

次成分提取信息准则的加权规则方向收敛分析

杜柏阳¹, 孔祥玉¹, 冯晓伟²

(1. 火箭军工程大学导弹工程学院, 陕西 西安 710025; 2. 火箭军工程大学核工程学院, 陕西 西安 710025)

摘 要: 对并行次成分提取算法在信号特征提取中实现方向收敛的过程进行了研究。采用比较未加权的 PAST 次子空间跟踪算法和加权的 PAST 并行次成分提取算法, 分析加权规则对次成分提取算法方向收敛行为的限定方式。理论分析表明, 加权规则对状态矩阵的各个向量和次成分之间的角度限定存在规律, 并给出了加权规则对状态矩阵各个向量方向的作用方式和定理和讨论。最后, Matlab 仿真实验验证了理论分析的结果。

关键词: 次子空间提取算法; 加权规则; 并行次成分分析; 信息准则

中图分类号: TP273

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2020014

Direction convergence analysis of weighted rule for minor component extraction information criteria

DU Boyang¹, KONG Xiangyu¹, FENG Xiaowei²

1. Department of Missile Engineering, the Rocket Force University of Engineering, Xi'an 710025, China

2. Department of Nuclear Engineering, the Rocket Force University of Engineering, Xi'an 710025, China

Abstract: The extraction of parallel minor components algorithm in directional convergence in signal feature extraction was studied. By comparing the unweighted projection approximation subspace tracking (PAST) algorithm with the weighted PAST parallel minor component extraction algorithm, the evolution method of the minor component extraction algorithm was analyzed. Theoretical analysis illustrated that the weighted rule was able to guide the angle evolution between the vectors of the state matrix and minor components. Finally, Matlab simulation verifies the validity of the proposed theory.

Key words: minor subspace extraction algorithm, weighted rule, parallel minor component analysis, information criterion

1 引言

在信息处理领域, 次成分 (MC, minor component) 是指输入信号的自相关矩阵中与最小特征值对应的特征向量, 次子空间 (MS, minor subspace) 是指由多个次成分张成的子空间。提取次成分和跟踪次子空间的算法分别被称为次成分分析 (MCA, MC analysis) 算法和次子空间跟踪 (MSA, MS analysis) 算法。MCA 算法和 MSA 算法在总体最小二乘^[1]、自适应波达方向估计^[2]、曲面拟合^[3]、稳健

波束分析^[4]等问题中具有重要应用。

传统的代数次成分批处理法只能处理离线的、静态的信号, 且具有较高的计算复杂度。实际中的信号往往是在线获取的动态数据, 因而传统方法的应用十分受限。相比而言, 神经网络算法能够处理的信号不受上述 2 个条件的限制^[5], 并且许多神经网络算法具有较小的计算复杂度。近年来, 学者们提出了许多 MSA 和 MCA 的神经网络算法^[5-7]。在这些算法中, Möller^[5]算法只能提取单个次成分, PAST (projection approximation subspace tracking)

收稿日期: 2019-10-08; 修回日期: 2019-12-15

基金项目: 国家自然科学基金资助项目 (No.61673387, No.61374120, No.61903375, No.61833016)

Foundation Item: The National Natural Science Foundation of China (No.61673387, No.61374120, No. 61903375, No.61833016)

算法^[8]、Ojam 算法^[9]和 Kong 等^[2]所提的算法可以跟踪输入信号的次子空间, Jankovic 等^[10]所提的算法和 AMMD (adaptive multiple minor direction) 算法^[11]可以提取输入信号的多个次成分。上述研究都是单独针对 MSA 或者 MCA 问题开展的, 实际上, MSA 和 MCA 算法之间存在一定关联, 在一定条件下可以相互转换。例如, Jankovic 等^[10]通过 TOHM (time-oriented hierarchical method) 转化装置实现了 MSA 算法到 MCA 并行算法的转换, Thameri 等^[12]则通过 Givens 旋转实现了 MSA 算法到 MCA 算法的转变。这种通过转化转置实现算法转化的方法具有结构复杂、计算复杂度较高的缺点^[13]。另一种行之有效的转化方法是使用加权规则。然而, 目前关于加权规则的研究还比较少。文献[14]介绍了用于提取多个次成分的加权信息准则。目前, 对经过加权的次成分分析算法的性能分析与传统的单维次成分提取算法的性能分析大体相同, 主要集中于算法的收敛性、自稳定性等。而实际上, 经过加权的算法能够实现多个次成分同时提取, 该性能是传统单维次成分提取算法所不具备的, 需要研究多个次成分在提取过程中的方向收敛问题。目前, 研究并行多个提取问题的学者主要将兴趣集中于算法的提出, 极少有对方向收敛问题开展探讨。那么, 加权规则对于 MSA 信息准则转变为 MCA 信息准则的过程发挥了什么样的作用, 这样的作用过程对于其他子空间提取信息准则的转变是否具有普适性? 这个问题的解答对灵活转换各种先进的成分提取算法, 增加人们对提取算法本质和多种性质的理解具有十分重要的意义。

多个次成分可以张成一个次子空间, 因而通常认为多个次成分的提取算法是次子空间跟踪算法的进步。Toshihisa^[15]通过研究多个次成分提取算法的广义加权规则, 指出广义加权规则的参数对提取算法的收敛速度存在影响, 即加权规则的参数变化会引起算法的性质变化。当参数的取值沿着实数轴负方向变化时, 加权矩阵则逐渐近似为单位矩阵, 算法的提取能力逐渐由多个次成分提取退化为次子空间跟踪。

实际上, 只有一部分次子空间提取算法在使用加权规则后可以转化为多个次成分的提取算法。次子空间与次成分的偏离程度可以通过次子空间的组成向量与次成分之间的夹角表示。当提取算法逐渐稳定以后, 次子空间跟踪算法状态矩阵的夹角是

一个随机值, 而转化为多个次成分提取算法状态矩阵的夹角总是某个对角矩阵。这说明加权规则的加入改变了该夹角的变化规则。本质上, 信息准则函数在算法对信息的归类方式、算法提取信息的方式等方面存在隐性的规定, 因而信息准则对夹角的变化方式具有决定作用。因此, 通过理论分析信息准则对 MSA 和 MCA 算法状态矩阵夹角的作用过程, 能够反映算法的一些本质特性。

本文主要针对加权规则对次成分提取算法的信息准则的作用进行分析, 以 PAST 的次子空间跟踪算法为例, 通过构建提取算法的动力学表达, 对比有加权规则和无加权规则下的次子空间跟踪信息准则, 挖掘信息准则对状态矩阵与次成分的方向夹角的梯度差异。本文专门对一类次成分分析算法的方向收敛性能进行总结, 并同时采用理论和仿真的方法对该性能开展分析。

2 加权规则

Yang^[8]在研究子空间跟踪问题时提出次子空间 PAST 算法, 该算法的信息准则为

$$\mathbf{W}^* = \arg \min J_1(\mathbf{W})$$

$$J_1(\mathbf{W}) = \text{tr}(\mathbf{W}) - 2\text{tr}(\mathbf{W}^T \mathbf{R} \mathbf{W}) + \text{tr}(\mathbf{W}^T \mathbf{R} \mathbf{W} \mathbf{W}^T \mathbf{W}) \quad (1)$$

其中, $\mathbf{W} \in \mathbb{R}^{n \times r}$ 为算法的状态矩阵, n 为输入信号的维数, r 为需要提取的次成分个数, $\mathbf{R} = \text{E}[\mathbf{x}\mathbf{x}^T] \in \mathbb{R}^{n \times n}$ 为输入信号 \mathbf{x} 的自相关矩阵, $\text{E}[\cdot]$ 表示数学期望。假设 \mathbf{R} 是满秩矩阵且特征值互不相同, 那么, 输入信号有 n 个特征向量 ϕ_i , $i = 0, 1, \dots, n-1$ 。当存在一个向量 $\mathbf{W} = \mathbf{W}_n$, 使信息准则函数值 $J_1(\mathbf{W})$ 取得最小值时, \mathbf{W}_n 就等价于信号的次子空间, 也就是前 r 个特征向量 ϕ_i 张成的子空间。

PAST 信息准则对应的加权信息准则 (WPAST, weighted PAST) 为

$$\mathbf{W}^* = \arg \min J_2(\mathbf{W})$$

$$J_2(\mathbf{W}) = \text{tr}(\mathbf{W}) - 2\text{tr}(\mathbf{W}^T \mathbf{R} \mathbf{W}) + \text{tr}(\mathbf{W}^T \mathbf{R} \mathbf{W} \mathbf{W}^T \mathbf{W} \mathbf{D}) \quad (2)$$

其中, $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_r) \in \mathbb{R}^{r \times r}$ 为一个对角矩阵, 其作用是状态矩阵加权, 其各元素降序排列。类似于 PAST 算法的信息准则, 当 $\mathbf{W} = \mathbf{W}_n$ 时, $J_2(\mathbf{W})$ 取得最小值, 此时 \mathbf{W}_n 的各列向量恰好等于 \mathbf{R} 的前 r 个特征向量 ϕ 。显然, 当加权矩阵等于单位阵时, WPAST 信息准则退化为 PAST 信息准则。

以上 2 个信息准则求取相应跟踪提取算法是通

过梯度下降法实现的, 其基本的计算式为

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta \Delta \mathbf{W}(k) \quad (3)$$

其中, $\Delta \mathbf{W}(k)$ 为算法的搜索方向, 文献[15]指出, 该方向为信息准则的梯度下降方向, 即 $\Delta \mathbf{W}(k) = \frac{\partial J}{\partial \mathbf{W}}|_{\mathbf{W}=\mathbf{W}(k)}$; $\mathbf{W}(k)$ 为第 k 步迭代中的状态矩阵, 其维度与式(1)中 \mathbf{W} 相同; $\eta \in (0,1)$ 为次成分提取算法的学习因子。另外, 有些算法采用可变步长来加速算法的收敛过程, 但是只要步长设置合理就只影响算法的收敛速度而不影响其收敛行为。因此, 为了简化算法分析过程, 假设学习因子是固定长度的。

式(1)信息准则对 \mathbf{W} 的梯度为

$$\nabla J_1(\mathbf{W}) = (-2\mathbf{R} + \mathbf{R}\mathbf{W}\mathbf{W}^T + \mathbf{R}\mathbf{W}\mathbf{W}^T\mathbf{R})\mathbf{W} \quad (4)$$

根据牛顿梯度下降法, 可得

$$\frac{d\mathbf{W}}{dt} = -\nabla J_1(\mathbf{W}) \quad (5)$$

进而, 可以得到 PAST 算法的常微分方程形式为

$$\frac{d\mathbf{W}}{dt} = -2\mathbf{R}\mathbf{W} + \mathbf{R}\mathbf{W}\mathbf{W}^T\mathbf{W} + \mathbf{W}\mathbf{W}^T\mathbf{R}\mathbf{W} \quad (6)$$

文献[8]的收敛性分析表明, 该算法能够有效地在线跟踪信号的次子空间。定义状态向量和实际次成分方向的偏转矩阵为 $\mathbf{M}_1 \in \mathbb{R}^{n \times r}$, 实际次成分组成的矩阵为 $\mathbf{P} = [\phi_1, \phi_2, \dots, \phi_{n-1}]$ 。由于矩阵 \mathbf{M}_1 一定位于由 \mathbf{P} 张成的空间中, 于是状态矩阵 \mathbf{W} 可以改写为 $\mathbf{W} = \mathbf{P}\mathbf{M}_1$, 其中 \mathbf{M}_1 随着 \mathbf{W} 变化。当算法收敛后, \mathbf{M}_1 的前 r 行元素收敛为某个旋转矩阵, 后 $n-r$ 行元素收敛为 0。

同理, 式(2)信息准则经过梯度下降法推导, 可得 WPAST 算法的常微分方程形式, 如式(7)所示。

$$\frac{d\mathbf{W}}{dt} = -2\mathbf{R}\mathbf{W} + \mathbf{R}\mathbf{W}\mathbf{W}^T\mathbf{W}\mathbf{D} + \mathbf{W}\mathbf{W}^T\mathbf{R}\mathbf{W}\mathbf{D} \quad (7)$$

该算法能够有效地在线并行提取多个次成分。类似地, 存在一个 $\mathbf{M}_2 \in \mathbb{R}^{n \times r}$, 使 $\mathbf{W} = \mathbf{P}\mathbf{M}_2$ 。总体而言, 2 种算法的稳定点 $\mathbf{P}\mathbf{M}_1$ 和 $\mathbf{P}\mathbf{M}_2$ 都包含信号的次成分, 说明这 2 种算法都提取出信号中能量较小的部分信息。不同之处在于, \mathbf{M}_1 的前 r 行和前 r 列是一个旋转矩阵, 这代表 PAST 算法跟踪的次子空间与信号的真实次成分方向存在偏角, 而且这个偏角随跟踪过程的变化而变化。在 WPAST

算法提取多个次成分时, \mathbf{M}_2 的前 r 行和前 r 列元素最终收敛到一个对角矩阵 $\mathbf{D}^{-\frac{1}{2}}$ 。这说明与 PAST 算法相比, WPAST 算法因为加权而附加了角度变化的属性, 这个属性决定了 WPAST 算法具备多个次成分提取的能力, 而 PAST 算法则只能提取次子空间。本文通过分析 \mathbf{M}_1 和 \mathbf{M}_2 变化的差别, 从微观角度探究加权规则对算法的动力学特性的影响。

3 次成分提取过程动力学分析

不同维数提取过程有不同的角度数量。一维特征提取时, 次子空间退化为次特征向量, 此时 2 种算法效果一致。二维特征提取时, 旋转矩阵包含一个旋转角度。三维特征提取时, 旋转矩阵包含 2 个角度。依次类推, 多维次成分分析算法提取 r 维特征时, 旋转矩阵包含的角度数量为 $\frac{r(r-1)}{2}$ 个。

需要说明的是, 因一维特征提取算法不具有旋转角度信息, 此时只需要考虑模值收敛性, 此种情况已被确定离散时间 (DDT, determinate discrete time) 方法^[16]广泛分析。因此, 本文仅针对二维及以上的信号特征进行分析。下面从二维特征情况入手分析, 并以二维特征为基础, 逐步分析高维特征提取过程。

定理 1 考虑二维旋转矩阵 \mathbf{M}_1 只有一个旋转角度, 那么对信息准则 $J_1(\mathbf{W})$ 而言, 其对该旋转角度的二阶导数为 0。

证明 已知二维旋转矩阵只有一个旋转角, 假设 \mathbf{M}_1 的旋转角度为 θ , 则 \mathbf{M}_1 可以表示为

$$\mathbf{M}_1 = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \quad (8)$$

此时, 信息准则式(1)则表示为

$$\begin{aligned} J_1(\mathbf{W}) &= \text{tr}(\mathbf{R}) - 2\text{tr}(\mathbf{M}_1^T \mathbf{P}^T \mathbf{R} \mathbf{P} \mathbf{M}_1) + \\ &\text{tr}(\mathbf{M}_1^T \mathbf{P}^T \mathbf{R} \mathbf{P} \mathbf{M}_1 \mathbf{M}_1^T \mathbf{P}^T \mathbf{P} \mathbf{M}_1) = \\ &\text{tr}(\mathbf{R}) - 2\text{tr}(\mathbf{M}_1^T \mathbf{A} \mathbf{M}_1) + \text{tr}(\mathbf{M}_1^T \mathbf{A} \mathbf{M}_1 \mathbf{M}_1^T \mathbf{M}_1) \end{aligned} \quad (9)$$

信息准则对角度的梯度可以通过梯度下降法求得, 如式(10)所示。

$$\mathbf{P} \frac{d\theta}{dt} \frac{d\mathbf{M}_1}{d\theta} = \frac{\partial J_1}{\partial (\mathbf{P}\mathbf{M}_1)} \quad (10)$$

其中, 有

$$\frac{\partial J_1}{\partial \mathbf{M}_1} = \frac{\partial (\text{tr}(\mathbf{R}) - 2\text{tr}(\mathbf{M}_1^T \mathbf{A} \mathbf{M}_1) + \text{tr}(\mathbf{M}_1^T \mathbf{A} \mathbf{M}_1 \mathbf{M}_1^T \mathbf{M}_1))}{\partial \mathbf{M}_1} = -4\mathbf{A} \mathbf{M}_1 + 2\mathbf{A} \mathbf{M}_1 \mathbf{M}_1^T \mathbf{M}_1 + 2\mathbf{M}_1^T \mathbf{A} \mathbf{M}_1 \mathbf{M}_1 \quad (11)$$

$$\frac{d\mathbf{M}_1}{d\theta} = \begin{bmatrix} -\sin \theta & \cos \theta \\ -\cos \theta & -\sin \theta \end{bmatrix} \quad (12)$$

因而可以得出

$$\frac{d\theta}{dt} = \frac{\partial J_1}{\partial \mathbf{M}_1} \left(\frac{\partial \mathbf{M}_1}{\partial \theta} \right)^{-1} = 0 \quad (13)$$

式(13)表明不存在某一点, 使在该点处状态向量与信号次成分方向的夹角最小。也就是说, 信息准则 $J_1(\mathbf{W})$ 不能确定状态矩阵的各个向量与信号次成分方向的变化。

证毕。

信息准则对角度的二阶导数为 0, 说明信息准则没有规定旋转角度的变化方式。具体地, 在信息准则的任何一点上, 状态矩阵中各个向量与对应次成分的夹角的导数以及其本身是不确定的。另外, 从导数的角度理解定理 1, 则有在信息准则式(1)的梯度算法下, 提取出的次子空间与信号次成分的夹角和算法设置的状态矩阵初值相关。实际上, 加权规则能够改变信息准则的这个性质。具体过程可以通过定理 2 说明。

定理 2 考虑二维旋转矩阵 \mathbf{M}_2 只有一个旋转角度, 那么对信息准则 $J_2(\mathbf{W})$ 而言, 其对该旋转角度的二阶导数不为 0。

证明 令 \mathbf{M}_2 的旋转角度为 θ , 那么 \mathbf{M}_2 可以表示为

$$\mathbf{M}_2 = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \quad (14)$$

此时, 信息准则式(2)则表示为

$$J_2(\mathbf{P} \mathbf{M}_2) = \text{tr}(\mathbf{R}) - 2\text{tr}(\mathbf{M}_2^T \mathbf{P}^T \mathbf{R} \mathbf{P} \mathbf{M}_2) + \text{tr}(\mathbf{M}_2^T \mathbf{P}^T \mathbf{R} \mathbf{P} \mathbf{M}_2 \mathbf{M}_2^T \mathbf{P}^T \mathbf{P} \mathbf{M}_2 \mathbf{D}) = \text{tr}(\mathbf{R}) - 2\text{tr}(\mathbf{M}_2^T \mathbf{A} \mathbf{M}_2) + \text{tr}(\mathbf{M}_2^T \mathbf{A} \mathbf{M}_2 \mathbf{M}_2^T \mathbf{M}_2 \mathbf{D}) \quad (15)$$

同样地, 该信息准则对角度的梯度也可通过梯度下降式求解, 可得

$$\mathbf{P} \frac{d\theta}{dt} \frac{d\mathbf{M}_2}{d\theta} = \frac{\partial J_2}{\partial (\mathbf{P} \mathbf{M}_2)} \quad (16)$$

其中, 有

$$\frac{\partial J_2}{\partial \mathbf{M}_2} = \frac{\partial (\text{tr}(\mathbf{R}) - 2\text{tr}(\mathbf{M}_2^T \mathbf{A} \mathbf{M}_2) + \text{tr}(\mathbf{M}_2^T \mathbf{A} \mathbf{M}_2 \mathbf{M}_2^T \mathbf{M}_2 \mathbf{D}))}{\partial \mathbf{M}_2} = -4\mathbf{A} \mathbf{M}_2 + 2\mathbf{A} \mathbf{M}_2 \mathbf{M}_2^T \mathbf{M}_2 \mathbf{D} + 2\mathbf{M}_2^T \mathbf{A} \mathbf{M}_2 \mathbf{M}_2 \mathbf{D} \quad (17)$$

$$\frac{d\mathbf{M}_2}{d\theta} = \begin{bmatrix} -\sin \theta & \cos \theta \\ -\cos \theta & -\sin \theta \end{bmatrix} \quad (18)$$

此时, 可得

$$\frac{d\theta}{dt} = \frac{\partial J_2}{\partial (\mathbf{M}_2)} \left(\frac{\partial \mathbf{M}_2}{\partial \theta} \right)^{-1} = 2\mathbf{A} \begin{bmatrix} 2d_1 \cos 2\theta - 2\cos \theta & 2d_2 \sin 2\theta - 2\sin \theta \\ d_1 \sin 2\theta + 2\sin \theta & d_2 \cos 2\theta - 2\cos \theta + d_2 \end{bmatrix} \quad (19)$$

显然, $\frac{d^2\theta}{dt^2} = 0$ 不恒成立, 而是需要一定的成立

条件。这说明式(2)信息准则对状态向量与信号次成分方向的变化敏感。经过进一步推导可知, 当 $\theta = \frac{k\pi}{2}$, $k \in \mathbf{N}^*$ 时, $J_2(\mathbf{W})$ 取最小值。

信息准则 $J_2(\mathbf{W})$ 对角度二阶导数不为 0, 说明信息准则规定了旋转角度的变化方式。具体地, 在信息准则的任何一点上, 状态矩阵中各个向量与对应次成分的夹角的导数是确定的, 而且在极小值点处, 状态矩阵中各个向量与对应次成分的夹角的导数以及其本身都是确定的。

证毕。

以上分析是在 $n=2$ 、 $r=2$ 这一特殊情况下进行的。为使结论更具有一般性, 下面进一步讨论 $n>2$ 、 $r=2$ 的情况。

推论 1 当 $n>2$ 、 $r=2$ 时, 二维旋转矩阵 \mathbf{M}_i , $i=1,2$ 各只有一个旋转角度, 信息准则 $J_1(\mathbf{W})$ 对旋转角度的二阶导数为 0, 而信息准则 $J_2(\mathbf{W})$ 对旋转角度的二阶导数不为 0。

证明 文献[8]已经证明, 对于 $n>2$ 、 $r=2$ 的情况下的 $J_1(\mathbf{W})$ 和 $J_2(\mathbf{W})$, 对应的梯度算法都能够跟踪次子空间。即 \mathbf{W} 的后 $n-r$ 列元素总是逐渐变化为 0。不妨认为, 当迭代次数大于某一个大数时, 状态矩阵可以表达为 $\mathbf{W} = \mathbf{P}[\mathbf{M}_i^T \quad \mathbf{0}]^T$, $i=1,2$ 。问题转化为 $n=2$ 、 $r=2$ 的情况, 此时结合定理 1 和定理 2 的结论可知, 信息准则 $J_1(\mathbf{W})$ 对旋转角度的二阶导数为 0, 而信息准则 $J_2(\mathbf{W})$ 对旋转角度的二阶导数不为 0。

分析推论 1 的结论可知，只要提取二维特征向量，信息准则 $J_1(\mathbf{W})$ 对角度二阶导数为 0，即在 $J_1(\mathbf{W})$ 的任何一点上，状态矩阵中各个向量与对应次成分的夹角的导数以及其本身是不确定的。信息准则 $J_2(\mathbf{W})$ 对角度二阶导数不为 0，即在 $J_2(\mathbf{W})$ 的任何一点上，状态矩阵中各个向量与对应次成分的夹角的导数是确定的，并且在极小值点处，状态矩阵中各个向量与对应次成分的夹角的导数以及其本身都是确定的。

在 $n=r, r>2$ 的情况下，主要考虑欧拉转角描述所有的角度变化。假设第 j 个相角变化为 θ_j ， $j=1,2,\dots,r$ ，根据欧拉转角规则，第 j 个角是在第 $j-1$ 个角的基础上进一步旋转的。例如 $r=3$ 时旋转矩阵 \mathbf{M}_* 表示为

$$\mathbf{M}_* = \mathbf{M}_{*1} \mathbf{M}_{*2} \mathbf{M}_{*3} \quad (20)$$

其中， \mathbf{M}_{*1} 和 \mathbf{M}_{*2} 分别为

$$\mathbf{M}_{*1} = \begin{bmatrix} \cos \theta_1 & \sin \theta_1 & 0 \\ -\sin \theta_1 & \cos \theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{M}_{*2} = \begin{bmatrix} \cos \theta_2 & 0 & \sin \theta_2 \\ 0 & 1 & 0 \\ -\sin \theta_2 & 0 & \cos \theta_2 \end{bmatrix}$$

依次类推，将 1 放在不动轴的位置，并且该元素所在行和列的其他元素均为 0，不动轴的数量为 $r-2$ ，则旋转次数为 $C_2^r = \frac{r(r-1)}{2 \times 1}$ 。

推论 2 当 $n=2, r>2$ 时，信息准则 $J_1(\mathbf{W})$ 对旋转角度的二阶导数为 0，而对信息准则 $J_2(\mathbf{W})$ 对旋转角度的二阶导数不为 0。

证明 类似于定理 1 和定理 2 的证明步骤，信息准则式(1)表示为

$$\frac{\partial J_1}{\partial \mathbf{M}_1} = \frac{\partial \left(\begin{matrix} \text{tr}(\mathbf{R}) - 2\text{tr}(\mathbf{M}_1^T \mathbf{A} \mathbf{M}_1) + \\ \text{tr}(\mathbf{M}_1^T \mathbf{A} \mathbf{M}_1 \mathbf{M}_1^T \mathbf{M}_1) \end{matrix} \right)}{\partial \mathbf{M}_1} =$$

$$-4\mathbf{A} \mathbf{M}_1 + 2\mathbf{A} \mathbf{M}_1 \mathbf{M}_1^T \mathbf{M}_1 + 2\mathbf{M}_1^T \mathbf{A} \mathbf{M}_1 \mathbf{M}_1 \quad (21)$$

不同的是此时的旋转矩阵变为

$$\mathbf{M}_1 = \mathbf{M}_{11} \mathbf{M}_{12} \cdots \mathbf{M}_{1r} \quad (22)$$

同时，信息准则对角度的梯度表达为多个计算式，如下所示。

$$\mathbf{P} \frac{d\theta_1}{dt} \frac{d\mathbf{M}_1}{d\theta_1} = \frac{\partial J_1}{\partial (\mathbf{P} \mathbf{M}_1)} \quad (23)$$

$$\mathbf{P} \frac{d\theta_2}{dt} \frac{d\mathbf{M}_1}{d\theta_2} = \frac{\partial J_1}{\partial (\mathbf{P} \mathbf{M}_1)} \quad (24)$$

$$\vdots$$

$$\mathbf{P} \frac{d\theta_r}{dt} \frac{d\mathbf{M}_1}{d\theta_r} = \frac{\partial J_1}{\partial (\mathbf{P} \mathbf{M}_1)} \quad (25)$$

其中， $\frac{\partial J_1}{\partial \mathbf{M}_1} = 2\mathbf{A} \mathbf{M}_1$ ， $\frac{d\mathbf{M}_1}{d\theta_1} = \frac{d\mathbf{M}_{11}}{d\theta_1} \mathbf{M}_{12} \cdots \mathbf{M}_{1r}$ ，且

$$\frac{d\mathbf{M}_{11}}{d\theta_1} = \begin{bmatrix} -\sin \theta & \cos \theta & 0 & \cdots & 0 \\ -\cos \theta & -\sin \theta & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

。因为 \mathbf{M}_{1*} 具有

旋转对称性，为了描述简便，此处仅以 θ_1 为例进行说明。

由式(25)可得

$$\frac{d\theta_1}{dt} = \frac{\partial J_1}{\partial \mathbf{M}_1} \left(\frac{d\mathbf{M}_1}{d\theta_1} \right)^{-1} = 2\mathbf{A} \begin{bmatrix} 0 & -1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (26)$$

显然，对于未加权的的信息准则 $J_1(\mathbf{W})$ ， $\frac{d^2\theta}{dt^2} \equiv 0$ 对于任意 θ 恒成立。类似地，对加权信息准则，有

$$\frac{d\theta_1}{dt} = \frac{\partial J_2}{\partial \mathbf{M}_2} \left(\frac{d\mathbf{M}_2}{d\theta_1} \right)^{-1} = 2\mathbf{A} \begin{bmatrix} \Omega_{11} & \Omega_{12} & 0 & \cdots & 0 \\ \Omega_{21} & \Omega_{22} & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (27)$$

其中，有

$$\Omega_{11} = 2d_1 \cos 2\theta_1 - 2\cos \theta_1$$

$$\Omega_{12} = 2d_2 \sin 2\theta_1 - 2\sin \theta_1$$

$$\Omega_{21} = d_1 \sin 2\theta_1 + 2\sin \theta_1$$

$$\Omega_{22} = d_2 \cos 2\theta_1 - 2\cos \theta_1 + d_2$$

显然， $\frac{d^2\theta_1}{dt^2} = 0$ 不是总能成立，需要一定的成立条件。由于 \mathbf{M}_* 矩阵存在旋转对称性，因此该结论同样适用于 $\theta_2, \dots, \theta_r$ 。这说明，当 $n=2, r>2$ 时，式(1)信息准则对旋转角度的二阶导数为 0，而式(2)信息准则对旋转角度的二阶导数不为 0。

对于 $n > r, r > 2$ 的情况,其基本思路同推论 1,即通过分析一定条件下状态矩阵各个向量的长度变化情况,最终转化为 $n = r, r > 2$ 的情况,并结合推论 2 的结论实现对该情况的讨论。

证毕。

通过上述定理和推论可知,加权规则对提取算法产生影响本质因素是加权矩阵改变信息准则在方向上的梯度变化。这对研究的启发是在考虑将次子空间提取算法改进成并行提取算法时,其核心步骤是使算法在状态矩阵与实际次成分的夹角方向上产生梯度。

4 数值分析与仿真校验

为直观展示加权算法的方向收敛属性,本文通过 2 组数值算例对比 $J_1(W)$ 和 $J_2(W)$ 这 2 种信息准则的方向收敛性能。实验中,输入信号的真实次成分 $P = [\phi_0, \phi_1, \dots, \phi_{n-1}]$ 是通过离线的特征值分解方法计算得到的。

4.1 2 个次成分提取算例

该算例展示 2 种信息准则对状态矩阵各个向量与信号次成分方向提取的不同表现。加权矩阵规定为 $D = \text{diag}(1, 2)$ 。令 θ 表示状态矩阵的单个向量与对应次成分方向的夹角,设置 θ 的变化区间为 $(0, 2\pi]$, 实验中均匀提取了 10 000 个点,点间最小间距为 $\frac{\pi}{5000}$ 。通过 θ 可以表示旋转矩阵

$$M_* = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix},$$

通过 M_* 可以表示状态向量

$W = PM_*$, 状态向量在第一次成分方向的投影为 X , 在第二次成分方向的投影为 Y , 又根据式(1)和式(2)可以分别表示 J_1 和 J_2 , 仿真结果分别如图 1~图 4 所示。

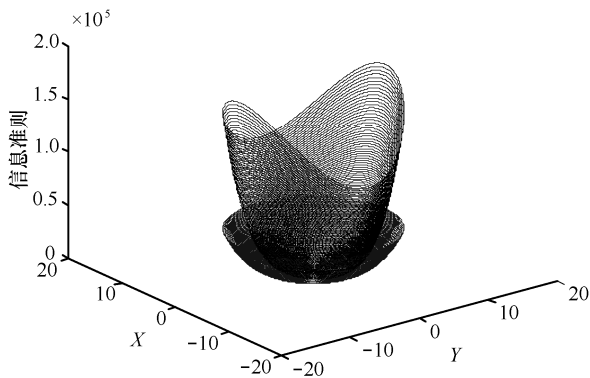


图 1 PAST 信息准则和 WPAST 信息准则在 θ 变化情况下的数值变化

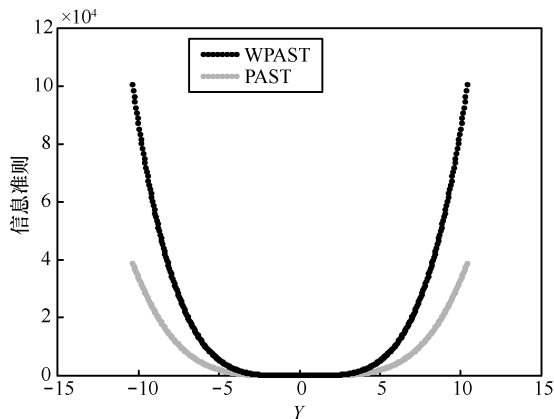


图 2 PAST 信息准则和 WPAST 信息准则在 $\theta = \frac{\pi}{2}$ 处的投影

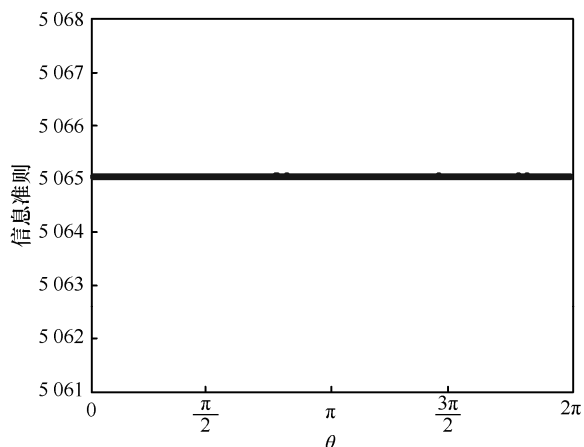


图 3 PAST 信息准则在 θ 变化情况下的数值变化

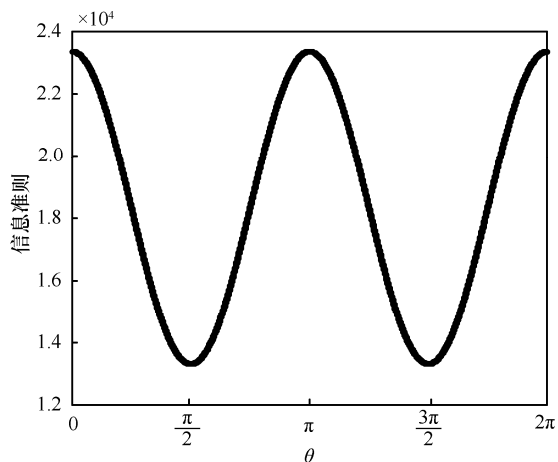


图 4 WPAST 信息准则在 θ 变化情况下的数值变化

图 1 中,外侧的曲面圆锥为 PAST 信息准则函数数值变化情况,内侧的曲面圆锥为 WPAST 信息准则函数数值变化情况。首先,函数值会随着状态矩阵中向量模值的变化而变化,为清晰地表征这个特点,该算例在某一个固定的角度投影 2 个信息准则

函数，如图 2 所示，这说明信息准则函数能够提取信号的次成分信息。其次，2 个曲面变化都具有对称性。其中，内侧的曲面圆锥是随着 θ 周期对称的，这印证了定理 2 的结论，而外侧的曲面圆锥是处处对称的。

另外，2 个曲面的变化也存在差别。首先，观察向量模长固定， θ 在区间 $(0, 2\pi]$ 变化时，2 个函数的变化情况。由图 1 中外侧的曲面圆锥可知，在 θ 变化的情况下，PAST 信息准则函数值没有变化。为了直观展示，在图 1 中做状态矩阵模长为 1.5、 θ 在区间 $(0, 2\pi]$ 上的圆柱面，将圆柱面与函数曲面圆锥的交线向底面投影并展开，展开的结果如图 3 所示。由图 1 中内侧的曲面圆锥可知，WPAST 信息准则函数值在 θ 变化的情况下，呈现三角函数的变化形态。为了直观展示，采用与外侧圆锥曲面相同的步骤得到投影的展开结果，并展示在图 4 中。其次，观察 θ 值一定时 2 个函数沿着模值长度的变化情况。如图 2 所示，WPAST 函数的斜率绝对值明显大于 PAST 函数的斜率绝对值。

上述实验分析表明，在提取 2 个次成分方向的情况下，PAST 信息准则不能确定状态矩阵各向量与对应次成分的方向关系，而 WPAST 信息准则能够确定状态矩阵各向量与对应次成分的方向关系。这验证了定理 1 和定理 2 对 PAST 和 WPAST 信息准则的理论分析结果。

4.2 多个次成分提取算例

多个次成分的信息准则对应的坐标系超过三维，不容易直观表示，因此该算例主要通过 PAST 和 WPAST 信息准则的提取算法在实际提取过程中的角度变化来展示加权规则对信息准则的角度关系的确定作用。假设提取状态矩阵的维数为 4，加权矩阵为 $D = \text{diag}(1, 2, 3, 4)$ ，状态矩阵各个向量与对应次成分的夹角分别为 θ_1 、 θ_2 、 θ_3 、 θ_4 ，状态矩阵的各向量与对应次成分的方向余弦为

$$Y(i, k) = \frac{\|\mathbf{w}_i^T(k)\mathbf{u}_i\|}{\|\mathbf{w}_i(k)\|\|\mathbf{u}_i\|} \quad (28)$$

其中， \mathbf{u}_i 表示输入信号的次成分，可通过特征值分解方法提前计算出来， $\mathbf{w}_i(k), i \in \{1, 2, 3, 4\}$ 表示算法在提取过程中的状态矩阵的单个向量， k 表示提取过程中的迭代次数， T 表示转置运算， $\|\cdot\|$ 表示求范数运算。

夹角的计算式为

$$\theta_i = \arccos(Y(i, k)) \quad (29)$$

需要说明的是，此处将方向余弦和方向夹角都作为实验结果展示。PAST 算法的结果分别如图 5 和图 6 所示，WPAST 算法的结果分别如图 7 和图 8 所示。

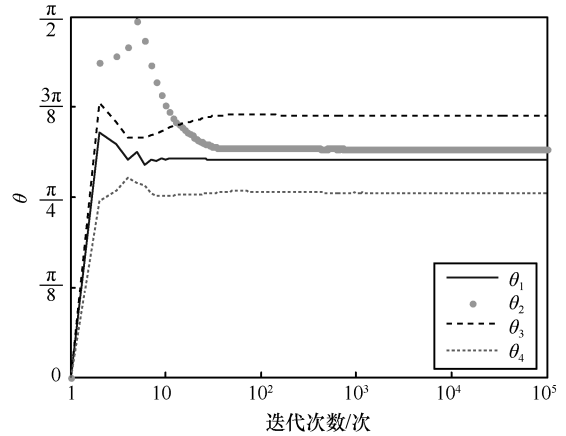


图 5 PAST 算法的状态矩阵各向量与对应次成分夹角 θ 变化情况

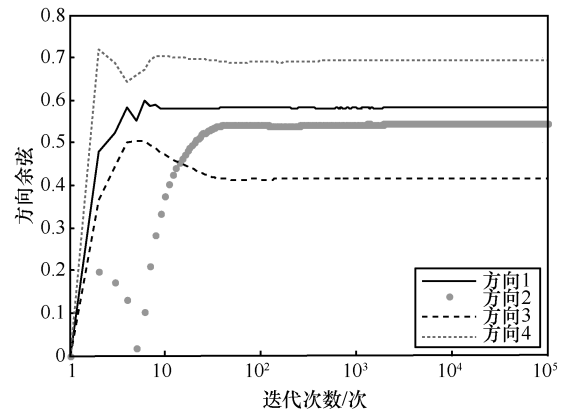


图 6 PAST 算法的状态矩阵各向量与对应次成分方向余弦变化情况

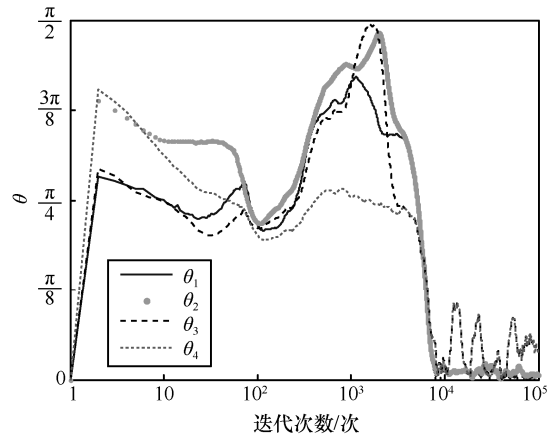


图 7 WPAST 算法的状态矩阵各向量与对应次成分夹角 θ 变化情况

如图 5 所示，横坐标为迭代次数，按照 10 的指数变化，纵坐标为方向夹角数值，其弧度值在区

间 $\left(0, \frac{\pi}{2}\right)$ 内变化。PAST 算法的 4 个方向夹角在经过振荡过程后, 均呈现稳定状态。但是, 稳定的常数各不相同。该特点在图 6 中也有体现, 方向余弦经过一段振荡过程, 都稳定在一定的数值上。相比 PAST 算法, WPAST 算法的方向夹角收敛过程比较长, 且最终都稳定在 0 附近。如图 7 所示, 4 个方向夹角最后都以不同的形式收敛到 0 附近, 这说明 WPAST 算法状态矩阵的 4 个向量最后都收敛到对应次成分的方向。图 8 展示的方向余弦也印证了 WPAST 算法的方向收敛性。在 5×10^3 次迭代后, 状态矩阵各向量的方向余弦都接近 1。需要说明的是, 由于三角函数的非线性, 在图 7 中接近 10^5 处的方向夹角振荡较剧烈, 且都在区间 $\left(0, \frac{\pi}{8}\right)$ 内, 图 8 中对应的方向余弦都在区间(0.95,1.00) 内振荡, 振荡范围非常小。

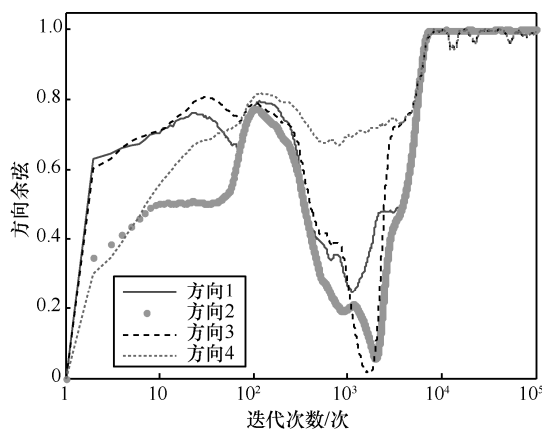


图 8 WPAST 算法的状态矩阵各向量与对应次成分方向余弦变化情况

5 结束语

本文通过分析 PAST 信息准则和 WPAST 信息准则的差异, 从二维到多维逐步推导出加权规则能够改变状态矩阵和信号次成分的旋转矩阵梯度的结论。一方面, 从理论上解释了加权规则对信息准则的作用规律, 提高人们在信息准则层次认识影响算法性能的因素; 另一方面, 对于不能通过加权规则转化为多个次成分提取算法的次子空间跟踪算法, 研究成果为推进未来并行提取多个次成分分析算法发展提供了研究思路和改进方向。

参考文献:

[1] GAO Y B, KONG X Y, HU C H, et al. Convergence analysis of Möller algorithm for estimating minor component [J]. Neural Processing Let-

ters, 2015,42(2): 268-355.
 [2] KONG X Y, HU C H, HAN C Z. A self-stabilizing MSA algorithm in high-dimensional data stream [J]. Neural Networks, 2010, 23(7): 865-871.
 [3] KONG X Y, HU C H, HAN C Z. A dual purpose principal and minor subspace gradient flow [J]. IEEE Transactions on Signal Processing, 2012, 60(1): 197-210.
 [4] NGUYEN D T, YAMADA I. A unified convergence analysis of normalized PAST algorithms for estimating principal and minor components [J]. Signal Processing, 2013, 93(1): 176-184.
 [5] MÖLLER R. A self-stabilizing learning rule for minor component analysis[J]. International Journal of Neural Systems, 2004, 14(1): 1-8.
 [6] CIRRINCIONE G, CIRRINCIONE M. Neural-based orthogonal data fitting: the EXIN neural networks[J]. Aevum, 2010(2):368-371.
 [7] GAO Y B, KONG X Y, ZHANG H H, et al. A weighted information criterion for multiple minor components and its adaptive extraction algorithms[J]. Neural networks, 2017, 89(5): 1- 10.
 [8] YANG B. Projection approximation subspace tracking [J]. IEEE transactions on signal processing, 1995, 43 (1): 95-107.
 [9] FENG D Z, ZHENG W, JIA Y. Neural network learning algorithms for tracking minor subspace in high-dimensional data stream[J]. IEEE Transactions on Neural Networks, 2005, 16(3): 513-521.
 [10] JANKOVIC M, OGAWA H. Time-oriented hierarchical method for computation of minor components[M]. Berlin: Springer, 2005.
 [11] TAN K K, LYU J, ZHANG Y, et al. Adaptive multiple minor directions extraction in parallel using a PCA neural network[J]. Theoretical Computer Science, 2010, 411(48): 4200-4215.
 [12] THAMERI M, ABED-MERAIM K, BELOUCHRANI A. Low complexity adaptive algorithms for principal and minor component analysis[J]. Digital Signal Processing, 2013, 23(1): 19-29.
 [13] OUYANG S, BAO Z. Fast principal component extraction by a weighted information criterion [J]. IEEE Transactions on Signal Processing, 2002, 50(8): 1994-2002.
 [14] BARTELMAOS S, ABED-MERAIM K. Fast adaptive algorithms for minor component analysis using Householder transformation [J]. Digital Signal Processing, 2011, 21(6): 667-678.
 [15] TOSHIIHISA T. Generalized weighted rules for principal components tracking [J]. IEEE Transactions on Signal Processing, 2005, 53(4): 1243-1253.
 [16] ZHANG Y, YE M, LYU J C, et al. Convergence analysis of a deterministic discrete time system of Oja's PCA learning algorithm [J]. IEEE Transactions on Neural Networks, 2005, 16(6): 1318-1328.

[作者简介]



杜柏阳 (1990-), 男, 山东滨州人, 火箭军工程大学博士生, 主要研究方向为信号特征提取、大型工业过程故障诊断。

孔祥玉 (1967-), 男, 山西临汾人, 博士, 火箭军工程大学教授、博士生导师, 主要研究方向为信号特征提取、复杂非线性系统建模、大型工业过程故障监测与故障诊断。

冯晓伟 (1986-), 男, 四川绵阳人, 博士, 火箭军工程大学讲师, 主要研究方向为信号特征提取、大型工业过程故障监测与故障诊断。