Convergence Analysis of ARD Algorithm

CHENG Guang-Hui[1]

Abstract A convergence analysis of the altering row diagonalization (ARD) algorithm is made in this paper. Based on the convergence analysis, we present some advice on how to choose the initial matrix, and give a new terminal condition of the algorithm. For cross validating our analysis, three examples are also given.

Key words Blind source separation (BSS), ARD algorithm, eigenvalue and eigenvector, joint diagonalization

Citation Cheng Guang-Hui. Convergence analysis of ARD algorithm. *Acta Automatica Sinica*, 2014, **40**(5): 980−982

DOI 10.3724/SP.J.1004.2014.00980

Joint diagonalization algorithms of a set of matrices have attracted much attention for the last twenty years within the fields of principle components estimation[1], blind beamforming[2, 3], blind source separation (BSS)[4, 5], frequency estimate[6], independent component analysis (ICA)[7], and so on. Many algorithms[8−15] were proposed for the so-called joint diagonalization problem. This problem can be described as follows.

Given a set of matrices $\mathcal{A} = \{A_1, A_2, \cdots, A_K\}$, where $A_k \in \mathbf{C}^{M\times M}$, $1 \le k \le K$, the approximate joint diagonalization problem is to seek a nonsingular diagonalizing matrix $V \in \mathbf{C}^{N\times M}$ and K associated diagonal matrices $\Lambda_1, \Lambda_2, \cdots, \Lambda_K \in \mathbf{C}^{M\times M}$ such that the following forms are best fitted:

$$A_k = V\Lambda_k V^{\mathrm{H}} \text{ or } V^{\mathrm{H}} A_k V = \Lambda_k,\ k = 1, 2, \cdots, K \tag{1}$$

The best fit is computed by minimization of some measures. One mainly used measure is defined as follows:

$$O(V) = \sum_{k=1}^{K} \mathrm{off}(VA_kV^{\mathrm{H}}) \tag{2}$$

where

$$\mathrm{off}(A) \equiv \sum_{1\le i\ne j\le N} |a_{ij}|^2$$

a_{ij} denotes the i-th row and j-th column entry of matrix A. Definition (2) was proposed by Cardoso and Souloumiac in the joint approximate diagonalization of eigen-matrices (JADE) algorithm[16].

Based on criterion (2), Wang et al.[10] proposed the ARD algorithm. In the ARD algorithm, it is assumed that the diagonalizing matrix V has the following structure:

$$V = (V_1, V_2, \cdots, V_N)^{\mathrm{H}} = \begin{bmatrix} V_1^{\mathrm{H}} \\ V_2^{\mathrm{H}} \\ \vdots \\ V_N^{\mathrm{H}} \end{bmatrix}$$

Manuscript received April 11, 2012; accepted November 1, 2013
Supported by National Natural Science Foundation of China (11301056)
Recommended by CHEN Jing-Dong
1. School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China

where vectors $V_i \in \mathbf{C}^{M\times 1}$, $i = 1, 2, \cdots, N$ satisfy $b_i = \|V_i\|_F > 0$, $i = 1, 2, \cdots, N$. The ARD algorithm is given in Table 1.

Algorithm 1. The ARD algorithm

Initialization: b, V.
$L = 0$.
(All rows of the diagonalizing matrix are updated once every iteration).
Do while (if not convergent)
    $L = L + 1$.
    Do $i = 1, \cdots, N$
        $B = (V_1, \cdots, V_{i-1}, V_{i+1}, \cdots, V_N)$.
        $R = \sum_{k=1}^{K}(A_k BB^{\rm H} A_k^{\rm H} + A_k^{\rm H} BB^{\rm H} A_k)$.
        Find the unit eigenvector h associated with the least eigenvalue of R.
        Update $\tilde{V}_i = b_i h$, $\tilde{V} = (V_1, \cdots, V_{i-1}, \tilde{V}_i, V_{i+1}, \cdots, V_N)^{\rm H}$.
        $V = \tilde{V}$.
    Enddo
Enddo

However, the authors have not considered the convergence of the ARD algorithm nor the influences of the initial matrix V on the ARD algorithm[10]. In this letter, we will discuss these issues and propose a new terminal condition.

1 Convergence of ARD algorithm

Without loss of generality, we only need consider the case of one updated V_i. For convenience, vector h denotes the unitary eigenvector corresponding to the least eigenvalue of R in the ARD algorithm.

Theorem 1. If $\widetilde{V} = (V_1, \cdots, V_{i-1}, \widetilde{V}_i, V_{i+1}, \cdots, V_N)^{\rm H}$, where $\widetilde{V}_i = b_i h$, vector h is the unitary eigenvector corresponding to the least eigenvalue of R in the ARD algorithm, then

$$O(V) = \sum_{k=1}^{K} \text{off}(VA_kV^{\rm H}) \geq O(\widetilde{V}) = \sum_{k=1}^{K} \text{off}(\widetilde{V}A_k\widetilde{V}^{\rm H}) \quad (3)$$

Proof. According to the respective forms of $VA_kV^{\rm H}$ and $\widetilde{V}A_k\widetilde{V}^{\rm H}$, we have

$$\begin{aligned} O(V) - O(\widetilde{V}) = &\sum_{k=1}^{K} \text{off}(VA_kV^{\rm H}) - \sum_{k=1}^{K} \text{off}(\widetilde{V}A_k\widetilde{V}^{\rm H}) = \\ &V_i^{\rm H}RV_i - \widetilde{V}_i^{\rm H}R\widetilde{V}_i \end{aligned} \quad (4)$$

where the Hermitian matrix $R = \sum_{k=1}^{K}(A_kBB^{\rm H}A_k^{\rm H} + A_k^{\rm H}BB^{\rm H}A_k)$ is positive definite (or positive semidefinite).

It is easy to show that $\|V_i\|_F = \|\widetilde{V}_i\|_F = b_i$. Furthermore, since vector h is the unitary eigenvector corresponding to the least eigenvalue of R and because of the equality (4), we have

$$O(V) - O(\widetilde{V}) \geq 0$$

i.e.,

$$O(V) \geq O(\widetilde{V})$$

This proof is completed. □

Remark 1. According to Theorem 1, the cost function $O(V)$ is a nonincreasing function of matrix V which is updated as in the ARD algorithm, and has a lower bound. Hence, the ARD algorithm is convergent.

Remark 2. Let $V_i = \widetilde{V}_i + \Delta V_i$. By (4), we have

$$\begin{aligned} \|O(V) - O(\widetilde{V})\| = &\|\Delta V_i^{\rm H}R\widetilde{V}_i + \widetilde{V}_i^{\rm H}R\Delta V_i + \Delta V_i^{\rm H}R\Delta V_i\| \leq \\ &\|\Delta V_i\|(2b_i\lambda_{\min}(R) + \|\Delta V_i\|\|R\|) \end{aligned}$$

where $\|A\|$ denotes any norm of A, and $\lambda_{\min}(R)$ denotes the minimum eigenvalue of R. In a manner similar to the JADE algorithm, when all $\|\Delta V_i\|, i = 1, 2, \cdots, N$, are less than some threshold, we can argue that the ARD algorithm can be terminated. In other words, $\|\Delta V_i\|, i = 1, 2, \cdots, N$ can be served as the terminal condition of the ARD algorithm.

In the ARD algorithm, we need an initial matrix V. By the proof of Theorem 1, the decrement of the cost function $O(V)$ and matrix V have a very tight relation. Hence, the initialization of matrix V is very important. But how to choose a good V is very difficult. In the next section, we will propose some advice on how to choose it.

2 Examples

In this section, we will give three examples for illustrating the importance of the initialization of matrix V and the validity of the new stopping criterion. All the numerical examples were performed with MATLAB 2009a. In the following, we denote some MATLAB function commands. The command diag(x) produces a diagonal matrix whose diagonal elements are the elements of vector x. The command randn(m, n) returns an $m \times n$ matrix containing pseudorandom values drawn from the standard normal distribution.

Example 1. Consider

$$A_1 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \ A_2 = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

$V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $W = \begin{bmatrix} -0.5257 & -0.8507 \\ -0.8507 & 0.5257 \end{bmatrix}$ are used as the initial matrices, respectively, where the columns of W are the unitary eigenvectors of A_2. After 40 sweeps, we have

$$VA_1V^{\rm H} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$
$$VA_2V^{\rm H} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

and

$$WA_1W^{\rm H} = \begin{bmatrix} 2.2761 & 0 \\ 0 & 0.3905 \end{bmatrix}$$
$$WA_2W^{\rm H} = \begin{bmatrix} 0.3905 & 0 \\ 0 & 2.2761 \end{bmatrix}$$

The log$O(V)$ and log$O(W)$ of each sweep are shown in Fig. 1.

From the above results, we can see that the initial matrix V is important in the ARD algorithm. In general, if the unitary eigenvectors of A_i, $1 \leq i \leq K$, can be used as the columns of the initial matrix V, then this choice is good.

Example 2[17]**.** Consider

$$A_1 = \begin{bmatrix} 1-\epsilon & 0 & 0 & 0 \\ 0 & 1+\epsilon & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad A_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1-\epsilon & 0 \\ 0 & 0 & 0 & 1+\epsilon \end{bmatrix} \quad (5)$$

where $\epsilon = 0.01$. Let $V = I_4$ and W be the initial matrices, respectively, where I_4 is a 4×4 identity matrix, and the columns of W are the unitary eigenvectors of A_1. In this example, we use $\max_{1 \le i \le 4} \|\triangle \widetilde{V}_i\|_F \le 10^{-6}$ as the stopping criterion. When the ARD algorithm is convergent, we have

$$VA_1V^{\rm H} = \begin{bmatrix} 0.9900 & 0.9900 & 0 & 0 \\ 0.9900 & 0.9900 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$VA_2V^{\rm H} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9900 & 0.9900 \\ 0 & 0 & 0.9900 & 0.9900 \end{bmatrix}$$

and

$$WA_1W^{\rm H} = \begin{bmatrix} 0.2294 & -0.0000 & -0.0000 & 0.0000 \\ -0.0000 & -0.2294 & -0.0000 & -0.0000 \\ -0.0000 & -0.0000 & 0.9999 & 0.0000 \\ 0.0000 & -0.0000 & 0.0000 & 0.9999 \end{bmatrix}$$

$$WA_2W^{\rm H} = \begin{bmatrix} -0.2294 & -0.0000 & 0.0000 & -0.0000 \\ -0.0000 & 0.2294 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.9999 & -0.0000 \\ -0.0000 & 0.0000 & -0.0000 & 0.9999 \end{bmatrix}$$

From Example 2, we can see that the condition $\|\triangle \widetilde{V}_i\|_F \le \epsilon$ can be used as the stopping criterion. A good initialization of matrix V can make $VA_iV^{\rm H},\ i = 1, \cdots, K$, simultaneously diagonal as far as possible. A bad initialization of the matrix V cannot make $VA_iV^{\rm H},\ i = 1, \cdots, K$ be of simultaneous diagonalization, even $VA_iV^{\rm H},\ i = 1, \cdots, K$, become degenerated matrices.

From the above two examples, the initialization of matrix V is very important. However, such deeper discussions are beyond the scope of this letter.

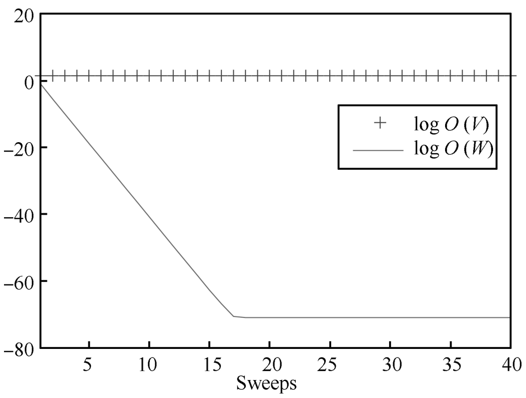

Fig. 1 Values of the cost function $O(V)$ for different initial matrices

Example 3. Consider $K = 100$ and $A_k = V^{\rm H} D_k V$, where $V = {\rm randn}(n, n)$, $D_k = {\rm diag}({\rm randn}(n, 1))$ and $n = 10$, $1 \le k \le K$. Let $V = I_n$ and W_i be the initial matrices, respectively, where I_n is an $n \times n$ identity matrix, and the columns of W_i are the unitary eigenvectors of A_i. In this example, we use $\max_{1 \le i \le n} \|\triangle \widetilde{V}_i\|_F \le 10^{-6}$ as the stopping criterion. In this example, we take the averages of 100 tests in sweeps for each initial matrix in Table 2.

Table 2 Averages of 100 tests in sweeps for different initial matrices

I_{10}	W_1	W_{10}	W_{20}	W_{50}	W_{60}	W_{90}	W_{100}
40.16	35.74	34.98	38.33	38.34	36.97	32.41	38.39

From Example 3, we can see that the proposed initial matrices are stable and effective.

3 Conclusion

In this paper, we performed some convergence analysis on the ARD algorithm. The importance of the initial matrix is also considered; it has great influence on the convergence of the ARD algorithm. By some examples, we can find that a good initial matrix can be obtained by the eigenvectors of $A_k,\ k = 1, \cdots, K$.

References

1 Flury B N. Common principal components in k groups. *Journal of the American Statistical Association*, 1984, **79**(388): 892−898
2 Cardoso J F, Souloumiac A. Blind beamforming for non-Gaussian signals. *IEEE Proceedings F, Radar and Signal Processing*, 1993, **140**(6): 362−370
3 Wax M, Sheinvald J. A least-squares approach to joint diagonalization. *IEEE Signals Processing Letters*, 1997, **4**(2): 52−53
4 Yeredor A. Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting. *IEEE Signal Processing Letters*, 2000, **7**(7): 197−200
5 Joho M, Mathis H. Joint diagonalization of correlation matrices by using gradient methods with application to blind signal separation. In: Proceedings of the Sensor Array and Multichannel Signal Processing Workshop. Rosslyn, VA, USA: IEEE, 2002. 273−277
6 Micka O J, Weiss A J. Estimating frequencies of exponentials in noise using joint diagonalization. *IEEE Transactions on Signal Processing*, 1999, **47**(2): 341−348
7 Feng D Z, Zhang X D, Bao Z. An efficient multistage decomposition approach for independent components. *Signal Processing*, 2003, **83**(1): 181−197
8 Cardoso J F. On the performance of orthogonal source separation algorithms. In: Proceedings of EUSIPCO. 1994. 776−779
9 Yeredor A. On using exact joint diagonalization for noniterative approximate joint diagonalization. *IEEE Signal Processing Letters*, 2005, **12**(9): 645−648
10 Wang F X, Liu Z K, Zhang J. A new joint diagonalization algorithm with application in blind source separation. *IEEE Signal Processing Letters*, 2006, **13**(1): 41−44
11 Vollgraf R, Obermayer K. Quadratic optimization for simultaneous matrix diagonalization. *IEEE Transactions on Signal Processing*, 2006, **54**(9): 3270−3278
12 Dégerine S, Kane E. A comparative study of approximate joint diagonalization algorithms for blind source separation in presence of additive noise. *IEEE Transactions on Signal Processing*, 2007, **55**(6): 3022−3031
13 Li X L, Zhang X D. Nonorthogonal joint diagonalization free of degenerate solution. *IEEE Transactions on Signal Processing*, 2007, **55**(5): 1803−1814
14 Tichavsky P, Yeredor A. Fast approximate joint diagonalization incorporating weight matrices. *IEEE Transactions on Signal Processing*, 2009, **57**(3): 878−891
15 Chabriel G, Barrére J. A direct algorithm for nonorthogonal approximate joint diagonalization. *IEEE Transactions on Signal Processing*, 2012, **60**(1): 39−47
16 Cardoso J F, Souloumiac A. Jacobi angles for simultaneous diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 1996, **17**(1): 161−164
17 Bunse-Gerstner A, Byers R, Mehrmann V. Numerical methods for simultaneous diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 1996, **14**(4): 927−949

CHENG Guang-Hui His research interest covers numerical linear algebra and blind source separation.
E-mail: ghcheng@126.com