "Design of near-allpass strictly stable minimal-phase real-valued rational IIR filters," *IEEE Transactions on Circuits and Systems—II: Express Briefs*, vol. 55, no. 8, pp. 781-785, 2008.

[23] Cagatay Candan, M. Alper Kutay and Haldun M. Ozaktas, "The discrete fractional Fourier transform," *IEEE Transactions on Signal Processing*, vol. 48, no. 5, pp. 1329-1337, 2000.

[24] Ahmet Serbes and Lutfiye Durak-Ata, "The discrete fractional Fourier transform based on the DFT matrix," *Signal Processing*, vol. 91, pp. 571-581, 2011.

[25] Yuanqing Li and Shun-ichi Amari, "Two conditions for equivalence of 0-norm solution and 1-norm solution in sparse representation," *IEEE Transactions on Neural Networks*, vol. 21, no. 7, pp. 1189-1196, 2010.

[26] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797-829, 2006.

[27] Sheldon Axler, Paul Bourdon and Wade Ramey, *Harmonic Function Theory*, Springer-Verlag New York, 2001.

[28] Antonio Lerario and Erik Lundberg, "On the zeros of random harmonic polynomials: the truncated model," *Journal of Mathematical Analysis and Applications*, vol. 438, pp. 1041-1054, 2016.

[29] Victor Y. Pan, "Solving a polynomial equation: some history and recent progress," *SIAM Review*, vol. 39, no. 2, pp. 187-220, 1997.