

多维广义次成分提取准则及自适应算法

董海迪^{1†}, 刘 刚¹, 何 兵¹, 郑建飞¹, 王世涛²

(1. 火箭军工程大学核工程学院, 西安 710025; 2. 海军装备研究院水下所, 北京 100161)

摘 要: 针对广义次成分在信号处理中没有与之相对应的信息准则, 而且只能提取单维广义次成分, 提出一种多维广义次成分提取准则, 并通过矩阵微分法证明了所提出信息准则只有唯一的全局极大值. 基于该信息准则, 采用梯度上升法导出一种多维广义次成分提取算法, 并采用李雅普诺夫函数法对所提出算法的全局收敛性进行证明. 与其他现有算法不同, 所提出算法可以并行提取多维次成分, 而并不需要模值归一化. 仿真实验表明, 所提出算法相比一些现有算法具有收敛速度快和估计精度高的优点.

关键词: 广义次成分; 信息准则; 李雅普诺夫函数; 多维提取

中图分类号: TP391

文献标志码: A

Multiple minor generalized eigenvectors extraction information and its adaptive algorithm

DONG Hai-di^{1†}, LIU Gang¹, HE Bing¹, ZHENG Jian-fei¹, WANG Shi-tao²

(1. Nuclear Engineering Institute, Rocket Force Engineering University, Xi'an 710025, China; 2. Underwater Institute, Naval Academy of Armament, Beijing 100161, China)

Abstract: Up to now, it is a hard work to find the corresponding information criterions for minor generalized components in many signal processing fields, and only one dimension minor generalized component can be extratred. In order to solve this problem, this paper proposes a multiple minor generalized components extraction information criterion, which is proved to have only one global maximum. A novel algorithm is also derived by using gradient ascent method to this information criterion. The global convergence of the proposed algorithm is analyzed by using the Lyapunov function method. Compared with existing algorithms, the proposed algorithm does not need normalization operation and can extract multiple minor generalized components in parallel. Simulation results show the accuracy and the speed advantages of the proposed algorithm.

Keywords: minor generalized component; information criterion; Lyapunov function; multiple extraction

0 引 言

广义特征值分解(GED)是现代信号处理领域重要的分析工具,可以应用在波达方向估计^[1]、特征提取^[2]和系统辨识^[3]等领域.在信号处理领域将矩阵束最小广义特征值对应的特征向量称为广义次成分.广义次成分已经应用于有限元分析^[4]、数据分类^[5]、目标识别^[6]和盲分离^[7]等领域.

为了求解矩阵束的次成分,学者们相继提出了很多数值算法^[8-10].这些数值算法需要矩阵束(R_y, R_x)是已知的,而在很多信号处理问题中矩阵束是未知的,只能从输入信号中进行估计得到,因此发展自适

应的广义次成分提取算法显得很有必要.近些年来,基于神经网络的广义次成分提取是国际上的一个研究热点,很多学者围绕这一方向展开了深入的研究.基于单层前向神经网络, Ye等^[11]提出了一种广义次成分提取算法,然而,该算法计算复杂度较大而且收敛速度慢,在处理时变信号时容易出现数值不稳定的情况; Nguyen等^[12]通过对幂迭代算法和Oja-Xu算法进行改进,提出了两种广义次成分提取算法,然而,该算法需要在每次迭代过程中增加模值归一化措施.上述算法均是基于启发式推理而提出的,并没有建立算法相对应的信息准则.信息准则在算法发展

收稿日期: 2017-08-01; 修回日期: 2017-10-10.

基金项目: 国家自然科学基金项目(61403399).

责任编委: 曹进德.

作者简介: 董海迪(1988—),男,博士生,从事自适应信号处理的研究; 刘刚(1964—),男,教授,博士生导师,从事自适应信号处理、神经网络等研究.

[†]通讯作者. E-mail: donghaidi123@163.com.

中具有重要意义^[13],因此发展广义次成分的信息准则是一个重要的研究方向.在文献[14]中,Nguyen等将CIC准则进行了扩展,发展了一种广义次成分提取准则,然而,基于此准则导出的算法仍然需要模值归一化操作.此外,上述算法均只能提取信号的单维广义次成分,并不能提取多维广义次成分.

本文提出一种新型的多维广义次成分提取准则,并采用矩阵微分法完成了该信息准则的极值点分析;基于该信息准则,采用梯度上升法导出了一个广义次成分提取算法,该算法具有计算复杂度低、收敛速度快和收敛精度高等优点;采用李雅普诺夫函数法对所提出算法的全局收敛性进行了证明.

1 新型信息准则的提出

GED问题就是求解向量 $\mathbf{v}$ 和标量 λ ,使得如下方程成立:

$$\mathbf{R}_y \mathbf{v} = \lambda \mathbf{R}_x \mathbf{v}, \quad (1)$$

其中 $\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^T]$ 和 $\mathbf{R}_y = E[\mathbf{y}\mathbf{y}^T]$ 分别是向量 $\mathbf{x}(k)$ 和向量 $\mathbf{y}(k)$ 的自相关矩阵.将满足式(1)的向量 $\mathbf{v}$ 和标量 λ 称为矩阵束 $(\mathbf{R}_y, \mathbf{R}_x)$ 的广义特征向量和广义特征值.

根据矩阵理论,矩阵束 $(\mathbf{R}_y, \mathbf{R}_x)$ 有 n 个正的广义特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n > 0$,并对应有 n 个广义特征向量 $\mathbf{v}_i (i = 1, 2, \dots, n)$,且有

$$\mathbf{R}_y \mathbf{v}_i = \lambda_i \mathbf{R}_x \mathbf{v}_i, \quad (2)$$

$$\mathbf{v}_i^T \mathbf{R}_x \mathbf{v}_j = \delta_{ij}, \quad i, j = 1, 2, \dots, n, \quad (3)$$

其中 δ_{ij} 是Kronecher函数.

考虑如下具有多输入多输出的神经网络:

$$\mathbf{q}(k) = \mathbf{W}^T \mathbf{p}(k), \quad k = 1, 2, \dots \quad (4)$$

其中: $\mathbf{q}_k \in \mathbf{R}^{r \times 1}$ 是神经网络的输出; $\mathbf{W} \in \mathbf{R}^{n \times r}$ 是神经网络的状态矩阵;输入信号 $\mathbf{p}_k \in \mathbf{R}^{n \times 1}$ 是一个零均值的随机过程,这里作为神经网络的输入; n 代表输入向量的维数; r 代表所需提取广义次成分的维数.基于神经网络的广义次成分提取算法本质上是构造合适的状态矩阵更新方程,使得状态矩阵能够最终收敛到矩阵束 $(\mathbf{R}_y, \mathbf{R}_x)$ 的广义次成分.

基于AMEX准则^[15]和加权矩阵思想^[16],本文提出一种新型的信息准则——WAMEX准则(Weighted AMEX),其表达式为

$$\mathbf{W}^* = \arg \max J(\mathbf{W}),$$

$$J(\mathbf{W}) = \frac{1}{2} \text{tr}[\ln(\mathbf{A}\mathbf{W}^T \mathbf{R}_x \mathbf{W} \mathbf{A}) - \mathbf{W}^T \mathbf{R}_y \mathbf{W}], \quad (5)$$

其中 $\mathbf{A} = \text{diag}(a_1, a_2, \dots, a_r)$ 是一个对角矩阵且其对角线元素满足 $a_1 > a_2 > \dots > a_r$.式(5)是一

个广义次成分提取准则,其研究对象是两个输入信号自相关矩阵构成的矩阵束;而文献[15-16]是普通的次成分提取准则,其研究对象是单个信号的自相关矩阵.所提出信息准则的研究对象更为复杂.如果令式(5)中 $\mathbf{R}_x = \mathbf{I}_n$ 和 $\mathbf{A} = \mathbf{I}_r$ ($\mathbf{I}_n$ 和 $\mathbf{I}_r$ 分别为 $n \times n$ 维和 $r \times r$ 维单位矩阵),则式(5)退化为文献[15]中的AMEX准则.因此可以说,式(5)是AMEX准则的广义模式.

由式(5)可得 $J(\mathbf{W})$ 是无下界的,且当 $\mathbf{W}$ 趋于无穷大时, $J(\mathbf{W})$ 也将趋于无穷大,因此研究 $J(\mathbf{W})$ 的极小值是没有意义的.这里将重点研究 $J(\mathbf{W})$ 的极大点,具体为以下几个问题:

- 1) $J(\mathbf{W})$ 有没有全局极大点;
- 2) 极大点与广义次成分之间的关系是什么;
- 3) $J(\mathbf{W})$ 有没有其他局部极值.

2 信息准则的平稳点分析

本节将采用矩阵微分法对所提出信息准则进行分析,回答上述问题.

定理1 在域 $\Omega = \{\mathbf{W} | 0 < \mathbf{W}^T \mathbf{R}_x \mathbf{W} < \infty\}$ 中,当且仅当 $\mathbf{W} = \mathbf{L}_r \mathbf{A}^{-1/2} \mathbf{Q}$ 时, $\mathbf{W}$ 是 $J(\mathbf{W})$ 准则的一个平稳点,其中 $\mathbf{L}_r$ 是一个由矩阵束 $(\mathbf{R}_y, \mathbf{R}_x)$ 的任意 r 个不同的特征向量组成的 $n \times r$ 维矩阵, $\mathbf{Q}$ 是任意一个 $r \times r$ 维交换矩阵.

证明 由于在域 Ω 内, $\mathbf{W}^T \mathbf{R}_x \mathbf{W}$ 是对称正定的,因此 $J(\mathbf{W})$ 对于矩阵 $\mathbf{W}$ 的一阶微分存在,即

$$\begin{aligned} dJ(\mathbf{W}) = & \frac{1}{2} d\{\text{tr}[\log(\mathbf{A}\mathbf{W}^T \mathbf{R}_x \mathbf{W} \mathbf{A})] - \text{tr}(\mathbf{W}^T \mathbf{R}_y \mathbf{W})\} = \\ & \text{tr}[(\mathbf{A}\mathbf{W}^T \mathbf{R}_x \mathbf{W} \mathbf{A})^{-1} \mathbf{A}\mathbf{W}^T \mathbf{R}_x d\mathbf{W} \mathbf{A}] - \\ & \text{tr}(\mathbf{W}^T \mathbf{R}_y d\mathbf{W}). \end{aligned} \quad (6)$$

进而可得

$$\nabla J(\mathbf{W}) = \mathbf{R}_x \mathbf{W} \mathbf{A}^2 (\mathbf{A}\mathbf{W}^T \mathbf{R}_x \mathbf{W} \mathbf{A})^{-1} - \mathbf{R}_y \mathbf{W}. \quad (7)$$

如果 $\mathbf{W} = \mathbf{L}_r \mathbf{A}^{-1/2} \mathbf{Q}$,则易得 $\nabla J(\mathbf{W}) = 0$.反之,根据定义,可得在 $J(\mathbf{W})$ 的平稳点处有 $\nabla J(\mathbf{W}) = 0$,进而有

$$\mathbf{R}_x \mathbf{W} \mathbf{A}^2 (\mathbf{A}\mathbf{W}^T \mathbf{R}_x \mathbf{W} \mathbf{A})^{-1} = \mathbf{R}_y \mathbf{W}. \quad (8)$$

将式(8)左右两边同时乘以 $\mathbf{W}^T$,可得

$$\mathbf{W}^T \mathbf{R}_x \mathbf{W} \mathbf{A}^2 (\mathbf{A}\mathbf{W}^T \mathbf{R}_x \mathbf{W} \mathbf{A})^{-1} = \mathbf{W}^T \mathbf{R}_y \mathbf{W}. \quad (9)$$

由于 $\mathbf{W}^T \mathbf{R}_y \mathbf{W}$ 是一个对称矩阵,有

$$\begin{aligned} \mathbf{W}^T \mathbf{R}_x \mathbf{W} \mathbf{A}^2 (\mathbf{A}\mathbf{W}^T \mathbf{R}_x \mathbf{W} \mathbf{A})^{-1} = \\ (\mathbf{A}\mathbf{W}^T \mathbf{R}_x \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^2 \mathbf{W}^T \mathbf{R}_x \mathbf{W}. \end{aligned} \quad (10)$$

经过适当简化,有

$$(AW^T R_x W A W^T R_x W A) A = A(AW^T R_x W A W^T R_x W A). \quad (11)$$

因 A 是对角矩阵, 故 $AW^T R_x W A W^T R_x W A$ 必定是对角矩阵. 定义如下矩阵:

$$A' = AW^T R_x W A W^T R_x W A, \quad (12)$$

A' 也是一个对角矩阵. 经过适当化简, 有

$$A^{-1/2} A' A^{-1/2} = (A^{1/2} W^T R_x W A^{1/2})^2. \quad (13)$$

A' 和 A 都是对角矩阵, 矩阵 $A^{1/2} W^T R_x W A^{1/2}$ 同样是对角矩阵, 进而可得 $W^T R_x W$ 是对角矩阵. 基于上述结论, 有

$$W^T R_x W A^2 = AW^T R_x W A. \quad (14)$$

将式(14)代入(8), 可得

$$W^T R_y W = I_r. \quad (15)$$

根据矩阵理论, 矩阵 R_y 的 GED 为

$$R_y = R_x U_1 \Lambda_1 U_1^T R_x + R_x U_n \Lambda_n U_n^T R_x. \quad (16)$$

其中: Λ_1 和 Λ_n 分别是对角线元素由矩阵束 (R_y, R_x) 的前 $n-r$ 个与剩余的 r 个广义特征值组成的对角矩阵, 而 U_1 和 U_n 则是由 Λ_1 和 Λ_n 中广义特征值对应的广义特征向量组成的矩阵. 将式(16)代入(15)可得 $Q^T Q = I_r$, 其中 $Q = A^{1/2} W^T R_x W$. 即在 $J(W)$ 准则的平稳点处, 矩阵 Q 的各列是彼此正交的. 因此可知 $W = L_r A^{1/2} Q$ 是 $J(W)$ 准则一个平稳点. $\square$

定理2 在域 Ω 内, 当且仅当 $W = L_n A^{-1/2} Q$ 时, $J(W)$ 准则达到全局最大点, 其中 $L_n = [v_{n-r+1}, v_{n-r+2}, \dots, v_n]$ 是一个 $n \times r$ 维矩阵, 而 $J(W)$ 准则的其他平稳点都是不稳定的鞍点. 在全局最大点处有

$$J(W) = \frac{1}{2} \left(\sum_{i=n-r+1}^n \frac{a_i^2}{\lambda_i - r} \right). \quad (17)$$

证明 由式(6)可以获得 $J(W)$ 准则对于矩阵 W 的二阶微分, 即

$$\begin{aligned} d^2 J(W) = & d\{\text{tr}[(AW^T R_x W A)^{-1} AW^T R_x dW A] - \\ & \text{tr}(W^T R_y dW)\} = \\ & - \text{tr}[(AW^T R_x W A)^{-1} d(AW^T R_x W A) \times \\ & (AW^T R_x W A)^{-1} AW^T R_x dW A] + \\ & \text{tr}[(AW^T R_x W A)^{-1} A d(W^T) R_x dW A] - \\ & \text{tr}[d(W^T) R_y dW] = \\ & - \text{tr}[(AW^T R_x W A)^{-1} A dW^T \times \\ & R_x W A (AW^T R_x W A)^{-1} AW^T R_x dW A] - \end{aligned}$$

$$\begin{aligned} & \text{tr}[(AW^T R_x W A)^{-1} AW^T R_x dW \times \\ & A (AW^T R_x W A)^{-1} AW^T R_x dW A] + \\ & \text{tr}[(AW^T R_x W A)^{-1} A d(W^T) R_x dW A] - \\ & \text{tr}[d(W^T) R_y dW] = \\ & - [A (AW^T R_x W A)^{-1} A] \otimes \\ & [R_x W A (AW^T R_x W A)^{-1} AW^T R_x] - \\ & K_{rn} [R_x W A (AW^T R_x W A)^{-1} A] \otimes \\ & [A (AW^T R_x W A)^{-1} AW^T R_x] + \\ & [A (AW^T R_x W A)^{-1} A] \otimes R_x - I_r \otimes R_y. \quad (18) \end{aligned}$$

令 $J(W)$ 向量 $\text{vec}(W) = [w_1^T, w_2^T, \dots, w_r^T]^T$ 的 Hessian 矩阵为 $HJ(W)$, 有

$$HJ(W) = \frac{\partial}{\partial(\text{vec}(W))^T} \left(\frac{\partial J(W)}{\partial(\text{vec}(W))^T} \right)^T. \quad (19)$$

由式(18)可得

$$\begin{aligned} HJ(W) = & - [A (AW^T R_x W A)^{-1} A] \otimes \\ & [R_x W A (AW^T R_x W A)^{-1} AW^T R_x] - \\ & K_{rn} [A (AW^T R_x W A)^{-1} AW^T R_x] \otimes \\ & [A (AW^T R_x W A)^{-1} AW^T R_x] + \\ & [A (AW^T R_x W A)^{-1} A] \otimes R_x - I_r \otimes R_y, \quad (20) \end{aligned}$$

其中 K_{rn} 是一个 M 维交换矩阵, 对于矩阵均有 $K_{rn} \text{vec}(M) = \text{vec}(M^T)$ 成立. K_{rn} 有一个非常好的性质, 就是对于任意的矩阵 $M_1 \in R^{m \times n}$ 和 $M_2 \in R^{p \times q}$, 有下式成立:

$$K_{pm}(M_1 \otimes M_2) = (M_2 \otimes M_1) K_{qn}. \quad (21)$$

在平稳点 $W = L_n A^{-1/2} Q$ 处, 有下式成立:

$$A(AW^T R_x W A)^{-1} A = Q^T \Lambda_n Q, \quad (22)$$

$$R_x W = R_x L_n A^{-1/2} Q. \quad (23)$$

将式(9)代入(20), 然后计算矩阵 $HJ(W)$ 在平稳点 $W = L_n A^{-1/2} Q$ 处的值, 即

$$\begin{aligned} H^* = HJ(W)|_{W=L_n A^{-1/2} Q} = & - [A (AW^T R_x W A)^{-1} A] \otimes \\ & [R_x W A (AW^T R_x W A)^{-1} AW^T R_x] - \\ & K_{rn} [R_x W A (AW^T R_x W A)^{-1} A] \otimes \\ & [A (AW^T R_x W A)^{-1} AW^T R_x] + \\ & [A (AW^T R_x W A)^{-1} A] \otimes R_x - I_r \otimes R_y = \\ & - [Q^T \Lambda_n Q] \otimes [R_x L_n L_n^T R_x] - \end{aligned}$$

$$K_{rn}[\mathbf{R}_x \mathbf{L}_n \mathbf{A}_n^{1/2} \mathbf{Q}] \otimes [\mathbf{Q}^T \mathbf{A}_n^{1/2} \mathbf{L}_n^T \mathbf{R}_x] +$$

$$[\mathbf{Q}^T \mathbf{A}_n \mathbf{Q}] \otimes \mathbf{R}_x - \mathbf{I}_r \otimes \mathbf{R}_y. \quad (24)$$

将式(16)代入(24),并化简得

$$\begin{aligned} \mathbf{H}^* = & -[\mathbf{Q}^T \mathbf{A}_n \mathbf{Q}] \otimes [\mathbf{R}_x \mathbf{L}_n \mathbf{L}_n^T \mathbf{R}_x] - \\ & K_{rn}[\mathbf{R}_x \mathbf{L}_n \mathbf{A}_n^{1/2} \mathbf{Q}] \otimes [\mathbf{Q}^T \mathbf{A}_n^{1/2} \mathbf{L}_n^T \mathbf{R}_x] + \\ & [\mathbf{Q}^T \mathbf{A}_n \mathbf{Q}] \otimes \mathbf{R}_x - \mathbf{I}_r \otimes \mathbf{R}_y = \\ & -[\mathbf{Q}^T \mathbf{A}_n \mathbf{Q}] \otimes [\mathbf{R}_x \mathbf{L}_n \mathbf{L}_n^T \mathbf{R}_x] - \\ & K_{rn}[\mathbf{R}_x \mathbf{L}_n \mathbf{A}_n^{1/2} \mathbf{Q}] \otimes [\mathbf{Q}^T \mathbf{A}_n^{1/2} \mathbf{L}_n^T \mathbf{R}_x] + \\ & [\mathbf{Q}^T \mathbf{A}_n \mathbf{Q}] \otimes [\mathbf{R}_x \mathbf{L}_1 \mathbf{L}_1^T \mathbf{R}_x + \mathbf{R}_x \mathbf{L}_n \mathbf{L}_n^T \mathbf{R}_x] - \\ & \mathbf{I}_r \otimes [\mathbf{R}_x \mathbf{L}_1 \mathbf{A}_1 \mathbf{L}_1^T \mathbf{R}_x + \mathbf{R}_x \mathbf{L}_n \mathbf{A}_n \mathbf{L}_n^T \mathbf{R}_x] = \\ & -K_{rn}[\mathbf{R}_x \mathbf{L}_n \mathbf{A}_n^{1/2} \mathbf{Q}] \otimes [\mathbf{Q}^T \mathbf{A}_n^{1/2} \mathbf{L}_n^T \mathbf{R}_x] + \\ & [\mathbf{Q}^T \mathbf{A}_n \mathbf{Q}] \otimes [\mathbf{R}_x \mathbf{L}_1 \mathbf{L}_1^T \mathbf{R}_x] - \\ & \mathbf{I}_r \otimes [\mathbf{R}_x \mathbf{L}_1 \mathbf{A}_1 \mathbf{L}_1^T \mathbf{R}_x] - \mathbf{I}_r \otimes [\mathbf{R}_x \mathbf{L}_n \mathbf{A}_n \mathbf{L}_n^T \mathbf{R}_x] = \\ & -K_{rn}[(\mathbf{R}_x \mathbf{L}_n) \otimes (\mathbf{Q}^T \mathbf{A}_n^{1/2})] \times \\ & [(\mathbf{A}_n^{1/2} \mathbf{Q}) \otimes (\mathbf{L}_n^T \mathbf{R}_x)] - \mathbf{I}_r \otimes [\mathbf{R}_x \mathbf{L}_1 \mathbf{A}_1 \mathbf{L}_1^T \mathbf{R}_x] + \\ & [(\mathbf{Q}^T \mathbf{A}_n^{1/2}) \otimes (\mathbf{R}_x \mathbf{L}_1)][(\mathbf{A}_n^{1/2} \mathbf{Q}) \otimes (\mathbf{L}_1^T \mathbf{R}_x)] - \\ & \mathbf{I}_r \otimes [\mathbf{R}_x \mathbf{L}_n \mathbf{A}_n \mathbf{L}_n^T \mathbf{R}_x]. \end{aligned} \quad (25)$$

将式(21)代入(25),右侧第1项有

$$K_{rn}[(\mathbf{R}_x \mathbf{L}_n) \otimes (\mathbf{Q}^T \mathbf{A}_n^{1/2})] =$$

$$[(\mathbf{Q}^T \mathbf{A}_n^{1/2}) \otimes (\mathbf{R}_x \mathbf{L}_n)] \mathbf{K}_{rr}, \quad (26)$$

其中 $\mathbf{K}_{rr}$ 是一个对称正交矩阵,即有 $\mathbf{K}_{rr}^T = \mathbf{K}_{rr}^{-1} = \mathbf{K}_{rr}$. 令其特征值分解为

$$\mathbf{K}_{rr} = \mathbf{V} \mathbf{S} \mathbf{V}^T. \quad (27)$$

其中: $\mathbf{S} = \text{diag}(s_1, s_2, \dots, s_{r^2})$ 是一个对角矩阵,其对角线元素由矩阵 $\mathbf{K}_{rr}$ 的 r^2 个特征值组成;而 $\mathbf{V}$ 是由其对应的特征向量组成的 $r^2 \times r^2$ 维正交矩阵. 应用 $\text{tr}(\mathbf{K}_{rr}) = \sum_{i=1}^{r^2} s_i = r$ 可得,矩阵 $\mathbf{S}$ 有 $(r^2 + r)/2$ 重“1”元素和 $(r^2 - r)/2$ 重“-1”元素. 将 $\mathbf{I}_r = \mathbf{Q}^T \mathbf{A}_n^{1/2} \mathbf{A}_n^{-1} \mathbf{A}_n^{1/2} \mathbf{Q}$ 代入式(25),有

$$\begin{aligned} \mathbf{H}^* = & -[(\mathbf{Q}^T \mathbf{A}_n^{1/2}) \otimes (\mathbf{R}_x \mathbf{L}_n)] \mathbf{K}_{rr}[(\mathbf{A}_n^{1/2} \mathbf{Q}) \otimes (\mathbf{L}_n^T \mathbf{R}_x)] + \\ & [(\mathbf{Q}^T \mathbf{A}_n^{1/2}) \otimes (\mathbf{R}_x \mathbf{L}_1)][(\mathbf{A}_n^{1/2} \mathbf{Q}) \otimes (\mathbf{L}_1^T \mathbf{R}_x)] - \\ & -(\mathbf{Q}^T \mathbf{A}_n^{1/2} \mathbf{A}_n^{-1} \mathbf{A}_n^{1/2} \mathbf{Q}) \otimes [\mathbf{R}_x \mathbf{L}_1 \mathbf{A}_1 \mathbf{L}_1^T \mathbf{R}_x] - \\ & (\mathbf{Q}^T \mathbf{A}_n^{1/2} \mathbf{A}_n^{-1} \mathbf{A}_n^{1/2} \mathbf{Q}) \otimes [\mathbf{R}_x \mathbf{L}_n \mathbf{A}_n \mathbf{L}_n^T \mathbf{R}_x] = \end{aligned}$$

$$\begin{aligned} & \begin{bmatrix} \mathbf{V}^T (\mathbf{A}_n^{1/2} \mathbf{Q}) \otimes (\mathbf{L}_n^T \mathbf{R}_x) \\ \mathbf{V}^T (\mathbf{A}_n^{1/2} \mathbf{Q}) \otimes (\mathbf{L}_1^T \mathbf{R}_x) \end{bmatrix}^T \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \end{bmatrix} \times \\ & \begin{bmatrix} \mathbf{V}^T (\mathbf{A}_n^{1/2} \mathbf{Q}) \otimes (\mathbf{L}_n^T \mathbf{R}_x) \\ \mathbf{V}^T (\mathbf{A}_n^{1/2} \mathbf{Q}) \otimes (\mathbf{L}_1^T \mathbf{R}_x) \end{bmatrix}. \end{aligned} \quad (28)$$

其中: $\mathbf{S}_1 = -\mathbf{S} - \mathbf{A}_n^{-1} \otimes \mathbf{A}_n$, $\mathbf{S}_2 = \mathbf{I}_r - \mathbf{A}_n^{-1} \otimes \mathbf{A}_1$. 式(28)实际上是矩阵 $\mathbf{H}^*$ 的特征值分解,而其特征值是矩阵 $\mathbf{S}_1$ 和 $\mathbf{S}_2$ 的对角线元素.

由式(28)可得:当且仅当 $\min\{\lambda_1, \lambda_2, \dots, \lambda_{n-r}\} \geq \max\{\lambda_{n-r+1}, \lambda_{n-r+2}, \dots, \lambda_n\}$ 时, $\mathbf{H}^*$ 是一个负定矩阵. 即在平稳点 $\mathbf{W} = \mathbf{L}_n \mathbf{A}^{1/2} \mathbf{Q}$ 处,有局部极大值. 在 $\mathbf{W}$ 趋于无穷时, $J(\mathbf{W})$ 是无界的,因此该局部极大值也是全局最大值. 除了 $\mathbf{W} = \mathbf{L}_n \mathbf{A}^{1/2} \mathbf{Q}$,所有其他平稳点都将使得矩阵 $\mathbf{H}^*$ 出现未定状态(即同时有正的和负的特征值),因此是不稳定的.

将 $\mathbf{W} = \mathbf{L}_n \mathbf{A}^{1/2} \mathbf{Q}$ 代入 $J(\mathbf{W})$,可得该信息准则的全局最大值为

$$J(\mathbf{W}) = \frac{1}{2} \left(\sum_{i=n-r+1}^n \frac{a_i^2}{\lambda_i - r} \right). \quad (29)$$

定理2得证. $\square$

3 多维广义次成分提取算法

由定理2可知 $J(\mathbf{W})$ 有一个全局最大值,而没有其他极值,因此可采用梯度上升算法求解 $J(\mathbf{W})$ 准则的全局最大点. 通过信息准则 $J(\mathbf{W})$ 对矩阵 $(\mathbf{W})$ 求一阶微分,可以获得如下的梯度上升算法:

$$\begin{aligned} \mathbf{W}(k+1) = & \mathbf{W}(k) - \eta [\mathbf{R}_y(k) \mathbf{W}(k) - \\ & \mathbf{R}_x(k) \mathbf{W}(k) \mathbf{A}^2 (\mathbf{A} \mathbf{W}^T(k) \mathbf{R}_x(k) \mathbf{W}(k) \mathbf{A})^{-1}]. \end{aligned} \quad (30)$$

在实际使用过程中,自相关矩阵 $\mathbf{R}_x$ 和 $\mathbf{R}_y$ 通常是未知的,需要从输入信号中进行实时估计. 因此,需要将算法(30)改造成为一种自适应广义主成分分析算法. 首先,采用加权窗函数法对自相关矩阵进行估计,即

$$\hat{\mathbf{R}}_y(k+1) = \alpha_1 \hat{\mathbf{R}}_y(k) + \mathbf{y}(k+1) \mathbf{y}^T(k+1), \quad (31)$$

$$\hat{\mathbf{R}}_x(k+1) = \alpha_2 \hat{\mathbf{R}}_x(k) + \mathbf{x}(k+1) \mathbf{x}^T(k+1), \quad (32)$$

其中 $\alpha_1, \alpha_2 \in (0, 1)$ 为遗忘因子,其作用是对数据施加遗忘窗. 该遗忘窗通过对不同时刻的采样样本赋予不同的权值来降低以往数据对于当前结果的影响. 遗忘因子的取值应根据具体问题来决定,一般而言,对于一个变化缓慢的样本,遗忘因子可以选择一个接近1的数以产生一个较大的遗忘窗;而对于快速

变化的输入样本,遗忘窗口宽度应比较小,此时 α 则应该在接近0处取值.

到此完成了基于所提出信息准则的多维广义次成分算法推导.下面给出一些对所提出信息准则及其导出算法的评价.

1) 由定理2可以得出, $J(\mathbf{W})$ 有一个全局最大点而没有局部极值,因此采用梯度上升法导出的算法(30)可以对 $J(\mathbf{W})$ 求取极值.由于在实际计算过程中随机干扰的存在,其他不稳定的鞍点不会对所提出算法的收敛性造成任何困难.

2) 以往为了确保神经网络状态向量的收敛性,一些算法不得不在每次迭代过程中添加状态矩阵归一化操作.然而,本节所提出的算法并不需要该操作.由于省去了状态矩阵归一化操作,所提出算法具有较低的计算复杂度.

4 多维广义次成分提取算法

本节中将主要对所提出算法进行全局收敛性分析,主要是采用李雅普诺夫函数法来确定所提出算法的收敛域.假定输入信号 $\mathbf{x}$ 和 $\mathbf{y}$ 是两个零均值的随机过程且学习因子 η 足够小,则算法(30)可以通过下述连续时间常微分方程来描述:

$$\frac{d\mathbf{W}(t)}{dt} = \mathbf{R}_x \mathbf{W} \mathbf{A}^2 (\mathbf{A} \mathbf{W}^T \mathbf{R}_x \mathbf{W} \mathbf{A})^{-1} - \mathbf{R}_y \mathbf{W}, \quad (33)$$

其中 $t = \eta k$.通过对式(33)的全局收敛性分析,可以完成对所提出算法的全局收敛性分析.为了方便起见,这里定义如下函数:

$$L(\mathbf{W}) = \frac{1}{2} \text{tr}(\mathbf{W}^T \mathbf{R}_y \mathbf{W}) - \frac{1}{2} \text{tr}[\log(\mathbf{A} \mathbf{W}^T \mathbf{R}_x \mathbf{W} \mathbf{A})]. \quad (34)$$

由式(34)可看出,在域 $\Omega = \{\mathbf{W} | 0 < \mathbf{W}^T \mathbf{R}_x \mathbf{W} < \infty\}$ 内, $L(\mathbf{W})$ 是有界的.根据李雅普诺夫函数的定义,证明 $L(\mathbf{W})$ 的一阶微分是非负的.通过矩阵微分的链式法则,有

$$\begin{aligned} \frac{dL(\mathbf{W})}{dt} &= \text{tr} \left[\mathbf{W}^T \mathbf{R}_y \frac{d\mathbf{W}}{dt} - (\mathbf{A} \mathbf{W}^T \mathbf{R}_x \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^2 \mathbf{W}^T \mathbf{R}_x \frac{d\mathbf{W}}{dt} \right] = \\ &= -\text{tr} \left(\frac{d\mathbf{W}^T}{dt} \frac{d\mathbf{W}}{dt} \right). \end{aligned} \quad (35)$$

为了书写方便,式(35)中省略了时刻标志 t .由式(35)可得,在域 $\mathbf{W} \in \Omega - \{\mathbf{W} | \mathbf{W} = \mathbf{L}_n \mathbf{A}^{-1/2} \mathbf{Q}\}$ 中有 $d\mathbf{W}/dt \neq 0$ 和 $dL(\mathbf{W})/dt < 0$;而当且仅当 $\mathbf{W} =$

$\mathbf{L}_n \mathbf{A}^{-1/2} \mathbf{Q}$ 时,有 $d\mathbf{W}/dt = 0$.这就是说,在域 Ω 内对于任意的初始值 $\mathbf{W}$, $L(\mathbf{W})$ 均是严格单调递减的.换言之,在域 Ω 内, $L(\mathbf{W})$ 是一个李雅普诺夫函数,且对于域内的任何初始状态矩阵 $\mathbf{W}(0) \in \Omega$,均有 $\mathbf{W}(t)$ 全局渐近收敛到所需要提取的广义次成分.

5 仿真实验

本节将通过3个仿真实验来证明所提出算法的有效性.第1个实验是对所提出算法多维广义次成分提取能力的验证;第2个实验是将所提出算法与一些现存算法进行对比;第3个实验是应用所提出算法求解盲分离问题.

为了衡量算法的收敛性能,在每次迭代中分别计算方向余弦

$$\text{Direction Cosine}_i(k) = \frac{|\mathbf{w}_i^T(k) \mathbf{v}_i|}{\|\mathbf{w}_i(k)\| \|\mathbf{v}_i\|}. \quad (36)$$

其中: $\mathbf{w}_i(k)$ 为 k 时刻状态矩阵的第 i 列, $\mathbf{v}_i$ 为矩阵束 $(\mathbf{R}_y, \mathbf{R}_x)$ 的第 i 个广义次成分.显然,如果 $\mathbf{w}_i(k)$ 能够收敛到广义次成分 $\mathbf{v}_i$ 的方向,则方向余弦的值应该收敛到1.

5.1 多维广义次成分提取实验

在本实验中通过Matlab命令随机生成如下两个对称正定矩阵:

$$\mathbf{R}_y = \begin{bmatrix} 0.5444 & -0.0596 & 0.1235 & 0.0197 \\ -0.0596 & 0.3892 & -0.0583 & -0.1300 \\ 0.1235 & -0.0583 & 0.5093 & 0.0570 \\ 0.0197 & -0.1300 & 0.0570 & 0.3229 \\ -0.0611 & 0.0984 & 0.0394 & -0.0350 \\ -0.1309 & 0.0138 & 0.0582 & 0.2035 \\ -0.0055 & 0.1919 & -0.0140 & -0.1035 \\ -0.0611 & -0.1309 & -0.0055 & 0.0984 \\ 0.0984 & 0.0138 & 0.1919 & 0.0394 \\ 0.0394 & 0.0582 & -0.0140 & -0.0350 \\ -0.0350 & 0.2035 & -0.1035 & 0.4960 \\ 0.4960 & -0.0191 & -0.1087 & -0.0191 \\ -0.0191 & 0.3148 & 0.0212 & -0.1087 \\ -0.1087 & 0.0212 & 0.2819 & 0.3979 \\ 0.3979 & 0.0633 & 0.0294 & 0.0805 \\ 0.0633 & 0.5778 & -0.0858 & -0.0067 \\ 0.0294 & -0.0858 & 0.5221 & -0.0413 \\ 0.0805 & -0.0067 & -0.0413 & 0.6070 \\ -0.0199 & 0.0971 & 0.1534 & 0.0896 \\ -0.1911 & 0.0691 & 0.1488 & -0.1212 \\ -0.1447 & -0.0046 & -0.1716 & 0.0724 \end{bmatrix} \rightarrow \mathbf{R}_x = \begin{bmatrix} 0.3979 & 0.0633 & 0.0294 & 0.0805 \\ 0.0633 & 0.5778 & -0.0858 & -0.0067 \\ 0.0294 & -0.0858 & 0.5221 & -0.0413 \\ 0.0805 & -0.0067 & -0.0413 & 0.6070 \\ -0.0199 & 0.0971 & 0.1534 & 0.0896 \\ -0.1911 & 0.0691 & 0.1488 & -0.1212 \\ -0.1447 & -0.0046 & -0.1716 & 0.0724 \end{bmatrix}, \quad (37)$$

$$\leftarrow \begin{bmatrix} -0.0199 & -0.1911 & -0.1447 \\ 0.0971 & 0.0691 & -0.0046 \\ 0.1534 & 0.1488 & -0.1716 \\ 0.0896 & -0.1212 & 0.0724 \\ 0.5633 & 0.1213 & 0.0399 \\ 0.1213 & 0.3171 & 0.1417 \\ 0.0399 & 0.1417 & 0.5541 \end{bmatrix}, \quad (38)$$

通过 Matlab 命令,容易得出矩阵束 $(\mathbf{R}_y, \mathbf{R}_x)$ 的广义特征值分别为 $\lambda_1 = 0.1021, \lambda_2 = 0.1612, \lambda_3 = 0.6464, \lambda_4 = 0.8352, \lambda_5 = 1.3525, \lambda_6 = 2.1360, \lambda_7 = 5.2276$; 然后利用算法(30)对由式(37)和(38)构成的矩阵束的前3个广义次成分进行提取. 算法的初始化参数设置如下: 加权矩阵 $\mathbf{A} = \text{diag}([3, 2, 1])$, 学习因子 $\eta = 0.2$, 算法的初始化状态矩阵随机产生. 所提出算法的仿真结果如图1所示, 该结果是100次独立仿真实验的平均值.

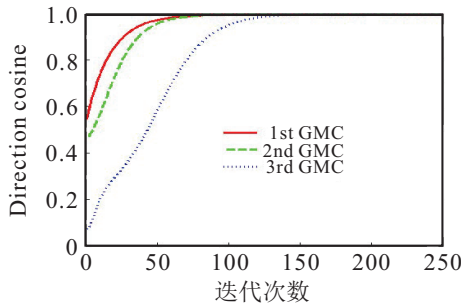


图1 多维广义次成分的提取结果

图1中实线代表着第1个广义次成分的方向余弦曲线;虚线为第2个广义次成分的曲线;点线为第3个广义次成分的方向余弦曲线. 由图1可得: 大约经过150次左右的迭代后, 3条方向余弦曲线都收敛到了1. 方向余弦代表着状态矩阵列向量的方向, 所以由图1可以得出结论: 所提出算法能够提取矩阵束的多维广义次成分.

5.2 算法性能对比实验

将所提出算法与 Ye 算法^[11]和 GMOX 算法^[12]进行对比. 5.1节已经对所提出算法多维广义次成分提取能力进行了验证, 且 Ye 算法和 GMOX 算法只能提取单维广义次成分, 因此本实验只针对单维广义次成分进行提取验证. 实验中输入数据由如下两式产生:

$$y(k) = \sqrt{2} \sin(0.62\pi k + \tau_1) + n_1(k), \quad (39)$$

$$x(k) = 10\sqrt{2} \sin(0.46\pi k + \tau_2) + 10\sqrt{2} \sin(0.74\pi k + \tau_3) + n_2(k). \quad (40)$$

其中: $\tau_i (i = 1, 2, 3)$ 称为初始相位, 服从区间 $[0, 2\pi]$ 内的一个均匀分布; $n_1(k)$ 和 $n_2(k)$ 是两个零均值的高

斯白噪声, 其方差分别为 $\sigma_1^2 = 0.01, \sigma_2^2 = 1$. 取方程(39)和(40)中8个连续的输出作为输入信号, 然后利用上述3个算法来提取式(39)和(40)产生的输入向量序列 $\{\mathbf{x}(k)\}$ 和 $\{\mathbf{y}(k)\}$ 的第1个广义次成分. 为了公平起见, 所有算法均采用相同的初始化参数: 初始状态向量为随机产生的, 学习因子 $\eta = 0.2$, 遗忘因子 $\alpha_1 = \alpha_2 = 0.998$. 仿真结果如图2所示, 该结果是100次独立仿真实验结果的平均值.

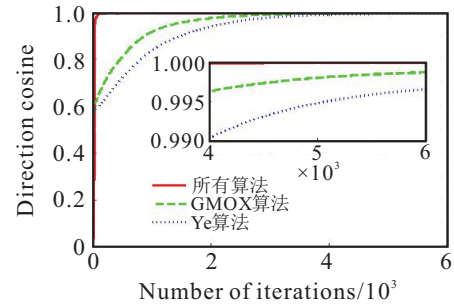


图2 3种算法对于广义次成分的提取结果

由图2可以看出, 经历了很少一些迭代运算后, 所提出算法的方向余弦曲线就已经收敛到1, 表明所提出算法具有从输入信号中提取广义次成分的能力. 通过与其他算法的对比可以发现, 所提出算法的收敛速度要远远快于其他两种算法, 将算法的收敛结果进行放大后可以发现, 所提出算法的收敛值与单位1最为接近, 即相比其他两个算法, 所提出算法具有更好的估计精度.

5.3 盲信号分离实验

广义次成分在很多领域均发挥着重要的应用. 本实验将通过盲信号分离问题对所提出算法的实际应用性进行验证. 考虑如下盲分离模型:

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{e}(k). \quad (41)$$

其中: $\mathbf{x}(k)$ 为 k 时刻的观测信号, $\mathbf{s}(k)$ 为未知的源信号, $\mathbf{e}(k)$ 为观测噪声, $\mathbf{A}$ 为混合信号矩阵. 盲信号分离问题可以表示为寻找一个分离矩阵 $\mathbf{W}$, 使得 n 维分离信号 $\mathbf{y} = \mathbf{W}^T \mathbf{x}$ 尽可能的相互独立.

在文献[17]中, Tomé 提出了一种基于 GED 的算法, 并证明了经过该方法获得的结果在大多数情况下要优于一般的盲信号分离算法, 而且该算法能够处理高维数据情况下盲信号分离问题. Tomé 算法首先是将混合信号 $\mathbf{x}(k)$ 通过一 FIR 滤波器并将获得滤波器的输出记为

$$\mathbf{z}(k) = \sum_{j=0}^l \eta(j) \mathbf{x}(k-j), \quad (42)$$

其中 $\eta(j)$ 为 FIR 滤波器系数. 令 $\mathbf{R}_x = E[\mathbf{x}(k)\mathbf{x}^T(k)]$

和 $\mathbf{R}_z = E[\mathbf{z}(k)\mathbf{z}^T(k)]$, 盲信号的分离矩阵 $\mathbf{W}$ 可以通过计算矩阵束 $(\mathbf{R}_z, \mathbf{R}_x)$ 的广义次成分来实现。

本实验将 ICALAB 中的 ABio7.mat 4 个信号作为源信号, 信号波形如图 3 所示。

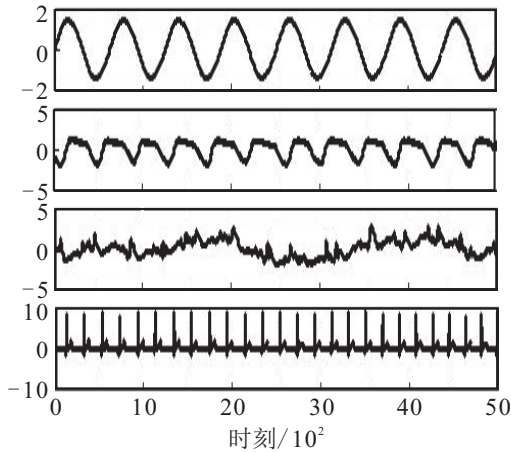


图 3 4 个源信号

本实验中混合矩阵随机产生, 有

$$\mathbf{A} = \begin{bmatrix} 2.7914 & -0.178 & -0.4945 & 0.3013 \\ 1.3225 & -1.7841 & -0.3669 & 0.446 \\ 0.0714 & -1.9163 & 0.4802 & -0.37 \\ -1.7396 & 0.1302 & 0.9249 & -0.4007 \end{bmatrix}. \quad (43)$$

观测噪声 $\mathbf{e}(k)$ 为零均值的高斯白噪声, 且其方差为 $10^{-5}\mathbf{I}$. 混合后的信号如图 4 所示。

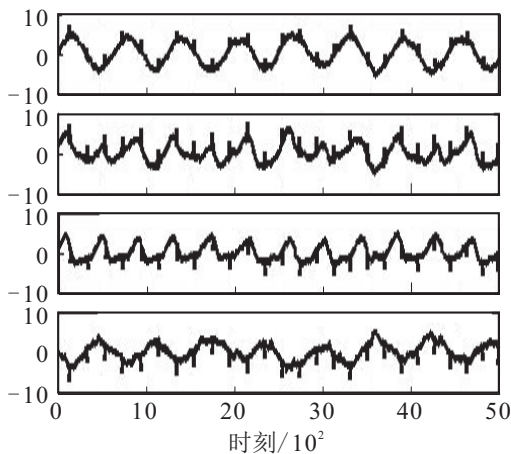


图 4 混合后的信号

利用 Tome 所提出算法对该盲分离问题进行求解. 选取 FIR 滤波器系数为 $\eta = [5 \ -10]^T$. 在计算过程中, 广义次成分分别由算法 (30) 和批处理 GED 算法计算. 为了定量衡量算法的分离结果, 这里分别计算源信号和分离信号之间的互相关系数. 图 5 和图 6 分别给出了两种算法获得的分离信号波形, 表 1 给出了两种算法的分离系数。

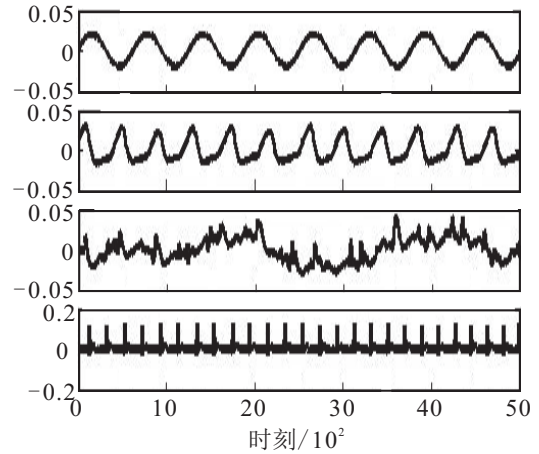


图 5 GED 获得的分离信号

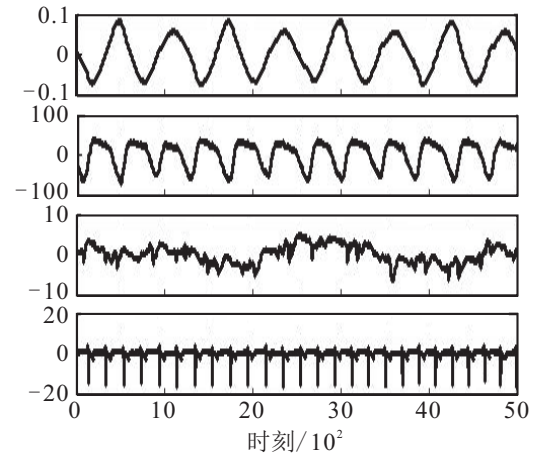


图 6 本文算法获得的分离信号

表 1 源信号和分离信号的互相关系数

方法	信号 1	信号 2	信号 3	信号 4
GED 算法	1.0000	0.9998	0.9997	0.9989
本文算法	0.9983	0.9978	0.9912	0.9984

由图 6 和图 3 的波形对比可得, 本文所提出算法分离获得的信号与源信号之间具有很强的相似性, 即所提出算法可用来解决盲分离问题. 由表 1 可看出, 虽然所提出算法的相似系数低于 GED 算法的结果, 但是所提出算法是一种迭代算法, 而 GED 算法是一种批处理算法. 由于本文算法可以实时在线, 在算法的实时性和计算复杂度方面更具有优势。

6 结 论

本文主要针对多维广义次成分提取算法进行研究. 首先基于 AMEX 准则, 提出了一种广义次成分提取准则, 并采用矩阵微分法对所提出准则的平稳点进行了分析, 结果表明: 当且仅当神经网络状态矩阵收敛到矩阵束的多维广义次成分时, 所提出信息准则取得全局极大值, 从而建立起了广义次成分与状态矩阵之间的关系. 基于梯度上升算法, 导出了一个自适应

广义次成分提取算法,并采用李雅普诺夫函数法对所提出算法的全局收敛性进行了证明.最后通过数值实验和实际应用验证了所提出算法的优点.

参考文献(References)

- [1] 徐保根, 谢伟, 万义和, 等. 一种基于广义特征向量的改进ESPRIT测向算法[J]. 电子科技大学学报, 2015, 44(2): 201-204.
(Xu B G, Xie W, Wan Y H, et al. An improved ESPRIT algorithm based on generalized eigenvectors[J]. J of University of Electronic Science and Technology of China, 2015, 44(2): 201-204.
- [2] 高迎彬, 孔祥玉, 胡昌华, 等. 一种广义主成分提取算法及其收敛性分析[J]. 电子与信息学报, 2016, 38(10): 2531-2537.
(Gao Y B, Kong X Y, Hu C H, et al. A generalized principal component extraction algorithm and its convergence analysis[J]. J of Electronics and Information Technology, 2016, 38(10): 2531-2537.
- [3] Batselier K, Wong N. Symmetric tensor decomposition by an iterative eigendecomposition algorithm[J]. J of Computational and Applied Mathematics, 2016, 308(3): 69-82.
- [4] Feng X, Kong X, Duan Z, et al. Adaptive generalized eigen-pairs extraction algorithms and their convergence analysis[J]. IEEE Trans on Signal Processing, 2016, 64(11): 2976-2989.
- [5] Wang Z, Li D M. Multiple-instance classification via generalized eigenvalue proximal SVM[J]. Advanced Materials Research, 2011, 143(32): 1235-1239.
- [6] Güvensen G M, Candan G, Koç S, et al. On generalized eigenvector space for target detection in reduced dimensions[C]. Radar Conf. Arlington, 2015: 1316-1321.
- [7] Karampatziakis N, Mineiro P. Discriminative features via generalized eigenvectors[C]. Proc of the 31th Int Conf on Machine Learning. Cambridge, 2014: 494-502.
- [8] Shougen W, Shuqin Z. An algorithm for $Ax = \lambda Bx$ with symmetric and positive-definite A and B [J]. SIAM J on Matrix Analysis and Applications, 1991, 12(2): 654-660.
- [9] Bunse-Gerstner A. An algorithm for the symmetric generalized eigenvalue problem[J]. Linear Algebra Application, 1984, 58(3): 43-68.
- [10] Liu L, Shao H, Nan D. Recurrent neural network model for computing largest and smallest generalized eigenvalue[J]. Neurocomputing, 2008, 71(3): 3589-3594.
- [11] Ye M, Liu Y, Wu H, et al. A few online algorithms for extracting minor generalized eigenvectors[C]. IEEE Int Joint Conf on Neural Networks. Hong Kong, 2008: 1714-1720.
- [12] Nguyen T D, Takahashi N, Yamada I. An adaptive extraction of generalized eigensubspace by using exact nested orthogonal complement structure[J]. Multidimensional Systems and Signal Processing, 2013, 24(7): 457-483.
- [13] Kong X, Hu C, Duan Z. Neural networks for principal component analysis[M]. Springer Singapore, 2017: 122-126.
- [14] Nguyen T D, Yamada I. Adaptive normalized quasi-newton algorithms for extraction of generalized eigen-pairs and their convergence analysis[J]. IEEE Trans on Signal Processing, 2013, 61(6): 1404-1418.
- [15] Ouyang S, Bao Z, Liao G, et al. Adaptive minor component extraction with modular structure[J]. IEEE Trans on Signal Processing, 2001, 49(9): 2127-2137.
- [16] Gao Y, Kong X, Zhang H, et al. A weighted information criterion for multiple minor components and its adaptive extraction algorithms[J]. Neural Networks, 2017, 89(3): 1-10.
- [17] Tomé A M. The generalized eigendecomposition approach to the blind source separation problem[J]. Digital Signal Processing, 2006, 16(2): 288-302.

(责任编辑: 孙艺红)