

主成分提取信息准则的加权规则

杜柏阳^{1,2} 孔祥玉¹ 罗家宇¹

摘要 并行主成分提取算法在信号特征提取中具有十分重要的作用, 采用加权规则将主子空间 (Principal subspace, PS) 提取算法转变为并行主成分提取算法是很有效的方式, 但研究加权规则对状态矩阵影响的理论分析非常少. 对加权规则影响的分析不仅可以提供加权规则下的主成分提取算法动力学的详细认知, 而且对于其他子空间跟踪算法转变为并行主成分提取算法的可实现性给出判断条件. 本文通过比较 Oja 的主子空间跟踪算法和加权 Oja 并行主成分提取算法, 通过两种算法的差异分析了加权规则对算法提取矩阵方向的影响. 首先, 针对二维输入信号, 研究了提取两个主成分时加权规则的信息准则对状态矩阵方向的作用方式. 进而, 针对大于二维输入信号的情况, 给出加权规则影响多个主成分提取方式的讨论. 最后, MATLAB 仿真验证了所提出理论的有效性.

关键词 主子空间提取算法, 加权规则, 并行主成分分析, 信息准则

引用格式 杜柏阳, 孔祥玉, 罗家宇. 主成分提取信息准则的加权规则. 自动化学报, 2021, 47(12): 2815–2822

DOI 10.16383/j.aas.c190226

Weighted Rules for Principal Components Extraction Information Criteria

DU Bo-Yang^{1,2} KONG Xiang-Yu¹ LUO Jia-Yu¹

Abstract The parallel principal component extraction algorithms play an important role in signal feature extraction. It is very effective and very useful to modify the principal subspace (PS) extraction algorithms into parallel principal components extraction algorithms by using weighted rules, but so far, few people theoretically analyze the variation of the state matrix under the extraction algorithm of weighted rules. The analysis of this variation can not only provide detailed knowledge of the dynamics of the principal component extraction algorithms under weighted rules, but also give the realization conditions for transferring subspace extraction algorithms into parallel principal components extraction algorithm. In this paper, by comparing the difference between Oja principal subspace extraction algorithm and weighted Oja parallel principal components extraction algorithm, the influence of weighted rules on the extraction direction is analyzed. Firstly, for the case of extracting two principal components, the influence of the information criterion under the weighted rule on the direction of the state matrix is studied. Furthermore, for the case of extracting more than two principal components, a discussion is given on the manner where the weighted rules are applied. Finally, the MATLAB simulation verifies the validity of the proposed theory.

Key words Principal subspace extraction algorithms, weighted rule, parallel principal components analysis, information criterion

Citation Du Bo-Yang, Kong Xiang-Yu, Luo Jia-Yu. Weighted rules for principal components extraction information criteria. *Acta Automatica Sinica*, 2021, 47(12): 2815–2822

在信息处理领域, 主成分分析 (Principal component analysis, PCA) 又称为 KL 变换 (Karhunen-Loève transform), 广泛应用于信号压缩^[1]、模式识别^[2]、图像处理^[3]、噪声估计^[4]、神经网络建模^[5] 等问题. 通常, 主成分 (Principal component, PC) 可以通过求解自相关矩阵的特征分解求得, 具体是指自

相关矩阵中前个最大特征值对应的特征向量. 由这些特征向量张成的特征子空间被称为主子空间 (Principal subspace, PS).

当前, 在大量跟踪输入信号主成分主子空间的算法中, 神经网络算法因为计算复杂度低等多种优良的性能而引起许多学者的兴趣^[6]. 学者们提出大量的神经网络学习算法, 用于提取主成分或者跟踪主子空间. 例如, 建立在启发推理机制上的 Oja 主子空间跟踪算法^[7]、对称误差校正算法^[8], 以及建立在信息准则基础上的最小均方误差重构 (Least mean square error reconstruction, LMSER) 算法^[9]、投影近似子空间跟踪 (Projection approximation subspace tracking, PAST) 算法^[10].

收稿日期 2019-03-21 录用日期 2019-07-30
Manuscript received March 21, 2019; accepted July 30, 2019
国家自然科学基金 (61374120, 61673387) 资助
Supported by National Natural Science Foundation of China (61374120, 61673387)
本文责任编辑 黎铭
Recommended by Associate Editor LI Ming
1. 西安高科技研究所 西安 710025 2. 96901 部队 北京 100094
1. The Xi'an Researching Institute of High Technology, Xi'an 710025 2. The 96901 Unit of PLA, Beijing 100094

早期的算法结构简单,但是在自稳定性、平滑性以及鲁棒性等方面还有待于优化.针对这些问题,Kong等^[11]提出了双目的的自稳定主成分提取神经网络算法,Kakimoto等^[12]提出了基于最小噪声准则的平滑自适应特征提取方法,Ouyang等^[13]提出了递归的鲁棒主成分分析算法.不仅如此,算法提取的对象还扩展到信号次成分^[14]和广义特征成分^[15].通过研究这些算法,可以知道提取信号多个成分和跟踪信号子空间的算法虽然实现提取的对象不同,但是在算法的原理上存在一定程度的关联.以主成分提取算法为例分析,多个主成分提取算法和主子空间跟踪算法都能够退化为单一主成分的提取算法,这反映出两个类型的算法具有某种相同的基本特性.反过来,如果存在一个能够提取单个主成分的提取算法,是否能遵循一定的原理或者规律,使得该算法能够转化为提取多个主成分或者跟踪主子空间的算法?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.

多个主成分可以张成一个主子空间,因而通常认为多个主成分的提取算法是主子空间跟踪算法的进步.一般而言,多个主成分的提取算法的提出方式有两种.一种是根据实际需求直接提出一种新的信息准则,通过推导信息准则的梯度函数得出对应的提取算法;另一种是在已有的主子空间跟踪算法的基础上,引入加权规则,使得转变后的算法能够提取多个主成分.Tanaka^[16]通过研究多个主成分提取算法的广义加权规则,分析出广义加权规则的参数对提取算法的收敛速度有影响.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阵,算法的提取能力逐渐由多个主成分提取退化为主子空间跟踪.

实际上,有的主子空间提取算法在使用加权规则后可以转化为多个主成分的提取算法,而有的算法则不具有这种能力.主子空间的组成向量与主成分之间的夹角能够说明主子空间与主成分的偏离程度.有的主子空间跟踪算法能够转化为多个主成分提取算法,原因在于这些跟踪算法在运行过程中能够减小夹角的大小.而本质上,确定这个夹角关系的是信息准则,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具有非直接的规定.因此,在信息准则的角度分析加权规则对主子空间跟踪算法和多个主成分提取算法的作用能够反映算法的一些本质特性.

本文主要针对加权规则对主成分提取算法的信

息准则的作用进行分析,以Oja主子空间跟踪算法为例,通过构建提取算法的动力学表达,对比存在和缺失加权规则下的主子空间跟踪信息准则,挖掘出信息准则对状态矩阵与主成分的方向夹角的梯度差异.

1 加权规则

根据文献[7]和文献[17]可知,Oja算法的信息准则为

$$J_1(W) = \text{tr}(W^T R W) \quad \text{s.t.} \quad W^T W = I_r \quad (1)$$

其中, $I_r \in \mathbf{R}^{r \times r}$ 表示单位阵, $W \in \mathbf{R}^{n \times r}$ 为算法的状态矩阵, n 为输入信号的维数, r 为需要提取的主成分个数, $R = E[\mathbf{x}\mathbf{x}^T] \in \mathbf{R}^{n \times n}$ 为随机输入信号 \mathbf{x} 的自相关矩阵, $E[\cdot]$ 表示计算方括号内的数学期望.假设 R 是满秩矩阵,特征值互不相同.那么,输入信号有 n 个特征向量 $\phi_i, i = 0, 1, \dots, n-1$,容易理解,当存在一个矩阵 $W = W_n$,使得信息准则函数值 $J_1(W)$ 能够取得最大时, W_n 就等价于信号的主子空间,也就是前 r 个特征向量 ϕ_i 张成的子空间.

另外,Oja还首先提出加权规则下的信息准则^[18]

$$J_2(W) = \text{tr}(D^{\frac{1}{2}} W^T R W D^{\frac{1}{2}}) \quad \text{s.t.} \quad W^T W = D^{-1} \quad (2)$$

其中, $D = \text{diag}\{d_1, d_2, \dots, d_r\} \in \mathbf{R}^{r \times r}$ 为一个对角矩阵,它的作用就是为状态向量加权.在主成分提取的算法中,加权矩阵的各个元素通常设置为 $d_1 > d_2 > \dots > d_r$.其余变量的含义与式(1)相同.类似于Oja算法的信息准则,当 $W = W_n$ 时, $J_2(W)$ 能够取得最大时,此时 W_n 不仅是前 r 个特征向量 ϕ_i 张成的子空间,而且组成 W_n 的各个向量与对应的特征向量 ϕ_i 相同.

通过以往研究可知,由以上两个信息准则求取相应跟踪或者提取算法通过梯度下降法实现,其基本的表达式为

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

其中, $\eta \Delta W(k)$ 为算法的搜索方向,根据文献[16]可知,该方向为信息准则的梯度下降方向 $\eta \Delta W(k) = \frac{\partial J}{\partial W} |_{W(k)} = W(k)$; $W(k)$ 为第 k 步迭代中的状态矩阵,维度与式(1)中 W 相同; $\eta \in (0, 1)$ 为算法的学习因子.实际上,有的加速算法中的学习因子并不是固定长度的,为了简化算法分析过程,此处假设学习因子是固定长度的.

相应地,由信息准则(1)推导得到的算法常微分方程形式^[7]为

$$\frac{dW}{dt} = RW - WW^T RW \quad (4)$$

该算法能够有效地在线跟踪信号的主子空间. 假设主子空间相对于多个主成分的旋转矩阵为 $M_1 \in \mathbf{R}^{n \times r}$, $P = [\phi_0, \phi_1, \dots, \phi_{r-1}]$, 于是有 $W = PM_1$, 主子空间的跟踪过程就是 M_1 的前 r 行元素变化为某个旋转矩阵, 后 $n-r$ 行元素变化为零的过程.

由信息准则 (2) 推导得到的算法常微分方程形式^[8]为

$$\frac{dW}{dt} = RW - WW^T RWD \quad (5)$$

该算法也能够有效地在线跟踪信号的主子空间. 与上文相似, 存在一个 $M_2 \in \mathbf{R}^{n \times r}$, $W = PM_2$. 这说明两种算法的信息准则都能够敏感到前 r 个特征值对应的特征信息. 而不同之处在于, 主子空间的跟踪过程中, M_2 的前 r 行元素变化为一个对角阵 $D^{-\frac{1}{2}}$. 本文主要针对 M_1 和 M_2 变化的差别展开分析, 从微观角度探究加权规则对算法作用的动力学变化.

2 主成分提取过程动力学分析

只考虑一维特征的情况可以通过确定离散时间 (Determinate discrete time, DDT) 方法分析^[9]. 直接分析高维特征提取过程非常复杂, 不容易说明清楚. 因而首先从二维特征情况开始.

定理 1. 信息准则 $J_2(W)$ 不能确定状态矩阵的各个向量与信号主成分方向之间的关系.

证明. 假设 M_1 的旋转角度为 θ , 那么 M_1 可以写为

$$M_1 = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \quad (6)$$

此时, 信息准则 (1) 表示为

$$J_1(PM_1) = \text{tr}(M_1^T P^T R P M_1) = \text{tr}(M_1^T A M_1) \quad (7)$$

那么, 信息准则对角度的梯度需要通过梯度下降公式解为

$$P \frac{d\theta}{dt} \frac{dM_1}{d\theta} = \frac{\partial J_1}{\partial P M_1} \quad (8)$$

其中,

$$\frac{\partial J_1}{\partial M_1} = \frac{\partial \text{tr}(M_1^T A M_1)}{\partial M_1} = 2A M_1 \quad (9)$$

$$\frac{dM_1}{d\theta} = \begin{bmatrix} -\sin(\theta) & \cos(\theta) \\ -\cos(\theta) & -\sin(\theta) \end{bmatrix} \quad (10)$$

因而可以计算得

$$\frac{d\theta}{dt} = \frac{\partial J_1}{\partial M_1} \left(\frac{\partial M_1}{\partial \theta} \right)^{-1} = 2A \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad (11)$$

上式很容易说明, $\frac{d^2\theta}{dt^2} \equiv 0$ 对于任意 θ 恒成立.

这说明不存在某一点内容使得在该点处状态向量与信号主成分方向的夹角最小. 也就是说, 信息准则 $J_1(W)$ 不能确定状态向量与信号主成分方向的关系. \square

注 1. 定理 1 可得的另外一个结论是, 信息准则 (1) 的梯度算法提取出的主子空间与信号主成分存在一个夹角, 并且该夹角与算法设置的状态矩阵初值有关.

注 2. 信息准则 $J_1(W)$ 并不能直接导出状态矩阵的变化方式, 通常是通过梯度下降法、牛顿法、Nestrove 加速下降法等等利用信息准则的一阶、二阶甚至是高阶导数或者偏导数得到迭代更新公式. 当迭代公式收敛时, 对信息准则 $J_1(W)$ 而言, 收敛结果只能确保状态矩阵的模值收敛到某固定值而不能确定状态矩阵的各个向量与输入信号自相关矩阵的主成分方向的关系. 信息准则在本质上规定了状态矩阵的变化方式.

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. 信息准则 $J_2(W)$ 能够确定状态矩阵的各个向量与信号主成分方向之间的关系.

证明. 假设 M_2 的旋转角度为 θ , 那么 M_2 可以写为

$$M_2 = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \quad (12)$$

此时, 信息准则 (2) 表示为

$$J_2(PM_2) = \text{tr}(D^{\frac{1}{2}} M_2^T P^T R P M_2 D^{\frac{1}{2}}) = \text{tr}(D^{\frac{1}{2}} M_2^T A M_2 D^{\frac{1}{2}}) \quad (13)$$

同样地, 该信息准则对角度的梯度也可通过梯度下降公式解为

$$P \frac{d\theta}{dt} \frac{dM_2}{d\theta} = \frac{\partial J_2}{\partial (P M_2)} \quad (14)$$

其中,

$$\frac{\partial J_2}{\partial M_2} = \frac{\partial \text{tr}(D^{\frac{1}{2}} M_2^T A M_2 D^{\frac{1}{2}})}{\partial M_2} = 2A M_2 D^{\frac{1}{2}} \quad (15)$$

$$\frac{dM_2}{d\theta} = \begin{bmatrix} -\sin(\theta) & \cos(\theta) \\ -\cos(\theta) & -\sin(\theta) \end{bmatrix} \quad (16)$$

此时, 可得

$$\frac{d\theta}{dt} = \frac{\partial J_2}{\partial M_2} \left(\frac{\partial M_2}{\partial \theta} \right)^{-1} = 2\Lambda \begin{bmatrix} \frac{d_2-d_1}{2} \sin(2\theta) & \frac{d_2-d_1}{2} \cos(2\theta) - \frac{d_2+d_1}{2} \\ \frac{d_2+d_1}{2} \cos(2\theta) + \frac{d_2-d_1}{2} & -\frac{d_2-d_1}{2} \sin(2\theta) \end{bmatrix} \quad (17)$$

显然, $\frac{d^2\theta}{dt^2} = 0$ 不是总能成立, 需要一定的条件. 这说明信息准则 (2) 可以确定状态向量与信号主成分方向的关系. 并且, 通过一些简单地推导可知当 $\theta = k\pi/2, k \in \mathbf{N}^*, J_2(W)$ 能够取得最小值. \square

以上分析本质上是在 $n = 2, r = 2$ 这种特殊情况情况的讨论. 下面需要对上述结论推广到 $n > 2, r = 2$ 的情况.

推论 1. 对于 $n > 2, r = 2$, 信息准则 $J_1(W)$ 不能规定状态向量与信号主成分方向的关系, 而信息准则 $J_2(W)$ 能够规定状态向量与信号主成分方向的关系.

证明. 显然对于 $n > 2, r = 2$ 情况下的 $J_1(W)$ 和 $J_2(W)$ 而言, 对应的梯度算法都能够跟踪主子空间, 这在原作者的论文中已有明确的证明. 也就是说, W 的后 $n - r$ 行元素总是逐渐变化为零. 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 $W = P[M_*^T \ 0]^T$, 其中, $*$ 位置为 1 或 2. 此时根据定理 1 和定理 2 的分析过程得知, 信息准则 $J_1(W)$ 不能敏感状态向量与信号主成分方向的变化, 而信息准则 $J_2(W)$ 能够敏感状态向量与信号主成分方向的变化. \square

对于 $n = r, r > 2$ 的情况而言, 主要考虑欧拉转角描述所有的角度变化. 第 1 个相角变化为 θ_1 , 第 2 个相角变化为 θ_2 , 以至于第 r 个相角变化为 θ_r . 每一个角总是在上一个旋转角度基础上继续旋转, 例如 $r = 3$ 时旋转矩阵 M_* 表示为

$$M_* = M_{*1} M_{*2} M_{*3} \quad (18)$$

其中,

$$M_{*1} = \begin{bmatrix} \cos(\theta_1) & \sin(\theta_1) & 0 \\ -\sin(\theta_1) & \cos(\theta_1) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$M_{*2} = \begin{bmatrix} \cos(\theta_2) & 0 & \sin(\theta_2) \\ 0 & 1 & 0 \\ -\sin(\theta_2) & 0 & \cos(\theta_2) \end{bmatrix}$$

$$M_{*3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_3) & \sin(\theta_3) \\ 0 & -\sin(\theta_3) & \cos(\theta_3) \end{bmatrix}$$

以此类推. 即将 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 $r - 2$. 旋转次数为 $C_2^r = \frac{r \times (r - 1)}{2 \times 1}$.

推论 2. 对于 $n = r, r > 2$, 信息准则 $J_1(W)$ 不能敏感状态向量与信号主成分方向的变化, 而信息准则 $J_2(W)$ 能够敏感状态向量与信号主成分方向的变化.

证明. 与定理 1 和定理 2 的证明步骤相似, 信息准则 (1) 表示为

$$J_1(PM_1) = \text{tr}(M_1^T P^T R P M_1) = \text{tr}(M_1^T \Lambda M_1) \quad (19)$$

此时的旋转矩阵变为

$$M_1 = M_{11} M_{12} \cdots M_{1r} \quad (20)$$

同时信息准则对角度的梯度表达为多个公式

$$P \frac{d\theta_1}{dt} \frac{dM_{11}}{d\theta_1} = \frac{\partial J_1}{\partial P M_{11}} \quad (21)$$

$$P \frac{d\theta_2}{dt} \frac{dM_{12}}{d\theta_2} = \frac{\partial J_1}{\partial P M_{12}} \quad (22)$$

直到

$$P \frac{d\theta_r}{dt} \frac{dM_{1r}}{d\theta_r} = \frac{\partial J_1}{\partial P M_{1r}} \quad (23)$$

其中, $\frac{\partial J_1}{\partial M_{11}} = 2\Lambda M_{11}, \frac{dM_1}{d\theta_1} = \frac{dM_{11}}{d\theta_1} M_{12} \cdots M_{1r}$, 且有

$$\frac{dM_{11}}{d\theta_1} = \begin{bmatrix} -\sin(\theta_1) & \cos(\theta_1) & 0 & \cdots & 0 \\ -\cos(\theta_1) & -\sin(\theta_1) & 0 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

因为 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.

由此可得,

$$\frac{d\theta_1}{dt} = \frac{\partial J_1}{\partial M_1} \left(\frac{\partial M_1}{\partial \theta} \right)^{-1} = 2\Lambda \begin{bmatrix} 0 & -1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (24)$$

显然, 对于未加权的信息准则 $J_1(W)$ 而言, $\frac{d^2\theta}{dt^2} \equiv 0$ 对于任意 θ_1 恒成立. 与此同时, 对于加权信息准则 $J_2(W)$ 而言, 经过与上述推导过程相似的步骤可得,

$$\frac{d\theta_1}{dt} = \frac{\partial J_1}{\partial M_2} \left(\frac{\partial M_2}{\partial \theta} \right)^{-1} = 2A \begin{bmatrix} \Omega_{11} & \Omega_{12} & 0 & \cdots & 0 \\ \Omega_{21} & \Omega_{22} & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (25)$$

其中,

$$\begin{aligned} \Omega_{11} &= \frac{d_2 - d_1}{2} \sin(2\theta_1) \\ \Omega_{12} &= \frac{d_2 - d_1}{2} \cos(2\theta_1) - \frac{d_2 + d_1}{2} \\ \Omega_{21} &= \frac{d_2 + d_1}{2} \cos(2\theta_1) + \frac{d_2 - d_1}{2} \\ \Omega_{22} &= -\frac{d_2 - d_1}{2} \sin(2\theta_1) \end{aligned}$$

显然, $\frac{d^2\theta_1}{dt^2} = 0$ 不是总能成立, 需要一定条件.

由于 M_* 矩阵存在旋转对称性, 所以该结论对 $\theta_2 \cdots \theta_r$ 也成立. 这说明信息准则 (1) 不能敏感状态向量与信号主成分方向的变化, 而信息准则 (2) 可以敏感状态向量与信号主成分方向的变化. \square

注 3. 对于 $n > r, r > 2$ 情况的讨论, 其基本思路同推论 1, 都是考虑在一定条件下, 状态矩阵各个向量的长度发生变化, 最终转化为 $n = r, r > 2$ 的情况, 结合推论 2 的结论实现对该情况的讨论.

注 4. 通过定理和推论的分析可知, 加权规则对提取算法产生影响的本质因素是加权算法改变信息准则在方向上的梯度变化. 这对后续该类算法研究的启发是, 在考虑将主子空间提取算法转变成并行提取算法时, 其核心步骤是使算法在状态矩阵与实际主成分的夹角方向上产生梯度. 为验证该思路对于 Oja 信息准则以外的信息准则也有效, 本文考察 Miao 信息规则在加权规则下的性质变化.

3 数值分析与仿真实验

为保证算法具有可比性, 在生成数据的时候对输入信号做出统一规定. 各个仿真实验的输入信号自相关矩阵的特征值为同一组数值, 在此考虑 $\Lambda = \text{diag} \{1.052, 2.335, 5.002, 5.251, 6.012, 6.123\}$. 通过算法提取出信息的主成分 $P = [\phi_0, \phi_1, \cdots, \phi_{n-1}]$. 另外, 加权矩阵的数值也是任意选取一组满足定义的数值, 此处选为 $P = \text{diag} \{1, 2\}$.

3.1 Oja 信息准则与 Oja 加权信息准则

该算例从角度变化方面展示两种信息准则的不同表现. 设置 θ 变化区间为 $(0, 2\pi]$, 不妨均匀提取

10 000 个点, 最小间距为 $\pi/5 000$. 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 $W = PM_*$, 又根据式 (1) 和式 (2) 可以分别表示 J_1 和 J_2 . 具体的仿真结果如图 1 和图 2 所示.

如图 1 所示, 外侧的曲面圆锥为 Oja 信息准则函数值变化, 内侧的曲面圆锥为加权 Oja 信息准则函数值变化. 首先, 两个函数值变化情况共同印证了信息准则能够良好地敏感状态矩阵的模值变化. 为更加清晰地表征这个特点, 该算例在某一个固定的角度投影两个信息准则函数, 如图 2. 其次, 两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, 印证了定理 2 中的一个结论, 外侧的曲面圆锥是处处对称的.

另外, 两个曲面的变化也是存在差别的. 这首先表现在 θ 从 $(0, 2\pi]$ 变化的过程中, 外侧的曲面圆

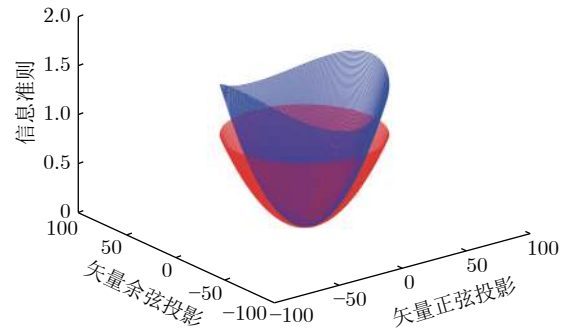


图 1 Oja 信息准则和加权 Oja 信息准则在 θ 变化情况下的数值变化

Fig. 1 Curves of the information criterion Oja and weighted Oja algorithms on θ

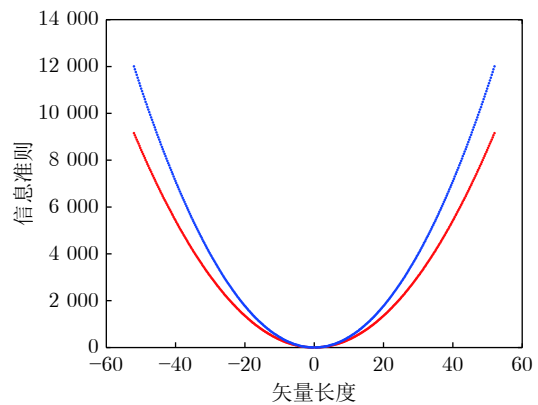


图 2 Oja 信息准则和加权 Oja 信息准则在 $\theta = \frac{\pi}{2}$ 处的投影

Fig. 2 Projection of the information criterion Oja and weighted Oja algorithms when $\theta = \frac{\pi}{2}$

锥在固定模值下, 数值没有变化, 如图 3 所示, 以 θ 为横坐标, 将图 1 中状态矩阵模长为 1.5 处的曲面函数值投影并展开, 作为纵坐标; 内侧的曲面圆锥在固定模值下, 数值一直处在变化中, 如图 4 所示, 以 θ 为横坐标, 将图 1 中状态矩阵模长为 1.5 处的曲面函数值投影并展开作为纵坐标. 其次, 在 θ 固定的时候, 两个曲面函数变化的梯度也不相同, 如图 2 所示, 明显外侧的曲面圆锥的下降坡度小于内侧的曲面圆锥的下降坡度. 这一点能够解释所有以 Oja 信息准则为基础的梯度下降算法, 加权规则下的信息准则具有更好的收敛速度.

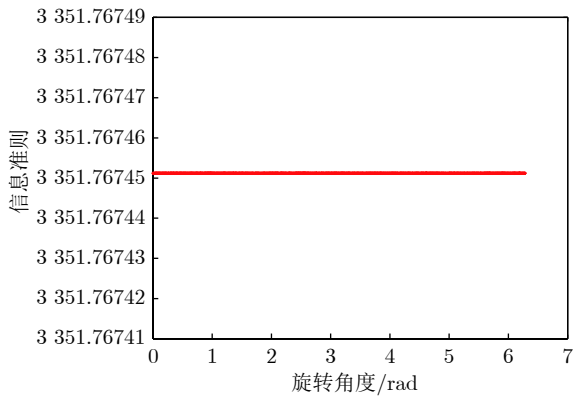


图 3 Oja 信息准则在 θ 变化情况下的数值变化
Fig.3 Curves of Oja information criterion on θ

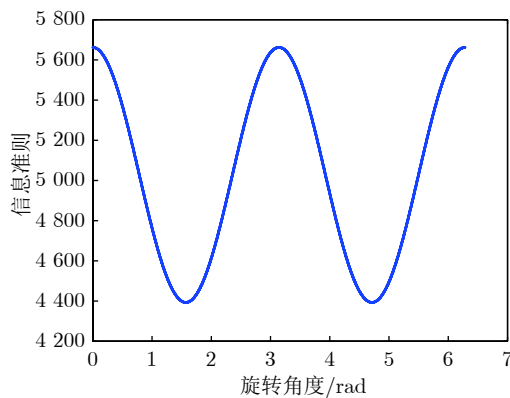


图 4 加权 Oja 信息准则在 θ 变化情况下的数值变化
Fig.4 Curves of weighted Oja information criterion on θ

综合上述分析, 该算例能够验证 Oja 信息准则与加权 Oja 信息准则对提取算法在理论上分析的结论.

3.2 Miao 信息准则与 Ouyang 信息准则

本文所研究的结论不仅适用于 Oja 信息准则, 还适用于其他信息准则, 此处考虑 Miao 信息准

则^[20] 和 Ouyang 信息准则^[21]. 的初始设置与第 3.1 节中的算例相同.

Miao 信息准则的基本表达式为

$$J_3(W) = \frac{1}{2}\text{tr}(\ln(W^T R W)) - \frac{1}{2}\text{tr}(W^T W) \quad (26)$$

Ouyang 信息准则的基本表达式为

$$J_4(W) = \frac{1}{2}\text{tr}(\ln(W^T R W D)) - \frac{1}{2}\text{tr}(W^T W) \quad (27)$$

从上述公式可以看出, Miao 信息准则在加权规则下的讨论就是 Ouyang 信息准则. 从图 5~8 中可以看出, Miao 和 Ouyang 信息准则的变化特性都能够呈现出第 3.1 节的三个特性, 这表明经过与 Oja 及加权 Oja 信息准则相似的动力学分析过程, 也能够得到一致的结论. 实际上, 不仅 Oja 算法、Miao 算法和 Ouyang 算法, 作者经过仿真验证, Yang^[10] 和 Lei^[9] 的 LMSER 算法也具有相似的动力

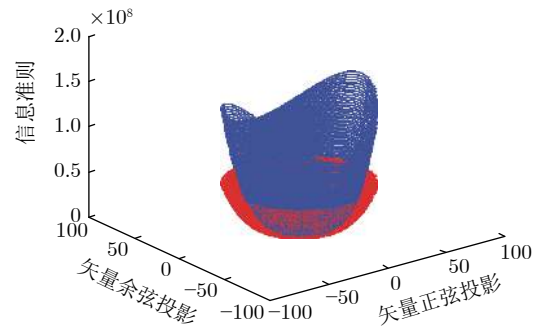


图 5 Miao 信息准则和 Ouyang 信息准则在 θ 变化情况下的数值变化
Fig.5 Curves of the information criterion Miao and Ouyang algorithms on θ

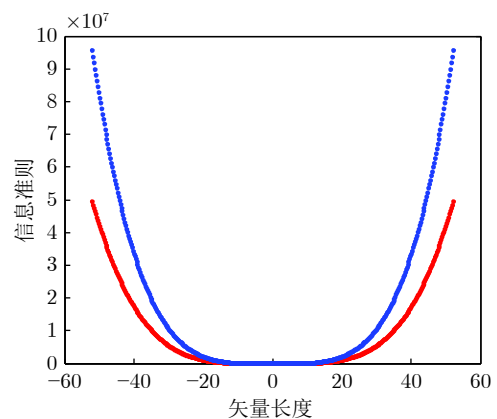


图 6 Miao 信息准则和 Ouyang 信息准则在 $\theta = \frac{\pi}{2}$ 处的投影
Fig.6 Projection of the information criterion Miao and Ouyang algorithms when $\theta = \frac{\pi}{2}$

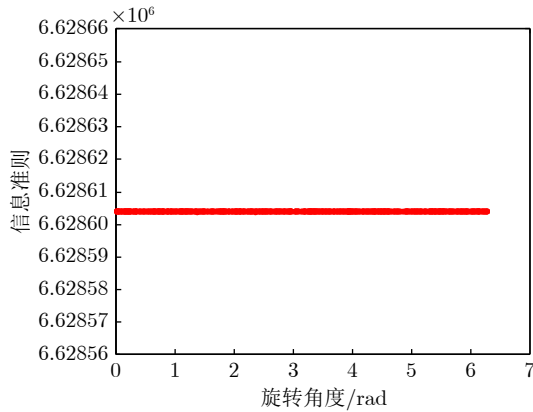


图 7 Miao 信息准则在 θ 变化情况下的数值变化
Fig.7 Curves of Miao information criterion on θ

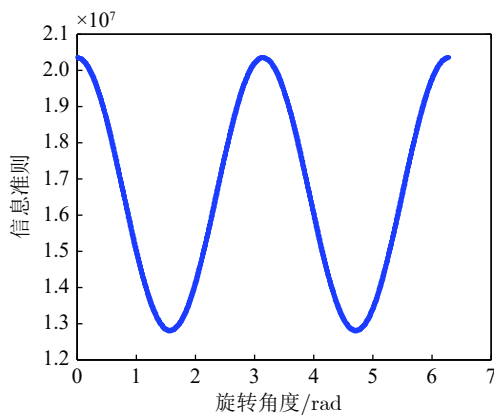


图 8 Ouyang 信息准则在 θ 变化情况下的数值变化
Fig.8 Curves of Ouyang information criterion on θ

学特性, 由于篇幅限制在这里不做展示. 这些算法的不同之处在下降的速度以及在平衡位置的信息准则数值上.

3.3 加权规则的应用案例

图像压缩是计算机图形图像学领域的热点问题. 通过图像压缩技术都可通过压缩图像数据实现更加高效地存储和传输. 基于主成分分析方法的压缩是图像压缩领域中的常用方法^[22]. 本节采用主成分分析对 Lena 图像进行压缩和重构. Lena 的原图如图 9 左上角所示, 分辨率为 512×512 像素. 这里将图像分解为 4×4 像素的不重叠小块. 这些小块是按照从左到右由上到下的顺序排列, 可得一个 $16 \times 16 \times 384$ 的数据向量. 该数据经过预处理后作为主成分分析算法的输入序列. 采用本文研究的加权 Oja 算法和 Ouyang 算法对图像压缩后重建.

加权 Oja 算法使用两次. 第 1 次只取一个主元, 学习因子为 0.01, 第 2 次并行提取两个主元, 学习



图 9 原始和重建的 Lena 图像
Fig.9 Original and reconstructed Lena images

因子也为 0.01, 主要对比单个主元和多个主元情况的重建误差. Ouyang 算法使用一次, 提取两个主元, 学习因子也取 0.01. 两个主元提取的算法加权矩阵为 D . 比较不同算法提取主元的精度. 图 9 右上角为一元 Oja 提取算法的重建效果, 重建误差为 0.0829. 图 9 左下角为二元 Oja 提取算法的重建效果, 重建误差为 0.0602. 图 9 右下角为二元 Ouyang 提取算法的重建效果, 重建误差为 0.0402. 对比实验效果可以发现, 多元并行提取算法相比单个主元提取算法具有优势, Ouyang 算法相比 Oja 算法具有优势. 这进一步表明先进算法需要通过加权算法确保算法具有并行提取多个主成分的能力.

4 结论

本文通过分析 Oja 信息准则和加权 Oja 信息准则的差异, 推导出加权规则能够改变状态矩阵和信号主成分的旋转矩阵梯度的结论. 一方面这对于其他主子空间跟踪算法的转变提供了理论支撑, 提高在信息准则层次影响算法性能的认识; 另一方面, 对于不能通过加权规则转化为多个主成分提取算法的主子空间跟踪算法, 本文也给出转变的基本思路. 研究为推进未来并行提取多个主成分分析算法发展提供了研究思路和转变方向, 最后通过数值实例仿真验证了理论的有效性.

References

- 1 Gersho A, Gray R M. *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.

- 2 Oja E. *Subspace Methods of Pattern Recognition*. Letchworth, UK: Research Studies Press, 1992.
- 3 Pan Zong-Xu, Yu Jing, Xiao Chuang-Bai, Sun Wei-Dong. Spectral similarity-based super resolution for hyperspectral images. *Acta Automatica Sinica*, 2014, **40**(12): 2797–2807
(潘宗序, 禹晶, 肖创柏, 孙卫东. 基于光谱相似性的高光谱图像超分辨率算法. *自动化学报*, 2014, **40**(12): 2797–2807)
- 4 Xiao Jin-sheng, Zhu li, Zhao Bo-qiang, Lei Jun-Feng, Wang Li. Block-based video noise estimation algorithm via principal component analysis. *Acta Automatica Sinica*, 2018, **44**(9): 1618–1625
(肖进胜, 朱力, 赵博强, 雷俊峰, 王莉. 基于主成分分析的分块视频噪声估计. *自动化学报*, 2018, **44**(9): 1618–1625)
- 5 Zhou Ping, Zhang Li, Li Wen-Peng, Dai Peng, Chai Tian-You. Autoencoder and PCA based RVFLNs modeling for multivariate molten iron quality in blast furnace ironmaking. *Acta Automatica Sinica*, 2018, **44**(10): 1799–1811
(周平, 张丽, 李温鹏, 戴鹏, 柴天佑. 集成自编码与PCA的高炉多元铁水质量随机神经网络建模. *自动化学报*, 2018, **44**(10): 1799–1811)
- 6 Kong X Y, Hu C H, Han C Z. A dual purpose principal and minor subspace gradient flow. *IEEE Transaction on Signal Processing*, 2012, **60**(1): 197–210
- 7 Oja E. Neural networks, principal components, and subspaces. *International Journal of Neural Systems*, 1989, **1**(1): 61–68
- 8 Williams R J. *Feature Discovery Through Error-Correction Learning*. San Diego, California: Institute of Cognitive Science, University of California, 1985.
- 9 Lei X. Least mean square error reconstruction principle for self-organizing neural-nets. *Neural Networks*, 1993, **6**(5): 627–648
- 10 Yang B. Projection approximation subspace tracking. *IEEE Transaction on Signal Processing*, 1995, **43**(1): 95–107
- 11 Kong X Y, Hu C H, Ma H G, Han C Z. A unified self-stabilizing neural network algorithm for principal and minor components extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, **23**(2): 185–198
- 12 Kakimoto K, Yamagishi M, Yamada I. Smoothing of adaptive eigenvector extraction in nested orthogonal complement structure with minimum disturbance principle. *Multidimensional Systems and Signal Processing*, 2018, **29**(1): 433–465
- 13 Ouyang S, Bao Z, Liao G S. Robust recursive least squares learning algorithm for principal component analysis. *IEEE Transaction on Neural Networks*, 2000, **11**(1): 215–221
- 14 Gao Y B, Kong X Y, Zhang H H, Hou L A. A weighted information criterion for multiple minor components and its adaptive extraction algorithms. *Neural Networks*, 2017, **89**: 1–10
- 15 Gao Y B, Kong X Y, Hu C H, Li H Z, Hou L A. A generalized information criterion for generalized minor component extraction. *IEEE Transactions on Signal Processing*, 2017, **65**(4): 974–959
- 16 Tanaka T. Generalized weighted rules for principal components tracking. *IEEE Transactions on Signal Processing*, 2005, **53**(4): 1243–1253
- 17 Kushner H J, Clark D S. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. New York, NY: Springer-Verlag, 1978.
- 18 Oja E, Ogawa H, Wangviattana J. Principal component analysis by homogeneous neural networks—Part I: Weighted subspace criterion. *IEICE Transactions on Information Systems*, 1992, **75**(3): 366–375
- 19 Zhang Y, Ye M, Lv J C, Tan K K. Convergence analysis of a deterministic discrete time system of Oja's PCA learning algorithm. *IEEE Transaction on Neural Networks*, 2005, **16**(6): 1318–1328
- 20 Miao Y F, Hua Y B. Fast subspace tracking and neural network learning by a novel information criterion. *IEEE Transactions on Signal Processing*, 1998, **46**(8): 1967–1979
- 21 Ouyang S, Bao Z. Fast principal component extraction by a weighted information Criterion. *IEEE Transactions on Signal Processing*, 2002, **50**(8): 1994–2002
- 22 Fang Wei-Tao, Ma Peng, Cheng Zheng-Bin, Yang Dan, Zhang Xiao-Hong. 2-dimensional projective non-negative matrix factorization and its application to face recognition. *Acta Automatica Sinica*, 2012, **38**(9): 1503–1512
(方蔚涛, 马鹏, 成正斌, 杨丹, 张小洪. 二维投影非负矩阵分解算法及其在人脸识别中的应用. *自动化学报*, 2012, **38**(9): 1503–1512)



杜柏阳 西安高科技研究所博士, 96901 部队助理研究员. 主要研究方向为信号特征提取.

E-mail: duboyangepgc@163.com

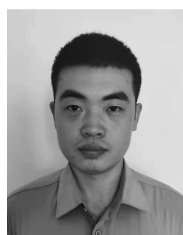
(DU Bo-Yang Ph.D. at the Xi'an Researching Institute of High Technology and assistant research fellow at the 96901 Unit of PLA. His main research interest is signal feature extraction.)



孔祥玉 西安高科技研究所副教授. 2004 年获得西安交通大学博士学位. 主要研究方向为多元信号分析, 信号处理. 本文通信作者.

E-mail: xiangyukong01@163.com

(KONG Xiang-Yu Associate professor at the Xi'an Researching Institute of High Technology. He received his Ph.D. degree from Xi'an Jiaotong University in 2004. His research interest covers multivariate signal analysis and signal processing. Corresponding author of this paper.)



罗家宇 西安高科技研究所硕士研究生. 2017 年获得湖南大学学士学位. 主要研究方向为偏最小二乘分析.

E-mail: luojiayuepgc@163.com

(LUO Jia-Yu Master student at the Xi'an Researching Institute of High Technology. He received his bachelor degree from Hunan University in 2017. His main research interest is partial least square analysis.)