

劣勾配射影に基づく適応的更新を用いた位相復元*

© 赤石夏輝, 山田宏樹, 矢田部浩平 (農工大)

1 はじめに

位相復元とは、振幅スペクトログラムのみから位相の情報を復元する技術である。音声合成や音声強調への応用が期待され、様々な手法が提案されている。中でも交互方向乗数法 (ADMM) を用いた手法は、少ない反復回数で良い性能を示すが、振幅が劣化している場合には更新が不安定になるという課題がある。この課題を改善するために、本稿では ADMM を用いた手法に対し、劣勾配射影に基づいて実現できる適応的更新を組み込んだ手法を提案する。

2 ADMM を用いた位相復元

音響分野の位相復元手法では、二つの射影作用素

$$P_A(\mathbf{X}) = \mathbf{A} \odot \mathbf{X} \oslash |\mathbf{X}| \quad (1)$$

$$P_C(\mathbf{X}) = \mathcal{G}_w \mathcal{G}_w^\dagger \mathbf{X} \quad (2)$$

を用いる。ただし、 \mathbf{X} はスペクトログラム、 \mathbf{A} は与えられた振幅、 \odot , \oslash , $|\cdot|$ はそれぞれ要素ごとの乗算、除算、絶対値、 \mathcal{G}_w は短時間フーリエ変換 (STFT)、 \mathcal{G}_w^\dagger は \mathcal{G}_w の疑似逆、 A は振幅が \mathbf{A} となるスペクトログラムの集合、 C は無矛盾なスペクトログラムの集合を表す。ここで、スペクトログラムが無矛盾であるとは、スペクトログラムが \mathcal{G}_w の像空間に属していることを表す。広く用いられる位相復元手法である GLA では、式 (1) と式 (2) を交互に射影する [1]。

交互射影よりも一般に効率的である ADMM を適用し、GLA を拡張した位相復元手法 (ADMM-PR) が提案されている [2]。ADMM-PR の更新は

$$\mathbf{X}^{[k+1]} = P_A(\mathbf{Z}^{[k]} - \mathbf{U}^{[k]}) \quad (3)$$

$$\mathbf{Z}^{[k+1]} = \text{prox}_{d_C^2/2\rho}(\rho\mathbf{X}^{[k+1]} + \mathbf{U}^{[k]}) \quad (4)$$

$$\mathbf{U}^{[k+1]} = \mathbf{U}^{[k]} + \mathbf{X}^{[k+1]} - \mathbf{Z}^{[k+1]} \quad (5)$$

で与えられる。ただし、 d_C は集合 C への距離

$$d_C(\mathbf{X}) = \|\mathbf{X} - P_C(\mathbf{X})\|_{\text{Fro}} \quad (6)$$

を、 $\|\cdot\|_{\text{Fro}}$ はフロベニウスノルムを表す。式 (4) は

$$\text{prox}_{d_C^2/2\rho}(\mathbf{V}) = \frac{1}{1+\rho}(\rho\mathbf{V} + P_C(\mathbf{V})) \quad (7)$$

で計算する。ただし、 $\rho > 0$ は更新の度合いを定めるパラメータである。式 (7) は \mathbf{V} と $P_C(\mathbf{V})$ を按分しており、 ρ が小さいほど射影 P_C に近い更新となる。射

影 P_C のみの更新では与えられた振幅 \mathbf{A} によっては解集合が空になり、アルゴリズムが不安定になることがあるが、式 (7) の更新によりこれを回避できる [2]。

ADMM-PR の性能は式 (4) の更新度合いに依存する。 ρ が小さい場合、式 (7) は射影 P_C に近い更新を行うため、アルゴリズムが不安定になる場合がある。一方 ρ が大きい場合には、収束までにより多くの反復が必要となる場合がある。特に、与えられた振幅が劣化している場合、少ない反復で良い解を得るためには式 (4) の更新度合いの調整が重要になる。これを反復ごとに調整できれば、性能の改善が期待できる。

3 提案手法

本稿では、ADMM-PR の改善のために、式 (4) の更新の度合いを適応的に調整する手法を提案する。そのために、適応的更新を実現できることで知られる劣勾配射影に基づく更新式を導入する。

更新式を導入するために、まず劣勾配射影について説明する。劣勾配射影は、 $f(\mathbf{V}') \leq \xi$ となる点 \mathbf{V}' を含む集合 H に点 \mathbf{V} を射影する作用素である。ただし、 $\xi \geq \mu_f$ 、 μ_f は f の最小値である。本稿で劣勾配射影を適用する関数 d_C^2 は微分可能なため、微分可能な関数 f に対する劣勾配射影を考える。関数 f と ξ に関する劣勾配射影は、 ∇f を f の勾配として

$$G_H(\mathbf{V}, \xi) = \mathbf{V} - \lambda \frac{(f(\mathbf{V}) - \xi)_+}{\|\nabla f(\mathbf{V})\|_{\text{Fro}}^2} \nabla f(\mathbf{V}) \quad (8)$$

で与えられる [3]。ただし、 $(\cdot)_+ = \max\{0, \cdot\}$ 、 $\lambda \in (0, 2)$ である。式 (8) を反復的に適用するアルゴリズムは、 $\lambda(f(\mathbf{V}) - \xi)_+ / \|\nabla f(\mathbf{V})\|_{\text{Fro}}^2$ をステップサイズとしてみたとき、 $f(\mathbf{V})$ 、 $\nabla f(\mathbf{V})$ 、 ξ に応じてステップサイズが変化する最急降下ステップとみなせる。すなわち、式 (8) は適応的な更新を可能にする。

提案手法では、式 (4) の代わりに劣勾配射影に基づく適応的な更新式を用いることで性能向上を目指す。この更新式の置き換えの妥当性を示すために、劣勾配射影に基づく更新式は式 (4) と対応がとれることを確認する。 $f = (1/2)d_C^2$ として式 (8) を整理すると

$$G_H(\mathbf{V}, \xi) = (1 - h(\mathbf{V}, \xi)) \mathbf{V} + h(\mathbf{V}, \xi) P_C(\mathbf{V}) \quad (9)$$

となる。ただし、簡単のために関数

$$h(\mathbf{V}, \xi) = \frac{\lambda}{2} \left(1 - \frac{\xi}{(1/2)d_C^2(\mathbf{V})} \right)_+ \quad (10)$$

* Audio phase retrieval with adaptive updates based on subgradient projection. By Natsuki AKAIISHI, Koki YAMADA, and Kohei YATABE (Tokyo University of Agriculture and Technology).

Algorithm 1 提案手法のアルゴリズム

Input: $\mathbf{X}^{[0]}, \lambda \in (0, 2), c \in (0, 1)$
Initialize: $\xi^{[0]} = \frac{c}{2} d_C^2(P_A(\mathbf{X}^{[0]})), \mathbf{Z}^{[0]} = \mathbf{X}^{[0]}, \mathbf{U}^{[0]} = \mathbf{O}$
Output: $\mathbf{X}^{[k+1]}$
1: **for** $k = 0, 1, 2, \dots$ **do**
2: $\mathbf{X}^{[k+1]} = P_A(\mathbf{Z}^{[k]} - \mathbf{U}^{[k]})$
3: $\mathbf{Z}^{[k+1]} = G_H(\mathbf{X}^{[k+1]} + \mathbf{U}^{[k]}, \xi^{[k]})$
4: $\mathbf{U}^{[k+1]} = \mathbf{U}^{[k]} + \mathbf{X}^{[k+1]} - \mathbf{Z}^{[k+1]}$
5: $\xi^{[k+1]} = \frac{c}{2} d_C^2(\mathbf{X}^{[k+1]} + \mathbf{U}^{[k]})$
6: **end for**

を置いた。 $\lambda \in (0, 2)$, $(\cdot)_+ \geq 0$ であるため、 $h(\mathbf{V}, \xi)$ は 0 より大きく 1 より小さい値を返す関数となる。すなわち、式 (9) は \mathbf{V} と $P_C(\mathbf{V})$ を按分している。これは式 (7) の計算と同じである。実際に、 $\xi = 0$, $\lambda = 2/(1 + \rho)$ のとき式 (9) は式 (7) と一致する。すなわち、式 (9) は $d_C^2(\mathbf{V})$ と ξ に応じて更新度合いが変化する、式 (7) に対応する更新式であると言える。

式 (9) の更新は、 $\xi > 0$ のとき $(1/2)d_C^2(\mathbf{V}) = 0$ または $(1/2)d_C^2(\mathbf{V}) = \xi$ となったときに停止する。このとき、 ξ の設定が悪いと、復元が不十分なまま更新が停止してしまう場合がある。これを防ぐために、提案手法では ξ を反復ごとに更新する。式 (8) が集合 H への射影であることを踏まえると、 ξ を前反復の $(1/2)d_C^2(\mathbf{V})$ の値より小さい値に設定することで $(1/2)d_C^2(\mathbf{V})$ を反復ごとに減少させることができる。そこで、 $k + 1$ 反復目の $\xi^{[k+1]}$ を

$$\xi^{[k+1]} = \frac{c}{2} d_C^2(\mathbf{V}^{[k]}) \quad (11)$$

のように定める。 $c \in (0, 1)$ は調整パラメータであり、 c を小さくするほど $d_C^2(\mathbf{V})$ をより減少させる働きがある。これにより、継続的な更新が可能になる。

提案手法のアルゴリズムを Algorithm 1 に示す。3 行目で式 (4) を置き換えた式 (9) の更新、5 行目で式 (11) の ξ の更新をそれぞれ行っている。

4 数値実験

TIMIT データセットの 200 個の音声の振幅スペクトログラムからの位相復元を行った。サンプリング周波数は 16 kHz である。STFT ではハン窓を用い、窓幅及びシフト幅は 32 ms, 16 ms とした。本実験では、GLA, ADMM-PR と提案アルゴリズムを比較する。反復回数は 300 回とした。位相の初期値は 0 とした。入力する振幅には SNR が 0 dB となるノイズを加えた。評価には PESQ と STOI を用いた。

反復ごとの更新度合いの違いが c と ξ のみに依存するように条件を揃えるために、提案手法の λ はすべて、ADMM-PR で $\rho = 0.1$ とした場合の設定に揃えた ($\lambda = 2/(1 + 0.1)$)。実験では、提案手法の c を 0.05, 0.1, 0.3 と設定した。また、 ρ を同程度に設定したときの ADMM-PR と提案手法を比較するために、

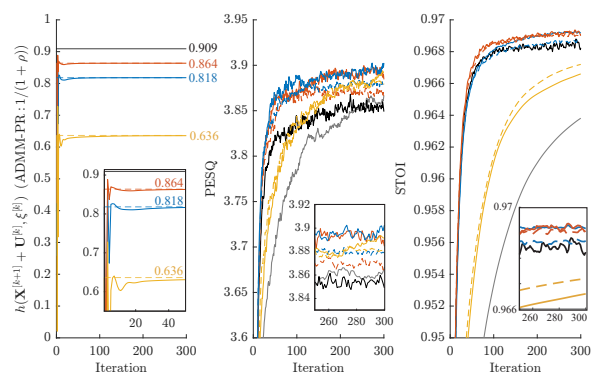


図-1 反復ごとの更新度合い (左), PESQ (中央), STOI (右) の比較。GLA を灰色の線、 $\rho = 0.1$ とした ADMM-PR を黒線で示した。更新度合いとして、式 (9) の h の値を示した。また、 $c = 0.05, 0.1, 0.3$ としたときの提案手法をそれぞれ赤、青、黄色の実線で示し、 $c = 0.05, 0.1, 0.3$ としたときの 300 反復目での ρ で固定した ADMM-PR を対応する色の破線で示した。このときの ADMM-PR の $1/(1 + \rho)$ の値を更新度合いとして表示している。更新度合いの図では 50 反復目までの、PESQ と STOI の図では 250 反復目から 300 反復目までの拡大図を示した。

それぞれの c を設定した提案手法の 300 反復目での ρ の値で固定した ADMM-PR とも比較を行った。

図-1 に反復ごとの更新度合い、PESQ、STOI の全データの平均値を示した。提案手法 (色付き実線) の PESQ と STOI は、 $\rho = 0.1$ としたときの ADMM-PR (黒線) をすべての条件で上回った。また、 $c = 0.05, 0.1$ のとき、提案手法 (実線) の PESQ と STOI が、提案手法の 300 反復目の ρ で固定した ADMM-PR (破線) を上回った。反復ごとの更新度合いの値を見ると、特に 50 反復目までの提案手法の更新度合いが変化していることが分かる。以上から、提案手法は反復初期の更新を安定させ、以降の反復あたりの性能を向上する働きがあると考えられる。一方で、 $c = 0.3$ とした提案手法 (黄色実線) の STOI が ADMM-PR よりも低くなった。これは、更新度合いが余分に小さくなるのが原因だと考えられる。

5 むすび

本稿では、劣勾配射影に基づく適応的更新を ADMM に組み込んだ位相復元手法を提案した。今後は他の位相復元手法や、加速した ADMM に劣勾配射影に基づく更新を組み込んだときの性能変化を調査する。

参考文献

- [1] D. Griffin and J. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust. Speech Signal Process.*, **32**(2), 236–243 (1984).
- [2] Y. Masuyama, K. Yatabe and Y. Oikawa, "Griffin–Lim like phase recovery via alternating direction method of multipliers," *IEEE Signal Process. Lett.*, **26**(1), 184–188 (2019).
- [3] N. Akaiishi, K. Yamada, K. Yatabe and Y. Takayama, "Subgradient-projection-based stable phase retrieval algorithm for X-ray ptychography," *J. Appl. Crystallogr.*, **57**, 13 pages (2024).