



Griffin–Lim Like Phase Recovery via Alternating Direction Method of Multipliers

Yoshiki Masuyama , Kohei Yatabe , *Member, IEEE*, and Yasuhiro Oikawa, *Member, IEEE*

Abstract—Recovering a signal from its amplitude spectrogram, or phase recovery, exhibits many applications in acoustic signal processing. When only an amplitude spectrogram is available and no explicit information is given for the phases, the Griffin–Lim algorithm (GLA) is one of the most utilized methods for phase recovery. However, GLA often requires many iterations and results in low perceptual quality in some cases. In this letter, we propose two novel algorithms based on GLA and the alternating direction method of multipliers (ADMM) for better recovery with fewer iteration. Some interpretation of the existing methods and their relation to the proposed method are also provided. Evaluations are performed with both objective measure and subjective test.

Index Terms—Non-convex optimization, short-time Fourier transform (STFT), spectrogram consistency, phaseless spectrogram inversion, STFT-based speech synthesis.

I. INTRODUCTION

IN RECENT years, phase recovery of spectrograms gains much attention in acoustical applications. This is because, on the one hand, recent studies have proven that recovering the phase spectrogram can improve the quality of speech signals [1]–[5] whose amplitude spectrograms are enhanced by some processing [6]–[9]. On the other hand, recent statistical speech synthesis methods attempt to synthesize amplitude spectrogram without phase [10]–[12]. Phase recovery is essential for the latter situation because phases are indispensable to generate the waveform and listen to the synthesized speech. The focus of this letter is mainly on such situation where no information is available for the phase to be recovered.

For phase recovery without any information for phase, the Griffin–Lim algorithm (GLA) has been utilized as the standard tool [13]. GLA consists of two projections (see Section II-B) which aim to find the *consistent* spectrogram [see Eq. (7)] with the given amplitude, where a spectrogram is said to be consistent when its bins retain the neighborhood relation owing to the overlapping window in time domain [14]–[16]. Although GLA has been successfully applied, its slow convergence may require a significant number of iterations until the quality of reconstruction becomes satisfactory. Therefore, some modifications of GLA have been presented for acceleration [17]–[20].

Manuscript received September 7, 2018; revised November 14, 2018; accepted November 23, 2018. Date of publication November 29, 2018; date of current version December 10, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Peter Jax. (*Corresponding author: Yoshiki Masuyama.*)

The authors are with the Department of Intermedia Art and Science, Waseda University, Tokyo 69-8050, Japan (e-mail: mas-03151102@akane.waseda.jp; k.yatabe@asagi.waseda.jp; yoikawa@waseda.jp).

Digital Object Identifier 10.1109/LSP.2018.2884026

The importance of accelerating GLA is not only in the computational aspect but also in the improvement of performance. For instance, synthesized sound with the fast GLA (FGLA) [17] is often better than that with GLA. This should be because such acceleration can avoid poor local-minima thanks to the stronger preference towards a global minimum. Meanwhile, the ability of the alternating direction method of multipliers (ADMM) to avoid poor local-minima has been shown in many applications [21], [22]. This suggests that it might be possible to improve the performance of GLA by applying ADMM.

In this letter, we propose two ADMM algorithms for phase recovery. Firstly, GLA is interpreted as an optimization algorithm, and ADMM is adopted for solving the same problem. Furthermore, its relaxation problem, which can result in better reconstruction than the original one, is proposed. The novelty of this letter comes from the application of ADMM to the phase recovery problem in Eq. (16) and the proposal of its relaxed version in Eq. (17) together with the algorithm for solving it. Some interpretation of the existing and proposed methods is also provided as a part of the contributions.

II. PRELIMINARIES

A. Phase Recovery From Amplitude Spectrogram

Let the short-time Fourier transform (STFT) of a signal $\mathbf{x} = (x[0], \dots, x[N-1])^T \in \mathbb{R}^N$ with a window $\mathbf{w} \in \mathbb{R}^L$ be

$$\mathcal{G}_{\mathbf{w}}(\mathbf{x})[\xi, \tau] = \sum_{l=1}^L x[l + a\tau]w[l]e^{-2\pi j\xi bl/L} = X[\xi, \tau], \quad (1)$$

where $j = \sqrt{-1}$, a and b are time and frequency shifting steps, and $\tau = 0, \dots, T-1$ and $\xi = 0, \dots, K-1$ denote time and frequency indices, respectively. The adjoint operator of STFT $\mathcal{G}_{\mathbf{w}}^*$ (inverse STFT) is given by

$$\mathcal{G}_{\mathbf{w}}^*(\mathbf{X})[n] = \sum_{\tau=0}^{T-1} \sum_{\xi=0}^{K-1} X[\xi, \tau]w[n - a\tau]e^{2\pi j\xi bn/L}. \quad (2)$$

The pseudo inverse of $\mathcal{G}_{\mathbf{w}}$ is $\mathcal{G}_{\mathbf{w}}^\dagger = (\mathcal{G}_{\mathbf{w}}^* \mathcal{G}_{\mathbf{w}})^{-1} \mathcal{G}_{\mathbf{w}}^*$, and it is easily implemented as $\mathcal{G}_{\mathbf{w}}^\dagger = \mathcal{G}_{\tilde{\mathbf{w}}}^*$ where $\tilde{\mathbf{w}}$ is the canonical dual window associated with \mathbf{w} [23]–[25], i.e., $\mathbf{x} = \mathcal{G}_{\tilde{\mathbf{w}}}^* \mathcal{G}_{\mathbf{w}} \mathbf{x}$.

Let an amplitude spectrogram $\mathbf{A} \in \mathbb{R}_+^{K \times T}$ be given without phase. By denoting the image of $\mathcal{G}_{\mathbf{w}}$ as $\text{Im}(\mathcal{G}_{\mathbf{w}})$, the phase recovery problem is formulated as follows:

$$\text{Find } \mathbf{X} \text{ s.t. } \mathbf{X} \in \text{Im}(\mathcal{G}_{\mathbf{w}}), \quad |X[\xi, \tau]| = A[\xi, \tau]. \quad (3)$$

This problem aims to find a complex-valued spectrogram \mathbf{X} whose amplitude is the given one \mathbf{A} . Its phase is recovered through the consistency which requires \mathbf{X} to be within $\text{Im}(\mathcal{G}_w)$.

B. Griffin-Lim Algorithm (GLA)

The first constraint in Eq. (3), $\mathbf{X} \in \text{Im}(\mathcal{G}_w)$, requires \mathbf{X} to be consistent, while the amplitude of \mathbf{X} must coincide with \mathbf{A} as requested in the second constraint, $|X[\xi, \tau]| = A[\xi, \tau]$. GLA implements these constraints directly as an alternating projection algorithm [13], [26], [27]

$$\mathbf{X}^{[m+1]} = P_{C_1}(P_{C_2}(\mathbf{X}^{[m]})), \quad (4)$$

where m is the iteration index, P_C is the projection,

$$P_C(\mathbf{Y}) = \arg \min_{\mathbf{X}} \iota_C(\mathbf{X}) + \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\|_{\text{Fro}}^2, \quad (5)$$

ι_C is the indicator function with respect to a set C ,

$$\iota_C(\mathbf{X}) = \begin{cases} 0 & (\mathbf{X} \in C) \\ \infty & (\text{otherwise}) \end{cases}, \quad (6)$$

and $\|\cdot\|_{\text{Fro}}$ is the Frobenius norm. Here, C_1 is the set of the STFT coefficients \mathbf{X} which satisfy the consistency criterion,

$$C_1 = \{ \mathbf{X} \in \mathbb{C}^{K \times T} \mid \exists \mathbf{x} \in \mathbb{R}^N \text{ s.t. } \mathbf{X} = \mathcal{G}_w \mathbf{x} \}. \quad (7)$$

The projection onto C_1 is given by [25]

$$P_{C_1}(\mathbf{X}) = \mathcal{G}_w \mathcal{G}_w^\dagger \mathbf{X}, \quad (8)$$

which ensures $P_{C_1}(\mathbf{X}) \in \text{Im}(\mathcal{G}_w)$. C_2 is the set of STFT coefficients whose amplitude coincides with the given one \mathbf{A} :

$$C_2 = \{ \mathbf{X} \in \mathbb{C}^{K \times T} \mid |X[\xi, \tau]| = A[\xi, \tau] \}. \quad (9)$$

The projection onto C_2 is simply given by the Hadamard product \odot of the given amplitude and normalized coefficients:

$$P_{C_2}(\mathbf{X}) = \mathbf{A} \odot \hat{\mathbf{X}}, \quad (10)$$

where $\hat{X}[\xi, \tau] = X[\xi, \tau]/|X[\xi, \tau]|$ if $X[\xi, \tau] \neq 0$, and 0 if $|X[\xi, \tau]| = 0$. While GLA has been successfully utilized, it may require many iterations as usual for the alternating projection algorithm. FGLA is a variant of GLA which employs Nesterov's technique to accelerate it [17].

C. Alternating Direction Method of Multipliers (ADMM)

Many optimization problems arising in the application can be interpreted as the following minimization problem:

$$\min_{\mathbf{X}} f(\mathbf{X}) + g(\mathbf{X}), \quad (11)$$

where f and g are real-valued functions often related to cost or loss. ADMM is a well-known algorithm for solving the above problem by iterating the following procedure [21], [28]:¹

$$\mathbf{X}^{[m+1]} = \text{prox}_{f/\rho}(\mathbf{Z}^{[m]} - \mathbf{U}^{[m]}), \quad (12)$$

$$\mathbf{Z}^{[m+1]} = \text{prox}_{g/\rho}(\mathbf{X}^{[m+1]} + \mathbf{U}^{[m]}), \quad (13)$$

$$\mathbf{U}^{[m+1]} = \mathbf{U}^{[m]} + \mathbf{X}^{[m+1]} - \mathbf{Z}^{[m+1]}, \quad (14)$$

¹This ADMM algorithm is a simplified version for easier explanation of the proposed method. See [21] for the general form which can handle a wide range of problems which are more complicated than Eq. (11).

Algorithm 1: Proposed Phase Recovery With the Exact Consistency in Eq. (16).

Input: $\mathbf{X}^{[0]}$

Initialization: $\mathbf{Z}^{[0]} = \mathbf{X}^{[0]}$, $\mathbf{U}^{[0]} = \mathbf{0}$

Output: $\mathbf{X}^{[m+1]}$

for $m = 0, 1, \dots$ **do**

$\mathbf{X}^{[m+1]} = P_{C_2}(\mathbf{Z}^{[m]} - \mathbf{U}^{[m]})$ using Eq. (10)

$\mathbf{Z}^{[m+1]} = P_{C_1}(\mathbf{X}^{[m+1]} + \mathbf{U}^{[m]})$ using Eq. (8)

$\mathbf{U}^{[m+1]} = \mathbf{U}^{[m]} + \mathbf{X}^{[m+1]} - \mathbf{Z}^{[m+1]}$

end for

where $\rho > 0$, \mathbf{Z} and \mathbf{U} are the auxiliary variables. One of the key points of ADMM is that the auxiliary variable \mathbf{U} accumulates the difference of the variables $\mathbf{X} - \mathbf{Z}$ which may be seen as the element-wise (signed) distance, and it is added back to \mathbf{X} and \mathbf{Z} . The proximity operator defined by [29]

$$\text{prox}_f(\mathbf{Y}) = \arg \min_{\mathbf{X}} f(\mathbf{X}) + \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\|_{\text{Fro}}^2, \quad (15)$$

is also important as it eases the problem via the proximity term. The effectiveness of ADMM has been investigated in many applications [21], [22].

III. PROPOSED PHASE RECOVERY ALGORITHMS

In this section, we propose two phase recovery algorithms utilizing ADMM, which can obtain better phase in fewer iterations than GLA. The consistency of spectrograms is considered either exactly or relaxedly in the proposed algorithms, and the existence of the solution is guaranteed for the relaxed case.

A. ADMM Phase Recovery With Exact Consistency

The phase recovery in Eq. (3) requires to find a spectrogram \mathbf{X} satisfying both constraints: $\mathbf{X} \in C_1$ and $\mathbf{X} \in C_2$. That is, \mathbf{X} should exist in $C_1 \cap C_2$ and should not be outside it. This requirement can be recast into the minimization problem,

$$\min_{\mathbf{X}} \iota_{C_2}(\mathbf{X}) + \iota_{C_1}(\mathbf{X}), \quad (16)$$

because minimization refuses ∞ of the indicator function ι_C in Eq. (6). A solution to Eq. (16) must satisfy both $\mathbf{X} \in C_1$ and $\mathbf{X} \in C_2$ so that $\iota_{C_1}(\mathbf{X}) + \iota_{C_2}(\mathbf{X}) = 0$ and not ∞ .

This problem is in the form of Eq. (11) where the ADMM algorithm in Eqs. (12)–(14) can be applied. Fortunately, the proximity operators $\text{prox}_{\iota_{C_1}}$ and $\text{prox}_{\iota_{C_2}}$ in Eqs. (12) and (13) can be calculated easily by Eqs. (8) and (10) because a proximity operator of an indicator function reduces to the projection in Eq. (5). By applying ADMM to the above problem, we obtain the first algorithm summarized in Algorithm 1, where the relation to GLA will be discussed later in Section III-C.

B. ADMM Phase Recovery With Relaxed Consistency

In the first algorithm, two constraints in Eq. (3) are directly imposed by the indicator functions. However, an artificial amplitude spectrogram may not admit a consistent spectrogram, and thus $C_1 \cap C_2$ can be empty. For such case, strictly speaking, a solution to Eq. (16) does not exist, and Algorithm 1 returns $\mathbf{X}^{[m+1]}$ satisfying only $\mathbf{X} \in C_2$. Although the effect of such a

Algorithm 2: Proposed Phase Recovery With the Relaxed Consistency in Eq. (17).

Input: $\mathbf{X}^{[0]}$, $\rho > 0$

Initialization: $\mathbf{Z}^{[0]} = \mathbf{X}^{[0]}$, $\mathbf{U}^{[0]} = \mathbf{0}$

Output: $\mathbf{X}^{[m+1]}$

for $m = 0, 1, \dots$ **do**

$\mathbf{X}^{[m+1]} = P_{C_2}(\mathbf{Z}^{[m]} - \mathbf{U}^{[m]})$ using Eq. (10)

$\mathbf{Z}^{[m+1]} = \text{prox}_{d_{C_1}^2/2\rho}(\mathbf{X}^{[m+1]} + \mathbf{U}^{[m]})$ using Eq. (19)

$\mathbf{U}^{[m+1]} = \mathbf{U}^{[m]} + \mathbf{X}^{[m+1]} - \mathbf{Z}^{[m+1]}$

end for

situation may not be significant in practice, we propose a relaxed formulation so that its solution always exists.

To ensure the existence of a solution, either of the two constraints can be relaxed. In this letter, relaxation of ι_{C_2} is not considered because the given amplitude spectrogram \mathbf{A} is an essential information which should be preserved, especially in speech synthesis. In addition, the convex relaxation of ι_{C_2} [30], [31] often leads to excessive size of the problem in acoustical applications, which makes it difficult to be solved. Therefore, only the consistency constraint ι_{C_1} is relaxed here, and the relaxed problem is formulated as

$$\min_{\mathbf{X}} \iota_{C_2}(\mathbf{X}) + \frac{1}{2}d_{C_1}^2(\mathbf{X}), \quad (17)$$

where $d_{C_1}^2$ is the squared distance to the set C_1 ,

$$\frac{1}{2}d_{C_1}^2(\mathbf{X}) = \frac{1}{2} \|\mathbf{X} - P_{C_1}(\mathbf{X})\|_{\text{Fro}}^2, \quad (18)$$

and $P_{C_1}(\mathbf{X}) \in C_1$ is a closest point to \mathbf{X} within the set C_1 . With this relaxation, a solution which satisfies $\mathbf{X} \in C_2$ and is closest to C_1 always exists.

For applying ADMM to this problem, the proximity operator of the squared distance, $\text{prox}_{d_{C_1}^2/2\rho}$, is necessary. By plugging Eqs. (18) and (8) into Eq. (15), one can see that $\text{prox}_{d_{C_1}^2/2\rho}$ is a differentiable quadratic function. Thus, its explicit form can be obtained by setting the derivative to 0:

$$\text{prox}_{d_{C_1}^2/2\rho}(\mathbf{Y}) = \frac{1}{1+\rho}(\rho\mathbf{Y} + P_{C_1}(\mathbf{Y})). \quad (19)$$

Our second phase recovery algorithm is summarized in Algorithm 2, where the projection in Algorithm 1 is replaced by the weighted average of current value and its projected one. Note that this proximity operator in Eq. (19) is related to the projection in Eq. (8) as the limit of ρ tending to 0, and thus the proposed algorithms are closely related to each other.

C. Relation to Conventional Algorithms

In this subsection, relation between our algorithms and the conventional algorithms are discussed.

1) *GLA and Proposed Algorithm 1:* GLA in Eq. (4) can be regarded as the alternating projection algorithm for Eq. (16),

$$\mathbf{X}^{[m+1]} = P_{C_2}(\mathbf{Z}^{[m]}), \quad (20)$$

$$\mathbf{Z}^{[m+1]} = P_{C_1}(\mathbf{X}^{[m+1]}), \quad (21)$$

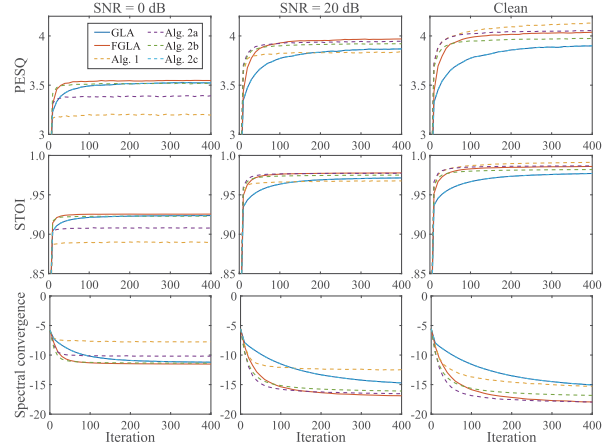


Fig. 1. Average PESQ, STOI, and spectral convergence over iterations. Each row corresponds to each objective measure, while each column represents the type of amplitude spectrograms inputted to the algorithms.

where Eq. (4) is obtained by omitting the intermediate variable. Comparing this representation with Algorithm 1, the proposed algorithm and GLA differ in the auxiliary variable \mathbf{U} . This variable accumulates not only \mathbf{X} but also the intermediate state \mathbf{Z} , which balances the effects of the projections so that the update results in a better outcome for both \mathbf{X} and \mathbf{Z} .

2) *GLA and Proposed Algorithm 2:* In Algorithm 2, the constraint on C_1 is relaxed to the squared distance $d_{C_1}^2/2$ which penalizes \mathbf{X} outside C_1 and becomes 0 when $\mathbf{X} \in C_1$. That is, $d_{C_1}^2/2$ can be considered as a smoothed version of the indicator function ι_{C_1} [32]. Such smoothing allows us to apply the projected gradient algorithm to the relaxed problem in Eq. (17), which reveals an interesting relation to GLA.

The projected gradient algorithm for Eq. (17) is given by

$$\mathbf{X}^{[m+1]} = P_{C_2}(\mathbf{X}^{[m]} - (\mu/2)\nabla d_{C_1}^2(\mathbf{X}^{[m]})), \quad (22)$$

where $\mu > 0$ is a step size. Since the gradient of $d_{C_1}^2$ is [32]

$$\nabla d_{C_1}^2(\mathbf{X}) = 2(\mathbf{X} - P_{C_1}(\mathbf{X})), \quad (23)$$

plugging this into Eq. (22) gives a variant of GLA when $\mu = 1$:

$$\mathbf{X}^{[m+1]} = P_{C_2}(P_{C_1}(\mathbf{X}^{[m]})). \quad (24)$$

Thus, Eq. (4) corresponds to the projected gradient method for Eq. (17) except the order of the projections.

3) *Relation to Consistent Wiener Filtering (CWF) [16]:* We briefly mention here that P_{C_1} and $\text{prox}_{d_{C_1}^2/2\rho}$ can be regarded as the hard- and soft-CWF [16], respectively, if the covariance matrix is identity. This point of view might be interesting when one considers a modification of the proposed algorithms.

IV. EXPERIMENTS

For investigating the performance, the proposed algorithms were applied to speech synthesis.² Amplitude spectrograms of 200 utterances from TIMIT database [33] (100 males and 100 females) were inputted to the algorithms without phase, and the recovered speeches were evaluated by objective measures:

²Our MATLAB code is available at Code Ocean: <http://doi.org/cxck>

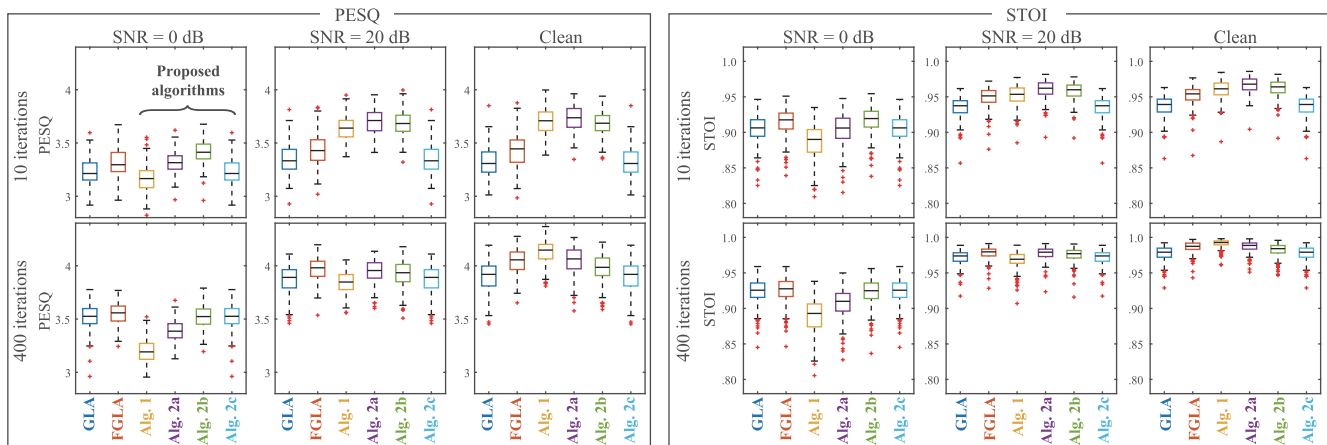


Fig. 2. Boxplots of PESQ and STOI for synthesized speech signals from amplitude spectrograms of 200 utterances in TIMIT database. The boxes indicate the first and third quartiles. The top row represents the performances after 10 iteration, while the bottom row indicates that of 400 iterations.

PESQ [34], STOI [35], and the spectral convergence [36],

$$\text{Spectral convergence} = \|\mathbf{A} - |\mathcal{G}_w(\mathbf{x})|\|_{\text{Fro}} / \|\mathbf{A}\|_{\text{Fro}}. \quad (25)$$

To imitate imperfection of generated amplitude spectrograms in speech synthesis, babble noise [37] was added to the speech signals, and then the oracle Wiener filter was applied for obtaining degraded amplitude spectrograms. The signal-to-noise ratio (SNR) before filtering was set to 20 and 0 dB.

The proposed algorithms were compared with the strongly related algorithms: GLA [13] and FGLA [17]. Algorithm 2 with different parameters, $\rho = 0.1, 0.3, 1.0$, will be abbreviated as Alg. 2a, 2b, and 2c, respectively, while the parameter of FGLA was set to the default value in [17]. In all conditions, initial phases were set to zero in the time-frequency domain. STFT was implemented with the Hann window whose length was 32 ms, and the overlap length was 13.5 ms.

A. Objective Evaluation

The objective performances averaged over 200 speeches are illustrated in Fig. 1. In all cases, the curves of GLA and Alg. 2c ($\rho = 1$) coincide. Recalling that Alg. 1 corresponds to $\rho = 0$, the proposed algorithm seems to interpolate the performance between Alg. 1 and GLA by varying ρ from 0 to 1. While the proposed algorithm achieved higher scores with smaller ρ (Alg. 1 and Alg. 2a) for perfect (clean) amplitude spectrograms (rightmost column), higher ρ (Alg. 2b) seems preferable for imperfect (noisy) amplitude spectrograms. For every case, the proposed algorithm tended to outperform GLA in the beginning (say, less than 100 iterations), while it was comparable to FGLA with an appropriate parameter ρ .

Boxplots of PESQ and STOI are shown in Fig. 2. As the deviations of the results are similar for all methods, the same trend as in Fig. 1 can be observed. Although the proposed method with small ρ (Alg. 1 and Alg. 2a) seems worse than the conventional algorithms when the amplitude spectrograms are not perfect, this is perceptually insignificant as shown next.

B. Subjective Test

To evaluate the naturalness of synthesized speeches, subjective evaluation by MUSHRA (Multiple Stimuli with Hidden Reference and Anchor) test was conducted. 10 speech signals

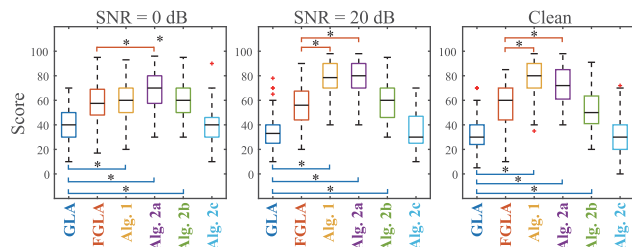


Fig. 3. Boxplots of MUSHRA scores for reconstructed speech signals at 10th iteration that corresponds to the top row of Fig. 2. Scores of the hidden references are not shown here as they were always 0 or 100.

with 3 conditions (SNR = 0 dB, SNR = 20 dB, and clean) were evaluated by 10 listeners. For each speech signal, the original signal and the initial value of the algorithms (zero-phase signal) were included in the test as hidden references.

Fig. 3 shows boxplots summarizing the MUSHRA scores with the results of t -test ($p < 0.001$). For the cases of clean and SNR = 20 dB, Alg. 1 and Alg. 2a significantly outperformed both GLA and FGLA, which are consistent with the objective evaluations. In contrast, when SNR = 0 dB, the scores of Alg. 1 were higher than GLA, and those of Alg. 2a were higher than FGLA, which contradicts with Fig. 2. This is because the speeches synthesized by GLA and FGLA were buzzy, while the proposed algorithms generated smooth speech signals. As the computational cost per iteration is similar to that of the conventional methods, the proposed algorithms are beneficial especially when a smaller number of iterations is preferable.

V. CONCLUSION

In this letter, we proposed two consistency-based phase recovery algorithms based on ADMM. The proposed algorithms effectively utilize the auxiliary variable of ADMM for modifying phase. The ADMM algorithms were able to avoid buzziness of the GLA reconstructions, which was reflected in the subjective evaluation. More specific relation between quality of reconstructed signals and optimization algorithms should be investigated in future works.

REFERENCES

- [1] K. Paliwal, K. Wójcicki, and B. Shannon, "The importance of phase in speech enhancement," *Speech Commun.*, vol. 53, no. 4, pp. 465–494, Apr. 2011.
- [2] T. Gerkmann, M. Krawczyk-Becker, and J. Le Roux, "Phase processing for single-channel speech enhancement: History and recent advances," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 55–66, Mar. 2015.
- [3] P. Mowlaee, R. Saeidi, and Y. Stylianou, "Advances in phase-aware signal processing in speech communication," *Speech Commun.*, vol. 81, pp. 1–29, Jul. 2016.
- [4] K. Yatabe, Y. Masuyama, and Y. Oikawa, "Rectified linear unit can assist Griffin–Lim phase recovery," in *Proc. Int. Workshop Acoust. Signal Enhancement*, Sep. 2018, pp. 555–559.
- [5] Y. Masuyama, K. Yatabe, and Y. Oikawa, "Model-based phase recovery of spectrograms via optimization on Riemannian manifolds," in *Proc. Int. Workshop Acoust. Signal Enhancement*, Sep. 2018, pp. 126–130.
- [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.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.
- [7] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 443–445, Apr. 1985.
- [8] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 113–116, Apr. 2002.
- [9] F. Weninger *et al.*, "Speech enhancement with LSTM recurrent neural networks and its application to noise-robust ASR," in *Proc. 12th Int. Conf. LVA/ICA*, Cham, Aug. 2015, pp. 91–99.
- [10] Y. Wang *et al.*, "Tacotron: A fully end-to-end text-to-speech synthesis model," in *Proc. INTERSPEECH*, Aug. 2017, pp. 4006–4010.
- [11] S. Takaki, H. Kameoka, and J. Yamagishi, "Direct modeling of frequency spectra and waveform generation based on phase recovery for DNN-based speech synthesis," in *Proc. INTERSPEECH*, Aug. 2017, pp. 1128–1132.
- [12] T. Kaneko, S. Takaki, H. Kameoka, and J. Yamagishi, "Generative adversarial network-based postfilter for STFT spectrograms," in *Proc. INTERSPEECH*, Aug. 2017, pp. 3389–3393.
- [13] D. Griffin and J. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 2, pp. 236–243, Apr. 1984.
- [14] J. Le Roux, N. Ono, and S. Sagayama, "Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction," in *Proc. ISCA Workshop Statistical Perception Audit*, Sep. 2008, pp. 23–28.
- [15] J. Le Roux, H. Kameoka, N. Ono, and S. Sagayama, "Fast signal reconstruction from magnitude STFT spectrogram based on spectrogram consistency," in *Proc. 13th Int. Conf. Digit. Audio Effects*, Sep. 2010, pp. 397–403.
- [16] J. Le Roux and E. Vincent, "Consistent Wiener filtering for audio source separation," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 217–220, Mar. 2013.
- [17] N. Perraudin, P. Balazs, and P. L. Søndergaard, "A fast Griffin–Lim algorithm," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, Oct. 2013, pp. 1–4.
- [18] X. Zhu, G. T. Beauregard, and L. L. Wyse, "Real-time signal estimation from modified short-time Fourier transform magnitude spectra," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 5, pp. 1645–1653, Jul. 2007.
- [19] X. Zhu, G. T. Beauregard, and L. Wyse, "Real-time iterative spectrum inversion with look-ahead," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2006, pp. 229–232.
- [20] Z. Průša, P. Balazs, and P. L. Søndergaard, "A noniterative method for reconstruction of phase from STFT magnitude," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 5, pp. 1154–1164, May 2017.
- [21] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.
- [22] S. Diamond, R. Takapoui, and S. Boyd, "A general system for heuristic minimization of convex functions over non-convex sets," *Optim. Methods Softw.*, vol. 33, no. 1, pp. 165–193, 2018.
- [23] H. G. Feichtinger and T. Strohmer, *Gabor Analysis and Algorithms: Theory and Applications*. Cambridge, MA, USA: Birkhäuser Boston, 1997.
- [24] K. Gröchenig, *Foundations of Time-Frequency Analysis*. Cambridge, MA, USA: Birkhäuser, 2001.
- [25] O. Christensen, *An Introduction to Frames and Riesz Bases*. Cambridge, MA, USA: Birkhäuser, 2016.
- [26] R. W. Grechberg and W. O. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.
- [27] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, Aug. 1982.
- [28] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*. New York, NY, USA: Elsevier, 1983.
- [29] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, Jan. 2014.
- [30] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, Nov. 2013.
- [31] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Math. Program.*, vol. 149, no. 1, pp. 47–81, Feb. 2015.
- [32] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Berlin, Germany: Springer, 2011.
- [33] P. Mowlaee, J. Kulmer, J. Stahl, and F. Mayer, *Single Channel Phase-Aware Signal Processing in Speech Communication: Theory and Practice*. Hoboken, NJ, USA: Wiley, 2016.
- [34] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ) — A new method for speech quality assessment of telephone networks and codecs," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2001, vol. 19, pp. 2125–2136.
- [35] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2155–2136, Sep. 2011.
- [36] N. Strumel and L. Daudet, "Signal reconstruction from STFT magnitude: a state of the art," in *Proc. Int. Conf. Digit. Audio Effects*, Sep. 2011, pp. 375–386.
- [37] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition II: NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 12, no. 3, pp. 247–251, Jul. 1993.