

適応フィルタにおける適応学習率アルゴリズムの検討*

© 内田蓮, 矢田部浩平 (農工大), 泉悠斗, 高橋祐, 近藤多伸 (ヤマハ株式会社)

1 はじめに

システム同定やエコーキャンセリングなどの音響信号処理において、適応フィルタは重要な役割を果たしている。適応アルゴリズムの一つである Normalized Least Mean Squares 法 (NLMS) はステップサイズの設定が容易であり、広く用いられている。また、NLMS を含む様々な適応アルゴリズムを一般化した Variable-metric Adaptive Projected Subgradient Method (V-APSM) [1] が提案されている。V-APSM では、計量行列を通じてパラメータ空間の幾何構造を適応的に更新することで、収束性能の向上を実現する。前回、我々は観測信号の周波数特性に基づいた計量行列の設計手法を提案し、NLMS よりも収束が早いことを実験的に確認した [2]。しかし、この手法による計量行列の設計は経験的なものであった。本稿では、深層学習で用いられる適応学習率アルゴリズムを計量行列の設計に導入し、収束性能を検証する。適応学習率アルゴリズムの前処理行列が V-APSM における計量行列として解釈できることを明らかにし、理論的背景に基づいた計量行列の構成を可能にする。

2 V-APSM

$t \in \mathbb{N}$ 反復目において、計量行列と呼ばれる正定値対称行列 $\mathbf{Q}_t \in \mathbb{R}^{N \times N}$ を用いて計量内積 $\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbf{Q}_t} = \mathbf{a}^T \mathbf{Q}_t \mathbf{b}$ ($\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$) を定義する。V-APSM は、計量内積に関する劣勾配射影を用いた適応アルゴリズムである。具体的には、初期点 $\mathbf{w}_1 \in \mathbb{R}^N$ から更新式

$$\mathbf{w}_{t+1} = T_t \left(\mathbf{w}_t - \lambda_t \frac{f_t(\mathbf{w}_t)}{\mathbf{g}_t^T \mathbf{Q}_t^{-1} \mathbf{g}_t} \mathbf{Q}_t^{-1} \mathbf{g}_t \right)$$

によって、 t 反復目での目的関数 f_t を最小化する。ここで、 $f_t: \mathbb{R}^N \rightarrow [0, \infty)$ は連続な凸関数、 $T_t: \mathbb{R}^N \rightarrow \mathbb{R}^N$ は解に関する先験情報を反映するための強アトラクト写像、 $\mathbf{g}_t \in \mathbb{R}^N$ は f_t の劣勾配、 $\lambda_t \in (0, 2)$ はステップサイズ、 $(\cdot)^T$ は転置である。ただし、 $f_t = 0$ では $\mathbf{w}_{t+1} = T_t(\mathbf{w}_t)$ と定義する。V-APSM では、 $\mathbf{Q}_t$ を通じて幾何構造を適応的に変化させることで、収束性能を改善できる。V-APSM を適応フィルタに用いる場合、 f_t は観測信号 $\mathbf{x}_t \in \mathbb{R}^N$ に対する適応フィルタの出力信号 $\mathbf{w}_t^T \mathbf{x}_t$ と所望信号 $d_t \in \mathbb{R}$ の誤差を評価する。特に、 $f_t = |d_t - \mathbf{w}_t^T \mathbf{x}_t|$ 、 $T = \text{Id}$ 、 $\mathbf{Q}_t = \mathbf{I}$ とし

た場合、V-APSM は NLMS と一致する。ここで、 Id は恒等写像、 $\mathbf{I}$ は単位行列である。

3 適応学習率アルゴリズム

深層学習では、目的関数を最小化するように高次元のパラメータを最適化するが、次元ごとに適切な学習率が大きく異なる。この課題を解決するために、適応的な学習率を用いた勾配降下法が提案されている。この更新式は

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \alpha \mathbf{V}_t^{-1} \mathbf{g}_t \quad (1)$$

と定式化でき、前処理行列 $\mathbf{V}_t \in \mathbb{R}^{N \times N}$ によって $\mathbf{g}_t$ が適応的に補正される。ここで、 $\alpha > 0$ は大域的な学習率である。 $\mathbf{V}_t$ を更新するための様々な適応学習率アルゴリズムが提案されているが、多くの手法は Adagrad [3] に基づいている。

Adagrad はオンライン勾配降下法の収束性能を改善するために提案され、前処理行列を

$$\mathbf{G}_t = \mathbf{G}_{t-1} + \mathbf{g}_t \mathbf{g}_t^T, \quad \mathbf{V}_t = \sqrt{\mathbf{G}_t} + \epsilon \mathbf{I}$$

により更新する。ここで、 ϵ は数値安定性のための定数であり、初期値 $\mathbf{G}_0$ は零行列に設定される。

Adagrad の派生手法である Adam [4] は深層学習において標準的に用いられている。Adam では、更新に伴い学習率が減少し続けるという Adagrad の問題に対処するために、 $\mathbf{g}_t \mathbf{g}_t^T$ の累積でなく指数移動平均 (EMA) を用いる。また、計算量削減のために $\mathbf{g}_t \mathbf{g}_t^T$ の対角成分のみを用いる。具体的には、前処理行列を

$$\mathbf{r}_t = (1 - \rho) \mathbf{r}_{t-1} + \rho \text{diag}(\mathbf{g}_t \mathbf{g}_t^T) \quad (2)$$

$$\mathbf{r}_t^{(\text{corr})} = \frac{\mathbf{r}_t}{1 - (1 - \rho)^t} \quad (3)$$

$$\mathbf{V}_t = \sqrt{\text{diag}(\mathbf{r}_t^{(\text{corr})})} + \epsilon \mathbf{I}$$

により更新する。ここで、 ρ は EMA の減衰率であり、初期値 $\mathbf{r}_0$ は零ベクトルに設定される。行列に対する $\text{diag}(\cdot)$ は対角成分を抽出し、ベクトルに対する $\text{diag}(\cdot)$ はその要素を対角成分とする対角行列を構成する。また、式 (3) は EMA の更新初期に生じるバイアスを補正する役割を持つ。Adam では、各次元の勾配の大きさに応じて学習率を調整することで、パラメータ更新の振動や停滞が抑制される。

*Investigation of adaptive learning-rate algorithms for adaptive filtering. By Ren UCHIDA, Kohei YATABE (Tokyo University of Agriculture and Technology), Yuto IZUMI, Yu TAKAHASHI and Kazunobu KONDO (Yamaha Corporation).

4 周波数領域 Adam の提案

式 (2) から, Adam では t 反復目における劣勾配の自己相関行列の EMA による推定値

$$\mathbf{R}_t = (1 - \rho)\mathbf{R}_{t-1} + \rho \mathbf{g}_t \mathbf{g}_t^\top$$

を対角行列と仮定している. よって, Adam の $\mathbf{V}_t$ では $\mathbf{R}_t$ の非対角成分に現れる次元間の相関を反映した $\mathbf{Q}_t$ を構成できない.

そこで, 時間的に隣接する次元間の相関を考慮可能にした Adam の派生手法を提案し, これを SpecAdam と呼ぶ. SpecAdam では t 反復目の前処理行列を

$$\begin{aligned} \hat{\mathbf{r}}_t &= (1 - \rho)\hat{\mathbf{r}}_{t-1} + \rho |\mathbf{F} \mathbf{g}_t|^2 \\ \hat{\mathbf{r}}_t^{(\text{corr})} &= \frac{\hat{\mathbf{r}}_t}{1 - (1 - \rho)^t} \\ \mathbf{V}_t &= \mathbf{F}^{-1} \sqrt{\text{diag}(\hat{\mathbf{r}}_t^{(\text{corr})})} \mathbf{F} + \epsilon \mathbf{I} \end{aligned}$$

と定義する. ここで, $|\cdot|$ は要素ごとの絶対値, $(\cdot)^H$ は複素共役転置, $\mathbf{F}$ は離散フーリエ変換行列であり, $\mathbf{F}^{-1} = \mathbf{F}^H$ が成り立つ. また, 式 (1) に含まれる $\mathbf{V}_t^{-1} \mathbf{g}_t$ は $\mathbf{F} \mathbf{g}_t$ を $\sqrt{\hat{\mathbf{r}}_t^{(\text{corr})}} + \epsilon$ で要素ごとに除算し, 逆フーリエ変換することで効率的に計算できる.

SpecAdam の導出は, 次元間の相関の考慮と計算効率を両立させるための近似に基づく. 一般に, 次元間の相関を考慮するには $\mathbf{R}_t$ を密行列で扱う必要があるが, これは逆行列演算に伴う計算量が膨大となり現実的でない. そこで, $\mathbf{R}_t$ を巡回行列と仮定することで, 次元間の相関を十分保持しつつ, フーリエ変換による対角化によって効率的な計算を可能にする. この仮定により, $\mathbf{R}_t$ と $\hat{\mathbf{R}}_t = \text{diag}(\hat{\mathbf{r}}_t)$ には

$$\begin{aligned} \mathbf{F} \mathbf{R}_t \mathbf{F}^{-1} &= \mathbf{F} ((1 - \rho)\mathbf{R}_{t-1} + \rho \mathbf{g}_t \mathbf{g}_t^\top) \mathbf{F}^H \\ &= (1 - \rho)\mathbf{F} \mathbf{R}_{t-1} \mathbf{F}^H + \rho \mathbf{F} \mathbf{g}_t (\mathbf{F} \mathbf{g}_t)^\top \\ &\approx (1 - \rho)\hat{\mathbf{R}}_{t-1} + \rho \text{diag}(|\mathbf{F} \mathbf{g}_t|^2) = \hat{\mathbf{R}}_t \end{aligned}$$

という関係が成り立つ. ただし, $\mathbf{F} \mathbf{g}_t (\mathbf{F} \mathbf{g}_t)^\top$ は対角行列ではないため, その非対角成分を棄却する近似を行う. この関係から, $\hat{\mathbf{R}}_t$ は $\mathbf{R}_t$ を周波数領域で近似していると解釈できる. また, $\hat{\mathbf{r}}_t$ を Adam と同様にバイアス補正と平方根の計算を適用し, 時間領域に戻すことで SpecAdam が導出できる.

5 V-APSM への適応学習率の導入

適応学習率アルゴリズムの $\mathbf{V}_t$ は, パラメータ空間を幾何学的に変化させるという点で, V-APSM の $\mathbf{Q}_t$ と同様の役割を持つ. そこで, Adam と SpecAdam の $\mathbf{V}_t$ を $\mathbf{Q}_t$ として用いた V-APSM を提案し, これらをそれぞれ Adam-APSM, SpecAdam-APSM と呼ぶ.

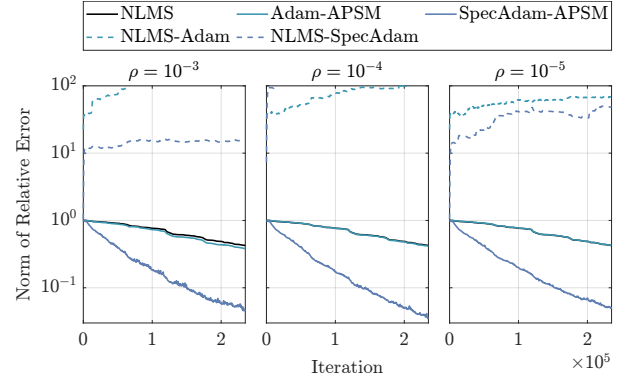


図-1 反復回数に対するインパルス応答の相対誤差の推移.

Adam-APSM と SpecAdam-APSM によるインパルス応答推定性能を従来手法である NLMS と比較した. また, V-APSM と適応学習率アルゴリズムの別の組み合わせ方として, NLMS の更新を式 (1) の $\mathbf{g}_t$ として導入した Adam (NLMS-Adam) と SpecAdam (NLMS-SpecAdam) も比較した. 入力音声およびインパルス応答は, 文献 [2] と同一の条件とした. パラメータは $\lambda = 1$, $\epsilon = 10^{-8}$, $\rho \in \{10^{-3}, 10^{-4}, 10^{-5}\}$ とした. また, $f_t = |d_t - \mathbf{w}_t^\top \mathbf{x}_t|$, $T_t = \text{Id}$ とした.

図-1 にインパルス応答の相対誤差の推移を示す. Adam-APSM は適切な ρ の調整によって NLMS をわずかに上回る性能を発揮した. SpecAdam-APSM はすべての ρ で Adam-APSM よりも優れた収束性能を示すことが確認された. この結果から, $\mathbf{Q}_t$ の設計では劣勾配の次元間の相関を考慮することが重要であると示唆された. また, NLMS-Adam と NLMS-SpecAdam はすべての ρ で誤差が増加していた. この結果から, V-APSM と適応学習率アルゴリズムを効果的に統合するためには, V-APSM に $\mathbf{Q}_t$ として $\mathbf{V}_t$ を導入する必要がある.

6 むすび

本稿では, 次元間の相関を考慮可能な Adam の派生手法を導出し, 適応学習率アルゴリズムの前処理行列を計量行列に用いた V-APSM を提案した. 実験結果から, 本手法の有効性と SpecAdam の優れた収束性能を示した. 今後は, SpecAdam のフーリエ変換を一般の直交行列に拡張し, 計算効率の向上を図る.

参考文献

- [1] M. Yukawa and I. Yamada, "A unified view of adaptive variable-metric projection algorithms," *EURASIP J. Adv. Signal Process.*, **2009**, (2009).
- [2] 村田佳斗, 松本和樹, 矢田部浩平, 泉悠斗, 高橋祐, 近藤多伸, "観測信号の周波数特性に基づく時変計量を用いた適応フィルタの実験的検討," 日本音響学会講演論文集, pp. 267–268 (2025.9).
- [3] J. Duchi, E. Hazan and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. mach. learn. res.*, **12**, 2121–2159 (2011).
- [4] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *ICLR* (2015).