

自稳定的双目的特征对提取算法及其收敛性分析

冯晓伟^{1†}, 孔祥玉¹, 马红光², 安秋生³

(1. 火箭军工程大学 三系, 西安 710025; 2. 北京理工大学 珠海学院, 广东 珠海 519088;
3. 山西师范大学 数学与计算机科学学院, 山西 临汾 041004)

摘要: 提出一种自稳定的双目的算法用以提取信号自相关矩阵的特征对. 该算法可以通过仅仅改变一个符号实现主/次特征向量估计的转化, 并且可以通过估计的特征向量的模值信息估计对应的特征值, 从而实现特征对的提取. 基于确定性离散时间方法对所提出的算法进行收敛性分析, 并确定算法收敛的边界条件. 与已有算法对比的仿真实验验证了所提出算法的收敛性能.

关键词: 主成分分析; 次成分分析; 自稳定性; 确定性离散时间; 神经网络; 特征对

中图分类号: TP183 **文献标志码:** A

Unified self-stabilizing eigen-pairs extraction algorithm and its convergence analysis

FENG Xiao-wei^{1†}, KONG Xiang-yu¹, MA Hong-guang², AN Qiu-sheng³

(1. Department 3, Xi'an Research Institute of High Technology, Xi'an 710025, China; 2. Zhuhai College, Beijing Institute of Technology, Zhuhai 519088, China; 3. School of Mathematics and Computer Science, Shanxi Normal University, Linfen 041004, China)

Abstract: A self-stabilizing and unified algorithm is introduced to extract the eigen-pairs of a covariance matrix. The proposed algorithm is available for both principal component analysis(PCA) and minor component analysis(MCA) by simply altering the sign, and can estimate the corresponding eigenvalue from the norm of the estimated eigenvector, which can realize the eigen-pair extraction. The convergence analysis of the proposed algorithm based on the deterministic discrete time(DDT) approach is evaluated and the condition for convergence is also given. Finally, compared with existing algorithms, the convergence of the proposed algorithm is verified by the simulations.

Keywords: PCA; MCA; self-stabilizing property; DDT; neural networks; eigen-pair

0 引言

主成分分析(PCA)和次成分分析(MCA)已被广泛应用于信号处理领域^[1]. PCA主要用于抽取向量空间中对应于最大特征值的特征向量^[2], 目前已被广泛应用于如降维^[3]、图像压缩^[4]以及故障诊断^[5]等领域. 与PCA相反, MCA主要用于抽取向量空间中对应于最小特征值的特征向量^[6], 目前已被广泛应用于如最小均方误差^[7]、曲线及曲面拟合^[8]以及数字波束形成^[9]等领域. 传统的PCA/MCA算法是一种批处理方法, 具有较高计算复杂度且不能用于在线跟踪. 相比于传统算法, 基于神经网络的PCA/MCA算法提供了一种有效的在线特征提取方法^[10]. 早期仅有 Mathew 等^[11-12]提出了基于 Hopfield 网络的

层非线性神经网络算法. 然而, 在该算法中, 网络模型的选择必须基于具体问题, 并且为了选择合适收敛因子, 必须要计算期望矩阵的迹. 除此之外的神经网络特征提取算法都是基于单层神经网络的算法^[13]. Xu 等^[8]最早提出了基于 Hebbian 规则的神经网络算法. 到目前为止, 在 Xu 等的工作基础上, 文献 [13-18] 提出了大量算法; 文献 [19-20] 给出了关于神经网络算法的具体讨论; 文献 [21] 给出了关于神经网络算法的一些具体应用的综述. 近年来, 尽管神经网络方法已经被提出多年, 学者们仍不断致力于寻求性能更优的新算法^[22].

关于神经网络算法的分类: 就被提取成分而言, 可以分为 PCA 和 MCA 算法; 就提取维数而言, 可以分

收稿日期: 2016-03-17; 修回日期: 2016-10-29.

基金项目: 国家自然科学基金项目(61673387, 61374120, 11405267); 陕西省自然科学基金项目(2016JM6015).

作者简介: 冯晓伟(1986—), 男, 博士, 从事随机信号处理、故障诊断的研究; 孔祥玉(1967—), 男, 副教授, 博士生导师, 从事自适应信号处理、非线性系统建模与应用、故障诊断等研究.

†通讯作者. E-mail: xiaowei121416@163.com

为单维提取、多维提取和子空间跟踪算法;就多维信号提取方式而言,可以分为并行提取和序贯提取算法;就是否提取特征值而言,可以分为耦合和非耦合算法;就是否同时提取主成分(主子空间)和次成分(次子空间)而言,可以分为统一(双目的)和单目的算法;就是否具有自稳定特性而言,可以分为自稳定和非自稳定算法。

值得注意的是,近年来,有一些各具优势的算法,如双目的算法、自稳定算法和耦合算法等被相继提出来。双目的算法可以通过简单的符号转换,或者取原矩阵的逆矩阵实现主、次成分提取的转换,这十分有利于减小计算复杂度和实际应用中的硬件开销^[4]。理论上,任何一个PCA算法都可以通过某种转换机制得到一个对应的MCA算法^[23],反之亦然。然而,通过这种转换机制得到的算法往往具有不同的表现形式,因此,有必要寻求新的具有统一形式的双目的算法。近年来,文献[4, 15, 18, 24-26]提出了许多双目的算法。除了双目的算法,自稳定算法^[27-28]因具有良好的稳定性而被广泛研究。自稳定算法的权向量模值在迭代中会自动收敛到一个固定长度,这有利于算法保持稳定。目前,自稳定性已被认为是算法应当具备的一个重要性质,因为不具有自稳定性的算法都是潜在发散的^[13, 27]。可以发现,上述算法都仅仅停留在特征向量的提取上。实际上,特征值在许多信号处理应用中也具有十分重要的作用。为了解决这一问题,文献[22]提出了一套耦合算法,该算法通过使用一套耦合公式同时提取特征向量及其对应的特征值。

在众多算法中,双目的、自稳定以及耦合算法各具特点,而目前已有算法大都最多只具备上述3个特点中的两个。那么,是否有一种算法能够同时具备这3种特点,即它既是双目的又是自稳定的,并且还能提取特征值。为了实现这一目的,本文提出一种自稳定的双目的特征对提取算法,并分析算法的自稳定特性。分析表明,不论是PCA算法还是MCA算法都是自稳定的。采用确定性离散时间(DDT)方法对所提出算法进行收敛性分析,并确定所提出算法收敛的边界条件,通过对比仿真实验验证所提出算法在提取特征对上的有效性。

1 算法的提出

考虑一个具有如下输入输出关系的单层线性神经网络:

$$y(k) = \mathbf{w}^T(k)\mathbf{x}(k), \quad k = 0, 1, 2, \dots \quad (1)$$

其中: $y(k)$ 是神经网络的输出; $\mathbf{x}(k) \in \mathbf{R}^{n \times 1}$ 是神经网络的输入,它是一个零均值的随机过程向量; $\mathbf{w}(k) \in$

$\mathbf{R}^{n \times 1}$ 是神经网络的权向量, n 是神经网络的维数。定义 $\mathbf{C} = E[\mathbf{x}(k)\mathbf{x}^T(k)]$ 是输入信号 \mathbf{x} 的自相关矩阵,并定义 λ_i 和 \mathbf{v}_i ($i = 1, 2, \dots, n$) 是矩阵 \mathbf{C} 的特征值和特征向量。显然,矩阵 \mathbf{C} 的特征值分解可以表示为 $\mathbf{C} = \sum_{i=1}^n \lambda_i \mathbf{v}_i \mathbf{v}_i^T = \mathbf{L}\mathbf{L}^T$, 其中 $\mathbf{L} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ 且 $\mathbf{L} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]^T$ 。所有特征向量都是单位化、相互正交的,即 $\mathbf{v}_i^T \mathbf{v}_j = \delta_{ij}$ ($i, j = 1, 2, \dots, n$), 其中 δ_{ij} 是狄拉克函数。不妨假设特征向量按照 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 的顺序排列,从而满足 $(0 <) \lambda_1 < \lambda_2 < \dots < \lambda_n$ 。最大(小)的特征值及其对应的特征向量分别称为主(次)特征值和主(次)特征向量,而 $(\mathbf{v}_i, \lambda_i)$ 称为矩阵 \mathbf{C} 的第 i 个特征对。

提出一个新的双目的主/次成分分析算法,即

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \gamma\eta\{y(k)\mathbf{x}(k) - [\mathbf{w}^T(k)\mathbf{w}(k)]^\gamma \mathbf{w}(k)\}, \quad (2)$$

其中 $\eta \in (0, 1)$ 是学习因子。当取 $\gamma = 1$ 时,该算法为PCA算法(简称为算法1);当取 $\gamma = -1$ 时,该算法为MCA算法(简称为算法2)。注意到,算法2与文献[29]中的AMEX(adaptive minor component extraction)算法一致,因此AMEX算法可以看作是该双目的算法的一个特例。

在对算法进行自稳定性和收敛性分析时,一般是针对其DDT系统进行的。DDT系统是通过原离散系统应用条件期望算子而得到的。这里先给出DDT系统的形式,具体分析及讨论将在后面进行。算法的DDT系统为

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \gamma\eta\{\mathbf{C}\mathbf{w}(k) - [\mathbf{w}^T(k)\mathbf{w}(k)]^\gamma \mathbf{w}(k)\}. \quad (3)$$

2 算法的自稳定性分析

自稳定性是指在神经网络迭代过程中,权向量的模值是确定的,并且能够收敛到一个固定值,而与初始权向量的选择无关^[13]。由于缺乏自稳定特性的算法都是潜在发散的,自稳定性已经成为神经网络算法应当具备的一个十分重要的特性。因此,有必要对算法的自稳定性进行研究。算法的自稳定性可以通过其权向量的模值的变化规律研究,由式(3)可得

$$\frac{\|\mathbf{w}(k+1)\|^2}{\|\mathbf{w}(k)\|^2} = \frac{\mathbf{w}^T(k+1)\mathbf{w}(k+1)}{\mathbf{w}^T(k)\mathbf{w}(k)} = 1 + \frac{2\gamma\eta}{\mathbf{w}^T(k)\mathbf{w}(k)} \{\mathbf{w}^T(k)\mathbf{C}\mathbf{w}(k) - [\mathbf{w}^T(k)\mathbf{w}(k)]^{\gamma+1}\}. \quad (4)$$

考虑到 η 足够小,式(4)中忽略了关于 η 的高阶无穷小项。对于算法1,有

$$\frac{\|\mathbf{w}(k+1)\|^2}{\|\mathbf{w}(k)\|^2} =$$

$$1 + \frac{2\eta}{\mathbf{w}^T(k)\mathbf{w}(k)} \{ \mathbf{w}^T(k)\mathbf{C}\mathbf{w}(k) - [\mathbf{w}^T(k)\mathbf{w}(k)]^2 \} = \begin{cases} \text{Eq}_1 > 1, [\mathbf{w}^T(k)\mathbf{w}(k)]^2 < \mathbf{w}^T(k)\mathbf{C}\mathbf{w}(k); \\ \text{Eq}_1 = 1, [\mathbf{w}^T(k)\mathbf{w}(k)]^2 = \mathbf{w}^T(k)\mathbf{C}\mathbf{w}(k); \\ \text{Eq}_1 < 1, [\mathbf{w}^T(k)\mathbf{w}(k)]^2 > \mathbf{w}^T(k)\mathbf{C}\mathbf{w}(k). \end{cases} \quad (5)$$

这表明算法1具有自稳定性^[27,30]:当 $[\mathbf{w}^T(k)\mathbf{w}(k)]^2 < \mathbf{w}(k)^T\mathbf{C}\mathbf{w}(k)$ 时,算法权向量的模值在下一步迭代中增加,而当 $[\mathbf{w}^T(k)\mathbf{w}(k)]^2 > \mathbf{w}(k)^T\mathbf{C}\mathbf{w}(k)$ 时,在下一步迭代中减小.因此,在每一步迭代中,权向量的模值会趋向于 $\mathbf{w}^T(k)\mathbf{w}(k) = \sqrt{\mathbf{w}^T(k)\mathbf{C}\mathbf{w}(k)}$.

对于算法2,有

$$\frac{\|\mathbf{w}(k+1)\|^2}{\|\mathbf{w}(k)\|^2} = 1 - \frac{2\eta}{\mathbf{w}^T(k)\mathbf{w}(k)} \{ \mathbf{w}^T(k)\mathbf{C}\mathbf{w}(k) - 1 \} = \begin{cases} \text{Eq}_2 > 1, \mathbf{w}^T(k)\mathbf{C}\mathbf{w}(k) < 1; \\ \text{Eq}_2 = 1, \mathbf{w}^T(k)\mathbf{C}\mathbf{w}(k) = 1; \\ \text{Eq}_2 < 1, \mathbf{w}^T(k)\mathbf{C}\mathbf{w}(k) > 1. \end{cases} \quad (6)$$

这表明算法2也具有自稳定特性,算法权向量的 \mathbf{C} 范数(定义为 $\|\mathbf{w}\|_C = \sqrt{\mathbf{w}^T\mathbf{C}\mathbf{w}}$)在迭代过程中会趋向于 $\|\mathbf{w}(k)\|_C = 1$.

3 收敛性分析

为了分析算法的收敛特性,学者们先后提出了多种分析方法.然而,这些方法都存在一些问题,例如一些算法被证明是收敛的,而实际应用却表明这些算法是发散的.后来,Zufiria^[31]提出了DDT分析方法分析算法的收敛性能,文献[18]指出,从实际应用角度出发,DDT方法更适合于算法性能分析.文献[29]虽然提出了AMEX算法(即算法2),但并没有对其进行收敛性分析.后来,文献[32]给出了AMEX算法的详细收敛性分析.因此,此处只针对算法1进行详细的收敛性分析.关于算法2的收敛性分析,可参见文献[32].算法1的DDT系统为

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \eta[\mathbf{C}\mathbf{w}(k) - \mathbf{w}^T(k)\mathbf{w}(k)\mathbf{w}(k)]. \quad (7)$$

考虑到向量集 $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ 是空间 \mathbf{R}^n 的一组正交基,故空间 \mathbf{R}^n 中的任意向量都可以表示为矩阵 \mathbf{C} 的所有特征向量的加权合成.于是,对于所有 $k \geq 0$, $\mathbf{w}(k)$ 和 $\mathbf{C}\mathbf{w}(k)$ 可以分别表示为

$$\mathbf{w}(k) = \sum_{i=1}^n z_i(k)\mathbf{v}_i, \quad (8)$$

$$\mathbf{C}\mathbf{w}(k) = \sum_{i=1}^n \lambda_i z_i(k)\mathbf{v}_i, \quad (9)$$

其中 $z_i(k) = \mathbf{w}^T(k)\mathbf{v}_i$ 为权向量 $\mathbf{w}(k)$ 在特征向量 \mathbf{v}_i 方向上的投影($i = 1, 2, \dots, n$).

将式(8)和(9)代入(7),可得

$$z_i(k+1) = \{1 + \eta[\lambda_i - \mathbf{w}^T(k)\mathbf{w}(k)]\}z_i(k), \quad (10)$$

其中 $i = 1, 2, \dots, n$.下面证明在一些约束条件下,算法1收敛到矩阵 \mathbf{C} 的主特征向量的方向上来,于是存在如下定理.

定理1 假定 $\eta < 0.3$ 和 $\eta\lambda_n < 0.35$.如果初始权向量有界并满足 $\mathbf{w}^T(0)\mathbf{v}_n \neq 0$,则对于所有的 $k \geq 0$,权向量的模值满足 $\|\mathbf{w}(k)\| \leq 1 + \eta(\lambda_n - 1)$.

证明 由式(8)和(10)可得

$$\begin{aligned} \|\mathbf{w}(k+1)\|^2 &= \sum_{i=1}^n z_i^2(k+1) = \\ &= \sum_{i=1}^n \{1 + \eta[\lambda_i - \mathbf{w}^T(k)\mathbf{w}(k)]\}^2 z_i^2(k) \leq \\ &= \{1 + \eta[\lambda_n - \mathbf{w}^T(k)\mathbf{w}(k)]\}^2 \sum_{i=1}^n z_i^2(k) = \\ &= [1 + \eta(\lambda_n - \|\mathbf{w}(k)\|^2)]^2 \|\mathbf{w}(k)\|^2. \end{aligned} \quad (11)$$

由此可以推出

$$\|\mathbf{w}(k+1)\|^2 \leq [1 + \eta(\lambda_n - \|\mathbf{w}(k)\|^2)]^2 \|\mathbf{w}(k)\|^2. \quad (12)$$

在区间 $[0,1]$ 上,定义一个可微方程,即

$$f(s) = [1 + \eta(\lambda_n - s)]^2 s, \quad (13)$$

其中 $s = \|\mathbf{w}(k)\|^2$,因此 $f(s) = \|\mathbf{w}(k+1)\|^2$.从式(13)可以得到对于所有的 $0 < s < 1$,有

$$\dot{f}(s) = (1 + \eta\lambda_n - \eta s)(1 + \eta\lambda_n - 3\eta s). \quad (14)$$

可以求得函数 $f(s) = 0$ 的根为

$$s_1 = \eta^{-1} + \lambda_n, s_2 = \frac{1}{3}(\eta^{-1} + \lambda_n). \quad (15)$$

显然,当 $\eta < 0.3$ 时,有 $s_1 > s_2 = \frac{1 + \eta s}{3\eta} > 1$.这意味着在区间 $[0,1]$ 内,函数 $f(s)$ 是单调增加的.这样,对于所有 $0 < s < 1$,有

$$f(s) \leq f(1) = [1 + \eta(\lambda_n - 1)]^2. \quad (16)$$

另外,当 $\eta < 0.3$ 时,有 $1 + \eta(\lambda_n - 1) = (1 - \eta) + \eta\lambda_n > 0$,从而有 $\|\mathbf{w}(k)\| \leq 1 + \eta(\lambda_n - 1)$ 对所有 $k \geq 0$ 成立.□

定理1证明了权向量的有界性.接下来证明在一些温和条件下,权向量将会收敛到 $\lim_{k \rightarrow \infty} \mathbf{w}(k) = \pm\sqrt{\lambda_n}\mathbf{v}_n$.为此,首先需要证明几个引理.

引理1 假定有 $\eta < 0.3$ 和 $\eta\lambda_n < 0.35$.如果初始权向量有界并满足 $\mathbf{w}^T(0)\mathbf{v}_n \neq 0$,则 $1 + \eta[\lambda_i - \mathbf{w}^T(k)\mathbf{w}(k)] > 0$ ($i = 1, 2, \dots, n$)对于所有的 $k > 0$ 都

成立.

证明 由定理1可知

$$\begin{aligned} 1 + \eta[\lambda_i - \mathbf{w}^T(k)\mathbf{w}(k)] &> 1 - \eta\mathbf{w}^T(k)\mathbf{w}(k) \geq \\ 1 - \eta[1 + \eta(\lambda_n - 1)]^2 &> 1 - \eta(1 + \eta\lambda_n)^2 > \\ 1 - 0.3 \times (1 + 0.35)^2 &= 0.4533 > 1. \end{aligned} \quad (17)$$

由此引理1得证. \square

引理1表明, $z_i(k)$ 在迭代中对于所有的 $k > 0$ 都不变号. 因此, 当 $\mathbf{w}^T(0)\mathbf{v}_n \neq 0$ 时, $\mathbf{w}^T(k)\mathbf{v}_n \neq 0$ 对于所有 $k > 0$ 都满足. 由式(8)可得

$$\mathbf{w}(k) = \sum_{i=1}^{n-1} z_i(k)\mathbf{v}_i + z_n(k)\mathbf{v}_n. \quad (18)$$

显然, 权向量 $\mathbf{w}(k)$ 的收敛性完全由 $z_i(k)$ ($i = 1, 2, \dots, n$) 决定. 鉴于此, 给出如下两个引理.

引理2 假定 $\eta < 0.3$ 和 $\eta\lambda_n < 0.35$. 如果初始权向量有界并满足 $\mathbf{w}^T(0)\mathbf{v}_n \neq 0$, 则 $\lim_{k \rightarrow \infty} z_i(k) = 0$, 其中 $i = 1, 2, \dots, n - 1$.

证明 对于 $i = 1, 2, \dots, n - 1$, 有

$$\begin{aligned} \left[\frac{z_i(k+1)}{z_n(k+1)} \right]^2 &= \\ \left[\frac{1 + \eta[\lambda_i - \mathbf{w}^T(k)\mathbf{w}(k)]}{1 + \eta[\lambda_n - \mathbf{w}^T(k)\mathbf{w}(k)]} \right]^2 \frac{z_i^2(k)}{z_n^2(k)} &= \\ \left[1 - \frac{\eta(\lambda_n - \lambda_i)}{1 + \eta[\lambda_n - \mathbf{w}^T(k)\mathbf{w}(k)]} \right]^2 \frac{z_i^2(k)}{z_n^2(k)}. \end{aligned} \quad (19)$$

定义 $\sigma_k = \frac{\eta(\lambda_n - \lambda_i)}{1 + \eta[\lambda_n - \mathbf{w}^T(k)\mathbf{w}(k)]}$, 显然有 $\eta(\lambda_n - \lambda_i) > 0$, 且有

$$\begin{aligned} 1 + \eta[\lambda_n - \mathbf{w}^T(k)\mathbf{w}(k)] &> 1 - \eta\mathbf{w}^T(k)\mathbf{w}(k) \geq \\ 1 - \eta[1 + \eta(\lambda_n - 1)]^2 &> 0, \end{aligned} \quad (20)$$

即 σ_k 的分子分母都为正数, 因此有 $\sigma_k > 0$, 即 $1 - \sigma_k < 1$. 另外, 有

$$\begin{aligned} \sigma_k &= \frac{\eta(\lambda_n - \lambda_i)}{1 + \eta[\lambda_n - \mathbf{w}^T(k)\mathbf{w}(k)]} < \frac{\eta\lambda_n}{1 - \eta\mathbf{w}^T(k)\mathbf{w}(k)} \leq \\ \frac{\eta\lambda_n}{1 - \eta[1 + \eta(\lambda_n - 1)]^2} &< \frac{\eta\lambda_n}{1 - \eta(1 + \eta\lambda_n)^2} < \\ \frac{0.35}{1 - 0.3[1 + 0.35]^2} &= 0.77 < 1. \end{aligned} \quad (21)$$

式(21)意味着 $1 - \sigma_k > 0$. 由式(20)和(21)可知 $0 < (1 - \sigma_k)^2 < 1$. 定义一个常数 $\theta = \max\{(1 - \sigma_k)^2\}$, 显然有 $0 < \theta < 1$. 因为 $z_i(0)$ 和 $z_n(0)$ 有界, 则

$$\begin{aligned} \left[\frac{z_i(k+1)}{z_n(k+1)} \right]^2 &\leq \theta \frac{z_i^2(k)}{z_n^2(k)} \leq \dots \leq \\ \theta^{k+1} \frac{z_i^2(0)}{z_n^2(0)} &\rightarrow 0 \text{ as } k \rightarrow \infty. \end{aligned} \quad (22)$$

已经证明, $\mathbf{w}(k)$ 对于所有 $k \geq 0$ 有界, 这也意味着 $z_n(k)$ 对于所有 $k \geq 0$ 必然有界, 从而有

$$\lim_{k \rightarrow \infty} z_i(k) = 0, \quad i = 1, 2, \dots, n - 1. \quad (23)$$

由此引理2得证. \square

引理3 假定有 $\eta < 0.3$ 和 $\eta\lambda_n < 0.35$. 如果初始权向量有界并且满足 $\mathbf{w}^T(0)\mathbf{v}_n \neq 0$, 则 $\lim_{k \rightarrow \infty} z_n(k) = \pm\sqrt{\lambda_n}$.

证明 由引理2可知, $\mathbf{w}(k)$ 收敛到主特征向量 \mathbf{v}_n 对应的方向. 现假设 $\mathbf{w}(k)$ 在某一时刻 k_0 已经收敛 \mathbf{v}_n 的方向, 即当 $k \geq k_0$ 时有 $\mathbf{w}(k) = z_n(k)\mathbf{v}_n$. 由式(10)可得

$$z_n(k+1) = \{1 + \eta[\lambda_n - z_n^2(k)]\}z_n(k). \quad (24)$$

由式(24)可知

$$\begin{aligned} z_n(k+1) - \sqrt{\lambda_n} &= \\ \{1 + \eta[\lambda_n - z_n^2(k)]\}z_n(k) - \sqrt{\lambda_n} &= \\ z_n(k) - \sqrt{\lambda_n} - & \\ \eta z_n(k)[z_n(k) + \sqrt{\lambda_n}][z_n(k) - \sqrt{\lambda_n}] &= \\ [z_n(k) - \sqrt{\lambda_n}]\{1 - \eta z_n(k)[z_n(k) + \sqrt{\lambda_n}]\}. \end{aligned} \quad (25)$$

当 $\eta < 0.3$ 和 $\eta\lambda_n < 0.35$ 时, 对于所有 $k \geq k_0$, 有

$$\begin{aligned} 1 - \eta z_n(k)[z_n(k) + \sqrt{\lambda_n}] &> \\ 1 - \eta[1 + \eta(\lambda_n - 1)][1 + \eta(\lambda_n - 1) + \sqrt{\lambda_n}] &> \\ 1 - \eta(1 + \eta\lambda_n)(1 + \eta\lambda_n + \sqrt{\lambda_n}) &= \\ 1 - (1 + \eta\lambda_n)(\eta + \eta \times \eta\lambda_n + \sqrt{\eta}\sqrt{\eta\lambda_n}) &> \\ 1 - (1 + 0.35)(0.3 + 0.3 \times 0.35 + \sqrt{0.3}\sqrt{0.35}) &= \\ 0.0158 &> 0. \end{aligned} \quad (26)$$

定义 $\xi = |1 - \eta z_n(k)[z_n(k) + \sqrt{\lambda_n}]|$, 显然有 $0 < \xi < 1$. 当 $k \geq k_0$ 时, 由式(25)和(26)可得

$$|z_n(k+1) - \sqrt{\lambda_n}| \leq \xi |z_n(k+1) - \sqrt{\lambda_n}| \leq \dots \leq \xi^{k+1} |z_n(0) - \sqrt{\lambda_n}|. \quad (27)$$

定义 $\psi = -\ln \xi$ 以及 $\Pi_1 = |z_n(0) - \sqrt{\lambda_n}|$, 并代入式(27), 可得

$$|z_n(k+1) - \sqrt{\lambda_n}| \leq (k+1)\Pi_1 e^{-\psi(k+1)}. \quad (28)$$

对于任意正实数 $\varepsilon > 0$, 必然存在一个常数 $K > 1$, 使得

$$\frac{\Pi_2 K e^{-\psi K}}{(1 - e^{-\psi})^2} < \varepsilon. \quad (29)$$

对于任意 $k_1 > k_2 > K$, 由式(10)和(28)可得

$$|z_n(k_1) - z_n(k_2)| = \left| \sum_{r=k_2}^{k_1-1} [z_n(r+1) - z_n(r)] \right| \leq$$

$$\begin{aligned}
& \left| \sum_{r=k_2}^{k_1-1} \{ \eta z_n(r) [\lambda_n - z_n^2(r)] \} \right| \leq \\
& \sum_{r=k_2}^{k_1-1} [\eta |z_n(r)| \times |z_n(r) + \sqrt{\lambda_n}| \times |z_n(r) - \sqrt{\lambda_n}|] \leq \\
& \eta [1 + \eta(\lambda_n - 1)] [1 + \eta(\lambda_n - 1) + \sqrt{\lambda_n}] \times \\
& \sum_{r=k_2}^{k_1-1} |z_n(r) - \sqrt{\lambda_n}| \leq \\
& \Pi_2 \sum_{r=k_2}^{k_1-1} r e^{-\psi r} \leq \Pi_2 \sum_{r=k_2}^{\infty} r e^{-\psi r} \leq \\
& \Pi_2 K e^{-\psi K} \sum_{r=0}^{\infty} r (e^{-\psi})^{r-1} \leq \frac{\Pi_2 K e^{-\psi K}}{(1 - e^{-\psi})^2} < \varepsilon. \quad (30)
\end{aligned}$$

其中 $\Pi_2 = \eta[1 + \eta(\lambda_n - 1)][1 + \eta(\lambda_n - 1) + \sqrt{\lambda_n}] \times |z_n(0) - \sqrt{\lambda_n}|$. 这意味着 $\{z_n(k)\}$ 是一个 Cauchy 序列. 根据 Cauchy 收敛原理, 必然存在一个常数 z^* 使得

$$\lim_{k \rightarrow \infty} z_n(k) = z^*.$$

由式(18)可知 $\lim_{k \rightarrow \infty} \mathbf{w}(k) = z^* \mathbf{v}_n$. 基于式(7)的自稳定性可得 $\lim_{k \rightarrow \infty} \mathbf{w}(k+1)/\mathbf{w}(k) = 1$, 从而有 $\lim_{k \rightarrow \infty} z_n(k+1)/z_n(k) = 1$. 由式(10)可推出 $1 = 1 + \eta(\lambda_n - z^{*2})$, 从而有 $z^* = \pm \sqrt{\lambda_n}$. \square

基于引理2和引理3可得如下定理.

定理2 假定有 $\eta < 0.3$ 和 $\eta\lambda_n < 0.35$. 如果初始权向量有界并且满足 $\mathbf{w}^T(0)\mathbf{v}_n \neq 0$, 则 $\lim_{k \rightarrow \infty} \mathbf{w}(k) = \pm \sqrt{\lambda_n} \mathbf{v}_n$.

证明 由式(18)、引理2和引理3可知

$$\begin{aligned}
\lim_{k \rightarrow \infty} \mathbf{w}(k) &= \lim_{k \rightarrow \infty} \left[\sum_{i=1}^{n-1} z_i(k) \mathbf{v}_i + z_n(k) \mathbf{v}_n \right] = \\
& \lim_{k \rightarrow \infty} \left[\sum_{i=1}^{n-1} z_i(k) \mathbf{v}_i \right] + \lim_{k \rightarrow \infty} [z_n(k) \mathbf{v}_n] = \\
& \pm \sqrt{\lambda_n} \mathbf{v}_n. \quad (31)
\end{aligned}$$

由此定理2得证. \square

至此, 算法1的收敛性分析完成.

推论1 由定理2可以得到 $\lim_{k \rightarrow \infty} \mathbf{w}^T(k)\mathbf{w}(k) = \lambda_n \mathbf{v}_n^T \mathbf{v}_n = \lambda_n$, 这意味着可以从算法1的权向量的模值信息中提取主特征值, 即

$$\lambda(k+1) = \mathbf{w}^T(k+1)\mathbf{w}(k+1). \quad (32)$$

推论2 由文献[32]可知, 对于算法2, 如果初始权向量有界并满足 $\mathbf{w}^T(0)\mathbf{v}_1 \neq 0$, 则当 $\eta < 0.25$ 和 $\eta\lambda_1 < 0.25$ 时有 $\lim_{k \rightarrow \infty} \mathbf{w}(k) = \pm \lambda_1^{-\frac{1}{2}} \mathbf{v}_1$, 于是可以推出 $\lim_{k \rightarrow \infty} \mathbf{w}^T(k)\mathbf{w}(k) = \lambda_1^{-1} \mathbf{v}_1^T \mathbf{v}_1 = \lambda_1^{-1}$, 这意味着可以从算法2的权向量的模值信息中提取次特征值, 即

$$\lambda(k+1) = [\mathbf{w}^T(k+1)\mathbf{w}(k+1)]^{-1}. \quad (33)$$

值得注意的是, 文献[32]只给出了算法2的收敛性分析, 但没有指出可以从算法权向量模值中提取特征值这一结论.

推论3 由推论1和推论2可知, 主、次特征值提取公式可以写成统一形式, 即

$$\lambda(k+1) = [\mathbf{w}^T(k+1)\mathbf{w}(k+1)]^\gamma. \quad (34)$$

考虑到 $\mathbf{w}^T(k)\mathbf{w}(k)$ 在计算 $\mathbf{w}(k+1)$ 时已经被计算过, 于是改写为下式可以减小计算复杂度:

$$\lambda(k+1) = [\mathbf{w}^T(k)\mathbf{w}(k)]^\gamma. \quad (35)$$

算法1中要求 $\eta\lambda_n < 0.35$, 而算法2中要求 $\eta\lambda_1 < 0.25$. 在实际应用中, 由于特征值是未知的, 满足这一条件存在一定困难. 然而, 不难发现学习因子的选择其实是与最大特征值的上界有关系. 文献[30]指出, 尽管最大特征值不能被提前确定, 但是它的上界是可以被估计的. 因此, 选择合适的学习因子在实际操作中是可行的. 另外, 算法1要求 $\mathbf{w}^T(0)\mathbf{v}_n \neq 0$, 算法2要求 $\mathbf{w}^T(0)\mathbf{v}_1 \neq 0$, 在实际应用中, 由于初始向量一般是随机产生的, 很容易满足这一条件.

4 仿真实验

这里通过一个仿真实验说明所提出的算法在提取主、次特征对的能力, 并与以下3种算法进行比较: 1) ALA (adaptive learning algorithm), 文献[22]中提出的耦合PCA算法; 2) 文献[15]中提出的双目的算法, 将其PCA和MCA算法分别命名为pPCA和pMCA算法; 3) 文献[18]中提出的双目的算法, 将其PCA和MCA算法分别命名为kPCA和kMCA算法.

实验使用算法1、ALA算法、pPCA算法和kPCA算法提取信号的主特征向量, 并使用算法1和ALA算法提取主特征值; 使用算法2、pMCA算法和kMCA算法提取信号的次特征向量, 使用算法2提取信号的次特征值. 上述算法都提取了一个时间序列的主、次成分, 该时间序列由下式产生:

$$\mathbf{x}(k) = \mathbf{B} \times \mathbf{y}(k). \quad (36)$$

其中: $\mathbf{B} \in \mathbf{R}^{10 \times 10}$ 是一个方差为0.1的高斯随机矩阵, $\mathbf{y} \in \mathbf{R}^{10 \times 1}$ 是随机产生的高斯信号. 在每一步迭代中都计算了权向量的模值及算法1、ALA算法提取的主特征值和算法2提取的次特征值. 另外, 为了说明算法中权向量的对特征向量在方向上的估计精度, 在每一步迭代中还计算了权向量与主/次特征向量的方向余弦, 即

$$\psi(k) = \frac{|\mathbf{w}(k)\mathbf{v}|}{\|\mathbf{w}(k)\| \times \|\mathbf{v}\|}. \quad (37)$$

其中: $\mathbf{v} = \mathbf{v}_1$ 用于MCA算法, $\mathbf{v} = \mathbf{v}_n$ 用于PCA算

法. 显然, 如果 $\psi(k)$ 收敛到 1, 则说明权向量收敛到主/次特征向量的方向. 实验中总共进行了 $k_{\max} = 10\ 000$ 次迭代. 对于所有算法, 设置学习因子 η 从 0.1 指数减小到 0.001, 实验结果见图 1~图 4.

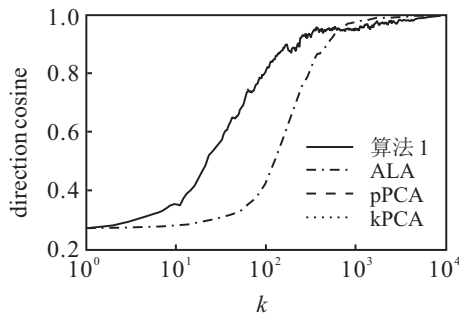


图 1 PCA 算法权向量方向余弦收敛情况

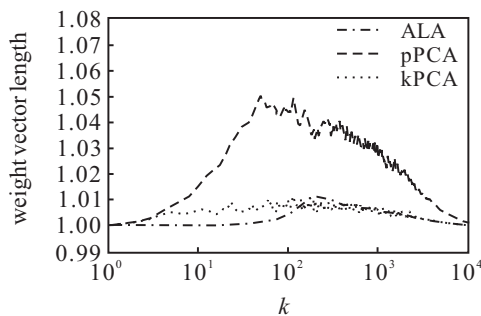


图 2 PCA 算法权向量模值收敛情况

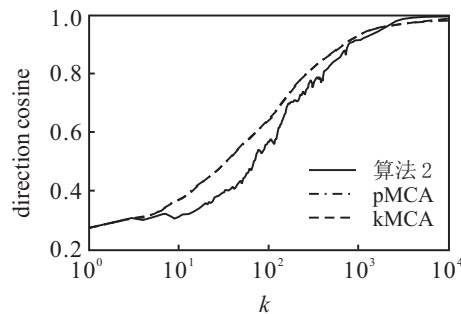


图 3 MCA 算法权向量方向余弦收敛情况

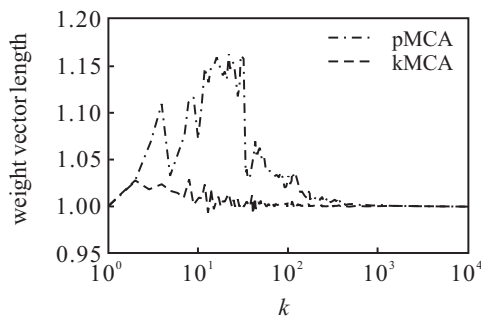


图 4 MCA 算法权向量模值收敛情况

图 1 显示了所有 PCA 算法权向量和主特征向量的方向余弦的收敛情况, 图 1 中除 ALA 算法外其他算法的曲线几乎重合. 由图 1 可知, 所有算法的方向余弦都收敛到 1, 表明这些算法都可以有效提取主特征向量, 并且具有相似的收敛速度.

图 2 显示了除算法 1 外的其他 PCA 算法的权向

量模值收敛情况, 可以发现, 这些算法的权向量模值都收敛到 1.

图 3 显示了所有 MCA 算法权向量和次特征向量的方向余弦的收敛情况, 图 3 中 pMCA 与 kMCA 算法的曲线重合. 由图 3 可知, 所有算法的方向余弦都收敛到 1, 这表明所有算法都可以有效提取主特征向量, 并且具有相似的收敛速度. 另外, 算法 2 的曲线收敛到更接近于 1 的位置, 说明算法 2 的估计精度稍好于其他算法.

图 4 显示了 pMCA 和 kMCA 算法的权向量模值收敛情况, 可以发现, 这些算法的权向量模值都收敛到 1.

图 2 和图 4 没有给出算法 1 和算法 2 的权向量模值收敛情况, 这是因为算法 1 和算法 2 的权向量模值分别收敛到与主、次特征值相关的长度, 从而可以分别从中提取主、次特征值信息. 图 5 显示了算法 1 和 ALA 算法提取主特征值的估计偏差和算法 2 提取次特征值的估计偏差, 该偏差定义为

$$\xi(k) = |\lambda(k) - \lambda|. \quad (38)$$

其中: $\lambda = \lambda_1$ 用于 MCA 算法, $\lambda = \lambda_n$ 用于 PCA 算法. 显然, 该偏差越小说明特征值估计越准确. 由图 5 可知, 所有算法都能较好估计主/次特征值.

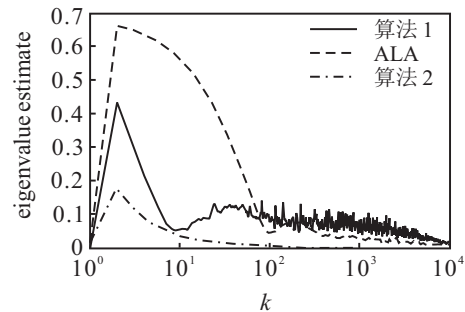


图 5 算法 1、算法 2 和 ALA 算法的特征值估计偏差

5 结 论

本文提出了一种自稳定的主、次特征对提取算法, 并且特征值提取不会增加计算复杂度. 性能分析表明, 所提出算法是自稳定的. 另外, 通过 DDT 方法给出了所提出算法的收敛性分析及收敛因子的选取边界条件. 通过一个对比仿真实验验证了所提出算法对主、次特征对提取的有效性. 就仿真结果而言, 该算法在收敛速度上与其他算法类似, 但不同于已有算法的是, 新算法同时集成了自稳定、双目的和特征对提取 3 个特点. 自稳定性保证了算法在收敛过程中的稳定性, 双目的算法可以同时提取主、次特征信息, 而特征对提取算法则可以同时提取特征值和特征向量, 这有利于一些需要特征值信息的实际工程应用.

参考文献(References)

- [1] Lai Z, Jin Z, Yang J, et al. Dynamic transition embedding for image feature extraction and recognition[J]. *Neural Computing & Applications*, 2012, 21(8): 1905-1915.
- [2] Abdi H, Williams L J. *Principal component analysis*[J]. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2010, 2(4): 433-459.
- [3] Kambhatla N, Leen T K. Dimension reduction by local principal component analysis[J]. *Neural Computation*, 1997, 9(7): 1493-1516.
- [4] Kong X Y, Hu C H, Han C Z. A dual purpose principal and minor subspace gradient flow[J]. *IEEE Trans on Signal Processing*, 2012, 60(1): 197-210.
- [5] Harmouche J, Delpha C. Incipient fault detection and diagnosis based on kullback-leibler divergence using principal component analysis: Part I[J]. *Signal Processing*, 2014, 94(1): 278-287.
- [6] Luo F L, Unbehauen R, Cichocki A. A minor component analysis algorithm[J]. *Neural Networks*, 1997, 10(2): 291-297.
- [7] Gao K, Ahmad M O, Swamy M. Learning algorithm for total least-squares adaptive signal processing[J]. *Electronics Letters*, 1992, 28(4): 430-432.
- [8] Xu L, Oja E, Suen C Y. Modified Hebbian learning for curve and surface fitting[J]. *Neural Networks*, 1992, 5(3): 441-457.
- [9] Griffiths J. Adaptive array processing: A tutorial[J]. *IEE Proc H - Microwaves, Optics and Antennas*, 1983, 130(1): 1-3.
- [10] Lü J C, Tan K K, Yi Z, et al. A family of fuzzy learning algorithms for robust principal component analysis neural networks[J]. *IEEE Trans on Fuzzy Systems*, 2010, 18(1): 217-226.
- [11] Mathew G, Reddy V U. Development and analysis of a neural network approach to pisarenko's harmonic retrieval method[J]. *IEEE Trans on Signal Processing*, 1994, 42(3): 663-667.
- [12] Mathew G, Reddy V U. Orthogonal eigensubspace estimation using neural networks[J]. *IEEE Trans on Signal Processing*, 1994, 42(7): 1803-1811.
- [13] Cirrincione G, Cirrincione M, Héroult J, et al. The MCA EXIN neuron for the minor component analysis[J]. *IEEE Trans on Neural Networks*, 2002, 13(1): 160-187.
- [14] Luo F L, Unbehauen R. A generalized learning algorithm of minor component[C]. 1997 IEEE Int Conf on Acoustics, Speech, and Signal Processing(ICASSP-97). Munich: IEEE Computer Society Press, 1997: 3229-3232.
- [15] Peng D, Yi Z, Xiang Y. A unified learning algorithm to extract principal and minor components[J]. *Digital Signal Processing*, 2009, 19(4): 640-649.
- [16] Ouyang S, Bao Z, Liao G S. Adaptive minor component extraction with modular structure[J]. *IEEE Trans on Signal Processing*, 2001, 49(9): 2127-2137.
- [17] Ye M, Yi Z, Lü J. A globally convergent learning algorithm for PCA neural networks[J]. *Neural Computing & Applications*, 2005, 14(1): 18-24.
- [18] Kong X, Hu C, Ma H. A unified self-stabilizing neural network algorithm for principal and minor components extraction[J]. *IEEE Trans on Neural Networks and Learning Systems*, 2012, 23(2): 185-198.
- [19] Diamantaras K I, Kung S Y. *Principal component neural networks: Theory and applications*[M]. New York: John Wiley & Sons, Inc, 1996: 74-75.
- [20] Du K L, Swamy M. *Neural networks and statistical learning*[M]. London: Springer, 2014: 355-405.
- [21] Qiu J, Wang H, Lu J, et al. Neural network implementations for PCA and its extensions[J]. *ISRN Artificial Intelligence*, 2012(4): 1-19.
- [22] Möller R, Köries A. Coupled principal component analysis[J]. *IEEE Trans on Neural Networks*, 2004, 15(1): 214-222.
- [23] Chen T, Amari S I, Lin Q. A unified algorithm for principal and minor components extraction[J]. *Neural Networks*, 1998, 11(3): 385-390.
- [24] Chen T, Amari S I. Unified stabilization approach to principal and minor components extraction algorithms[J]. *Neural Networks*, 2001, 14(10): 1377-1387.
- [25] Manton J H, Helmke U, Mareels I M. A dual purpose principal and minor component flow[J]. *Systems & Control Letters*, 2005, 54(8): 759-769.
- [26] Hasan M. Self-normalizing dual systems for minor and principal component extraction[C]. *IEEE Int Conf on Acoustics, Speech and Signal Processing*. Honolulu: IEEE, 2007: IV885-IV888.
- [27] Möller R. A self-stabilizing learning rule for minor component analysis[J]. *Int J of Neural Systems*, 2004, 14(1): 1-8.
- [28] Ye M, Fan X Q, Li X. A class of self-stabilizing mca learning algorithms[J]. *IEEE Trans on Neural Networks*, 2006, 17(6): 1634-1638.
- [29] Ouyang S, Bao Z, Liao G. Adaptive step-size minor component extraction algorithm[J]. *Electronics Letters*, 1999, 35(6): 443-444.
- [30] Feng D Z, Zheng W X, Jia Y. Neural network learning algorithms for tracking minor subspace in high-dimensional data stream[J]. *IEEE Trans on Neural Networks*, 2005, 16(3): 513-521.
- [31] Zufiria P J. On the discrete-time dynamics of the basic hebbian neural network node[J]. *IEEE Trans on Neural Networks*, 2002, 13(6): 1342-1352.
- [32] Peng D Z, Zhang Y. Global convergence of an adaptive minor component extraction algorithm[J]. *Chaos, Solitons & Fractals*, 2008, 35(3): 550-561.