PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Designing nearly tight window for improving time-frequency masking

Tsubasa KUSANO[(1)]; Yoshiki MASUYAMA[(2)]; Kohei YATABE[(3)]; Yasuhiro OIKAWA[(4)]

[(1)]Waseda University, Japan, tsubasa.k@suou.waseda.jp
[(2)]Waseda University, Japan, mas-03151102@akane.waseda.jp
[(3)]Waseda University, Japan, k.yatabe@asagi.waseda.jp
[(4)]Waseda University, Japan, yoikawa@waseda.jp

Abstract

Many audio signal processing methods are formulated in the time-frequency (T-F) domain which is obtained by the short-time Fourier transform (STFT). The properties of the STFT are fully characterized by window function, number of frequency channels, and time-shift. Thus, designing a better window is important for improving the performance of the processing especially when a less redundant T-F representation is desirable. While many window functions have been proposed in the literature, they are designed to have a good frequency response for analysis, which may not perform well in terms of signal processing. The window design must take the effect of the reconstruction (from the T-F domain into the time domain) into account for improving the performance. In this paper, an optimization-based design method of a nearly tight window is proposed to obtain a window performing well for the T-F domain signal processing.

Keywords: Discrete Gabor transform (DGT), Short-time Fourier transform (STFT), Window design, Speech enhancement, Non-convex optimization.

1 INTRODUCTION

Many audio signal processing methods are formulated as modifications of the signal in the time-frequency (T-F) domain, which is often called T-F masking. For converting the signal into the T-F domain, the short-time Fourier transform (STFT) [1][1] is usually utilized owing to its simplicity and easily understandable structure [2–9]. While most of the research has concentrated on the method of modification in the T-F domain (the way how to construct a T-F mask), the method of converting a signal into the T-F domain is also important for improving the performance of processing.

When STFT is considered as the conversion to the T-F domain, its property is fully characterized by the window function since STFT is a highly structured transform. Aiming to obtain a better T-F representation, many window functions have been proposed to improve their frequency responses [10–16]. For example, the Hann window is one popular window which has a good sidelobe decay. The Nuttall window was proposed to achieve a better sidelobe decay, while the Kaiser window was proposed so that its frequency response was adjustable by a tuning parameter. Although such research on window functions has provided a better T-F representation, most research only has considered the *analysis* side. That is, there is little research on window functions considering the *reconstruction* point of view.

To realize processing in the T-F domain, the signal must be reconstructed back into the time domain after the T-F-domain processing. The reconstruction side of STFT is achieved by the (pseudo-) inverse STFT which also involves a window function. Therefore, for the T-F domain signal processing, a window function must be chosen in accordance with not only STFT but also the inverse STFT. Indeed, incorrect choice of the pair of window functions (for STFT and the inverse STFT) makes reconstruction impossible. To allow a reasonable window for the reconstruction, the T-F representation is usually chosen to be *redundant*, and the error-minimizing window called the canonical dual window is often used for the reconstruction (see Section 2).

For some applications favoring less redundant T-F representation, choice of the window function is more critical for the reconstruction (and thus, critical for processing). One example is T-F masking in low-power devices which allow a little computation [17, 18]. In such cases, redundancy should be lowered because higher redundancy directly results in higher computational cost. Another very important example is speech enhancement based on deep learning. Recent study has shown that non-redundant T-F representation can improve

[1] STFT is also often called the Gabor transform based on [1]. Note that some literature strictly distinguishes STFT from the Gabor transform by their mapping properties [2], while others do not. In this paper, we may utilize the term "STFT" in the sense of the discrete Gabor transform (DGT), which is a common habit especially in the acoustical signal processing community.

the performance of enhancement using deep neural networks (DNN) [19]. This is because less redundant T-F representation reduces the number of parameters to be learned, which makes the training easier. For those applications, redundancy of STFT should be lowered by increasing the window shifting width. However, the inverse transform becomes more sensitive to the error of signal processing when the redundancy is reduced (see Section 2.3), which also degrades the performance. Although there exists a type of window insensitive to such processing error, called *tight* window, it has a drawback that its frequency response is often poor (sidelobe level is high). Therefore, a window function which is less sensitive to processing error and, at the same time, has a good frequency response is desired for realizing a better processing in a less redundant situation.

In this paper, we propose a window design method to simultaneously meet both requirements. It aims to make a window function closer to a tight window, while its frequency response is constrained to be better. Since the designed window is not strictly tight, we call it *nearly tight* window. The proposed method is formulated as an optimization problem so that it can easily control the trade-off between the two requirements, and it is solved by the linearized alternating direction method of multipliers (ADMM).

2 PRELIMINARIES

While the discrete and downsampled T-F transform is called "STFT" in acoustical signal processing, the literature of T-F analysis calls it the discrete Gabor transform (DGT) [2]. Hereafter, we will utilize the language used in DGT to express the T-F representation because it will be easier for explaining the proposed method.

2.1 Gabor system and discrete Gabor transform (DGT)

Let a window be denoted by $\mathbf{g} = [\mathbf{g}[0], \mathbf{g}[1], \ldots, \mathbf{g}[L-1]]^T \in \mathbb{R}^L$. DGT is a T-F transform based on a collection of windowed sinusoids,

$$\mathcal{G}(\mathbf{g}, a, M) = \{\mathbf{g}_{m,n}\}_{m=0,\ldots,M-1,\ n=0,\ldots,N-1}\,, \tag{1}$$

which is called the Gabor system, where $a \in \mathbb{N}$ is the time-shifting width, $M \in \mathbb{N}$ is the number of frequency channels,

$$\mathbf{g}_{m,n}[l] = \mathrm{e}^{\mathrm{i}\frac{2\pi ml}{M}}\mathbf{g}[l - an], \tag{2}$$

is a windowed complex sinusoid, and $\mathrm{i} = \sqrt{-1}$. DGT of a discrete signal $\mathbf{f} \in \mathbb{R}^L$ is defined by the following inner product:

$$(\mathbf{G}_\mathbf{g}\mathbf{f})[m + nM] = \langle \mathbf{f}, \mathbf{g}_{m,n}\rangle = \sum_{l=0}^{L-1} \mathbf{f}[l]\,\overline{\mathbf{g}_{m,n}[l]}, \tag{3}$$

where $\overline{x}$ is the complex conjugate of x, and $\mathbf{G}_\mathbf{g} \in \mathbb{C}^{MN\times L}$ is the matrix consisting of all the elements in the Gabor system in Eq. (1). That is, multiplying $\mathbf{G}_\mathbf{g}$ to a signal obtains the vectorized version of its T-F representation which is often called "spectrogram."

2.2 Reconstruction of time-domain signal from T-F domain

A system $\mathcal{G}(\mathbf{g}, a, M)$ is said to be a *frame* [20, 21] if there exist $0 < A, B < \infty$ such that

$$A\,\|\mathbf{f}\|_2^2 \le \sum_{m,n} |\langle \mathbf{f}, \mathbf{g}_{m,n}\rangle|^2 \le B\,\|\mathbf{f}\|_2^2\,, \tag{4}$$

for all $\mathbf{f} \in \mathbb{R}^L$, where $\|\cdot\|_p$ is the ℓ_p norm. A and B are called the lower and upper frame bound, respectively. If the Gabor system is a frame, a time-domain signal can be reconstructed from its T-F domain representation. The inverse DGT, reconstructing a signal from its coefficients $\mathbf{c} \in \mathbb{C}^{MN}$, with respect to $\mathcal{G}(\mathbf{g}, a, M)$ is defined by

$$\mathbf{f}_{\text{syn}} = \sum_{n,m} \mathbf{c}[m + nM]\,\mathbf{g}_{m,n} = \mathbf{G}_\mathbf{g}^*\mathbf{c}, \tag{5}$$

where $\mathbf{G}_\mathbf{g}^*$ denotes the complex-conjugate transpose of $\mathbf{G}_\mathbf{g}$. If a Gabor system $\mathcal{G}(\mathbf{g}, a, M)$ is a frame, then there exists the corresponding dual Gabor frame $\mathcal{G}(\mathbf{h}, a, M) = \{\mathbf{h}_{m,n}\}$ which satisfies

$$\mathbf{f} = \sum_{n,m} \langle \mathbf{f}, \mathbf{g}_{m,n}\rangle \mathbf{h}_{m,n}, \tag{6}$$

where $\mathbf{h}_{m,n}[l] = \mathrm{e}^{\mathrm{i}\frac{2\pi ml}{M}}\mathbf{h}[l-an]$, and $\mathbf{h}$ is a dual window of $\mathbf{g}$. That is, a time-domain signal can be reconstructed if (1) $\mathcal{G}(\mathbf{g}, a, M)$ is a frame, and (2) $\mathbf{h}$ is a dual window of $\mathbf{g}$. These conditions are decided by the window pair $\mathbf{g}, \mathbf{h}$, the time-shifting width a, and the number of frequency channels M.

When a Gabor system $\mathcal{G}(\mathbf{g}, a, M)$ is redundant, the corresponding dual window $\mathbf{h}$ is not unique, and infinitely many variation of $\mathbf{h}$ can satisfy the reconstruction formula, Eq. (6). One standard choice among all possible dual windows is the canonical dual window

$$\tilde{\mathbf{g}} = \mathbf{S}_{\mathbf{g}}^{-1}\mathbf{g}, \tag{7}$$

where $\mathbf{S}_{\mathbf{g}} = \mathbf{G}_{\mathbf{g}}^{*}\mathbf{G}_{\mathbf{g}}$ is the so-called frame operator defined as

$$\mathbf{S}_{\mathbf{g}}\mathbf{f} = \sum_{m,n}\langle \mathbf{f}, \mathbf{g}_{m,n}\rangle \mathbf{g}_{m,n} = \left(\mathbf{G}_{\mathbf{g}}^{*}\mathbf{G}_{\mathbf{g}}\right)\mathbf{f}. \tag{8}$$

The canonical dual window is optimal in the sense that its synthesis operator corresponds to the Moore–Penrose pseudo-inverse:

$$\sum_{n,m}\mathbf{c}[m+nM]\,\tilde{\mathbf{g}}_{m,n} = \sum_{n,m}\mathbf{c}[m+nM]\,\mathbf{S}_{\mathbf{g}}^{-1}\mathbf{g}_{m,n} = \left(\mathbf{G}_{\mathbf{g}}^{*}\mathbf{G}_{\mathbf{g}}\right)^{-1}\mathbf{G}_{\mathbf{g}}^{*}\mathbf{c}. \tag{9}$$

In this paper, the canonical dual window is considered for inverse DGT as it is the standard choice in acoustical signal processing. One reason for such popularity should be because of the optimality to the following least squares signal reconstruction problem:

$$\underset{\mathbf{x}}{\text{minimize}} \quad \|\mathbf{G}_{\mathbf{g}}\mathbf{x} - \hat{\mathbf{c}}\|_2^2, \tag{10}$$

whose solution is $\mathbf{G}_{\tilde{\mathbf{g}}}^{*}\hat{\mathbf{c}}$ as can be confirmed from the fact in Eq. (9).

2.3 Influence of window functions on signal processing

A signal processing framework in T-F domain is illustrated in Figure 1. In words, some processing is performed in the T-F domain to modify the Gabor coefficient $\mathbf{c}$ to $\hat{\mathbf{c}}$, and then the inverse DGT is applied to obtain the processed result $\hat{\mathbf{f}}$. While the quality of the processing is important for obtaining a good result, the transformation pair, DGT and inverse DGT, is also important since it decides the coefficient $\mathbf{c}$.

To see the effect of the window pair in terms of T-F domain signal processing, a preliminary experiment was performed. 200 speech signals [22] from TIMIT database [23] were degraded by adding the Gaussian noise in the time domain so that the signal-to-noise ratio (SNR) became 0 dB. They were enhanced by the Wiener filter (T-F masking based on the power ratio of noisy and clean signals) with a minimum mean-square error (MMSE) estimator of noise power [8] and the decision-directed approach [6]. The redundancy was changed by changing a while fixing the window length to 256 and $M = 256$. Its performance was compared with the ideal Wiener filter and the condition number of $\mathbf{G}_{\mathbf{g}}$,

$$\kappa(\mathbf{G}_{\mathbf{g}}) = \sigma_{\max}(\mathbf{G}_{\mathbf{g}}) \,/\, \sigma_{\min}(\mathbf{G}_{\mathbf{g}}) = \sqrt{B\,/A}, \tag{11}$$

which is the standard measure of numerical stability of Eq. (9), where $\sigma_{\max}(\mathbf{G}_{\mathbf{g}})$ and $\sigma_{\min}(\mathbf{G}_{\mathbf{g}})$ denote the maximum and minimum singular value of $\mathbf{G}_{\mathbf{g}}$, respectively.

Three types of window functions were utilized for comparison: the Hann window, the Kaiser window ($\alpha = 10$), and the canonical tight window of the Kaiser window. A window $\mathbf{g}_{\mathrm{T}}$ is said to be *tight* if its canonical dual window is itself (i.e., self-dual) [24]. Then,

$$\mathbf{S} = \mathbf{G}_{\mathbf{g}_{\mathrm{T}}}^{*}\mathbf{G}_{\mathbf{g}_{\mathrm{T}}} = A\,\mathbf{I} \tag{12}$$

holds, where $\mathbf{I}$ is the identity. Thus, the condition number of a tight window is always 1. Particularly, a tight window with $A = 1$ is called the Parseval tight window. The canonical tight window of a window $\mathbf{g}$ can be obtained by inverting square root of the frame operator:

$$\mathbf{g}_{\mathrm{T}} = \mathbf{S}_{\mathbf{g}}^{-\frac{1}{2}}\mathbf{g}, \tag{13}$$

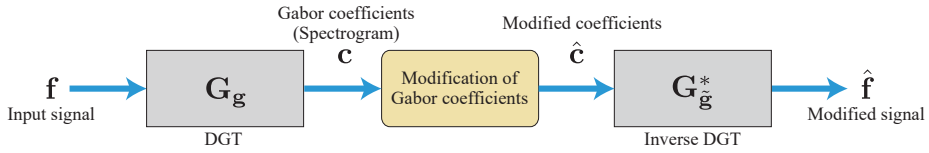

Figure 1. Framework of the signal processing in the T-F domain.

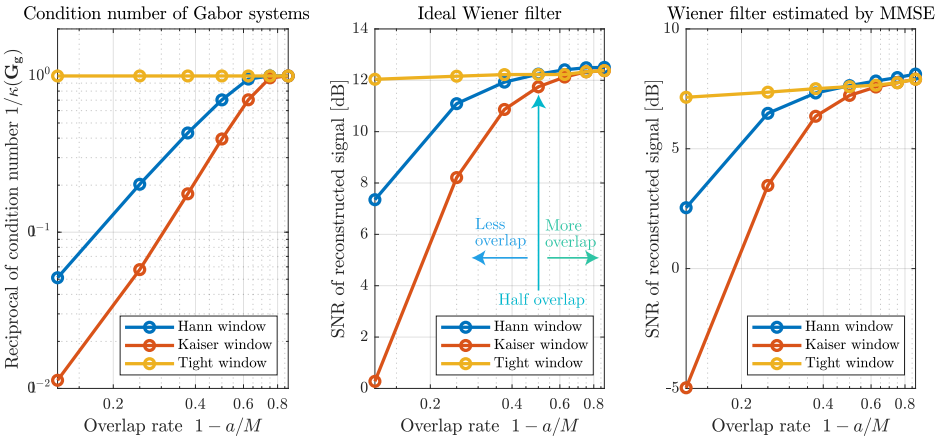

Figure 2. (left) Condition number of the DGT matrix $\kappa(\mathbf{G_g})$. (center) Denoising result using the ideal Wiener filter. (right) Denoising result using the Wiener filter with MMSE noise power estimation.

which corresponds to the solution of the following problem [25]:

$$\underset{\mathbf{x}\in\mathcal{T}}{\text{minimize}} \quad \|\mathbf{g}-\mathbf{x}\|_2 \,, \tag{14}$$

where $\mathcal{T}$ is the set of all Parseval tight windows. Thus, the canonical tight window is the closest Parseval tight window from the window $\mathbf{g}$.

Results of the experiment are shown in Figure 2, where SNR is the average among all speech signals. For both ideal and realistic Wiener filters (center and right), the processing performances of the Hann and Kaiser windows were degraded as the redundancy decreased (horizontal axes are related to the redundancy). In contrast, the performance of the tight Kaiser window was not degraded much. These results can be predicted from the condition numbers (left). Based on this experiment, a window $\mathbf{g}$ should be designed so that the condition number $\kappa(\mathbf{G_g})$ becomes lower. A tight window is the best window in this sense because its condition number is the lowest. However, as in the figure, a tight window is not always the best in terms of processing, which is because the frequency response of a tight window is usually not better than one of a non-tight window (see Figure 3). Therefore, in this paper, a design method of a *nearly tight* window is proposed so that the condition number is lowered while its frequency response is kept well.

2.4 Related works on Gabor window design

For designing a low-condition-numbered window, design methods of tight windows have been proposed [26,27]. These methods aim to find a tight window with better frequency responses. However, since the constraint to the tight window greatly limits the set of variables, desired characteristics may not be obtained.

On the other hand, some methods of nearly-tight window design have been proposed [28–30]. One approach of this research is to minimize the difference between the frame operator and identity operator using the gradient-based optimization [28, 29]. These methods minimize the distance to the set of tight windows by the gradient method, whereby they have a possibility of falling into the local minima. Another approach is to replace the non-convex cost of measuring the distance to the tight window with convex functions [30]. Since that method is formulated as convex optimization, it is guaranteed that globally optimal solutions can be obtained, so a trade-off between the condition number and the frequency response can be easily considered. However,

as a result of approximating the cost function, the obtained solutions may not be close to the original solution which is tight. The cost should be reduced strictly without approximation, while the trade-off should be easily adjusted.

3 PROPOSED METHOD

In this section, we propose a design method of nearly tight window that can easily control the trade-off between the desired frequency response and the condition number. At first, we formulate the nearly tight Gabor window design as a constrained minimization problem. Then, an algorithm solving this problem through the proximal operators is introduced. Since a window whose support is shorter than the signal length is used in most signal processing, the formulation considers $\mathbf{g}[l] = 0$ for $l = K, \dots, L-1$, i.e., only $\mathbf{g}[l]$ $(l = 0, \dots, K-1)$ are treated as the variables in this paper.

3.1 Problem formulation for designing nearly tight window

To propose an easily adjustable window function design, the desired frequency response is considered as a constraint, and the window is made closer to a tight window as possible. Its direct formulation is

$$\underset{\mathbf{g}\in\mathcal{C}}{\text{minimize}} \quad \frac{1}{2} d_{\mathcal{T}}^2(\mathbf{g}), \tag{15}$$

where $d_{\mathcal{T}}(\mathbf{g})$ is the distance to the set of Parseval tight windows $\mathcal{T}$,

$$d_{\mathcal{T}}(\mathbf{g}) = \min_{\mathbf{x}\in\mathcal{T}} \|\mathbf{g} - \mathbf{x}\|_2, \tag{16}$$

and $\mathcal{C}$ is the set of windows satisfying the desired frequency response. Since the magnitude response should be considered in decibels for audio applications, a popular choice for the set $\mathcal{C}$ to constrain the frequency response into desired one, in filter design [31], is

$$\mathcal{C} = \{\mathbf{g} \in \mathbb{R}^K \mid \| \log_{10} |\tilde{\mathbf{F}}\mathbf{g}| - \log_{10} \mathbf{d}\|_\infty \le \log_{10} \beta\}, \tag{17}$$

where $\mathbf{d} \in \mathbb{R}_+^{\tilde{K}}$ $(\tilde{K} \ge K)$ is magnitude of the desired frequency response, $\tilde{\mathbf{F}} \in \mathbb{C}^{\tilde{K}\times K}$ is the zero-padded discrete Fourier transform,

$$\tilde{\mathbf{F}}[m,n] = \frac{1}{\sqrt{\tilde{K}}} \mathrm{e}^{-\mathrm{i}\frac{2\pi mn}{\tilde{K}}}, \tag{18}$$

and $\beta \ge 1$ is a parameter for controlling the amount of error. However, directly treating this constraint is not easy because taking difference after absolute value results in the non-convex set.

Since the requirement in window design (in contrast to filter design) is to lower the sidelobe level towards zero (i.e., increasing the magnitude of sidelobe is usually not desired), it should be sufficient to constrain only the upper bound. Based on this observation,

$$\tilde{\mathcal{C}} = \{\mathbf{g} \in \mathbb{R}^K \mid |(\tilde{\mathbf{F}}\mathbf{g})[n]| \le \beta \mathbf{d}[n] \ \text{ for } n = 0, \dots, N-1\}, \tag{19}$$

is considered as the constraint set instead of Eq. (17). Consequently, our formulation becomes a minimization problem on the convex set:

$$\underset{\mathbf{g}\in\tilde{\mathcal{C}}}{\text{minimize}} \quad \frac{1}{2} d_{\mathcal{T}}^2(\mathbf{g}). \tag{20}$$

This model directly handles the distance function $d_{\mathcal{T}}$ instead of approximation as in [30], while the desired frequency response is strictly imposed by the constraint $\tilde{\mathcal{C}}$ as opposed to [28, 29].

3.2 Algorithm for solving problem using linearized ADMM

To solve Eq. (20), linearized ADMM [32–34] is utilized in this paper. It is an algorithm solving problems written in the following form:

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathscr{F}(\mathbf{x}) + \mathscr{G}(\mathbf{A}\mathbf{x}), \tag{21}$$

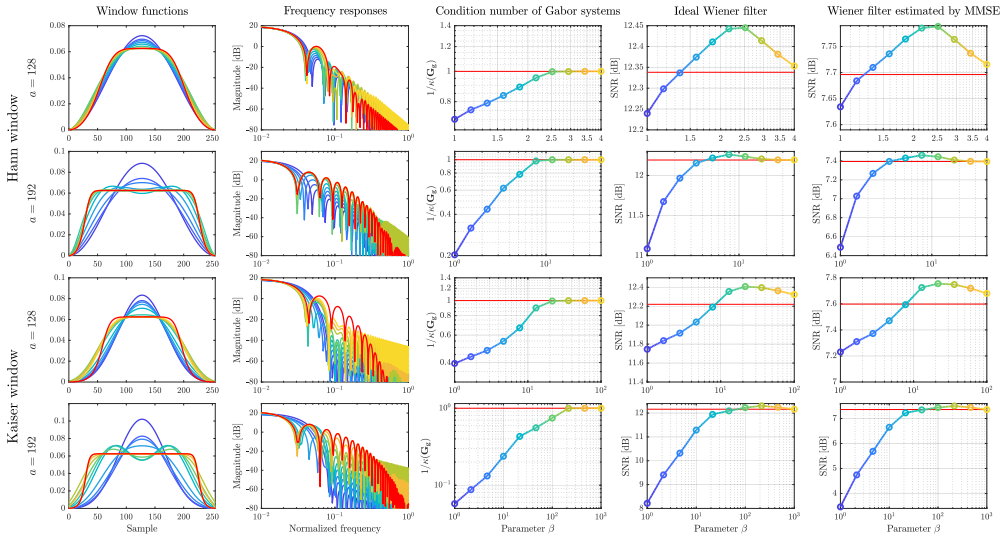

Figure 3. Designed nearly tight windows by the proposed method. Each column shows (from left to right) the obtained window shapes, their frequency responses, their condition numbers, denoising results for the ideal Wiener filter, and those for the Wiener filter with MMSE noise power estimation. Each row shows (from top to bottom) the results of the Hann-based windows for $a = 128$, 192, and those of the Kaiser-based windows ($\alpha = 10$) for $a = 128$, 192. The transition of colors from blue to yellow represents a change in parameter β of the proposed method, where the blue represents $\mathbf{g}^{\mathrm{o}}$ and brighter color (larger β) means closer to tight. Red lines indicate the canonical tight window of $\mathbf{g}^{\mathrm{o}}$.

where $\mathscr{F}(x)$ and $\mathscr{G}(x)$ are proper lower semi-continuous functions, and $\mathbf{A}$ is a linear operator. By using the proximity operator [34],

$$\mathrm{prox}_{\rho\mathscr{F}}(\mathbf{x}) = \underset{\mathbf{y}}{\mathrm{argmin}} \left\{ \mathscr{F}(\mathbf{y}) + \frac{1}{2\rho}\|\mathbf{y}-\mathbf{x}\|_2^2 \right\}, \tag{22}$$

the linearized ADMM algorithm is given as the following procedure:

$$\mathbf{x}^{[k+1]} = \mathrm{prox}_{\mu\mathscr{F}}\Big(\mathbf{x}^{[k]} - \frac{\mu}{\lambda}\mathbf{A}^*(\mathbf{A}\mathbf{x}^{[k]} - \mathbf{z}^{[k]} + \mathbf{u}^{[k]})\Big), \tag{23}$$

$$\mathbf{z}^{[k+1]} = \mathrm{prox}_{\lambda\mathscr{G}}(\mathbf{A}\mathbf{x}^{[k+1]} + \mathbf{u}^{[k]}), \tag{24}$$

$$\mathbf{u}^{[k+1]} = \mathbf{u}^{[k]} + \mathbf{A}\mathbf{x}^{[k+1]} - \mathbf{z}^{[k+1]}, \tag{25}$$

where λ and μ are real numbers satisfying $0 < \mu \leq \lambda/\|\mathbf{A}\|_{\mathrm{op}}^2$, and $\|\cdot\|_{\mathrm{op}}$ is the operator norm.

For applying this linearized ADMM algorithm to Eq. (20), it is rewritten as the equivalent problem having the form of Eq. (21):

$$\underset{\mathbf{g}}{\text{minimize}} \quad \frac{1}{2}d_{\mathcal{T}}^2(\mathbf{g}) + \iota(\tilde{\mathbf{F}}\mathbf{g}), \tag{26}$$

where $\iota(\mathbf{z})$ is the indicator function corresponding to Eq. (19),

$$\iota(\mathbf{z}) = \begin{cases} 0 & (|\mathbf{z}[n]| \leq \beta\mathbf{d}[n] \ \text{ for } n = 0, \ldots, N-1) \\ \infty & (\text{otherwise}) \end{cases}. \tag{27}$$

Then, Eq. (26) is solved by iterating the following procedure:

$$\mathbf{g}^{[k+1]} = \mathrm{prox}_{\frac{\mu}{2}d_{\mathcal{T}}^2}\left(\mathbf{g}^{[k]} - \frac{\mu}{\lambda}\tilde{\mathbf{F}}^*(\tilde{\mathbf{F}}\mathbf{g}^{[k]} - \mathbf{z}^{[k]} + \mathbf{u}^{[k]})\right), \tag{28}$$

$$\mathbf{z}^{[k+1]} = \mathrm{prox}_{\iota}(\tilde{\mathbf{F}}\mathbf{g}^{[k+1]} + \mathbf{u}^{[k]}), \tag{29}$$

$$\mathbf{u}^{[k+1]} = \mathbf{u}^{[k]} + \tilde{\mathbf{F}}\mathbf{g}^{[k+1]} - \mathbf{z}^{[k+1]}, \tag{30}$$

where $\mathrm{prox}_{\frac{\mu}{2}d_{\mathcal{T}}^2}(\cdot)$ and $\mathrm{prox}_{\iota}(\cdot)$ in Eqs. (28) and (29) are given by

$$\mathrm{prox}_{\frac{\mu}{2}d_{\mathcal{T}}^2}(\mathbf{g}) = \frac{1}{1+\mu}\mathbf{g} + \frac{\mu}{1+\mu}\mathbf{S}_{\mathbf{g}}^{-\frac{1}{2}}\mathbf{g}, \tag{31}$$

$$\mathrm{prox}_{\iota}(\mathbf{z})[n] = \min\left\{\beta\frac{\mathbf{d}[n]}{|\mathbf{z}[n]|}, 1\right\}\mathbf{z}[n]. \tag{32}$$

Thanks to the property of the canonical tight window in Eq. (14), Eq. (31) can be expected to give an appropriate descent direction even though the cost function $d_{\mathcal{T}}^2$ is non-convex. Therefore, this algorithm should be able to effectively manage the difficulty associated with the non-convexity of $d_{\mathcal{T}}^2$.

4 NUMERICAL EXPERIMENTS

The shapes, frequency responses and condition numbers of the windows designed by the proposed method were compared with the denoising performance provided by the same experiment in Section 2.3 using ideal and MMSE Wiener filters. For the initial window inputted to the algorithm, the Hann and Kaiser windows, whose energies were normalized to a/M, were chosen in accordance with Section 2.3. By iterating the algorithm from these windows denoted by $\mathbf{g}^{\mathrm{o}}$, the designed windows are expected to have characteristics similar to $\mathbf{g}^{\mathrm{o}}$ with a better condition number. The frequency responses $\mathbf{d}$ for the constraint set $\tilde{\mathcal{C}}$ were constructed by interpolating the maxima of $\log_{10}|\tilde{\mathbf{F}}\mathbf{g}^{\mathrm{o}}|$ by the cubic C^2-splines.

The obtained nearly tight windows by the proposed method and the denoising results for $a = 128, 192$ are summarized in Figure 3. When the parameter β was set to a higher value (brighter color), then the obtained windows got closer to a tight window, which can be confirmed by the condition numbers. Note that the canonical tight window has the highest level of the first side lobe which may prevent a denoising method to be work correctly. It can be seen that some windows obtained by the proposed method outperformed both the original window (blue) and the canonical tight window (red) in terms of the denoising results. These results indicate that the proposed method can design a window having better characteristics for T-F domain signal processing than the original and the canonical tight window. The performance was adjustable by the single parameter β, which enables to look for a better window by a simple line search.

5 CONCLUSION

In this paper, the nearly tight window designing method for signal processing in the T-F domain is proposed. The proposed method can obtain nearly tight windows having desired frequency responses, which can result in a better performance of T-F masking than those of original and canonical tight windows. Future work includes the automatic adjustment of β as well as the generalization of the method.

REFERENCES

[1] D. Gabor, "Theory of communication," J. Inst. Electr. Eng. **93**, 429–457 (1946).

[2] H. G. Feichtinger and T. Strohmer, *Gabor Analysis and Algorithms: Theory and Applications* (Birkhäuser Boston, Boston, MA, 1998).

[3] D. F. Walnut, "Continuity properties of the Gabor frame operator," J. Math. Anal. Appl. **165**, 479–504 (1992).

[4] P. L. Søndergaard, "Efficient algorithms for the discrete Gabor transform with a long f window," J. Fourier Anal. Appl. **18**, 456–470 (2012).

[5] S. Moreno-Picot, F. J. Ferri, M. Arevalillo-Herráez, and W. Díaz-Villanueva, "Efficient analysis and synthesis using a new factorization of the Gabor frame matrix," IEEE Trans. Signal Process. **66**, 4564–4573 (2018).

[6] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **32**, 1109–1121 (1984).

[7] Ö. Yılmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," IEEE Trans. Signal Process. **52**, 1830–1847 (2004).

[8] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," IEEE Trans. Audio, Speech, Lang. Process. **20**, 1383–1393 (2012).

[9] K. Yatabe, Y. Masuyama, T. Kusano, and Y. Oikawa, "Representation of complex spectrogram via phase conversion," Acoust. Sci. Technol. **40**, 170–177 (2019).

[10] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—I," Bell Syst. Tech. J. **40**, 43–63 (1961).

[11] J. Kaiser and R. Schafer, "On the use of the I_0-sinh window for spectrum analysis," IEEE Trans. Acoust., Speech, Signal Process. **28**, 105–107 (1980).

[12] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," Proc. IEEE **66**, 51–83 (1978).

[13] A. Nuttall, "Some windows with very good sidelobe behavior," IEEE Trans. Acoust., Speech, Signal Process. **29**, 84–91 (1981).

[14] J. W. Adams, "A new optimal window (signal processing)," IEEE Trans. Signal Process. **39**, 1753–1769 (1991).

[15] K. F. C. Yiu, M. J. Gao, T. J. Shiu, S. Y. Wu, T. Tran, and I. Claesson, "A fast algorithm for the optimal design of high accuracy windows in signal processing," Optim. Method. Softw. **28**, 900–916 (2013).

[16] H. Kawahara, K. Sakakibara, M. Morise, H. Banno, T. Toda, and T. Irino, "A new cosine series antialiasing function and its application to aliasing-free glottal source models for speech and singing synthesis," in "Interspeech 2017," (2017), pp. 1358–1362.

[17] M. Jeub, C. Herglotz, C. Nelke, C. Beaugeant, and P. Vary, "Noise reduction for dual-microphone mobile phones exploiting power level differences," in "IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)," (2012), pp. 1693–1696.

[18] M. Parchami, W.-P. Zhu, B. Champagne, and E. Plourde, "Recent developments in speech enhancement in the short-time Fourier transform domain," IEEE Circuits Syst. Mag. **16**, 45–77 (2016).

[19] Y. Koizumi, N. Harada, Y. Haneda, Y. Hioka, and K. Kobayashi, "End-to-end sound source enhancement using deep neural network in the modified discrete cosine transform domain," in "IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)," (2018), pp. 706–710.

[20] I. Daubechies, *Ten Lectures on Wavelets* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992).

[21] O. Christensen, *An introduction to frames and Riesz bases* (Birkhäuser Boston, Boston, MA, 2003).

[22] P. Mowlaee, J. Kulmer, J. Stahl, and F. Mayer, *Single Channel Phase-Aware Signal Processing in Speech Communication: Theory and Practice* (Wiley, Hoboken, NJ, USA, 2016).

[23] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, *DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM* (NIST, 1993).

[24] Z. Cvetković and M. Vetterli, "Tight Weyl–Heisenberg frames in $\ell^2(\mathbb{Z})$," IEEE Trans. Signal Process. **46**, 1256–1259 (1998).

[25] A. J. E. M. Janssen and T. Strohmer, "Characterization and computation of canonical tight windows for Gabor frames," J. Fourier Anal. Appl. **8**, 1–28 (2002).

[26] Z. Cvetkovic, "On discrete short-time Fourier analysis," IEEE Trans. Signal Process. **48**, 2628–2640 (2000).

[27] N. Perraudin, N. Holighaus, P. L. Søndergaard, and P. Balazs, "Designing Gabor windows using convex optimization," Appl. Math. Comput. **330**, 266–287 (2018).

[28] W.-S. Lu, T. Saramäki, and R. Bregović, "Design of Practically Perfect-Reconstruction Cosine-Modulated Filter Banks: A Second-Order Cone Programming Approach," IEEE Trans. Circuits Syst. I **51**, 552–563 (2004).

[29] J. Jiang, S. Ouyang, and F. Zhou, "Design of NPR DFT-modulated filter banks via iterative updating algorithm," Circuits, Syst., Signal Process. **32**, 1351–1362 (2013).

[30] M. R. Wilbur, T. N. Davidson, and J. P. Reilly, "Efficient design of oversampled NPR GDFT filterbanks," IEEE Trans. Signal Process. **52**, 1947–1962 (2004).

[31] S.-P. Wu, S. Boyd, and L. Vandenberghe, *FIR Filter Design via Spectral Factorization and Convex Optimization* (Birkhäuser Boston, Boston, MA, 1999), pp. 215–245.

[32] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," SIAM J. Imag. Sci. **3**, 253–276 (2010).

[33] J. Yang and X. Yuan, "Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization," Math. Comput. **82**, 301–329 (2012).

[34] N. Parikh and S. Boyd, "Proximal algorithms," **1**, 123–231 (2014).