F-024

単調方程式のニュートン型解法に基づく新たな非負値行列因子分解アルゴリズム
A Novel Nonnegative Matrix Factorization Algorithm based on Newton-Type Method for Monotone Equations

佐野 雄大† 右田 剛史† 高橋 規一†
Takehiro Sano Tsuyoshi Migita Norikazu Takahashi

1 はじめに

非負値行列因子分解 (NMF: Nonnegative Matrix Factorization) [1, 2] とは，与えられた非負値行列を二つの非負値因子行列の積で近似する手法である．NMF によって非負値の基本構成要素の抽出やデータの次元削減が行われるため，顔画像処理，生体信号処理，ネットワーク分析などの様々な分野に応用されている．

NMF は，二つの非負値因子行列のすべての要素が非負であるという制約の下で，与えられた非負値行列と因子行列の積との間の誤差を最小化する制約付き最適化問題として定式化される．この問題の解法として，乗法型更新 [3, 4] とよばれる反復解法が広く知られている．乗法型更新には，制約条件が自動的に満たされる，実装が容易である，多様な誤差関数に適用可能である，といった利点があるが，その一方で収束が遅いという問題がある．そのため NMF の高速な計算法が求められている．

中津と高橋は，α ダイバージェンス [5] で誤差を評価する場合の NMF の計算法として，ある種のニュートン型アルゴリズムを提案し，その大域収束性を証明した [6]．大域収束性とは，任意の初期値からアルゴリズムによって生成される変数の値の列が少なくとも一つ収束部分列をもち，かつ任意の収束部分列の極限が対応する最適化問題の停留点であることを意味する [7]．彼らの解法は多くの場合に乗法型更新よりも高速であるが，α の値によっては乗法型更新よりも大幅に遅くなる．そこで本研究では，大域収束性を有し，かつ α の値によらず高速な解法として，単調方程式のニュートン型解法に基づく新たなアルゴリズムを提案する．

2 最適化問題と既存解法

α ダイバージェンスで誤差を評価する NMF は次の制約付き最適化問題として定式化される．

$$\begin{array}{ll}\text{minimize} & D_\alpha(\boldsymbol{X}\|\boldsymbol{W}\boldsymbol{H}^T)\\ \text{subject to} & \boldsymbol{W} \geq \epsilon\boldsymbol{1}_{M\times K},\ \boldsymbol{H} \geq \epsilon\boldsymbol{1}_{N\times K}\end{array} \tag{1}$$

ここで $\boldsymbol{X} \in \mathbb{R}_+^{M\times N}$ は与えられた非負値行列であり，$\boldsymbol{W} \in \mathbb{R}_+^{M\times K}$ と $\boldsymbol{H} \in \mathbb{R}_+^{N\times K}$ は求める非負値因子行列である．また，ϵ は十分に小さい正定数である．目的関数は

$$D_\alpha(\boldsymbol{X}\|\boldsymbol{W}\boldsymbol{H}^T) = \frac{1}{\alpha(1-\alpha)}\sum_{i=1}^{M}\sum_{j=1}^{N}\Big[\alpha X_{ij} + (1-\alpha)(\boldsymbol{W}\boldsymbol{H}^T)_{ij} - X_{ij}^{\alpha}(\boldsymbol{W}\boldsymbol{H}^T)_{ij}^{1-\alpha}\Big] \quad (\alpha \neq 0, 1) \tag{2}$$

で定義される．尚，$\alpha < 0$ かつ $X_{ij} = 0$ のとき X_{ij}^{α} は定義されないため，この場合には X_{ij} の値を十分に小さい正定数（例えば ϵ）で置き換える．

最適化問題 (1) を解いて得られる行列 $\boldsymbol{W},\boldsymbol{H}$ は正値行列であるため密行列である．しかし，一般には最適解または局所最適解では多くの要素が下限値をとる [1] ので，それらを 0 で置き換えると疎行列になると予想される．

中津と高橋は，問題 (1) の解法として，1 変数ずつ値を更新するニュートン型アルゴリズムを開発した [6]．他のすべての変数の値を固定して W_{ik} の値のみを最適化する問題は

$$\begin{array}{ll}\text{minimize} & g_{ik}(W_{ik})\\ \text{subject to} & W_{ik} \geq \epsilon\end{array} \tag{3}$$

と表される．ここで $g_{ik}(W_{ik})$ は，(2) において W_{ik} 以外のすべての変数を固定して得られる 1 変数関数である．同様に，他のすべての変数の値を固定して H_{jk} の値のみを最適化する問題は

$$\begin{array}{ll}\text{minimize} & h_{jk}(H_{jk})\\ \text{subject to} & H_{jk} \geq \epsilon\end{array} \tag{4}$$

と表される．ここで $h_{jk}(H_{jk})$ は，(2) において H_{jk} 以外のすべての変数を固定して得られる 1 変数関数である．最適化問題 (3), (4) の目的関数 $g_{ik}(W_{ik}), h_{jk}(H_{jk})$ の導関数は狭義単調増加関数であり，$\alpha \leq -1$ のとき凸関数，$\alpha > -1$ のとき凹関数となる．これらの性質に着目し，中津と高橋はニュートン法と線形補間を用いて W_{ik} と H_{jk} の値を更新する方法を提案した．変数 W_{ik} の更新では

$$d \leftarrow -g'_{ik}(W_{ik})/g''_{ik}(W_{ik}) \tag{5}$$

$$W_{ik}^{\mathrm{n}} \leftarrow W_{ik} + d \tag{6}$$

$$W_{ik}^{\mathrm{i}} \leftarrow W_{ik} - \frac{W_{ik}^{\mathrm{n}} - W_{ik}}{g'_{ik}(W_{ik}^{\mathrm{n}}) - g'_{ik}(W_{ik})}g'_{ik}(W_{ik}) \tag{7}$$

によって W_{ik}^{i} を求め W_{ik} の値を W_{ik}^{i} に更新する．変数 H_{jk} の更新規則も同様である．

中津と高橋は，彼らのアルゴリズムが大域収束性を有することや，問題 (1) の最適性条件を適切に緩めたものを終了条件に用いれば有限回の反復で終了することを証明した [6]．また，多くの場合に計算時間が乗法型更新よりも大幅に短縮されることや，α の値によっては計算時間が乗法型更新よりも長くなることを実験的に示した．

3 提案法

本節では，Solodov と Svaiter によって提案された単調方程式のニュートン型解法 [8] に基づく新たなアルゴリズムを提案する．Solodov と Svaiter の解法は，図 1 に示されるように，単調方程式 $g'_{ik}(W_{ik}) = 0$ の解と現在の点 W_{ik} の間の点 W_{ik}^{s} を

$$W_{ik}^{\mathrm{s}} \leftarrow W_{ik} + \beta^{p^*} d \tag{8}$$

によって求め，W_{ik} の値を W_{ik}^{s} に更新するものである．ただし，β は $0 < \beta < 1$ を満たす定数であり，d は式 (5)

† 岡山大学大学院自然科学研究科 Graduate School of Natural Science and Technology, Okayama University

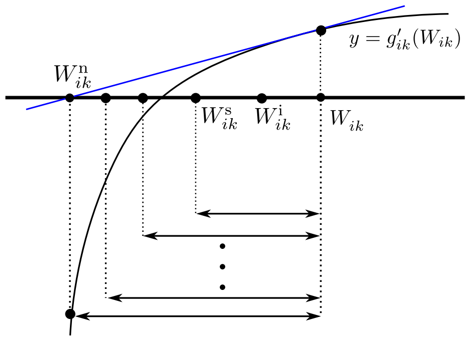

図 1　提案法における W_{ik} の更新

で与えられ，p^* は

$$p^* = \min\{p \in \mathbb{Z}_+ \mid g'_{ik}(W_{ik} + \beta^p d)d \leq 0\}$$

で与えられる．このようにして求めた W_{ik}^{s} と (7) によって得られる W_{ik}^{i} を比較し，方程式 $g'_{ik}(W_{ik}) = 0$ の解に近い方を更新後の値として採用することで中津と高橋の解法の計算時間を短縮できると期待される．提案法における変数 W_{ik} の更新手順をアルゴリズム 1 に示す（変数 H_{jk} の更新手順も同様である）．

アルゴリズム 1 W_{ik} の更新

入力: $\boldsymbol{X} \in \mathbb{R}_+^{M\times N}$, $\boldsymbol{W} \in [\epsilon,\infty)^{M\times K}$, $\boldsymbol{H} \in [\epsilon,\infty)^{N\times K}$, $\alpha \in \mathbb{R}\setminus\{0,1\}$, $\beta \in (0,1)$, $\epsilon \in \mathbb{R}_{++}$

出力: $W_{ik}^{\mathrm{new}} \in [\epsilon,\infty)$

1: W_{ik} が問題 (3) の緩和最適性条件 [6] を満たすならば $W_{ik}^{\mathrm{new}} \leftarrow W_{ik}$ としてステップ 5 へ進む．
2: 式 (5), (8) を用いて d と W_{ik}^{s} を求める．
3: 式 (7) を用いて W_{ik}^{i} を求める．$d > 0$ ならば $y \leftarrow \max\{W_{ik}^{\mathrm{s}}, W_{ik}^{\mathrm{i}}\}$ とし，そうでなければ $y \leftarrow \min\{W_{ik}^{\mathrm{s}}, W_{ik}^{\mathrm{i}}\}$ とする．
4: $W_{ik}^{\mathrm{new}} \leftarrow \max\{y,\epsilon\}$ とする．
5: W_{ik}^{new} を出力して終了する．

提案法は中津と高橋の方法と同じく大域収束性を有する（証明は紙面の都合上省略する）．したがって，(1) に対する最適性条件を適切に緩めたものを終了条件に用いることにより，有限回の反復で必ず終了する．

4　評価実験

提案法の有効性を検証するために，文書データ (CLUTO tr23[1]) から得られる 5832×204 の非負値行列 $\boldsymbol{X}$ に乗法型更新，中津と高橋の方法，提案法を適用する実験を行った．パラメータの値を $K = 6$, $\epsilon = 0.1$, $\beta = 0.95$ に固定し，α の値を -2.5 から 2.5 まで 0.2 刻みで変化させ，共通の終了条件を満たすまでに要する計算時間を測定した．尚，行列 $\boldsymbol{W}, \boldsymbol{H}$ の各要素の初期値を区間 $[\epsilon, 10]$ の一様乱数で与え，終了条件は文献 [9] と同じに設定した．数値実験に用いた計算機環境を表 1 に示す．

実験の結果を図 2 に示す．横軸は α の値を，縦軸は計算時間を表す．グラフは 5 種類の異なる初期値に対する各アルゴリズムの計算時間の平均値をプロットしたものである．MUR, NT, Proposed はそれぞれ乗法型更新，中津と高橋の方法，提案法を表す．図 2 より，提案法の計算時間は α の値によって大きく変化しないことがわかる．さらに，多くの場合で乗法型更新より短く，中津と高橋の方法と同程度であることもわかる．

1)　http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview/

表 1　計算機環境

CPU	Intel(R) Pentium(R) CPU G3258
Memory	2GB
OS	Ubuntu 18.04 LTS 64bit
Language	C
Library	reference BLAS

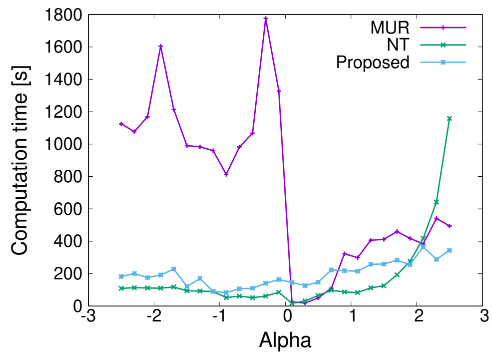

図 2　CLUTO データセットに対する計算時間

5　おわりに

α ダイバージェンスで誤差を評価する NMF のための新たなニュートン型アルゴリズムを提案し，それが大域収束性を有することや既存手法よりも高速に動作することを示した．今後の課題は，スパースさや滑らかさを制御するための正則化項が目的関数に含まれる場合のアルゴリズムの開発である．

参考文献

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, p. 788 (1999).
[2] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126 (1994).
[3] A. Cichocki, H. Lee, Y. D. Kim, and S. Choi, "Non-negative matrix factorization with α-divergence," *Pattern Recognition Letters*, vol. 29, no. 9, pp. 1433–1440 (2008).
[4] N. Takahashi, J. Katayama, M. Seki, and J. Takeuchi, "A unified global convergence analysis of multiplicative update rules for non-negative matrix factorization," *Computational Optimization and Applications*, vol. 71, no. 1, pp. 221–250 (2018).
[5] S. I. Amari, *Differential-Geometrical Methods in Statistics*, Springer, New York (1985).
[6] S. Nakatsu and N. Takahashi, "A novel Newton-type algorithm for nonnegative matrix factorization with alpha-divergence," *Proceedings of 2017 International Conference on Neural Information Processing*, pp. 335–344 (2017).
[7] W. I. Zangwill, *Nonlinear Programming: A Unified Approach*, Englewood Cliffs, NJ, Prentice-Hall (1969).
[8] M. V. Solodov and B. F. Svaiter, "A globally convergent inexact Newton method for systems of monotone equations," *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pp. 355–369 (1998).
[9] T. Kimura and N. Takahashi, "Gauss-Seidel HALS algorithm for nonnegative matrix factorization with sparseness and smoothness constraints," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E100-A, no. 12, pp. 2925–2935 (2017).