

ALS と ICA のスペクトル分離法への応用 *Haematococcus pluvialis* 代謝物質の HPLC-DAD データの解析

○山本博之 大野弘 福田秀樹(神戸大学)

Application of alternating least squares and independent component analysis to curve resolution problem

Analysis of HPLC-DAD data of *Haematococcus pluvialis* metabolites

*H.Yamamoto, H.Ohno, H.Fukuda (Kobe University)

Abstract— We applied Alternating least squares (ALS) and Independent component analysis (ICA) to the HPLC-DAD data of *Haematococcus pluvialis* metabolites. In ALS, different solutions are produced depending on randomly chosen initial values, so we used the results of ICA as initial values of ALS. We obtained a high accuracy of peak detection, suggesting that these curve resolution methods are useful for identification of metabolites in metabolomics.

Keyword: Metabolomics, Chemometrics, HPLC-DAD, Alternating least squares, Independent component analysis

1. 緒言

ケモメトリックス分野においてスペクトル分離法として最もよく用いられる手法の一つである Alternating Least Squares (ALS)¹⁾は、近年様々なアルゴリズムが提案されている Non-negative matrix factorization (NMF)^{2,3)}の各手法に比べて計算時間が早い⁴⁾等の利点がある。

NMFの各アルゴリズムと同様に、ALSも初期値によって異なる結果が得られる。2 成分系のような単純な場合では、初期値に様々な値を用いた場合でも、規格化制約条件により比較的限定された解が得られる⁵⁾が、現実のデータへのALSの適用では、それだけではこの問題の解決としては必ずしも十分ではない。

本研究では、上記の初期値の問題を回避し、また化学的に意味のある結果が得られることを期待して、独立成分分析(ICA)あるいは非負制約条件下での独立成分分析の結果を ALS の初期値として用いた。

2. 理論

2.1 Non-negative matrix factorization

Non-negative matrix factorization(NMF)は、データ行列 X を非負の条件下で C と S に分解する手法である。

$$\min \|X - CS\|_F^2 \\ C_{ij} \geq 0, S_{ij} \geq 0 \quad \text{for all } i, j$$

Lee, Seungのmultiplicative update ruleによるアルゴリズム²⁾、制約条件付きのNMFアルゴリズム³⁾等、最近様々な研究報告がある。またこれらの手法は少なくともローカルミニマムに収束することが知られている²⁾。

2.1.1 Alternating least Squares

Alternating least squares(ALS)¹⁾は、適当な初期値 C を決め、以下の計算を繰り返し行うことにより、NMFと同様に非負の条件下で行列の分解を行う手法である。

$$S = (C^T C)^{-1} C^T X$$

$$C = X S^T (S S^T)^{-1}$$

ALS の繰り返し計算において、non-negativity の条件を満たす為に C, S それぞれの負の要素を強制的に 0 とする。

2.1.2 ACLS

ACLS^{4,5)}はALSにパラメータのノルムに関する制約を加えたアルゴリズムである。

$$S = (C^T C + \lambda I)^{-1} C^T X$$

$$C = X S^T (S S^T + \lambda I)^{-1}$$

I は単位行列である。最適なパラメータ λ は $\|X - CS\|_F$ が最小となる時とする⁵⁾。

2.2 独立成分分析

独立成分分析 (Independent component analysis, ICA) は、観測データを統計的に独立な成分の線形和として表現することが出来る。

本研究では独立性の基準として 4 次のキュムラントの一般形であるネグントロピーを非線形関数で近似し、その最大化を行うFastICAアルゴリズム⁶⁾、非負制約条件下で独立成分分析を行い、非負の独立成分を得ることの出来るNon-negative ICA⁷⁾を用いて計算を行った。

2.3 Unimodality 制約条件

各物質に対応するクロマトグラムが単峰であることから、ALSの繰り返し計算中でunimodality制約を科すことが多い。例として $[0.0 \ 1.0 \ 5.1 \ 2.1 \ 0]$ から $[0.0 \ 0.5 \ 0.5 \ 1.2 \ 1.0]$ へと修正される様子をFig.1 に示す。

本研究では C に関して unimodality 制約条件、 A に関して規格化制約条件を加えて ALS を計算した。

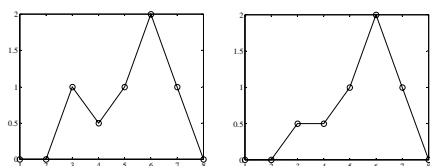


Fig.1 Example of unimodality constraint

3. データ

本研究で用いたデータは、淡水性単細胞緑藻である *Haematococcus pluvialis* の代謝物質の分析データである。*H. pluvialis* を 217 時間培養し、培養液から菌体を回収、メタノール溶媒中で破碎および抽出を行い、高速クロマトグラフィー(HPLC)とフォトダイオードアレイ検出器(DAD)を組み合わせた複合分析機器により分析されたデータ⁵⁾を用いた。

4. 結果と考察

測定データを前処理後、Non-negative ICA または FastICAによって得られた結果を初期値として用い、ALSを計算した結果をFigs2-4 に示す。クロマトグラムの保持時間とスペクトルのピーク位置、文献値⁸⁾と比較することにより各データに含まれる物質を同定することができる。結果をTable1 にまとめた。

また ACLS は ALS とほぼ同様の結果が得られた。

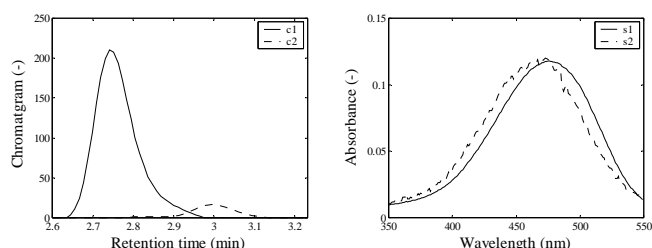


Fig.2 Results of ALS 2.6~3.2min

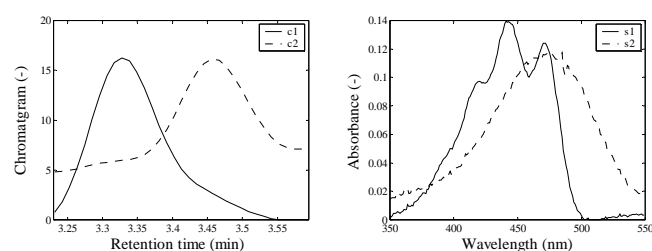


Fig.3 Results of ALS 3.2~3.6min

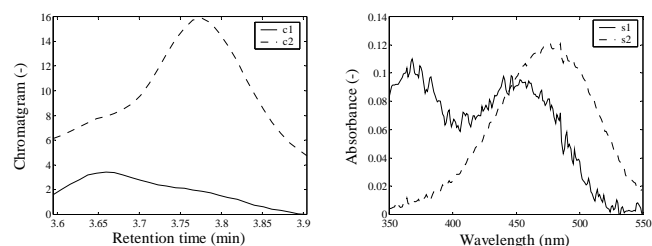


Fig.4 Results of ALS 3.6~3.9min

Table1 Identification of metabolites

Metabolite	Peak Wavelength (nm)	
	Experimental	ALS
(3S,3'S)- <i>trans</i> -astaxanthin	480	476
Adonirubin	472.8	471
Lutein	443.9, 472.8	441, 470
(3S,3'S)-9- <i>cis</i> -astaxanthin	472.8	475
(3S,3'S)-13- <i>cis</i> -astaxanthin	371.8, 472.8	368, 454
(3R,3'R)- <i>trans</i> -astaxanthin	480	478

5. 今後の展望

微生物細胞内の代謝物質の同定と定量を目的とするメタボロミクス研究にとって、ケモメトリックス分野で研究されてきたスペクトル分離法は有効な道具となりうる。本研究で用いた ALS は初期値を変えることで様々な結果が得られることから、結果の解釈が恣意的で実験研究者にとって必ずしも使いやすいとは言えない。今後は ALS の初期値として他の手法を用いた結果との比較などを行っていく。

参考文献

- 1) E. J. Karjalainen, "The spectrum reconstruction problem, use of alternating regression for unexpected spectral components in two-dimensional spectroscopies", Chemom. Intell. Lab. Syst., 7, pp. 31/38 (1989).
- 2) D. D. Lee and H. S. Seung, "Algorithms for Non-negative Matrix Factorization", Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference, pp. 556/562, MIT Press (2001).
- 3) A. Cichocki, R. Zdunek, and S. Amari, "Csiszar's Divergences for Non-Negative Matrix Factorization: Family of New Algorithms", 6th International Conference on Independent Component Analysis and Blind Signal Separation, Charleston SC, USA, March 5-8, Springer LNCS 3889, pp. 32/39 (2006).
- 4) R. Albright, J. Cox, D. Duling, A. Langville and C. D. Meyer, "Algorithms, Initializations, and Convergence for the Nonnegative Matrix Factorization", NCSU Technical Report Math 81706, to be submitted.
- 5) H. Yamamoto, K. Hada, H. Yamaji, T. Katsuda, H. Ohno and H. Fukuda, "Application of regularized alternating least squares and independent component analysis to HPLC-DAD data of *Haematococcus pluvialis* metabolites", Biochem. Eng. Journal, in press.
- 6) A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis.", IEEE Trans. Neural Netw., 10(3), pp. 626/634 (1999).
- 7) M. D. Plumbley, "Algorithms for non-negative independent component analysis", IEEE Trans. Neural Netw., 14, pp. 534/543 (2003).
- 8) J. P. Yuan and F. Chen, "Chromatographic separation and purification of trans-astaxanthin from the extracts of *Haematococcus pluvialis*", J. Agric. Food Chem., 46, pp. 3371/3375 (1998).