交互最小二乗法の加速化について[1]

データ科学研究系（計算機統計グループ）
客員教授　　森　裕　一　(岡山理科大学)

1　はじめに

非計量データが含まれるデータに対して主成分分析を適用する方法として，交互最小二乗法 (ALS: alternaing least squares algorithm) を利用した PRINCIPALS (principal components analysis by alternating least squares, Young *et al*, 1978) がある．これは，主成分得点と固有ベクトルの行列の推定 (モデルパラメータ推定) と，データの最適変換 (データパラメータ推定) の 2 つの推定に対して最小二乗法を交互に行うもので，データの構造によっては，その収束に多くの時間を要する．

そこで，PRINCIPALS の収束を加速するために，vector ε 法 (vε 法，Wynn, 1962) を組み込んだ加速化法を提案する．実際には，PRINCIPALS の中で行われている ALS の加速化を行うものであり，提案の手法は，他に ALS を利用する手法（PRINCALS など）にも適用できる．

2　PRINCIPALS によるパラメータ推定

n 個体，p 変数の測定値を標準化した $n \times p$ のデータ行列 $\mathbf{X}$ は，$r(\leq p)$ 個の主成分の得点行列 $\mathbf{Z}$ $(n \times r)$ と $\mathbf{X}^\top\mathbf{X}$ の固有ベクトルの行列 $\mathbf{A}$ $(p \times r)$ を用いて，$\hat{\mathbf{X}} = \mathbf{Z}\mathbf{A}^\top$ と書ける．$\mathbf{X}$ が最適変換された行列 $\mathbf{X}^*$ を r 個の主成分によって最もよく表現することは，$\theta = \mathrm{tr}(\mathbf{X}^* - \hat{\mathbf{X}})^\top(\mathbf{X}^* - \hat{\mathbf{X}})$ を最小化するデータパラメータ $\mathbf{X}^*$ とモデルパラメータ $\mathbf{Z}$，$\mathbf{A}$ の最小二乗推定問題に帰着する．

初期値 $\mathbf{X}^{*(0)}$ が与えられたもとで，PRINCIPALS は，収束するまで，次のステップを繰り返す．

- <u>モデルパラメータ推定ステップ:</u>　固有値問題 $\left[\dfrac{\mathbf{X}^{*(t)\top}\mathbf{X}^{*(t)}}{n}\right]\mathbf{A} = \mathbf{A}\mathbf{D}_r$ を解いて，$\mathbf{A}^{(t)}$ を求める．ここで，$\mathbf{D}_r$ は対角要素が固有値である $r \times r$ の対角行列，$\mathbf{A}^\top\mathbf{A} = \mathbf{I}_r$ とする．$\mathbf{Z}^{(t)} = \mathbf{X}^{*(t)}\mathbf{A}^{(t)}$ を計算する．
- <u>データパラメータ推定ステップ:</u>　$\hat{\mathbf{X}}^{(t)} = \mathbf{Z}^{(t)}\mathbf{A}^{(t)\top}$ を計算する．$\hat{\mathbf{X}}$ を固定し，最小二乗基準のもとで，$\mathbf{X}^{*(t+1)} = \arg\min_{\mathbf{X}^*} \mathrm{tr}(\mathbf{X}^* - \hat{\mathbf{X}}^{(t)})^\top(\mathbf{X}^* - \hat{\mathbf{X}}^{(t)})$ となる $\mathbf{X}^{*(t+1)}$ を求め，平均 0，分散 1 となるように列ごとに基準化する．

収束判定は，$|\theta^{(t+1)} - \theta^{(t)}| < \delta$ で行う．$\theta^{(t+1)} = \mathrm{tr}(\mathbf{X}^{*(t+1)} - \hat{\mathbf{X}}^{(t)})^\top(\mathbf{X}^{*(t+1)} - \hat{\mathbf{X}}^{(t)})$ である．

3　vε 法による PRINCIPALS の加速化法：vε.PRINCIPALS

PRINCIPALS の収束スピードを加速化するために vε 法を組み込むことを考える．

パラメータベクトルの推定列を $\{\mathbf{Y}^{(t)}\}_{t\geq 0}$ で表すとき，vε 法はその加速化列 $\{\dot{\mathbf{Y}}^{(t)}\}_{t\geq 0}$ を

$$\dot{\mathbf{Y}}^{(t-1)} = \mathbf{Y}^{(t)} + \left[\left[\mathbf{Y}^{(t-1)} - \mathbf{Y}^{(t)}\right]^{-1} + \left[\mathbf{Y}^{(t+1)} - \mathbf{Y}^{*(t)}\right]^{-1}\right]^{-1}$$

[1] 本研究は，黒田正博・榊原道夫（岡山理科大学）・飯塚誠也（岡山大学）との共同研究で，本報告は，2009 年度科学研究費シンポジウム「多変量データ解析法への計算機統計学・行動計量学的アプローチの新展開」(2009.12.4，大阪大学）での発表を再構成したものである．

で生成する．ただし，$\mathbf{Y}^{-1}=\mathbf{Y}/||\mathbf{Y}||^2$ である．このとき，もとの列 $\{\mathbf{Y}^{(t)}\}_{t\geq 0}$ に停留点が存在するならば，加速列 $\{\dot{\mathbf{Y}}^{(t)}\}_{t\geq 0}$ は $\{\mathbf{Y}\}_{t\geq 0}$ より速くこの停留点に収束する (Wynn, 1962)．また，上記の vε 法の収束率は，最良で準ニュートン法と同じく超 1 次収束になる．さらに，パラメータ次元数を d としたとき，1 回の反復で必要とされる計算量は $O(d)$ であり，PRINCIPALS で必要な固有値問題を解くための数値解法や逆行列の計算法と比べて計算量は少ない．

vε.PRINCIPALS は，次のステップを繰り返し，$\{\mathbf{X}^{*(t)}\}_{t\geq 0}$ の収束を加速する．

- **PRINCIPALS ステップ**: $\mathbf{X}^{*(t)}$ から $\mathbf{Z}^{(t)}$ と $\mathbf{A}^{(t)}$ を計算し，$\mathbf{X}^{*(t+1)}=\mathbf{Z}^{(\mathbf{t})}\mathbf{A}^{(\mathbf{t})^\top}$ より $\mathbf{X}^{*(t+1)}$ を推定する．
- **vε 加速ステップ**: 推定列 $\{\mathbf{X}^{*(t-1)},\ \mathbf{X}^{*(t)},\ \mathbf{X}^{*(t+1)}\}$ から

$$\mathrm{vec}\dot{\mathbf{X}}^{*(t-1)}=\mathrm{vec}\mathbf{X}^{*(t)}+\left[\left[\mathrm{vec}(\mathbf{X}^{*(t-1)}-\mathbf{X}^{*(t)})\right]^{-1}+\left[\mathrm{vec}(\mathbf{X}^{*(t+1)}-\mathbf{X}^{*(t)})\right]^{-1}\right]^{-1}$$

を用いて $\dot{\mathbf{X}}^{*(t-1)}$ を生成する．ただし，$\mathrm{vec}\mathbf{X}^*$ は $\mathbf{X}^*$ をベクトル化したものである．

収束判定は，$\left\|\mathrm{vec}(\dot{\mathbf{X}}^{*(t-1)}-\dot{\mathbf{X}}^{*(t-2)})\right\|^2<\delta$ により行う．

vε.PRINCIPALS は，**PRINCIPALS ステップ**から生成される $\{\mathbf{X}^{*(t)}\}_{t\geq 0}$ と独立に $\{\dot{\mathbf{X}}^{*(t)}\}_{t\geq 0}$ を **vε 加速ステップ**により生成する．このとき，この加速列の生成に用いるのは $\{\mathbf{X}^{*(t)}\}_{t\geq 0}$ であり，PRINCIPALS の推定方式は触っていないので，vε.PRINCIPALS は PRINCIPALS の安定した収束性を保持できている．したがって，$\{\mathbf{X}^{*(t)}\}_{t\geq 0}$ の停留点が存在するならば，$\{\mathbf{X}^{*(t)}\}_{t\geq 0}$ より速くその停留点に収束する $\{\dot{\mathbf{X}}^{*(t)}\}_{t\geq 0}$ の生成ができたことになる（Wang *et al*, 2008）．

4　数値実験

授業アンケートの回答結果 (51 名，5 段階評価の 13 項目）に両手法を適用したときの収束までの反復回数と CPU 時間を表 4.1 に示す（$\delta=10^{-8}$）．反復回数については ε.PRINCIPALS は，PRINCIPALS より 3 から 4 倍速く収束していることがわかる．特に，PRINCIPALS の反復回数が多い場合に，よく加速されていることがわかる．また，CPU 時間では，vε.PRINCIPALS は，PRINCIPALS より 2.5 から 3.5 倍速く収束している．したがって，CPU 時間から見ても，vε.PRINCIPALS は，十分に PRINCIPALS の収束を加速しているといえる．

表 4.1　反復回数と CPU 時間の比較

r	反復回数			CPU 時間		
	P	V	P/V	P	V	P/V
1	9	4	2.25	0.35	0.233	1.50
2	92	23	4.00	3.87	1.102	3.51
3	28	9	3.11	1.14	0.458	2.49
4	25	8	3.57	1.00	0.360	2.78
5	28	10	2.80	1.13	0.494	2.29
6	29	9	3.22	1.16	0.460	2.52
7	28	9	3.11	1.11	0.436	2.55
8	47	14	3.36	1.97	0.681	2.89
9	45	13	3.46	1.82	0.627	2.90
10	45	14	3.21	1.84	0.674	2.73
11	33	10	3.30	1.35	0.491	2.75
12	40	10	4.00	1.61	0.483	3.33

P: PRINCIPALS, V: vε.PRINCIPALS

参考文献

Young, F.W., Takane, Y., and de Leeuw, J. (1978). Principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika*, **43**, 279-281.

Wang, M., Kuroda, M., Sakakihara, M. and Geng, Z. (2008). Acceleration of the EM algorithm using the vector epsilon algorithm. *Comput. Stat.*, **23**, 469-486.

Wynn, P. (1962). Acceleration techniques for iterated vector and matrix problems. *Math. Comp.*, **16**, 301-322.