

振幅スペクトログラムからの非Griffin-Lim型音源信号復元手法

Non-Griffin-Lim Type Sound Recovery from Magnitude Spectrogram

中津 龍星 北原 大地 平林 晃
立命館大学 大学院情報理工学研究科

Ryusei NAKATSU Daichi KITAHARA Akira HIRABAYASHI
Graduate School of Information Science and Engineering, Ritsumeikan University

あらまし 近年、音声音響信号処理において、観測信号を短時間フーリエ変換して得られるスペクトログラムの振幅値を処理し、所望の音源信号を得る手法が数多く開発されている。一方、音楽等の解析に関しては、周波数成分を対数軸上でサンプリングする、定Q変換のスペクトログラムが有効である。しかし、定Q変換は逆変換が簡単な閉形式として表現できないため、Griffin-Lim型のアルゴリズムによって振幅スペクトログラムから音源信号を復元することが困難であった。本研究では、定Q変換等の振幅スペクトログラムからも、音源信号を復元できるようにするために、逆変換計算を必要としない非Griffin-Lim型のアルゴリズムを提案する。提案法では、SDCA-ADMMを利用することで、逆変換計算を回避しながらも高速な収束を実現する。数値実験により提案法の有効性を示す。

1 はじめに

信号の時間周波数解析手段として、スペクトログラムが広く用いられている。近年、音声音響信号処理を中心に、振幅スペクトログラムを加工することで所望の時間信号を生成する技術が注目されている。例えば音源分離の分野では、複数の音源が混在した混合音の振幅スペクトログラムに対して、適切な時間周波数マスクを乗じることで、特定音源のみからなる振幅スペクトログラムを推定する手法が提案されている [1]。推定した振幅スペクトログラムに適切な位相スペクトログラムを付与できれば、特定音源のみを含む時間信号が復元される。

適切な位相を求めるために、スペクトログラム空間の冗長性を利用することが多い。通常、スペクトログラムは短時間フーリエ変換 (Short Time Fourier Transform: STFT) によって計算される。このとき、時間フレームをオーバーラップさせながら周波数成分を計算するため、スペクトログラム空間の次元は、元の時間信号空間の次元 (時間信号の長さ) より大きくなる。したがって、スペクトログラムは元の信号に対し冗長な表現であり、STFTの値域はスペクトログラム空間内のある低次元部分空間となる。理想的な

状況下では、振幅に付与した位相が適切ならば、得られるスペクトログラムは上記の低次元部分空間に属するはずである。そこで、GriffinとLimは、所望の振幅値を持つスペクトログラム集合とSTFTの値域集合に交互に射影を行うことで、適切な位相が付与されたスペクトログラムを推定した [2]。この手法はGriffin-Lim Algorithm (GLA) という名称で広く知られている。GLAの高速化も行われており、PerraudinらはFast GLAと呼ばれる手法 [3] を、升山らは交互方向乗数法 (Alternating Direction Method of Multipliers: ADMM) [4] に基づく手法 [5] を提案している。2つのGriffin-Lim型高速位相復元手法 [3], [5] は、どちらも同等の優れた収束性能を有することが報告されている。

スペクトログラムの計算手法はSTFT以外にも複数提案されており [6], [7]、特に周波数成分を対数軸上で等間隔にサンプリングする定Q変換 (Constant-Q Transform: CQT) が音楽等の解析において有用である [8]。これは、音源信号の倍音成分をSTFTと比較して高精度に表現できるためである。しかしながら、STFTと異なり、CQTでは逆変換計算が容易ではなく、CQTの値域集合に射影を行うことも困難となる。結果としてGriffin-Lim型の位相復元手法が適用できないため、CQT振幅スペクトログラムを加工して所望の信号を生成する研究はあまり行われていない。

本研究では、STFT以外の振幅スペクトログラムの応用可能性を広げるために、逆変換計算不要な非Griffin-Lim型の時間信号復元手法を提案する。提案法では、スペクトログラムが探索変数として記述された従来の最適化問題を、時間信号が変数として記述された問題へと再定式化する。そして、この問題の双対問題をStochastic Dual Coordinate Ascent ADMM (SDCA-ADMM) [9] によって解くことで、時間信号を復元する。提案法では、SDCA-ADMMに基づいて、近似拡張ラグランジュ関数と確率的最適化を利用することで、逆変換計算を回避しながらも高速に信号を復元できる。STFT振幅スペクトログラムからの信号復元の数値実験により、提案法では逆変換を使用しないにも関わらず、ADMMに基づくGriffin-Lim型の高速手法 [5] と同程度の優れた収束性能が実現できることを示す。

2 Griffin-Lim 型位相復元手法

2.1 Griffin-Lim Algorithm

離散時間信号 $x[t]$ を、フレーム長 L 、フレームシフト量 $\xi (< L)$ 、窓関数 $w[\tau]$ で STFT して得られるスペクトログラムを行列 $X := (\hat{x}_{i,j}) \in \mathbb{C}_{\downarrow}^{L \times J}$ で表すと、各成分は

$$\hat{x}_{i,j} := \sum_{\tau=0}^{L-1} w[\tau] x[(j-1)\xi + \tau] e^{-i\frac{2\pi(i-1)\tau}{L}}$$

のように計算される。ここで、 $i \in \mathbb{C}$ は $i^2 = -1$ を満たす虚数単位、 $\mathbb{C}_{\downarrow}^{L \times J} := \{(\hat{x}_{i,j}) \in \mathbb{C}^{L \times J} \mid \hat{x}_{1,j} \in \mathbb{R} \text{ for all } j \text{ and } \hat{x}_{i,j} = \hat{x}_{L-i+2,j} \text{ for all } j \text{ and } i = 2, 3, \dots, \lfloor \frac{L+2}{2} \rfloor\}$ はスペクトログラム空間を表す。ここで、 $\bar{\hat{x}}_{i,j}$ は $\hat{x}_{i,j}$ の複素共役を、 $[\cdot]$ は床関数を表す。 $\xi < L$ から、スペクトログラム空間の次元 JL は離散時間信号のサンプル数 $(J-1)\xi + L$ より大きくなるため、STFT を線形写像 $\mathcal{S} : \mathbb{R}^{(J-1)\xi + L} \ni \mathbf{x} \mapsto \mathcal{S}(\mathbf{x}) \in \mathbb{C}_{\downarrow}^{L \times J}$ として表現すると、STFT の値域集合 $\mathcal{R}_S := \{X \in \mathbb{C}_{\downarrow}^{L \times J} \mid \exists \mathbf{x} X = \mathcal{S}(\mathbf{x})\}$ は、スペクトログラム空間 $\mathbb{C}_{\downarrow}^{L \times J}$ 内のある低次元部分空間となる。

GLA [2] では上記の次元の大小関係を利用して、非負値行列 $M \in \mathbb{R}_{+\downarrow}^{L \times J}$ ($\subset \mathbb{C}_{\downarrow}^{L \times J}$) で表される振幅スペクトログラムから、位相を含むスペクトログラム $X \in \mathbb{C}_{\downarrow}^{L \times J}$ を

$$X^{(l+1)} = \mathcal{P}_{\mathcal{R}_S}(\mathcal{P}_{\mathcal{M}}(X^{(l)})) \quad (1)$$

の反復で推定する。式 (1) は、所望の振幅値を持つスペクトログラム集合 $\mathcal{M} := \{X \in \mathbb{C}_{\downarrow}^{L \times J} \mid |X| := (|\hat{x}_{i,j}|) = M\}$ と STFT の値域集合 \mathcal{R}_S の共通部分を求めるために、交互射影法を適用しているにすぎない。各集合への射影は

$$\begin{cases} \mathcal{P}_{\mathcal{M}}(X) = M \odot \Theta_X \\ \mathcal{P}_{\mathcal{R}_S}(X) = \mathcal{S} \circ (\mathcal{S}^H \circ \mathcal{S})^{-1} \circ \mathcal{S}^H(X) = \text{STFT}(\text{ISTFT}(X)) \end{cases}$$

のように簡単に計算できる。ここで、 \odot はアダマール積、 $\Theta_X := (e^{i\angle \hat{x}_{i,j}}) \in \mathbb{C}_{\downarrow}^{L \times J}$ は偏角 $\angle \hat{x}_{i,j} \in (-\pi, \pi]$ を用いて定義される X の位相行列、 $(\cdot)^H$ は随伴作用素、 \circ は写像の合成、 $(\cdot)^{-1}$ は逆写像を表し、 $(\mathcal{S}^H \circ \mathcal{S})^{-1} \circ \mathcal{S}^H$ は逆短時間フーリエ変換 (Inverse STFT: ISTFT) となる。式 (1) は単純かつ強力ではあるものの、位相を高精度に復元するためには多くの更新回数を必要とする。GLA の高速化も研究されており、以下で ADMM に基づく手法 [5] を説明する。

2.2 Alternating Direction Method of Multipliers

凸関数 $f : \mathbb{R}^m \rightarrow \mathbb{R}_{\infty}$ と $g : \mathbb{R}^k \rightarrow \mathbb{R}_{\infty}$ の和の最小化

$$\underset{\mathbf{x} \in \mathbb{R}^m, \mathbf{z} \in \mathbb{R}^k}{\text{minimize}} \quad f(\mathbf{x}) + g(\mathbf{z}) \quad \text{s.t. } A\mathbf{x} + B\mathbf{z} = \mathbf{c} \quad (2)$$

を考える。ここで、 $A \in \mathbb{R}^{n \times m}$ 、 $B \in \mathbb{R}^{n \times k}$ 、 $\mathbf{c} \in \mathbb{R}^n$ である。ADMM [4] は式 (2) の問題に対する拡張ラグランジュ関数

$$\begin{aligned} \mathcal{L}_{\rho}(\mathbf{x}, \mathbf{z}, \mathbf{u}) := & f(\mathbf{x}) + g(\mathbf{z}) - \mathbf{u}^T (A\mathbf{x} + B\mathbf{z} - \mathbf{c}) \\ & + \frac{\rho}{2} \|A\mathbf{x} + B\mathbf{z} - \mathbf{c}\|_2^2 \end{aligned}$$

を構築する。ここで、 $\rho > 0$ である。そして、 \mathcal{L}_{ρ} を \mathbf{x} と \mathbf{z} に関して交互に最小化し、双対変数 $\mathbf{u} \in \mathbb{R}^n$ に関して最大化することによって、式 (2) の問題を解く。結果として、

$$\begin{cases} \mathbf{x}^{(l+1)} = \underset{\mathbf{x} \in \mathbb{R}^m}{\text{argmin}} \quad f(\mathbf{x}) + \frac{\rho}{2} \|A\mathbf{x} + B\mathbf{z}^{(l)} - \mathbf{c} - \frac{1}{\rho}\mathbf{u}^{(l)}\|_2^2 \\ \mathbf{z}^{(l+1)} = \underset{\mathbf{z} \in \mathbb{R}^k}{\text{argmin}} \quad g(\mathbf{z}) + \frac{\rho}{2} \|A\mathbf{x}^{(l+1)} + B\mathbf{z} - \mathbf{c} - \frac{1}{\rho}\mathbf{u}^{(l)}\|_2^2 \\ \mathbf{u}^{(l+1)} = \mathbf{u}^{(l)} - \rho(A\mathbf{x}^{(l+1)} + B\mathbf{z}^{(l+1)} - \mathbf{c}) \end{cases} \quad (3)$$

のような更新アルゴリズムとなる。ADMM を用いて問題を解く際には、式 (3) において 1 行目及び 2 行目の argmin が簡単に求まるように注意して定式化を行う必要がある。

2.3 ADMM に基づいた位相復元手法

GLA では、 $\mathcal{M} \cap \mathcal{R}_S \neq \emptyset$ であると仮定して、交互射影法により共通部分に属するスペクトログラムを求めようとしていた。升山らは、 $\mathcal{M} \cap \mathcal{R}_S = \emptyset$ となる場合も考慮して、より一般化された以下の非凸最適化問題

$$\underset{X \in \mathcal{M}}{\text{minimize}} \quad \frac{1}{2} \|X - \mathcal{P}_{\mathcal{R}_S}(X)\|_{\text{F}}^2 \quad (4)$$

を解くことで位相復元を行った [5]。式 (4) は非凸な問題ではあるが、式 (2) の \mathbf{x} を $X \in \mathbb{C}_{\downarrow}^{L \times J}$ 、 \mathbf{z} を $Z \in \mathbb{C}_{\downarrow}^{L \times J}$ 、 f を集合 \mathcal{M} の指示関数、 g を式 (4) の損失関数、 A を恒等写像、 B を符号反転写像、 \mathbf{c} を零行列とすれば、形式的に

$$\begin{cases} X^{(l+1)} = \mathcal{P}_{\mathcal{M}}(Z^{(l)} + \frac{1}{\rho}U^{(l)}) \\ Z^{(l+1)} = \frac{1}{\rho+1} [\rho X^{(l+1)} - U^{(l)} + \mathcal{P}_{\mathcal{R}_S}(X^{(l+1)} - \frac{1}{\rho}U^{(l)})] \\ U^{(l+1)} = U^{(l)} - \rho(X^{(l+1)} - Z^{(l+1)}) \end{cases} \quad (5)$$

のような更新式が得られる。この手法は Fast GLA [3] と同程度の優れた収束性能を示すことが報告されている。

3 逆変換計算を回避した非 Griffin-Lim 型信号復元手法

Griffin-Lim 型の位相復元手法 [2], [3], [5] では、いずれも逆変換計算、すなわち $(\mathcal{S}^H \circ \mathcal{S})^{-1}$ の計算が必要となる。STFT の場合には、 $\mathcal{S}^H \circ \mathcal{S}$ が簡単な対角行列となるため逆行列の計算も容易となるが、 \mathcal{S} が CQT 等を表す場合には、 $(\mathcal{S}^H \circ \mathcal{S})^{-1}$ の計算が困難となり、式 (1) や式 (5) の更新式が適用できなくなる。更には、式 (4) のようにスペクトログラムが探索変数として記述された最適化問題を考えた場合には、仮に最適解 \hat{X} が $(\mathcal{S}^H \circ \mathcal{S})^{-1}$ の計算を回避して求めたとしても、最終的に時間信号 $\mathbf{x} := x[t]$ を求める際に、後処理として逆変換 $(\mathcal{S}^H \circ \mathcal{S})^{-1} \circ \mathcal{S}^H$ が必要となる。

本研究ではまず、後処理としての逆変換 $(\mathcal{S}^H \circ \mathcal{S})^{-1} \circ \mathcal{S}^H$ を不要にするために、式 (4) の最適化問題を、時間信号が探索変数として記述されたほぼ等価な最適化問題

$$\underset{\mathbf{x} \in \mathbb{R}^{(J-1)\xi + L}}{\text{minimize}} \quad \frac{1}{2} \|M - |\mathcal{S}(\mathbf{x})|\|_{\text{F}}^2 \quad (6)$$

へと再定式化する。そして、式 (6) の問題の解を $(S^H \circ S)^{-1}$ の計算を回避しながらも高速に求める、非 Griffin-Lim 型の信号復元手法を提案する。 $(S^H \circ S)^{-1}$ の回避と高速な収束という一見相反する 2 つの事象を両立させるために、提案法では SDCA-ADMM に基づいて非凸な問題を解く。

3.1 Stochastic Dual Coordinate Ascent ADMM

鈴木は、式 (3) の変数 \mathbf{u} が双対問題の最適解に収束する性質と、確率的最適化理論を融合させた SDCA-ADMM を提案した [9]。SDCA-ADMM では以下の凸最適化問題

$$\underset{\mathbf{x} \in \mathbb{R}^m}{\text{minimize}} \sum_{i=1}^n f_i(\mathbf{a}_i^T \mathbf{x}) + g(B^T \mathbf{x}) \quad (7)$$

を考える。ここで、 $A := (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times r}$, $f: \mathbb{R}^n \ni \mathbf{u} \mapsto \sum_{i=1}^n f_i(u_i) \in \mathbb{R}_\infty$ がデータ整合性かつ $g: \mathbb{R}^r \rightarrow \mathbb{R}_\infty$ が正則化関数である状況を想定している。式 (7) の主問題に対する双対問題は

$$\underset{\mathbf{u} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^r}{\text{minimize}} \sum_{i=1}^n f_i^*(u_i) + g^*(\mathbf{v}) \quad \text{s.t. } A\mathbf{u} + B\mathbf{v} = \mathbf{0} \quad (8)$$

のように与えられる。ここで、 f_i^* と g^* はそれぞれ f_i と g の共役関数である。式 (8) の問題を ADMM により解けば、式 (3) における \mathbf{x} と \mathbf{u} の関係が逆転するため、双対変数 \mathbf{x} が式 (7) の問題の解に収束することになる。

しかしながら、上記のアプローチは多くの場合に、式 (3) の 1 行目及び 2 行目の argmin が簡単に計算できないことにより実現できなくなってしまう。そこで SDCA-ADMM では、近接項を加えた以下の近似拡張ラグランジュ関数 $\hat{\mathcal{L}}_\rho^{(l)}(\mathbf{u}, \mathbf{v}, \mathbf{x}) := f^*(\mathbf{u}) + g^*(\mathbf{v}) - \mathbf{x}^T(A\mathbf{u} + B\mathbf{v})$

$$+ \frac{\rho}{2} \|A\mathbf{u} + B\mathbf{v}\|_2^2 + \frac{1}{2} \|\mathbf{u} - \mathbf{u}^{(l)}\|_{G_A}^2 + \frac{1}{2} \|\mathbf{v} - \mathbf{v}^{(l)}\|_{G_B}^2$$

を \mathbf{u} と \mathbf{v} に関して交互に最小化し、 \mathbf{x} に関して最大化する。ここで、 $G_A := \rho(\eta_A I_n - A^T A)$ と $G_B := \rho(\eta_B I_r - B^T B)$ は、 $(A^T A)^{-1}$ と $(B^T B)^{-1}$ の計算を回避するために導入された正定値行列である。更には、式 (8) の \mathbf{u} に関する損失関数は成分 u_i 毎に分離されており、確率的最適化を簡単に導入できる。結果として、インデックス集合 $\mathcal{I} := \{1, 2, \dots, n\}$ を b 個に分割した際の更新アルゴリズムは

$$\left[\begin{array}{l} \text{Choose one index } d \in \{1, 2, \dots, b\} \text{ uniformly at random} \\ \mathbf{q}^{(l+1)} = \mathbf{v}^{(l)} + \frac{1}{\rho\eta_B} B^T [\mathbf{x}^{(l)} - \rho(A\mathbf{u}^{(l)} + B\mathbf{v}^{(l)})] \\ \mathbf{v}^{(l+1)} = \mathbf{q}^{(l+1)} - \frac{1}{\rho\eta_B} \text{prox}_{\rho\eta_B g}(\rho\eta_B \mathbf{q}^{(l+1)}) \\ \mathbf{p}_{\mathcal{I}_d}^{(l+1)} = \mathbf{u}_{\mathcal{I}_d}^{(l)} + \frac{1}{\rho\eta_{A_{\mathcal{I}_d}}} A_{\mathcal{I}_d}^T [\mathbf{x}^{(l)} - \rho(A\mathbf{u}^{(l)} + B\mathbf{v}^{(l+1)})] \\ \mathbf{u}_{\mathcal{I}_d}^{(l+1)} = \mathbf{p}_{\mathcal{I}_d}^{(l+1)} - \frac{1}{\rho\eta_{A_{\mathcal{I}_d}}} \text{prox}_{\rho\eta_{A_{\mathcal{I}_d}} f_{\mathcal{I}_d}}(\rho\eta_{A_{\mathcal{I}_d}} \mathbf{p}_{\mathcal{I}_d}^{(l+1)}) \\ \mathbf{x}^{(l+1)} = \mathbf{x}^{(l)} - \gamma\rho [A\mathbf{u}^{(l+1)} + B\mathbf{v}^{(l+1)} \\ \quad - (1 - \phi)(A\mathbf{u}^{(l)} + B\mathbf{v}^{(l)})] \end{array} \right. \quad (9)$$

と与えられる。ここで、 $\mathcal{I}_d \neq \emptyset$ は、 $\bigcup_{d=1}^b \mathcal{I}_d = \mathcal{I}$ かつ $\mathcal{I}_d \cap \mathcal{I}_{d'} = \emptyset$ ($d \neq d'$) を満たす。 $\gamma > 0$, $\eta_{A_{\mathcal{I}_d}} > \|A_{\mathcal{I}_d}\|_{\text{op}}^2$, $\eta_B > \|B\|_{\text{op}}^2$ で、 $\phi = 1/b$ は各 u_i が選ばれる確率である。また、prox は近接写像を表しており、 f_i と g の近接写像が簡単に計算可能であれば、式 (9) の更新式を実行できる。

3.2 SDCA-ADMM に基づいた時間信号復元手法

式 (6) の非凸な最適化問題を SDCA-ADMM により形式的に解く。式 (7) の $f(\mathbf{u}) = \sum_i f_i(u_i)$ を

$$\sum_{i,j} f_{i,j}(\hat{u}_{i,j}) := \sum_{i,j} \frac{(m_{i,j} - |\hat{u}_{i,j}|)^2}{2} = \frac{1}{2} \|M - |U|\|_F^2$$

とし、 g を常に 0 を返す関数、 A を S^H とすれば、式 (6) の問題の双対問題は

$$\underset{U \in \mathbb{C}_\dagger^{L \times J}}{\text{minimize}} \sum_{i,j} f_{i,j}^*(\hat{u}_{i,j}) \quad \text{s.t. } S^H(U) = \mathbf{0} \quad (10)$$

と表される。式 (10) の問題を SDCA-ADMM に基づいて解く際に重要となるのが、更新位置を示すインデックス集合 $\mathcal{I}_d \subset \mathcal{I} := \{(i, j)\}_{j=1,2,\dots,J}^{i=1,2,\dots,L}$ をどのように決定するかである。従来のように固定された b 個のブロックから 1 つのみを選ぶ方法では、収束速度に限界があるだけでなく、非凸性に起因して局所解に留まってしまう可能性がある。

本研究では、式 (9) の更新式の収束は、パラメータ ϕ の値と \mathbf{u} の各成分を選択する確率が一致していれば保証されることに着目し、重複のない b 個のブロックのうち 1 つを更新するのではなく、それぞれのブロックを確率 $\phi \in (0, 1]$ で更新することを提案する。つまり、本研究では複数のブロックを同時に選択して更新する以下の更新式

$$\left[\begin{array}{l} \text{Choose multiple indices } \mathcal{D} \in 2^{\{1,2,\dots,b\}} \text{ with probability } \phi \\ P_{\mathcal{I}_{\mathcal{D}}}^{(l+1)} = U_{\mathcal{I}_{\mathcal{D}}}^{(l)} + \frac{1}{\rho\eta} S_{\mathcal{I}_{\mathcal{D}}}(\mathbf{x}^{(l)} - \rho S^H(U^{(l)})) \\ U_{\mathcal{I}_{\mathcal{D}}}^{(l+1)} = \frac{1}{\rho\eta+1} (\rho\eta P_{\mathcal{I}_{\mathcal{D}}}^{(l+1)} - M_{\mathcal{I}_{\mathcal{D}}} \odot \Theta_{P_{\mathcal{I}_{\mathcal{D}}}^{(l+1)}}) \\ \mathbf{x}^{(l+1)} = \mathbf{x}^{(l)} - \gamma\rho S^H(U^{(l+1)} - (1 - \phi)U^{(l)}) \end{array} \right. \quad (11)$$

を提案する。ここで、 $\eta > \|S\|_{\text{op}}^2$ である。ブロックの作成方法は、数値実験で示すように、周波数成分で $\{(1, j)\}_{j=1}^J$, $\{(i, j), (L-i+2, j)\}_{j=1}^J$ ($i = 2, 3, \dots, \lfloor \frac{L+2}{2} \rfloor$) と分ける方法を提案する。式 (11) から、提案法は逆変換 $(S^H \circ S)^{-1} \circ S^H$ を計算せずに時間信号を復元可能であることが分かる。

4 数値実験

STFT 振幅スペクトログラムからの時間信号復元の数値実験により、提案法の収束性能と確率的最適化の妥当性を評価する。音源には、公開されている 7 秒のピアノ音を用いた。サンプリング周波数 16 kHz, フレーム長 64 ms, フレームシフト量 32 ms で、ハニング窓を用いて STFT を行った。式 (1), (5), (11) をそれぞれ 10,000 回反復した。

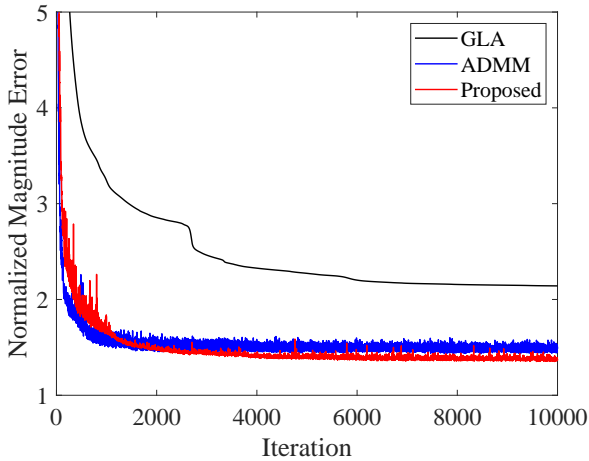


図 1: 各手法における正規化誤差曲線

4.1 提案法と従来法の性能評価

$\mathcal{M} \cap \mathcal{R}_S \neq \emptyset$ を満足する理想的な状況下で、提案法と従来法の性能を評価する。理想的な振幅スペクトログラム M から、提案法と従来法 [2], [5] で元のピアノ音 x を復元し、振幅スペクトログラムの正規化誤差 $100 \frac{\|M - |S(x)|\|_F}{\|M\|_F}$ の値を比較した。ADMM 法 [5] のパラメータは $\rho = 0.08$, 提案法のパラメータは $\gamma = 1, \rho = 0.01, \eta = 1.001L, \phi = 0.75$ とし、インデックス集合 \mathcal{I} は周波数成分で分割した (第 3.2 節参照)。これらは、結果が最良となるように実験的に調整した値である。図 1 に各手法の正規化誤差曲線を示す。図 1 から、ADMM 法が GLA よりも優れた収束性能を有していることが確認できる。そして、驚くべきことに、提案法は逆変換計算を使用しないにも関わらず、ADMM 法と同程度の収束性能を実現していることが分かる。本結果より、CQT 等の振幅スペクトログラムに対しても、提案法を用いれば、高速かつ高精度な時間信号復元が可能となることが期待できる。

4.2 提案法における確率的最適化の妥当性評価

提案法における確率的最適化の妥当性を検証するため、インデックス集合 \mathcal{I} の分割方法と選択確率 ϕ を複数用意し、それぞれの場合で ρ を適切に調整したときの正規化誤差の値を比較する。 \mathcal{I} の分割方法は、(a) 完全分割 $\{(1, j)\} (j = 1, 2, \dots, J), \{(i, j), (L - i + 2, j)\} (i = 2, 3, \dots, \lfloor \frac{L+2}{2} \rfloor; j = 1, 2, \dots, J)$, (b) フレーム分割 $\{(i, j)\}_{i=1}^L (j = 1, 2, \dots, J)$, (c) 周波数分割 $\{(1, j)\}_{j=1}^J, \{(i, j), (L - i + 2, j)\}_{j=1}^J (i = 2, 3, \dots, \lfloor \frac{L+2}{2} \rfloor)$, の 3 種類とした。各ブロックの選択確率は $\phi = 0.5$ または $\phi = 0.75$ とし、比較のため $\phi = 1$ (非確率的最適化) の場合も検証した。図 2 にそれぞれの誤差曲線を示す (厳密には、曲線を見やすくするために 10 点の移動平均値を示している)。 $\phi = 0.5$ と $\phi = 0.75$ の多くの場合で、 $\phi = 1$ の場合よりも正規化誤差が低くなっていること

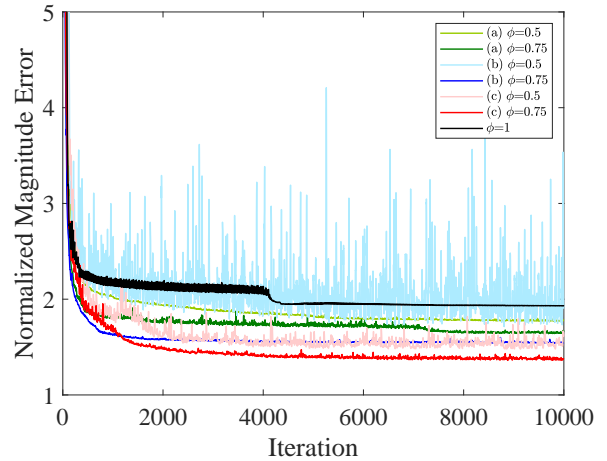


図 2: 提案法における各ブロック分割の正規化誤差曲線

から、確率的最適化によって収束性能が向上されたことが分かる。また、 \mathcal{I} の分割に関しては、周波数成分で分割した場合が、最も良好な結果を安定して復元していた。

5 おわりに

振幅スペクトログラムから時間信号を復元する際に、逆変換計算を行わない非 Griffin-Lim 型の手法を提案した。提案法では、SDCA-ADMM を用いることで、逆変換計算を回避しつつ高速な収束を可能とした。STFT 振幅スペクトログラムからの時間信号復元の実験によって、逆変換を使用する Griffin-Lim 型の手法と比較して、提案法が同等の優れた収束性能を実現できることを確認した。今後は、CQT 振幅スペクトログラムに対しても有効性を検証する。

参考文献

- [1] A. Jansson, E. Humphrey, N. Montecchio, R. Bittner, A. Kumar, and T. Weyde, "Singing voice separation with deep U-Net convolutional networks," in *Proc. ISMIR*, 2017, pp. 745–751.
- [2] D. W. Griffin and J. S. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 32, no. 2, pp. 236–243, 1984.
- [3] N. Perraudin, P. Balazs, and P. L. Søndergaard, "A fast Griffin-Lim algorithm," in *Proc. WASPAA*, 2013, 4 pages.
- [4] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.
- [5] Y. Masuyama, K. Yatabe, and Y. Oikawa, "Griffin-Lim like phase recovery via alternating direction method of multipliers," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 184–188, 2019.
- [6] R. Kronland-Martinet, J. Morlet, and A. Grossmann, "Analysis of sound patterns through wavelet transforms," *Int. J. Pattern Recogn. Artif. Intell.*, vol. 1, no. 2, pp. 273–302, 1987.
- [7] J. C. Brown, "Calculation of a constant Q spectral transform," *J. Acoust. Soc. Am.*, vol. 89, no. 1, pp. 425–434, 1991.
- [8] A. Nagathil and R. Martin, "Evaluation of spectral transforms for music signal analysis," in *Proc. WASPAA*, 2013, 4 pages.
- [9] T. Suzuki, "Stochastic dual coordinate ascent with alternating direction multiplier method," in *Proc. ICML*, 2014, 9 pages.