

一种广义主成分提取算法及其收敛性分析

高迎彬 孔祥玉* 胡昌华 张会会 侯立安
(第二炮兵工程大学 西安 710025)

摘要: 广义主成分分析在现代信号处理的诸多领域发挥着重要的作用。目前, 自适应广义主成分分析算法还并不多见。针对这一现状, 该文提出一种快速收敛的广义主成分分析算法, 并通过理论分析所提算法的确定性离散时间系统, 导出了保证算法收敛的学习因子和初始权向量模值等边界条件。仿真实验和实际应用验证了所提算法的正确性和有用性。仿真结果还表明, 所提算法比现有同类算法具有更快的收敛速度和更高的估计精度。

关键词: 广义主成分; 确定性离散时间; 收敛性分析; 神经网络

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2016)10-2531-07

DOI: 10.11999/JEIT151433

A Generalized Principal Component Extraction Algorithm and Its Convergence Analysis

GAO Yingbin KONG Xiangyu HU Changhua ZHANG Huihui HOU Li'an
(The Second Artillery Engineer University, Xi'an 710025, China)

Abstract: The generalized principal component analysis plays an important roles in many fields of modern signal processing. However, up to now, there are few algorithms, which can extract the generalized principal component adaptively. In this paper, a generalized principal component extraction algorithm, which has fast convergence speed, is proposed. The corresponding Deterministic Discrete Time (DDT) system of the proposed algorithm is analyzed and some conditions about the learning rate and initial weight vector are also obtained. Finally, computer simulation and practical application results show that compared with some existing algorithms, the proposed algorithm has faster convergence speed and higher estimation accuracy.

Key words: Generalized principal component; Deterministic Discrete Time (DDT); Convergence analysis; Neural networks

1 引言

广义主成分分析是现代信号处理和数据分析领域内的一门重要工具, 已经广泛应用于如雷达信号处理^[1], 移动通信^[2], 盲信号分离^[3], 医学检查^[4]等各个领域。因此发展广义主成分分析算法是非常有意义的工作。

为了求解上述问题, 一些学者提出了数值算法^[5]。然而这些算法均是批处理算法, 计算复杂度非常大, 而且这些算法还要求输入信号的自相关矩阵是已知的, 而在实时信号处理领域输入信号的自相关矩阵通常是不确定的, 只能根据输入信号来进行在线估计, 因此这些数值算法是难以满足实时信号

处理的需要。近些年来, 学者们相继提出了一些采用神经网络来进行广义主成分提取^[6-8]。相比传统的数值算法, 神经网络算法具有以下 3 个方面的优点: (1)可以对输入信号的自相关矩阵进行在线估计; (2)算法的计算复杂度较低; (3)能够处理非平稳的随机信号^[9]。因此基于神经网络的广义主成分分析算法研究成为了近些年来国内外一个研究热点。本文提出了一种新的广义主成分分析算法。相比一些现存算法, 该算法具有算法形式简单、收敛速度快、收敛精度高等特点。

神经网络算法的收敛性分析是一个非常值得研究的问题。通过对其进行收敛性分析, 给出神经网络算法收敛的边界条件, 这对于神经网络算法的实际应用十分有利。一般而言, 神经网络算法是由其对应的随机离散时间(Stochastic Discrete Time, SDT)来描述的, 这就使得直接对神经网络算法进行分析变得十分困难^[10]。长期以来, 学者们相继提出了一些间接分析方法, 主要有: 确定性连续时间(Deterministic

收稿日期: 2015-12-17; 改回日期: 2016-05-10; 网络出版: 2016-07-04

*通信作者: 孔祥玉 xiangyukong01@163.com

基金项目: 国家自然科学基金面上项目(61074072, 61374120), 国家杰出青年基金(61025014)

Foundation Items: The National Natural Science Foundation of China (61074072, 61374120), The National Science Fund for Distinguished Youth Scholars (61025014)

Continuous Time, DCT)方法和确定性离散时间(Deterministic Discrete Time, DDT)方法。相比 DCT 方法, DDT 方法可以保留神经网络算法的离散本质^[1], 而且可以允许神经网络算法的学习因子为一个常值, 十分符合实际使用的需要, 因此成为近些年来人们研究的一个热点。近些年来, 学者们相继采用 DDT 方法对一些神经网络算法进行了收敛性分析, 取得了很多成果^[10-12]。然而以上所提文献中的所有算法都针对的是一般主成分分析问题, 到目前为止, 还很少有文献对广义主成分分析算法采用 DDT 方法进行研究。由于相比一般主成分分析算法, 广义主成分分析算法形式复杂, 所以采用 DDT 方法对广义主成分分析算法进行分析时也会比一般主成分分析算法复杂。因此采用 DDT 方法对广义主成分分析算法进行收敛性分析就显得很有必要。因此本文将采用 DDT 方法对所提广义主成分分析算法进行收敛性分析, 确定其收敛的边界条件。

本文的结构安排如下: 在第 2 节中, 提出了一种新的广义主成分分析算法; 在第 3 节中采用 DDT 方法对所提算法进行收敛性分析, 确定了所提算法的收敛性条件; 第 4 节主要是通过两组仿真实验对所提算法的性能进行验证; 本文的总结结论安排在第 5 节。

2 一种广义主成分分析算法

根据矩阵理论, 广义特征值分解就是寻找一个向量 \mathbf{v} 和标量 λ , 使得式(1)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 $\mathbf{x}(k) \in \mathbb{R}^{n \times 1}$ 和向量 $\mathbf{y}(k) \in \mathbb{R}^{n \times 1}$ 的自相关矩阵, $\mathbf{x}(k)$ 和 $\mathbf{y}(k)$ 则是两个零均值的随机过程向量。通常将满足方程(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, 并对应有 n 个关于 \mathbf{R}_x 正交的广义特征向量 \mathbf{v}_i , 即对于任意的 $i=1, 2, \dots, n$, 始终有 $\mathbf{R}_y \mathbf{v}_i = \lambda_i \mathbf{R}_x \mathbf{v}_i$ 和 $\mathbf{v}_i^T \mathbf{R}_x \mathbf{v}_j = \delta_{ij}$ ($i, j = 1, 2, \dots, n$) 成立, 其中 δ_{ij} 是 Kronecher 函数。为了后续使用方便, 这里将广义特征值按照由大到小的方式排列, 即有 $\lambda_1 > \lambda_2 > \dots > \lambda_n$ 。在信号处理领域, 通常将最大广义特征值对应的广义特征向量称为信号的广义主成分, 也就是说, 广义特征值 λ_1 对应的广义特征向量 \mathbf{v}_1 即为矩阵束 $(\mathbf{R}_y, \mathbf{R}_x)$ 的广义主成分。

目前, 基于 Hebb 神经网络的广义主成分分析算法的研究是国际上的一个研究热点。Hebb 是一个具有线性形式的神经网络, 它在迭代过程中不引入

后项传播误差来更新权矩阵, 其权值更新只依靠相应神经元的输入输出乘积。Hebb 神经网络的神经元模型可以通过式(2)来描述:

$$p(k) = \mathbf{w}^T(k) \mathbf{q}(k) \quad (2)$$

其中, $\mathbf{q}(k) \in \mathbb{R}^{n \times 1}$ 是一个零均值的随机过程向量, 这里作为神经网络的输入; $\mathbf{w}(k) \in \mathbb{R}^{n \times 1}$ 是神经网络的权向量; $p(k)$ 为该神经网络的输出; k 为迭代次数变量(下同)。基于神经网络的广义主成分分析算法的目的是构造一个神经网络权向量更新方程, 使神经网络的权向量最终能够收敛到矩阵束 $(\mathbf{R}_y, \mathbf{R}_x)$ 的广义主成分。

近些年来, 基于 Hebb 神经网络学者们提出了一些的广义主成分分析算法。这些方法按照提出方式可以分为两类: 一类是基于某种信息准则导出的算法, 如 RLS 算法^[6]; 另一类是基于某种启发式推理而提出的, 如 R-GEVE 算法^[13]等。虽然第 2 类算法缺乏相对应的信息准则, 但是这并没有限制算法的使用, 也不会影响对算法进行收敛性分析。本文提出了一个基于启发式推理的神经网络权向量更新算法:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \eta \left[(\mathbf{R}_x)^{-1} \mathbf{R}_y \mathbf{w}(k) - \mathbf{w}^T(k) \mathbf{R}_y \mathbf{w}(k) \mathbf{w}(k) - \mathbf{w}(k) \right] \quad (3)$$

其中, η 是神经网络的学习因子, 取值范围为 $0 < \eta < 1$ 。式(3)本质上是一个基于 Hebb 神经网络权向量更新规则, 该神经网络的输入为两个观测信号序列, 即 $\mathbf{x}(k)$ 和 $\mathbf{y}(k)$, 输出是观测信号的一个线性表示, 通常没有具体含义。经过多次迭代使得上述算法收敛时, 式(3)中的神经网络权向量将收敛到矩阵束的广义主成分, 即有 $\mathbf{w}(k) \rightarrow \mathbf{v}_1$ 。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的在线估计。为了满足实际需要, 这里采用加权窗函数法对自相关矩阵 \mathbf{R}_y 和 \mathbf{R}_x 进行在线估计:

$$\left. \begin{aligned} \hat{\mathbf{R}}_y(k+1) &= \alpha_1 \hat{\mathbf{R}}_y(k) + \mathbf{y}(k+1) \mathbf{y}^T(k+1) \\ \hat{\mathbf{R}}_x(k+1) &= \alpha_2 \hat{\mathbf{R}}_x(k) + \mathbf{x}(k+1) \mathbf{x}^T(k+1) \end{aligned} \right\} \quad (4)$$

对式(4)应用矩阵求逆引理, 则可以获得 $\mathbf{Q}_x(k) = \mathbf{R}_x^{-1}(k)$ 的在线估计:

$$\mathbf{Q}_x(k+1) = \frac{1}{\alpha_2} \left[\mathbf{Q}_x(k) - \frac{\mathbf{Q}_x(k) \mathbf{x}(k+1) \mathbf{x}^T(k+1) \mathbf{Q}_x(k)}{\alpha_2 + \mathbf{x}^T(k+1) \mathbf{Q}_x(k) \mathbf{x}(k+1)} \right] \quad (5)$$

其中, $\alpha_1, \alpha_2 \in (0, 1)$ 为遗忘因子。将式(4)和式(5)所获得的 $\hat{\mathbf{R}}_x(k+1)$, $\hat{\mathbf{R}}_y(k+1)$ 和 $\mathbf{Q}_x(k+1)$ 分别代替式(3)中的矩阵 \mathbf{R}_x , \mathbf{R}_y 和 \mathbf{R}_x^{-1} , 就可以获得式(3)的自适应算法形式, 其具体计算过程如表 1 所示。

表 1 自适应广义主成分分析算法

初始化: 令 $k=0$, 设置矩阵 $\widehat{\mathbf{R}}_y(0), \widehat{\mathbf{R}}_x(0), \mathbf{Q}_x(0)$ 的初始值, 随机产生初始权向量 $\mathbf{w}(0)$;
迭代过程:
步骤 1 令 $k=k+1$, 根据式(4)和式(5)更新矩阵 $\widehat{\mathbf{R}}_y(k), \widehat{\mathbf{R}}_x(k), \mathbf{Q}_x(k)$;
步骤 2 根据下式更新权向量 $\mathbf{w}(k+1)$
$\mathbf{w}(k+1) = \mathbf{w}(k) + \eta \left[\mathbf{Q}_x(k+1) \widehat{\mathbf{R}}_y(k+1) \mathbf{w}(k) - \mathbf{w}^T(k) \widehat{\mathbf{R}}_y(k+1) \mathbf{w}(k) \mathbf{w}(k) - \mathbf{w}(k) \right]$ (6)

3 算法的收敛性分析

本节采用 DDT 方法对所提算法进行收敛性分析: 对式(6)施加条件期望因子 $E\{\mathbf{w}(k+1)/\mathbf{w}(0), \mathbf{x}(i), i < k\}$, 并将辨识的条件期望值作为下一次迭代值, 则可获得所提算法的 DDT 系统^[10], 该系统与式(3)形式相同。

由矩阵理论可知: 矩阵 \mathbf{R}_x 和 \mathbf{R}_y 是两个正定对称矩阵, 而且矩阵束 $(\mathbf{R}_y, \mathbf{R}_x)$ 的广义特征向量组成了空间 $\mathbb{R}^{n \times n}$ 内的一组关于 \mathbf{R}_x 正交的基。为了方便后续的使用, 这里将矩阵束 $(\mathbf{R}_y, \mathbf{R}_x)$ 的广义特征向量按照降序排列, 即有: $\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$ 。那么神经网络的权向量可以通过这组基表示出来。

$$\mathbf{w}(k) = \sum_{i=1}^n z_i(k) \mathbf{v}_i \quad (7)$$

其中, $z_i(k) = \mathbf{v}_i^T \mathbf{R}_x \mathbf{w}(k)$ ($i = 1, 2, \dots, n$) 是一个常数, 表示权向量 $\mathbf{w}(k)$ 在广义特征向量 \mathbf{v}_i 上投影的长度。将式(7)代入式(3)可得: 对于所有的 $k \geq 0$, 均有

$$z_i(k+1) = \left\{ 1 + \eta \left[\lambda_i - \mathbf{w}^T(k) \mathbf{R}_y \mathbf{w}(k) - 1 \right] \right\} z_i(k) \quad (8)$$

根据 Rayleigh 商的相关特性, 对于所有的 $k \geq 0$, 均有式(9)成立:

$$\begin{aligned} 0 < \lambda_n \mathbf{w}^T(k) \mathbf{R}_x \mathbf{w}(k) &\leq \mathbf{w}^T(k) \mathbf{R}_y \mathbf{w}(k) \\ &\leq \lambda_1 \mathbf{w}^T(k) \mathbf{R}_x \mathbf{w}(k) \end{aligned} \quad (9)$$

接下来, 通过定理 1-定理 6 来完成对所提算法进行收敛特性分析。

定理 1 假定学习因子满足 $\eta \lambda_1 \leq 0.3$ 和 $\eta \leq 0.3$, 如果初始权向量满足 $\mathbf{w}^T(0) \mathbf{R}_x \mathbf{v}_n \neq 0$ 和 $\|\mathbf{w}(0)\|_{\mathbf{R}_x} \leq 1$, 则有对于所有的 $k \geq 0$ 均有 $\|\mathbf{w}(k)\|_{\mathbf{R}_x} < 1 + \eta \lambda_1$ 成立。

证明 应用式(3), 可以得到

$$\begin{aligned} \|\mathbf{w}(k+1)\|_{\mathbf{R}_x}^2 &= \left(\sum_{i=1}^n z_i(k+1) \mathbf{v}_i^T \right) \mathbf{R}_x \left(\sum_{i=1}^n z_i(k+1) \mathbf{v}_i \right) \\ &= \sum_{i=1}^n z_i^2(k+1) \end{aligned} \quad (10)$$

将式(8)代入式(10)可得

$$\begin{aligned} &\|\mathbf{w}(k+1)\|_{\mathbf{R}_x}^2 \\ &= \sum_{i=1}^n \left\{ 1 + \eta \left[\lambda_i - \mathbf{w}^T(k) \mathbf{R}_y \mathbf{w}(k) - 1 \right] \right\}^2 z_i^2(k) \\ &< \sum_{i=1}^n \left\{ 1 + \eta \left[\lambda_1 - \lambda_n \mathbf{w}^T(k) \mathbf{R}_x \mathbf{w}(k) - 1 \right] \right\}^2 z_i^2(k) \\ &= \left\{ 1 + \eta \left[\lambda_1 - \lambda_n \|\mathbf{w}(k)\|_{\mathbf{R}_x}^2 - 1 \right] \right\}^2 \|\mathbf{w}(k)\|_{\mathbf{R}_x}^2 \end{aligned} \quad (11)$$

在区间 $[0, 1]$ 内, 定义如下一个连续可微函数:

$$f(s) = \left\{ 1 + \eta \left[\lambda_1 - \lambda_n s - 1 \right] \right\}^2 s \quad (12)$$

由连续函数的求导法则可知, 式(12)对于 s 的导数为

$$\dot{f}(s) = (1 + \eta \lambda_1 - \eta - \eta \lambda_n s)(1 + \eta \lambda_1 - \eta - 3\eta \lambda_n s) \quad (13)$$

求解方程 $\dot{f}(s) = 0$ 的根, 可得

$$s_1 = (1 + \eta \lambda_1 - \eta) / \eta \lambda_n, s_2 = (1 + \eta \lambda_1 - \eta) / 3\eta \lambda_n \quad (14)$$

由 $\lambda_1 > \lambda_n > 0$ 和 $0 < \eta < 1$ 可得: $s_1 > s_2 > 0$ 和 $s_1 > 1$ 。由于 $\dot{f}(s)$ 在区间 $[0, 1]$ 内的取值情况难以确定, 所以, 这里分两种情况来进行讨论:

情况 1 当 $1 + \eta \lambda_1 - \eta > 3\eta \lambda_n$, 即 $s_2 > 1$ 时。

根据式(13), 可得在区间 $[0, 1]$ 内有: $\dot{f}(s) > 0$, 即函数 $f(s)$ 在该区间内是一个单调递增函数, 函数 $f(s)$ 在边界 $s = 1$ 处取得极值。即对于所有的 $0 < s < 1$, 有

$$f(s) \leq f(1) = [1 + \eta(\lambda_1 - \lambda_n - 1)]^2 < (1 + \eta \lambda_1)^2 \quad (15)$$

情况 2 当 $1 + \eta \lambda_1 - \eta < 3\eta \lambda_n$, 即 $s_2 < 1$ 时。

此时, 根据式(13)可得: 当 $0 \leq s < s_2$ 时 $\dot{f}(s) > 0$, 而当 $s_2 < s \leq 1$ 时 $\dot{f}(s) < 0$, 所以在区间 $[0, 1]$ 内, 函数 $f(s)$ 在 $s = s_2 = (1 + \eta \lambda_1 - \eta) / 3\eta \lambda_n$ 时取得极大值。即对于所有的 $0 < s < 1$, 有

$$f(s) \leq f(s_2) = 4(1 + \eta \lambda_1 - \eta)^3 / 27\eta \lambda_n \quad (16)$$

由于 $1 + \eta \lambda_1 - \eta < 3\eta \lambda_n$, 可得

$$\begin{aligned} \frac{4(1 + \eta \lambda_1 - \eta)^3}{27\eta \lambda_n} &< \frac{4(1 + \eta \lambda_1 - \eta)^3}{9(1 + \eta \lambda_1 - \eta)} = \left[\frac{2}{3} (1 + \eta \lambda_1 - \eta) \right]^2 \\ &< (1 + \eta \lambda_1)^2 \end{aligned} \quad (17)$$

综合两种情况可得, 在区间 $[0, 1]$ 内, 有

$$f(s) < (1 + \eta \lambda_1)^2 \quad (18)$$

因此, 对于所有的 $k \geq 0$ 有: $\|\mathbf{w}(k)\|_{\mathbf{R}_x} < (1 + \eta \lambda_1)$ 。证毕

定理 1 主要完成的是权向量模值的上限证明, 定理 2 则提供了权向量模值下限的分析。

定理 2 假定学习因子满足 $\eta \lambda_1 \leq 0.3$ 和 $\eta \leq 0.3$, 如果初始权向量满足 $\mathbf{w}^T(0) \mathbf{R}_x \mathbf{v}_n \neq 0$ 和 $\|\mathbf{w}(0)\|_{\mathbf{R}_x} \leq 1$, 则有对于所有 $k \geq 0$ 均有 $\|\mathbf{w}(k)\|_{\mathbf{R}_x} > c$ 成立, 其中 $c = 1 - \eta \lambda_1 (1 + \eta \lambda_1)^2 - \eta$ 是一个常数。

证明 根据式(8), 可得

$$\begin{aligned} & \|\mathbf{w}(k+1)\|_{\mathbf{R}_x}^2 \\ &= \sum_{i=1}^n \left\{ 1 + \eta [\lambda_i - \mathbf{w}^T(k) \mathbf{R}_y \mathbf{w}(k) - 1] \right\}^2 z^2(k) \\ &< \sum_{i=1}^n \left\{ 1 - \eta \lambda_1 \mathbf{w}^T(k) \mathbf{R}_x \mathbf{w}(k) - \eta \right\}^2 z^2(k) \\ &= \left\{ 1 - \eta \lambda_1 \|\mathbf{w}(k)\|_{\mathbf{R}_x}^2 - \eta \right\}^2 \|\mathbf{w}(k)\|_{\mathbf{R}_x}^2 \end{aligned} \quad (19)$$

根据定理 1, 对于任意的 $k \geq 0$ 均有 $\|\mathbf{w}(k)\|_{\mathbf{R}_x} < (1 + \eta \lambda_1)$, 则

$$\|\mathbf{w}(k+1)\|_{\mathbf{R}_x}^2 > \left\{ 1 - \eta \lambda_1 (1 + \eta \lambda_1)^2 - \eta \right\}^2 \|\mathbf{w}(k)\|_{\mathbf{R}_x}^2 \quad (20)$$

令 $c = 1 - \eta \lambda_1 (1 + \eta \lambda_1)^2 - \eta$, 根据 $\eta \lambda_1 \leq 0.3$ 和 $\eta \leq 0.3$, 得

$$\begin{aligned} c &= 1 - \eta \lambda_1 (1 + \eta \lambda_1)^2 - \eta \\ &> 1 - 0.3 \times (1 + 0.3)^2 - 0.3 = 0.1930 > 0 \end{aligned} \quad (21)$$

由此, 有 $\|\mathbf{w}(k+1)\|_{\mathbf{R}_x} > c$. 证毕

通过定理 1 和定理 2 可以完成式(3)的边界性证明。下面, 将证明在一些温和的条件下, 式(3)中权向量将最终收敛到矩阵束的广义主成分。为了分析式(3)的收敛性, 这里通过定理 3 给出一些预备结论。

定理 3 假定学习因子满足 $\eta \lambda_1 \leq 0.3$ 和 $\eta \leq 0.3$, 如果初始权向量满足 $\mathbf{w}^T(0) \mathbf{R}_x \mathbf{v}_n \neq 0$ 和 $\|\mathbf{w}(0)\|_{\mathbf{R}_x} \leq 1$, 则有对于所有的 $k \geq 0$ 均有 $1 + \eta [\lambda_i - \mathbf{w}^T(k) \mathbf{R}_y \mathbf{w}(k) - 1] > 0$ 成立。

证明 从定理 1 有 $\|\mathbf{w}(k)\|_{\mathbf{R}_x} < (1 + \eta \lambda_1)$, 则

$$\begin{aligned} & 1 + \eta [\lambda_i - \mathbf{w}^T(k) \mathbf{R}_y \mathbf{w}(k) - 1] \\ & > 1 - \eta \lambda_1 (1 + \eta \lambda_1)^2 - \eta \end{aligned} \quad (22)$$

当 $\eta \lambda_1 \leq 0.3$ 时, 有

$$\begin{aligned} 1 - \eta \lambda_1 (1 + \eta \lambda_1)^2 - \eta &> 1 - 0.3 \times (1 + 0.3)^2 - 0.3 \\ &= 0.1930 > 0 \end{aligned} \quad (23)$$

即对于所有的 $k \geq 0$, 均有 $1 + \eta [\lambda_i - \mathbf{w}^T(k) \mathbf{R}_y \mathbf{w}(k) - 1] > 0$. 证毕

由式(7)可得: 对于所有的 $k \geq 0$, $\mathbf{w}(k)$ 可以表示为

$$\mathbf{w}(k) = z_1(k) \mathbf{v}_1 + \sum_{i=2}^n z_i(k) \mathbf{v}_i \quad (24)$$

根据式(24)可得, 神经网络权向量 $\mathbf{w}(k)$ 的收敛性将由 $z_i(k) (i = 1, 2, \dots, n)$ 的收敛性决定。下面的定理 4 和定理 5 则分别提供了 $z_i(k) (i = 2, 3, \dots, n-1)$ 和 $z_1(k)$ 的收敛性证明。

定理 4 假定学习因子满足 $\eta \lambda_1 \leq 0.3$ 和 $\eta \leq 0.3$, 如果初始权向量满足 $\mathbf{w}^T(0) \mathbf{R}_x \mathbf{v}_n \neq 0$ 和 $\|\mathbf{w}(0)\|_{\mathbf{R}_x} \leq 1$, 则有 $\lim_{k \rightarrow \infty} z_i(k) = 0$, 其中 $i = 2, 3, \dots, n$ 。

证明 应用定理 1 和定理 2 可得, 对于所有的

$k \geq 1$, 均有 $c < \|\mathbf{w}(k)\|_{\mathbf{R}_x} < 1 + \eta \lambda_1$ 。这样, 对于所有 $k \geq 1$, 有

$$\begin{aligned} \left[\frac{z_i(k+1)}{z_1(k+1)} \right]^2 &= \frac{\left\{ 1 + \eta [\lambda_i - \mathbf{w}^T \mathbf{R}_y \mathbf{w} - 1] \right\}^2 \left[\frac{z_i(k)}{z_1(k)} \right]^2}{\left\{ 1 + \eta [\lambda_1 - \mathbf{w}^T \mathbf{R}_y \mathbf{w} - 1] \right\}^2 \left[\frac{z_1(k)}{z_1(k)} \right]^2} \\ &= \left[1 - \frac{\eta (\lambda_1 - \lambda_i)}{1 + \eta [\lambda_1 - \mathbf{w}^T \mathbf{R}_y \mathbf{w} - 1]} \right]^2 \left[\frac{z_i(k)}{z_1(k)} \right]^2 \end{aligned} \quad (25)$$

假设 $\theta'_1 = \eta (\lambda_1 - \lambda_i) / [1 + \eta (\lambda_1 - \mathbf{w}^T \mathbf{R}_y \mathbf{w} - 1)]$ 接下来, 将证明 $0 < \theta'_1 < 1$ 。

首先, 由 $\lambda_1 > \lambda_n$ 和定理 3 可得: $\theta'_1 > 0$ 。接下来, 应用定理 1 和 $\eta \lambda_1 < 0.3$ 以及 $\eta < 0.3$, 有:

$$\begin{aligned} \theta'_1 &= \frac{\eta (\lambda_1 - \lambda_i)}{1 + \eta [\lambda_1 - \mathbf{w}^T(k) \mathbf{R}_y \mathbf{w}(k) - 1]} \\ &< \frac{\eta \lambda_1}{1 + \eta [\lambda_1 - \lambda_1 \|\mathbf{w}(k)\|_{\mathbf{R}_x}^2 - 1]} \\ &\leq \frac{0.3}{1 + 0.3 - 0.3 \times (1 + 0.3)^2 - 0.3} \\ &= 0.6085 < 1 \end{aligned} \quad (26)$$

将 $0 < \theta'_1 < 1$ 代入式(25)可得

$$\left[\frac{z_i(k+1)}{z_1(k+1)} \right]^2 < \theta_1 \left[\frac{z_i(k)}{z_1(k)} \right]^2 < \dots < \theta_1^{k+1} \left[\frac{z_i(0)}{z_1(0)} \right]^2 \quad (27)$$

其中, $\theta_1 = (1 - \theta'_1)^2$ 。显然有: $0 < \theta_1 < 1$, 则对于 $i = 2, 3, \dots, n$, 有

$$\lim_{k \rightarrow \infty} \left[\frac{z_i(k)}{z_1(k)} \right]^2 = 0 \quad (28)$$

根据定理 1 和定理 2, 对于所有的 $k \geq 0$, $z_1(k)$ 均是有界的, 则有

$$\lim_{k \rightarrow \infty} z_i(k) = 0, \quad i = 2, 3, \dots, n \quad (29)$$

证毕

定理 5 假定学习因子满足 $\eta \lambda_1 \leq 0.3$ 和 $\eta \leq 0.3$, 如果初始权向量满足 $\mathbf{w}^T(0) \mathbf{R}_x \mathbf{v}_n \neq 0$ 和 $\|\mathbf{w}(0)\|_{\mathbf{R}_x} \leq 1$, 则有 $\lim_{k \rightarrow \infty} z_1(k) = z^*$ 。

证明 从定理 4 可以得知, 当 $k \rightarrow \infty$ 时, 权向量 $\mathbf{w}(k)$ 将逐渐收敛到广义主成分 \mathbf{v}_1 的方向。假定在 k_0 时刻, $\mathbf{w}(k)$ 已经收敛到 \mathbf{v}_1 的方向, 则有 $\mathbf{w}(k_0) = z_1(k_0) \mathbf{v}_1$ 。

由式(8)可得, 当 $k \geq k_0$ 时, 有

$$z_1(k+1) = (1 - \eta) z_1(k) + \eta \lambda_1 [1 - z_1(k)] [z_1(k) + z_1^2(k)] \quad (30)$$

从式(30)可以得出, 对于任意的 $k \geq k_0$, 均有

$$\begin{aligned} z_1(k+1) - 1 &= -\eta z_1(k) + \eta \lambda_1 [1 - z_1(k)] [z_1(k) + z_1^2(k)] \\ &< [z_1(k) - 1] \left\{ 1 - \eta \lambda_1 [z_1(k) + z_1^2(k)] \right\} \end{aligned} \quad (31)$$

令 $\theta_2 = 1 - \eta \lambda_1 [z_1(k) + z_1^2(k)]$, 则由 $z_1(k) < \|\mathbf{w}(k)\|_{\mathbf{R}_x} < 1 + \eta \lambda_1$ 和 $\eta \lambda_1 \leq 0.3$ 可得

$$\begin{aligned} \theta_2 &= 1 - \eta\lambda_1 [z_1(k) + z_1^2(k)] \\ &> 1 - \eta\lambda_1 [1 + \eta\lambda_1 + (1 + \eta\lambda_1)^2] \\ &> 1 - 0.3 \times [1 + 0.3 + (1 + 0.3)^2] \\ &= 0.103 > 0 \end{aligned} \tag{32}$$

显然, 由式(32)可得 $0 < |\theta_2| < 1$ 。然后, 由式(31)和式(32)可得

$$|z_1(k+1) - 1| < \theta_2 [|z_1(k) - 1|] \tag{33}$$

即对于任意的 $k \geq k_0$ 均有

$$|z_1(k+1) - 1| < \theta_2^{k+1} [|z_1(0) - 1|] < (k+1)\Pi e^{-\delta(k+1)} \tag{34}$$

其中, $\Pi = |z_1(0) - 1|$, $\delta = -\ln(\theta_2)$ 。

给定任意小数 $\varepsilon > 0$, 均存在常数 K , 使得式 $\Pi_2 K e^{-\delta K} / (1 - e^{-\theta})^2 \leq \varepsilon$ 成立。则对于任意的 $k_1 > k_2 > K$, 有

$$\begin{aligned} &|z_1(k_1) - z_1(k_2)| \\ &= \left| \sum_{r=k_2}^{k_1-1} [z_1(r+1) - z_1(r)] \right| \\ &< \left| \sum_{r=k_2}^{k_1-1} \eta\lambda_1 [1 - z_1(r)] [z_1(r) + z_1^2(r)] \right| \\ &\leq \sum_{r=k_2}^{k_1-1} \eta\lambda_1 |z_1(r)| |z_n(r) + 1| |z_n(r) - 1| \\ &\leq \eta\lambda_1 (1 + \eta\lambda_1) (2 + \eta\lambda_1) \sum_{r=k_2}^{k_1} |z_n(r) - 1| \\ &\leq \prod_{2} \sum_{r=k_2}^{k_1} r e^{-\alpha r} \leq \frac{\prod K e^{-\alpha K}}{(1 - e^{-\alpha})^2} \leq \varepsilon \end{aligned} \tag{35}$$

其中, $\Pi_2 = \eta\lambda_1(1 + \eta\lambda_1)(2 + \eta\lambda_1)$ 。

由式(35)可得, 序列 $\{z_1(k)\}$ 是一个柯西序列。根据柯西序列的收敛特性, 则必然存在一常数 z^* , 使得当 $k \rightarrow \infty$ 时, $\lim_{k \rightarrow \infty} z_1(k) \rightarrow z^*$ 。证毕

下面将通过定理 6 来完成式(3)中权向量 $w(k)$ 的收敛性证明。

定理 6 假定学习因子满足 $\eta\lambda_1 \leq 0.3$ 和 $\eta \leq 0.3$, 如果初始权向量满足 $w^T(0)R_x v_n \neq 0$ 和 $\|w(0)\|_{R_x} \leq 1$, 则有 $\lim_{k \rightarrow \infty} w(k) = z^* v_1$, 其中 z^* 是一个常数。

证明 由式(24)和定理 4, 定理 5 可得

$$\begin{aligned} \lim_{k \rightarrow \infty} w(k) &= \lim_{k \rightarrow \infty} \left[z_1(k)v_1 + \sum_{i=2}^n z_i(k)v_i \right] \\ &= \lim_{k \rightarrow \infty} [z_1(k)]v_1 + \left[\sum_{i=2}^n v_i \lim_{k \rightarrow \infty} z_i(k) \right] \\ &= z^* v_1 \end{aligned} \tag{36}$$

证毕

到此就完成了对所提算法的收敛性证明。下面对所提算法收敛性条件进行讨论。从条件 $\eta\lambda_1 \leq 0.3$ 和 $\eta \leq 0.3$ 可得, 所提算法学习因子的选择与最大广义特征值有关。在实际应用过程中, 尽管最大广义特征值是未知的, 但是它的上限是可以被估计得出的^[12]。因此, 学习因子 η 的限制条件是可以达到的。此外, 当初始权向量为随机产生时, 初始权向量的条件 $w^T(0)R_x v_n \neq 0$ 是以概率 1 满足的。因此可以说, 通过 DDT 分析获得的所提算法的收敛性条件是符合实际使用需要的。

4 仿真实验

本节将提供两个实验是对所提算法的性能验证。第 1 个实验是考验算法对于两个随机输入向量的广义特征向量提取的能力以及和其他算法的性能比较; 第 2 个实验是应用所提算法解决盲信号分离问题。

4.1 广义主成分提取实验

本实验目的是对所提算法的广义主成分提取能力进行实验验证。这里将所提算法与 RLS 算法^[6]和 FPI-GED 算法^[7]进行对比。为了能够衡量所提算法的性能, 这里引入方向余弦作为评价函数:

$$DC(k) = \frac{|w^T(k)v_1|}{\|w(k)\| \|v_1\|} \tag{37}$$

在仿真实验过程中, 神经网络的输入向量由式(38)和式(39)产生^[6]:

$$y(k) = \sqrt{2} \sin(0.62\pi k + \tau_1) + n_1(k) \tag{38}$$

$$\begin{aligned} x(k) &= 10\sqrt{2} \sin(0.46\pi k + \tau_2) \\ &+ 10\sqrt{2} \sin(0.74\pi k + \tau_3) + n_2(k) \end{aligned} \tag{39}$$

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: $\sigma_1^2 = 0.01, \sigma_2^2 = 1$ 。取式(38)和式(39)的 8 个连续输出作为神经网络的输入向量。接下来分别采用 RLS 算法、FPI-GED 算法和式(6)对矩阵束 (R_y, R_x) 的广义主成分进行提取。为了确保公平比较, 这里对两种算法采用相同的初始化参数。两种算法的参数设置为: RLS 算法中, 遗忘因子 $\alpha = \beta = 0.9998$, $P(0) = 1$; FPI-GED 算法中, $\alpha = \beta = 0.98$, $P(0) = 0.01$; 本文所提算法中: 学习因子 $\eta = 0.1$, 遗忘因子 $\alpha_1 = 0.9998$, 矩阵 $Q_x(0) = 0.023I_n$ 。3 种算法采用相同的初始化权向量(权向量是随机产生的)。采用两种方法对同一组矩阵束 (R_y, R_x) 进行 100 次蒙特卡洛仿真, 仿真结果如图 1 所示。

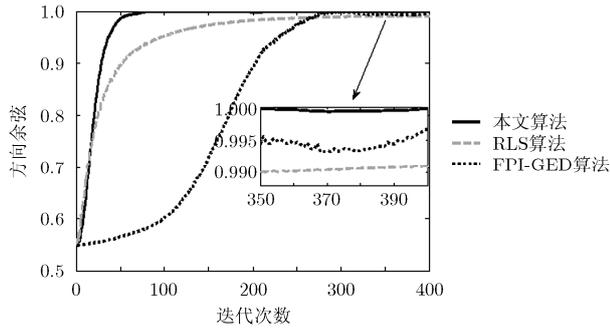


图 1 两种算法的方向余弦曲线

从图 1 中可以看出:相比 RLS 算法和 FPI-GED 算法,本文所提算法具有较快的收敛速率。此外,虽然两种算法的方向余弦最终均收敛到了 1 附近,但是当把仿真的最后 50 步进行放大后可以发现:本文所提算法与单位 1 的接近程度要优于 RLS 算法。由于方向余弦可以表征算法的收敛精度,所以可以得出结论:相比 RLS 算法和 FPI-GED 算法,本文所提算法不仅具有较快的收敛速度而且具有较高的收敛精度。

4.2 盲信号分离实验

本实验主要是应用所提算法解决盲信号分离问题。考虑式(40)所示的盲分离模型:

$$\mathbf{x}(k) = \mathbf{A} \mathbf{s}(k) + \mathbf{v}(k) \quad (40)$$

其中, $\mathbf{x}(k)$ 为观测信号, $\mathbf{s}(k)$ 为源信号, \mathbf{A} 为混合矩阵, $\mathbf{v}(k)$ 为观测噪声。则盲分离问题就是求解分离矩阵 \mathbf{W} , 使得分离信号 $\mathbf{y} = \mathbf{W}^T \mathbf{x}$ 尽可能地相互独立。文献[3]证明:盲信号的分离矩阵可以通过求取一个矩阵束 $(\mathbf{R}_z, \mathbf{R}_x)$ 的广义主成分来实现, 其中

$$\mathbf{R}_x = E[\mathbf{x}(k) \mathbf{x}^T(k)] \quad , \quad \mathbf{R}_z = E[\mathbf{z}(k) \mathbf{z}^T(k)] \quad , \quad \mathbf{z}(k) = \sum_{t=0}^l \eta(t) \mathbf{x}(k-t) \quad .$$

在本实验将 ICALAB 中的 ABio7.mat(获取地址为: <http://www.bsp.brain.riken.go.jp/ICALAB/>) 中的 4 个信号作为输入, 如图 2 所示, 图中横纵坐标为时刻, 纵坐标为幅值。盲信号的混合矩阵为

$$\mathbf{A} = \begin{bmatrix} 2.7914 & -0.1780 & -0.4945 & 0.3013 \\ 1.3225 & -1.7841 & -0.3669 & 0.4460 \\ 0.0714 & -1.9163 & 0.4802 & -0.3700 \\ -1.7396 & 0.1302 & 0.9249 & -0.4007 \end{bmatrix} \quad (41)$$

观测噪声为随机产生的, 且其方差为 $10^{-5} \mathbf{I}$, $\eta = [0.5 \quad -0.5]$ 。然后采用本文所提的广义主成分分析算法对该盲分离问题进行求解。为了衡量算法的分离结果, 这里对采用源信号和分离信号之间的互相关系数作为评价函数。采用本文算法获得的分离信号如图 3 所示, 表 2 则给出了两个信号之间的互相关系数。

从图 3 和表 2 中可以看出, 采用本文所提算法获得的分离信号与源信号之间具有很强的相似性。虽然本文算法的分离结果略逊于广义特征值分解算法, 但是本文算法是可以在线进行的, 而广义特征值分解算法是一种批处理算法, 而且具有很高的计算复杂度。

5 结论

本文提出了一种新型的广义主成分分析算法, 并采用一种 DDT 分析方法对所提算法进行了收敛性分析, 确定了保持该算法收敛性的边界条件, 为

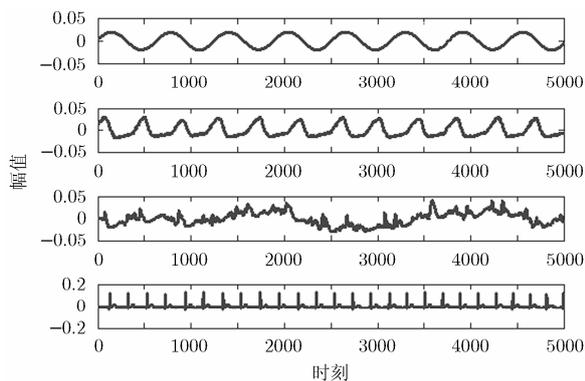


图 2 源信号曲线

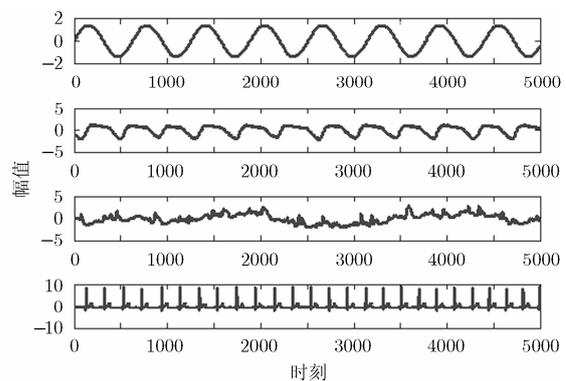


图 3 分离信号曲线

表 2 原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
广义特征值分解	1.0000	0.9998	0.9997	0.9989
本文算法	0.9994	0.9998	0.9998	0.9988

算法的实际应用奠定了基础。仿真实验和实际应用表明：相比一些现有广义主成分分析算法，本文所提算法不仅具有很快的收敛速度而且具有很高的估计精度。

参 考 文 献

- [1] 谢荣, 刘峥, 刘俊. 基于矩阵束的 MIMO 雷达低仰角快速估计方法[J]. 电子与信息学报, 2011, 33(8): 1833-1838. doi: 10.3724/SP.J.1146.2010.01242.
XIE Rong, LIU Zheng, and LIU Jun. Fast algorithm for low elevation estimation based on matrix pencil in MIMO radar[J]. *Journal of Electronics & Information Technology*, 2011, 33(8): 1833-1838. doi: 10.3724/SP.J.1146.2010.01242.
- [2] 蔡振浩, 赵昆, 陈文. TD-LTE-A 系统下行多用户 CoMP 联合预编码算法[J]. 北京邮电大学学报, 2015, 38(1): 67-70.
CAI Zhenhao, ZHAO Kun, and CHEN Wen. Research on CoMP joint transmission for downlink MU-MIMO in TD-LTE-A[J]. *Journal of Beijing University of Posts and Telecommunications*, 2015, 38(1): 67-70.
- [3] ANA Maria Tomé. The generalized eigen-decomposition approach to the blind source separation problem[J]. *Digital Signal Processing*, 2006, 16: 288-302.
- [4] ZHANG Weitao, LOU Shuntian, and FENG Dazheng. Adaptive quasi-newton algorithm for source extraction via CCA approach[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, 25(4): 677-689.
- [5] WANG Shougen and ZHAO Shuqin. An algorithm for $Ax = \lambda Bx$ with symmetric and positive-definite A and B [J]. *SIAM Journal on Matrix Analysis and Applications*, 1991, 12(4): 654-660.
- [6] YANG Jian, XI Hongsheng, and YANG Feng. RLS-based adaptive algorithms for generalized eigen-decomposition[J]. *IEEE Transactions on Signal Processing*, 2006, 54(4): 1177-1188.
- [7] YANG Jian, HU Han, and XI Hongsheng. Weighted non-linear criterion-based adaptive generalised eigen decomposition[J]. *IET Signal Processing*, 2013, 7(4): 285-295.
- [8] Tuan Duong Nguyen and Isao Yamada. Adaptive normalized quasi-Newton algorithms for extraction of generalized eigen-pairs and their convergence analysis[J]. *IEEE Transactions on Signal Processing*, 2013, 61(6): 1404-1418.
- [9] MÖLLER R. A self-stabilizing learning rule for minor component analysis[J]. *International Journal of Neural Systems*, 2004, 14(1): 1-8.
- [10] ZUFIRIA P J. On the discrete-time dynamics of the basic Hebbian neural-network node[J]. *IEEE Transactions on Neural Networks*, 2002, 13(6): 1342-1352.
- [11] GAO Yingbin, KONG Xiangyu, HU Changhua, et al. Convergence analysis of möller algorithm for estimating minor component[J]. *Neural Processing Letters*, 2015, 42(2): 355-368.
- [12] Tuan Duong Nguyen and Isao Yamada. Necessary and sufficient conditions for convergence of the DDT systems of the normalized PAST algorithms[J]. *Signal Processing*, 2014, 94: 288-299.
- [13] ATTALLAH S and ABED-MERAİM K. A fast adaptive algorithm for the generalized symmetric eigenvalue problem[J]. *IEEE Signal Processing Letters*, 2008, 15: 797-800.
- 高迎彬: 男, 1986 年生, 博士生, 研究方向为自适应信号处理.
孔祥玉: 男, 1967 年生, 教授、博士生导师, 研究方向为自适应信号处理、非线性系统建模和故障诊断.
胡昌华: 男, 1966 年生, 教授、博士生导师, 研究方向为导弹测试与故障诊断.
张会会: 男, 1986 年生, 博士生, 研究方向为故障诊断与寿命预测.
侯立安: 男, 1957 年生, 中国工程院院士, 博士生导师, 研究方向为污水处理、大气污染治理等.