MODEL-BASED PHASE RECOVERY OF SPECTROGRAMS VIA OPTIMIZATION ON RIEMANNIAN MANIFOLDS

Yoshiki Masuyama, Kohei Yatabe and Yasuhiro Oikawa

Department of Intermedia Art and Science, Waseda University, Tokyo, Japan

ABSTRACT

In acoustical signal processing, the importance of modifying the phase spectrogram has been shown. Recently, model-based phase recovery which is based on the sinusoidal model has been studied. Although their effectiveness has been proven, some of them deal with the phase in inflexible forms owing to the wrapping effect of phase. In addition, they need much pre-processing, including the estimation of the instantaneous frequency, which is not easy tasks. In order to overcome these issues, we propose a novel model-based phase recovery method which is formulated as an optimization over complex-valued phases. In the proposed method, the instantaneous frequency is not handled fixedly, which avoids the prior estimation of the instantaneous frequency. The technique of optimization on Riemannian manifolds is adopted for efficient computation. The proposed method is validated by noise reduction of audio signals.

Index Terms— Instantaneous frequency, phase derivative, sinusoidal modeling, non-convex optimization, gradient descent.

1. INTRODUCTION

Many noise reduction and source separation techniques are formulated in the time-frequency domain in which signals are represented by complex-valued STFT coefficients. Most of them [1–5] dedicated to only the modification of the amplitude spectrogram owing to the difficulty of phase modification such as wrapping effect. That is, the phase spectrogram remains intact in resynthesizing the signal. Recent studies, however, have proven the importance of the phase modification in speech enhancement [6–8], source separation [9], and other applications [10]. The phase spectrograms of harmonic signals contain distinctive structures, which characterizes their sound.

There exist two approaches for phase recovery: consistency-based approach and model-based approach. At first, consistency-based approach considers the redundancy of STFT [11–13]. The consistency describes the relationship among the STFT coefficients based on the procedure of STFT. Namely, the consistency-based approach considers the property of STFT. One of the most popular consistency-based phase recovery algorithms is the Griffin–Lim algorithm (GLA) which involves much computation [11]. In order to circumvent this issue, a recent study utilizes the technique of optimization on Riemannian manifolds, which is called non-convex phase cut (NCPC) [14]. Instead of the phase spectrogram, NCPC considers complex-valued phases represented by the Hadamard product of complex numbers whose absolute values restricted to 1. The complex-valued phases construct a Riemannian manifold, and thus the technique of optimization on Riemannian manifolds can be utilized for efficient calculation. Moreover, the representation by complex-valued phases is easy to incorporate with other prior knowledge of phase. However, signals resynthesized by the consistency-based approach cannot always achieve high performance, because they do not take into account the property of harmonic signals [15].

Meanwhile, model-based approach utilizes the idea of the sinusoidal model which considers the property of a sum of sinusoids [16–21].The sinusoidal model is compatible with harmonic signals, and the model-based phase recovery improves perceptual quality [17]. In the spectrogram of a sum of sinusoids, the phase evolution can be predictable from the instantaneous frequency of each sinusoid. In model-based phase recovery, the phase spectrogram is modified along the time-direction using the prediction of the phase evolution as described in Section 2.1. One of model-based phase recovery for audio signals is phase unwrapping (PU), which modifies the unwrapped phase spectrogram along the time-direction deterministically [16]. However, PU is implemented in inflexible form because it deals with the unwrapped phase spectrogram. In addition, it needs much pre-processing, including the prior estimation of the instantaneous frequency, which plays important roles. Its estimation error significantly corrupts the performance of phase recovery although the estimation of the instantaneous frequency is not easy.

In order to circumvent these issues, we propose a new model-based phase recovery method. We formulate phase recovery as an optimization over complex-valued phases, and the idea of the sinusoidal model is incorporated as the regularization term. Thanks to this, the proposed method does not need the prior estimation of the instantaneous frequency, and it is irrespective of the error of the prior estimation. The technique of optimization on Riemannian manifolds is adopted for the proposed formulation, and its effectiveness is shown in noise reduction of audio signals.

2. PRELIMINARIES

2.1. Model-based phase recovery

Let us denote a discrete signal by $\mathbf{x} = (x_1, \ldots, x_N)^T \in \mathbb{R}^N$, and the STFT of the discrete signal $\mathbf{x}$ with a window $\mathbf{g} \in \mathbb{R}^L$ by

$$\mathscr{F}(\mathbf{x})_{\xi,\tau} = \sum_{l=0}^{L-1} x_{l+a\tau}\, \overline{g_l\, \mathrm{e}^{2\pi j \xi b l/L}}, \tag{1}$$

where $\bar{z}$ is the complex conjugate of z, $j = \sqrt{-1}$, L is the frame length, a and b is the time and frequency shifting steps, and $\xi = 0, 1, \ldots, K$ and $\tau = 0, 1, \ldots, T$ denote the frequency and the time indices, respectively. The sum of sinusoids is given by

$$x_l = \sum_{h=0}^{H-1} \mathcal{E}_h\, e^{2\pi j f_h l + \phi_{h,0}}, \tag{2}$$

where $\mathcal{E}_h$, f_h and $\phi_{h,0}$ are the amplitude, the frequency, and the initial phase of hth sinusoid, respectively. The phase spectrogram of the sum of sinusoids has the following relationship:

$$\phi_{\xi,\tau+1} = \phi_{\xi,\tau} + 2\pi a v_{\xi,\tau}, \tag{3}$$

978-1-5386-8151-0/18/$31.00 ©2018 IEEE

Algorithm 1 Phase unwrapping (PU)

```
**Input**: onset phase $\phi_{:,\tau_{\text{on}}}$, amplitude spctrogram $\mathbf{A}$.
**Output**: $\boldsymbol{\phi}$
**while** existing harmonic signals **do**
    **Peak localization** $\xi_{h,\tau}$ from $A_{:,\tau}$.
    **Instantaneous frequency estimation** $f_{h,\tau}$ via QIFFT.
    **Regions of influence estimation** $I_{h,\tau}$ and $v_{\xi,\tau} = f_{h,\tau} \in I_{h,\tau}$.
    **Phase update**: $\phi_{\xi,\tau+1} = \phi_{\xi,\tau} + 2\pi a v_{\xi,\tau}$
**end while**
```

where $v_{\xi,\tau}$ is the instantaneous frequency of each bin which is determined from the instantaneous frequencies of the corresponding sinusoids. This analysis is applicable to harmonic signals such as speeches and audio signals, and it is often utilized in the phase vocoders [22, 23]. The model-based phase recovery is based on the relationship among successive time frames given by Eq. (3) [16]. The phase of the next time frame is predictable from the phase and the instantaneous frequency of the current time frame. Specifically, the instantaneous frequency of each bin is inferred from that of the sinusoids or only from the fundamental frequency. The model-based phase recovery considers the sinusoidal model, and hence it is suitable for phase recovery of harmonic signals.

One of model-based phase recoveries for audio signals is PU [16]. PU recursively modifies the unwrapped phase spectrogram as satisfying the relationship given by Eq. (3) from the onset τ_{on}. The specific algorithm of PU is shown in Algorithm 1, and the instantaneous frequency $v_{\xi,\tau}$ of each sinusoid is estimated in three steps. In order to estimate the instantaneous frequency, at first, the instantaneous frequency of each sinusoid is approximately estimated from the location of the amplitude spectrogram peaks. Next, the instantaneous frequency of each sinusoid is estimated by QIFFT [24]. The instantaneous frequency of each bin is estimated from neighbor sinusoids with the assumption of the region of influence [22]. The region of influence means that $v_{\xi,\tau} \approx f_{h,\tau}$ is satisfied in the hth region I_h, in which the hth sinusoid is dominant. In [16], the region of influence is defined by

$$I_{h,\tau} = \left[\frac{A_{h,\tau} f_{h-1} + A_{h-1,\tau} f_h}{A_{h-1,\tau} + A_{h,\tau}}, \frac{A_{h+1,\tau} f_h + A_{h,\tau} f_{h+1}}{A_{h,\tau} + A_{h+1,\tau}} \right], \quad (4)$$

where $\mathbf{A}$ is the amplitude spectrogram and $A_{h,\tau}$ is the approximated amplitude of the hth sinusoid estimated by the peak localization. Although the above procedure estimates the instantaneous frequency, it is not exactly accurate. The estimation error significantly corrupts its performance owing to its deterministic implementation as described in Algorithm 1.

2.2. Optimization on Riemannian manifolds

In this subsection, we briefly review the gradient descent algorithm for the optimization on Riemannian manifolds [25, 26]. We refer the readers to [25] for more information about it. Let us consider the following non-convex optimization:

$$\min_{\mathbf{u} \in \mathcal{M}} \mathcal{F}(\mathbf{u}), \quad (5)$$

where $\mathcal{M}$ is a Riemannian manifold, and $\mathcal{F}$ is an objective function. The Riemannian gradient $\operatorname{grad} \mathcal{F}(\mathbf{u})$ is given by the projection of the unconstrained gradient $\nabla \mathcal{F}(\mathbf{u})$ onto its tangent space $\mathrm{T}_{\mathbf{u}}\mathcal{M}$:

$$\operatorname{grad} \mathcal{F}(\mathbf{u}) = \mathcal{P}_{\mathbf{u}}(\nabla \mathcal{F}(\mathbf{u})), \quad (6)$$

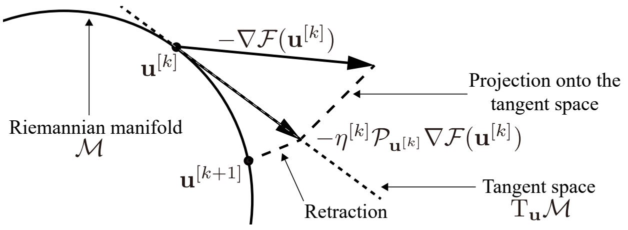
$-\nabla\mathcal{F}(\mathbf{u}^{[k]})$
$\mathbf{u}^{[k]}$
Projection onto the tangent space
Riemannian manifold $\mathcal{M}$
$-\eta^{[k]}\mathcal{P}_{\mathbf{u}^{[k]}}\nabla\mathcal{F}(\mathbf{u}^{[k]})$
$\mathbf{u}^{[k+1]}$
Tangent space $\mathrm{T}_{\mathbf{u}}\mathcal{M}$
Retraction

Fig. 1. An illustration of the Riemannian gradient decent algorithm.

where $\mathcal{P}_{\mathbf{u}}$ is the orthogonal projection onto the tangent space at $\mathbf{u}$. In each iteration of the Riemannian gradient descent algorithm, the map from the tangent space onto the manifold is needed in order to maintain the variable on the manifold $\mathcal{M}$. An usable map from the tangent space onto the manifold is the *retraction*:

$$\operatorname{Ret}_{\mathbf{u}}(\boldsymbol{\nu}) : \mathrm{T}_{\mathbf{u}}\mathcal{M} \ni \boldsymbol{\nu} \mapsto \mathbf{u} \in \mathcal{M}. \quad (7)$$

Then, the Riemannian gradient decent algorithm is given by

$$\mathbf{u}^{[k+1]} = \operatorname{Ret}_{\mathbf{u}^{[k]}}(-\eta^{[k]} \operatorname{grad} \mathcal{F}(\mathbf{u}^{[k]})), \quad (8)$$

where $\eta^{[k]}$ is an appropriate step size, and k is the iteration index. The procedure of the Riemannian gradient descent algorithm is illustrated in Fig. 1. Its global convergence is guaranteed when the objective function is smooth, the manifold is compact, and the step size is appropriate [25].

3. PROPOSED METHOD

In this section, we propose a flexible model-based phase recovery method which is formulated as an optimization over complex-valued phases. The proposed method does not need the prior estimation of the instantaneous frequency, and thus it is irrespective of the estimation error of the instantaneous frequency.

3.1. Proposed formulation

For flexible phase recovery, we do not consider the unwrapped phases but the complex-valued phases defined by $u_{\xi,\tau} = e^{j\phi_{\xi,\tau}}$ where $\phi_{\xi,\tau}$ is the complex argument of the STFT coefficient. We introduce a distance between two complex-valued phases:

$$1 - \operatorname{Re}\left(\frac{u_{\xi,\tau}}{u_{\zeta,\eta}} \right) = 1 - \cos(\phi_{\xi,\tau} - \phi_{\zeta,\eta}), \quad (9)$$

where $\operatorname{Re}(u)$ is the real part of u. The right hand side of Eq. (9) corresponds to the negative log-likelihood associated with the von Mises distribution.[1] The proposed phase recovery is formulated as the following optimization over complex-valued phases $\mathbf{u} \in \mathbb{C}^{K \times T}$:

$$\min_{\mathbf{u}} \mathcal{D}(\mathbf{u}, \mathbf{d}) + \mathcal{G}(\mathbf{u}), \ \text{ s.t. } |u_{\xi,\tau}| = 1, \quad \forall \xi, \tau, \quad (10)$$

where

$$\mathcal{D}(\mathbf{u}, \mathbf{d}) = \sum_{\xi,\tau} \lambda_{\xi,\tau} \left\{ 1 - \operatorname{Re}\Big(\frac{u_{\xi,\tau}}{d_{\xi,\tau}} \Big) \right\} \quad (11)$$

[1] The von Mises distribution $p(\cdot)$ is a probability distribution for a periodic variable ϕ whose probability density function given by $p(\phi; \mu, \kappa) = \exp\big(\kappa \cos(\phi - \mu)\big)/2\pi I_0(\kappa)$, where μ is a circular mean, κ is a concentration, and $I_0(\kappa)$ is the modified Bessel function of the first kind of order 0. The negative log-likelihood can be written as $-\kappa \cos(\phi - \mu) + C(\kappa)$, which is related to the right hand side of Eq. (9).

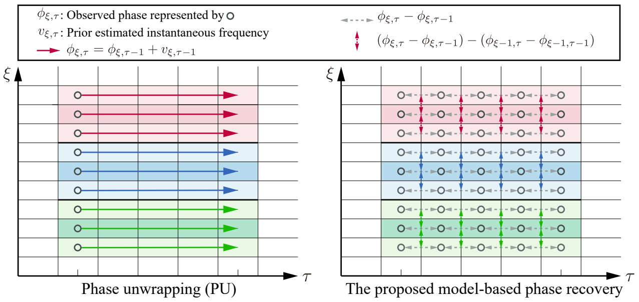
$\phi_{\xi,\tau}$: Observed phase represented by ○
$v_{\xi,\tau}$: Prior estimated instantaneous frequency
$\phi_{\xi,\tau} = \phi_{\xi,\tau-1} + v_{\xi,\tau-1}$
$\phi_{\xi,\tau} - \phi_{\xi,\tau-1}$
$(\phi_{\xi,\tau} - \phi_{\xi,\tau-1}) - (\phi_{\xi-1,\tau} - \phi_{\xi-1,\tau-1})$
ξ
τ
Phase unwrapping (PU)
The proposed model-based phase recovery

Fig. 2. Illustrations of PU and the proposed phase recovery.

indicates the data fidelity term with the parameter $\boldsymbol{\lambda} \in \mathbb{R}_{+}^{K\times T}$, $\mathbf{d} \in \mathbb{C}^{K\times T}$ is the complex-valued phases of the observed signal,

$$\begin{aligned}\mathcal{G}(\mathbf{u}) &= \sum_{\xi,\tau} \gamma_{\xi,\tau} \left\{ 1 - \mathrm{Re}\Big(\frac{u_{\xi,\tau}}{u_{\xi,\tau-1}} \frac{u_{\xi-1,\tau-1}}{u_{\xi-1,\tau}} \Big) \right\}, \\ &= \sum_{\xi,\tau} \gamma_{\xi,\tau} \left\{ 1 - \mathrm{Re}\Big(\frac{v_{\xi,\tau}}{v_{\xi-1,\tau}} \Big) \right\} \end{aligned} \tag{12}$$

indicates the regularization term with the complex-valued instantaneous frequency $v_{\xi,\tau} = u_{\xi,\tau}/u_{\xi,\tau-1}$, and the parameter $\boldsymbol{\gamma} \in \mathbb{R}_{+}^{K\times T}$. In Eq. (12), the idea of the sinusoidal model described in Section 2.1 is incorporated as the regularization term $\mathcal{G}(\mathbf{u})$. The regularization term expects that the instantaneous frequencies of adjacent bins along the frequency direction are the same when $\gamma_{\xi,\tau} > 0$.

The proposed formulation is based on the same assumption as PU, the sinusoidal model, but there exist significant advantages:

- The proposed method is formulated as an optimization problem over complex-valued phases in flexible form. Hence, it can incorporate knowledge of the target signal through adjustable parameters.
- The instantaneous frequency is not treated fixedly, and thus the prior estimation of the instantaneous frequency is not required. Hence, the proposed method is irrespective of the estimation error of the instantaneous frequency.
- The data fidelity is considered at not only the onset but all time frames.

Comparing to PU, these advantages of the proposed method are illustrated in Fig. 2, where circles are bins considering the data fidelity. Focusing on circles in Fig. 2, the proposed method considers the data fidelity at all time frames, while PU considers only at the onset. In addition, while PU modifies phases deterministically with the estimated instantaneous frequency (the solid allow in the left of Fig. 2), the proposed method considers minimization of the difference of the instantaneous frequency among adjacent bins along the frequency direction (the colored dotted allow in the right of Fig. 2). These differences enable the proposed method to be more flexible. The parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ play an important role in the proposed method, and examples of their choice are described later.

3.2. Proposed algorithm for solving Eq. (10)

In this subsection, we introduce the efficient optimization algorithm for the proposed optimization problem given by Eq. (10). It can be interpreted as the minimization of the smooth objective function over complex-valued phases which constructs the Riemannian manifold. The technique of optimization on Riemannian manifolds is suitable for solving it efficiently [25]. The Riemannian manifold constructed by complex-valued phases is given by

$$\mathcal{M} = \{ \mathbf{u} \in \mathbb{C}^{K\times T} : |u_{\xi,\tau}| = 1, \quad \forall \xi, \tau \}. \tag{13}$$

Hence, the proposed method can be reformulated as an unconstrained optimization on the Riemannian manifold:

$$\min_{\mathbf{u}\in\mathcal{M}} \mathcal{D}(\mathbf{u}, \mathbf{d}) + \mathcal{G}(\mathbf{u}). \tag{14}$$

Thanks to this reformulation, the Riemannian gradient decent algorithm can be applied to solve the proposed optimization problem. Considering the Riemannian manifold constructed by complex-valued phases, the tangent space $\mathrm{T}_{\mathbf{u}}\mathcal{M}$ is defined by:

$$\mathrm{T}_{\mathbf{u}}\mathcal{M} = \{ \boldsymbol{\nu} \in \mathbb{C}^{K\times T} : \mathrm{Re}(\boldsymbol{\nu} \odot \mathbf{u}) = \mathbf{0} \ \forall \xi, \tau \}, \tag{15}$$

where $\odot$ is the Hadamard product. The projection onto the tangent space at $\mathbf{u}$ is given by

$$\mathcal{P}_{\mathbf{u}}(\boldsymbol{v}) = \boldsymbol{v} - \mathrm{Re}(\bar{\mathbf{u}} \odot \boldsymbol{v}) \odot \mathbf{u}, \tag{16}$$

In addition, the retraction $\mathrm{Ret}_{\mathbf{u}}(\boldsymbol{\nu})$ at $\mathbf{u}$ is also given by [25]

$$\mathrm{Ret}_{\mathbf{u}}(\boldsymbol{\nu}) = \mathrm{phase}(\mathbf{u} + \boldsymbol{\nu}), \tag{17}$$

where $\mathrm{phase}(\mathbf{u}) = u_{\xi,\tau}/|u_{\xi,\tau}|$ for $u_{\xi,\tau} \neq 0$ and 0 otherwise. We solve the proposed optimization problem by Riemaniann gradient descent algorithm using the above operators as described in Section 2.2. We suppose the Riemaniann gradient descent is suitable for solving the proposed optimization problem for two reasons. First, Riemaniann gradient descent algorithm achieves efficient calculation, comparing to the projected gradient algorithm, thanks to the projection of the unconstrained gradient onto the tangent space. Second, when the step size is appropriate, its global convergence is guaranteed because the Riemannian manifold is compact, and the objective function $\mathcal{F} = \mathcal{D} + \mathcal{P}$ is smooth.

Next, we mention the unconstrained gradient of the objective function in Eq. (14). The objective function of Eq. (14) is a real valued-function over complex-valued variables, and thus Wirtinger calculus is a useful technique for calculation of its gradient [27–29]. In Wirtinger calculus, the objective function $\mathcal{F}(\mathbf{u})$ is considered as a bivariate function $\mathcal{F}(\mathbf{u}, \bar{\mathbf{u}})$, in which two variables, $\mathbf{u}$ and $\bar{\mathbf{u}}$, are treated independently. The derivatives $\partial\mathcal{F}/\partial\mathbf{u}$ and $\partial\mathcal{F}/\partial\bar{\mathbf{u}}$ are also calculated independently, and then the unconstrained gradient is given by $-2\partial\mathcal{F}/\partial\bar{\mathbf{u}}$. According to Wirtinger calculus [27], the partial derivative of the data fidelity term $\mathcal{D}(\mathbf{u}, \mathbf{d})$ is given by:

$$\frac{\partial\mathcal{D}}{\partial\bar{u}_{\xi,\tau}} = \lambda_{\xi,\tau} \frac{\partial}{\partial\bar{u}_{\xi,\tau}} \left\{ 1 - \mathrm{Re}\Big(\frac{u_{\xi,\tau}}{d_{\xi,\tau}} \Big) \right\} = -\frac{\lambda_{\xi,\tau}}{2} d_{\xi,\tau}, \tag{18}$$

where we utilized the relationship: $\mathrm{Re}(z) = (z + \bar{z})/2$. The partial derivative of the regularization term $\mathcal{G}$ is also calculated by:

$$\begin{aligned}\frac{\partial\mathcal{G}}{\partial\bar{u}_{\xi,\tau}} &= \sum_{\zeta=\xi}^{\xi+1} \sum_{\eta=\tau}^{\tau+1} \gamma_{\zeta,\eta} \frac{\partial}{\partial\bar{u}_{\xi,\tau}} \left\{ 1 - \mathrm{Re}\Big(\frac{u_{\zeta,\eta}}{u_{\zeta,\eta-1}} \frac{u_{\zeta-1,\eta-1}}{u_{\zeta-1,\eta}} \Big) \right\}, \\ &= \sum_{\zeta=\xi}^{\xi+1} \sum_{\eta=\tau}^{\tau+1} \gamma_{\zeta,\eta} \frac{\partial\mathcal{G}_{\zeta,\eta}}{\partial\bar{u}_{\xi,\tau}} \end{aligned} \tag{19}$$

where

$$\partial\mathcal{G}_{\xi,\tau}/\partial\bar{u}_{\xi,\tau} = -u_{\xi-1,\tau} u_{\xi,\tau-1} \bar{u}_{\xi-1,\tau-1}/2, \tag{20}$$

$$\partial\mathcal{G}_{\xi,\tau+1}/\partial\bar{u}_{\xi,\tau} = -u_{\xi-1,\tau} u_{\xi,\tau+1} \bar{u}_{\xi-1,\tau+1}/2, \tag{21}$$

$$\partial\mathcal{G}_{\xi+1,\tau}/\partial\bar{u}_{\xi,\tau} = -u_{\xi+1,\tau} u_{\xi,\tau-1} \bar{u}_{\xi+1,\tau-1}/2, \tag{22}$$

$$\partial\mathcal{G}_{\xi+1,\tau+1}/\partial\bar{u}_{\xi,\tau} = -u_{\xi,\tau+1} u_{\xi+1,\tau} \bar{u}_{\xi+1,\tau+1}/2. \tag{23}$$

Algorithm 2 The proposed model-based phase recovery

```
Input: complex-valued phases of the observed signal $\mathbf{d}$, the data fidelity parameter $\boldsymbol{\lambda}$, and the regularization parameter $\boldsymbol{\gamma}$
Output: $\mathbf{u}^{[k+1]}$
Set $\mathbf{u}^0 = \mathbf{d}$
for $k = 0, 1, \ldots$ do
    $\mathrm{grad}\, \mathcal{F}(\mathbf{u}^{[k]}) = \nabla \mathcal{F}(\mathbf{u}^{[k]}) - \mathrm{Re}\big(\bar{\mathbf{u}}^{[k]} \odot \nabla \mathcal{F}(\mathbf{u}^{[k]})\big) \odot \mathbf{u}^{[k]}$
    where $\nabla \mathcal{F}(\mathbf{u}^{[k]})$ is calculated by sum of Eq. (18) and (19).
    Compute a step size $\eta^{[k]}$ which satisfies the Almijo condition.
    Set $\mathbf{u}^{[k+1]} = \mathrm{phase}(\mathbf{u}^{[k]} - \eta^{[k]} \mathrm{grad}\, \mathcal{F}(\mathbf{u}^{[k]}))$.
end for
```

The unconstrained gradient $\nabla \mathcal{F}$ is calculated from Eq. (18) and (19), and then the Riemannian gradient is given by:

$$\mathrm{grad}\, \mathcal{F}(\mathbf{u}) = \nabla \mathcal{F}(\mathbf{u}) - \mathrm{Re}(\bar{\mathbf{u}} \odot \nabla \mathcal{F}(\mathbf{u})) \odot \mathbf{u}. \tag{24}$$

The Riemannian gradient decent algorithm for the proposed phase recovery is summarized in Algorithm 2. Thanks to the representation of the phase spectrogram by the Riemannian manifold, the technique of optimization on Riemannian manifolds and Wiltinger calculus gives the efficient solution of the proposed formulation.

3.3. Choice of regularization parameters

In the proposed method, the parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ play important roles. Here we introduce choices of these parameters. At first, we utilize the amplitude spectrogram $\mathbf{A}$ as the data fidelity parameter $\boldsymbol{\lambda}$. That is, the data fidelity at the large amplitude was emphasized. We assume the amplitude peaks correspond to the sinusoids, and the observed phases are relatively reliable. In contrast, the observed phases with the small amplitude are modified by the regularization term.

The regularization parameter $\boldsymbol{\gamma}$ should define the region of influence introduced in Section 2.1. The instantaneous frequency in the same region should be same, and thus $\gamma_{\xi,\tau} = \gamma_{\text{fix}} \in \mathbb{R}_+$ in each region. Namely, minimizing the regularization term corresponds to aligning the instantaneous frequency in the same region. In contrast, we assume that the instantaneous frequency in the different regions is different. In order to avoid that the instantaneous frequencies of bins in the different region are aligned, $\gamma_{\xi,\tau} = 0$ at the bound of the regions. In addition, $\gamma_{\xi,\tau} = 0$ at the onset, because the instantaneous frequency at the onset is difficult to estimate from the phase spectrogram of the former time frame.

4. NUMERICAL EXPERIMENTS

4.1. Condition

We applied the proposed phase recovery to noise reduction for audio signals (the clarinet and the piano sound) from songKitamura [30] corrupted by the additive Gaussian noise, and the Signal-to-Noise Ratio (SNR) was set to $\{-10,\ 0,\ 10\}$ dB. The audio signals were sampled at 44100 Hz, and the STFT was implemented with the Hann window whose length and the shift length were 1024 samples and 256 samples, respectively. Wiener filter was applied to the spectrogram of the noisy signals in the oracle condition (i.e., the amplitude spectrogram of the target signal and the noise is utilized for calculating Wiener filter), and then phase recovery was applied as post-processing. The performance was evaluated by Signal-to-Distortion Ratio (SDR) and Overall-Perceptual-Score (OPS) as perceptual quality by the PEASS toolbox [31] where we consider noise reduction as separation into the audio signal and the noise. As the target signal is not speech, we utilize OPS for the audio quality assessment instead of the methods for speech (e.g., PESQ).

Table 1. Comparison of the output SDR and OPS in each condition for the clarinet sound.

Input SNR	-10 [dB]		0 [dB]		10 [dB]	
	SDR	OPS	SDR	OPS	SDR	OPS
Wiener	11.5	35.7	18.0	44.3	25.5	37.3
GLA [11]	10.2	**39.9**	16.6	**45.9**	24.7	44.9
PU [16]	−1.5	20.0	−1.6	31.6	−0.6	34.0
Proposed	**11.7**	36.9	**18.4**	44.6	**25.8**	**45.0**
Oracle	15.3	42.9	22.2	47.6	29.7	30.2

Table 2. Comparison of the output SDR and OPS in each condition for the piano sound.

Input SNR	-10 [dB]		0 [dB]		10 [dB]	
	SDR	OPS	SDR	OPS	SDR	OPS
Wiener	**10.9**	36.7	**17.1**	49.1	**24.1**	48.2
GLA [11]	9.3	36.6	15.9	**49.4**	23.1	48.4
PU [16]	−1.3	20.5	1.1	38.9	1.2	41.6
Proposed	**10.9**	**37.9**	**17.1**	**49.4**	**24.1**	**48.6**
Oracle	14.7	47.2	21.2	40.2	28.2	30.6

The proposed method was compared with GLA [11] and PU [16]. In PU and the proposed method, the region of influence was estimated by Eq. (4), and the onset was estimated by the temporal QIFFT following [16]. In PU, the instantaneous frequency was estimated by the QIFFT [24], and the onset phase was set to the oracle phase instead of estimating the onset phase in order to eliminate the estimation error of the onset phase. The proposed method and GLA were iterated 500 times.

4.2. Results of phase recovery

Tables 1 and 2 show the output SDR and OPS achieved by each algorithm for the clarinet and the piano sounds, respectively. As shown in Tables 1 and 2, PU resulted in the lowest SDR and OPS for all conditions. This is because PU takes into account the phase of the observed signal only at the onset and ignores the phase of the observed signals without the onset. Thus, the performance is easily corrupted by the estimation error of the instantaneous frequency even if the oracle phase is known at the onset. While GLA improves OPS in some cases, it also could not improve SDR. In contrast, the proposed method simultaneously improved or maintained SDR and OPS in all conditions. Specifically, the proposed method achieved the highest SDR for the clarinet as illustrated in Table 1. On the other hand, it achieved the highest OPS for the piano as illustrated in Table 2. The proposed method is formulated in the flexible form, and it can incorporate with prior knowledge of the target signal.

These results suggest the effectiveness of the proposed flexible formulation comparing to PU. We suppose the proposed method is applicable to other tasks including source separation.

5. CONCLUSION

In this paper, we proposed a flexible model-based phase recovery method which is formulated as an optimization problem. The proposed method does not need the prior estimation of the instantaneous frequency, and thus it is irrespective of the estimation error of the instantaneous frequency. The advantages of the proposed method were shown through noise reduction of audio signals. As a future work, other regularization terms should be considered.

6. REFERENCES

[1] 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.

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. on Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 443–445, Apr. 1985.

[3] 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.

[4] P. Vary, "Noise suppression by spectral magnitude estimation -mechanism and theoretical limits-," *Signal Process.*, vol. 8, no. 4, pp. 387–400, July 1985.

[5] C. Févotte, N. Bertin, and J. L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural Comput.*, vol. 21, no. 3, pp. 793–830, Mar. 2009.

[6] P. Mowlaee and R. Saeidi, "On phase importance in parameter estimation in single-channel speech enhancement," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2013, pp. 7462–7466.

[7] 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.

[8] T. Gerkmann, M. Krawczyk, 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.

[9] P. Magron, R. Badeau, and B. David, "Phase recovery in NMF for audio source separation: An insightful benchmark," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2015.

[10] P. Mowlaee, R. Saeidi, and Y. Stylianou, "Advances in phase-aware signal processing in speech communication," *Speech Commun.*, vol. 81, pp. 1–29, July 2016.

[11] 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.

[12] 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, 2007.

[13] X. Zhu, G. T. Beauregard, and L. Wyse, "Real-time iterative spectrum inversion with look-ahead," in *IEEE Int. Conf. Multimed. Expo*, July 2006, pp. 229–232.

[14] T. Bendory, Y. C. Eldar, and N. Boumal, "Non-convex phase retrieval from STFT measurements," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 467–484, Jan. 2018.

[15] S. Arik, M. Chrzanowski, A. Coates, G. Diamos, A. Gibiansky, Y. Kang, X. Li, J. Miller, J. Raiman, S. Sengupta, and M. Shoeybi, "Deep voice 2: Multi-speaker neural text-to-speech," *arXiv:1705.08947*, 2017.

[16] P. Magron, R. Badeau, and B. David, "Phase reconstruction of spectrograms with linear unwrapping: Application to audio signal restoration," in *Eur. Signal Process.Conf. (EUSIPCO)*, Aug. 2015.

[17] M. Krawczyk and T. Gerkmann, "STFT phase reconstruction in voiced speech for an improved single-channel speech enhancement," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 12, pp. 19031–1940, Dec. 2014.

[18] P. Magron, R. Badeau, and B. David, "Model-based STFT phase recovery for audio source separation," *IEEE/ACM Trans. on Audio, Speech, Lang. Process.*, vol. 26, no. 6, pp. 1095–1105, June 2018.

[19] K. Yatabe and Y. Oikawa, "Phase corrected total variation for audio signals," in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 656–660.

[20] J. Kulmer and P. Mowlaee, "Harmonic phase estimation in single-channel speech enhancement using von mises distribution and prior SNR," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 5063–5067.

[21] F. Mayer, D. S. Williamson, P. Mowlaee, and D. Wang, "Impact of phase estimation on single-channel speech separation based on time-frequency masking," *J. Acoust. Soc. Am.*, vol. 141, no. 6, pp. 4668–4679, 2017.

[22] J. Laroche and M. Dolson, "Improved phase vocoder time-scale modification of audio," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 3, pp. 323–332, May 1999.

[23] M. Dolson, "The phase vocoder: A tutorial," *Compt. Music J.*, vol. 10, no. 4, pp. 14–27, 1986.

[24] M. Abe and J. O. III Smith, "Design criteria for simple sinusoidal parameter estimation based on quadratic interpolation of FFT magnitude peaks," in *Audio Eng. Soc. Conv. 117*, Oct. 2004.

[25] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton Univ. Press, Princeton, NJ, USA, 2008.

[26] W. Ring and B. Wirth, "Optimization methods on Riemannian manifolds and their application to shape space," *SIAM J. Optimi*, vol. 22, no. 2, pp. 596–27, 2012.

[27] Laurent S., Marc V. Bare., and Lieven D. L., "Unconstrained optimization of real functions in complex variables," *SIAM J. Opt.*, vol. 22, no. 3, pp. 879–898, 2012.

[28] T. Adali and P. J. Schreier, "Optimization and estimation of complex-valued signals: Theory and applications in filtering and blind source separation," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 112–128, Sept. 2014.

[29] K. Kreutz-Delgado, "The complex gradient operator and the CR-calculus," *arXiv:0906.4835*, 2009.

[30] D. Kitamura, H. Saruwatari, H. Kameoka, Y. Takahashi, K. Kondo, and S. Nakamura, "Multichannel signal separation combining directional clustering and nonnegative matrix factorization with spectrogram restoration," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 4, pp. 654–669, Apr. 2015, Available: http://d-kitamura.net/.

[31] V. Emiya, E. Vincent, N. Harlander, and V. Hohmann, "Subjective and objective quality assessment of audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2046–2057, Sept. 2011.