

主子空间跟踪信息准则及算法

高迎彬¹ 孔祥玉² 崔巧花¹ 侯立安²

摘要 信息准则在发展主子空间跟踪算法中具有十分重要的意义. 目前, 主子空间信息准则还不多见, 为此本文提出了一种新型的主子空间跟踪信息准则. 当且仅当神经网络权矩阵收敛到信号主子空间的一个正交基时, 该信息准则取得唯一全局极大值. 通过梯度上升法对该信息准则求极值导出了一个主子空间跟踪算法. 最后通过仿真实验和实际应用证明了所提算法的正确性和实用性.

关键词 主子空间跟踪, 信息准则, 梯度上升法, 神经网络

引用格式 高迎彬, 孔祥玉, 崔巧花, 侯立安. 主子空间跟踪信息准则及算法. 自动化学报, 2020, 46(10): 2214–2220

DOI 10.16383/j.aas.c180141

Principal Subspace Tracking Information Criterion and Algorithm

GAO Ying-Bin¹ KONG Xiang-Yu² CUI Qiao-Hua¹ HOU Li-An²

Abstract It is clear that a properly chosen information criterion is a very important part in developing principal subspace tracking algorithms. Up to now, the numbers of principal subspace tracking information criterion are not rich. Based on this fact, a novel information criterion is proposed in this paper. This information criterion exhibits a single global maximum attained if and only if the weight matrix spans the desired principal subspace. A subspace tracking algorithm is derived by using gradient ascent method to this information criterion. Numerical simulation and real application experiments are carried out to show the correctness and practicality of the proposed algorithm.

Key words Principal subspace tracking, information criterion, gradient ascent method, neural networks

Citation Gao Ying-Bin, Kong Xiang-Yu, Cui Qiao-Hua, Hou Li-An. Principal subspace tracking information criterion and algorithm. *Acta Automatica Sinica*, 2020, 46(10): 2214–2220

在信息处理领域, 通常将信号自相关矩阵最大特征值对应的特征向量称之为信号的主成分, 而由信号的多个主成分张成的空间称为信号的主子空间. 在很多信号处理问题中, 需要对信号的主子空间进行在线跟踪, 如视觉跟踪^[1]、波达方向估计^[2]、图像处理^[3]、谱分析^[4]等领域. 因此, 发展主子空间跟踪算法就成为了一件非常有意义的工作.

以往解决主子空间跟踪问题主要依靠矩阵特征值分解 (Eigenvalue decomposition, EVD) 和奇异值分解 (Singular value decomposition, SVD) 等, 然而该方法计算复杂度高, 而且难以满足实时信号处理的要求. 为了克服这些缺点, 学者们提出了基于

Hebbian 神经网络的主子空间跟踪方法. 相比传统的 EVD 和 SVD 方法, 神经网络方法具有以下 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^[4]. 基于上述优势, 神经网络方法已经成为近些年来国际上的一个研究热点.

基于单层线性神经网络, Oja 提出了著名的 Oja 算法^[5], 然而 Oja 算法是一个单维主成分提取算法. 为了能够实现对信号主子空间的跟踪, 学者们通过对 Oja 算法进行了改进提出了很多算法, 如 FDP (Fast data projection method) 算法^[6]、SOJA (Stable and orthonormal OJA algorithm) 算法^[7]、SDPM (Stable DPM algorithm) 算法^[8]等, 然而上述算法大多是基于启发式推理提出来的, 而并没有建立相对应的信息准则. 由于信息准则在算法发展中具有很重要的意义^[9], 因此研究主子空间准则是一件非常有意义的工作. 在文献 [10] 中, 基于最小均方误差 (Least mean squared error, LMSE) 准则, Yang 提出了投影近似子空间跟踪算法 (Projection approximation subspace tracking, PAST); 此后, Miao 等提出了 NIC (Novel information criterion) 准则^[11], 并分别导出了梯度和递归型主子空间跟踪算法; 基于 Rayleigh 商函数, Kong

收稿日期 2018-03-12 录用日期 2018-08-23
Manuscript received March 12, 2018; accepted August 23, 2018
国家自然科学基金 (61903375, 61673387, 61374120), 陕西省自然科学基金 (2016JM6015) 资助

Supported by National Natural Science Foundation of China (61903375, 61673387, 61374120), National Natural Science Foundation of Shaanxi Province (2016JM6015)

本文责任编辑 梅生伟
Recommended by Associate Editor MEI Sheng-Wei
1. 中国电子科技集团公司第五十四研究所 石家庄 050081 2. 火箭军工程大学控制工程系 西安 710025

1. The 54th Research Institute of China Electronics Technology Group Corporation, Shijiazhuang 050081 2. Department of Control Engineering, Rocket Force University of Engineering, Xi'an 710025

等提出了 UIC (Unified information criterion) 准则^[12], 仿真实验表明基于 UIC 准则导出的算法具有很快的收敛速度. 目前, 发展主子空间跟踪准则仍然具有很强的研究价值.

本文将提出一种新型的主子空间跟踪信息准则, 并通过梯度法导出快速的主子空间跟踪算法. 论文的结构安排如下: 第 1 节是提出一种新的主子空间信息准则; 第 2 节是对所提信息准则进行前景分析; 第 3 节主要是采用梯度上升法导出主子空间算法; 第 4 节通过两组仿真实验对所提算法的性能进行验证; 本文的结论在第 5 节.

1 新型主子空间跟踪信息准则

考虑一个具有如下形式的多输入多输出线性神经网络模型:

$$\mathbf{y} = W^T \mathbf{x} \quad (1)$$

其中, $W = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r] \in \mathbf{R}^{n \times r}$ 是神经网络的权矩阵, \mathbf{w}_i 是权矩阵 W 的第 i 列; $\mathbf{x} \in \mathbf{R}^{n \times 1}$ 是采样信号, 这里作为神经网络的输入; $\mathbf{y} \in \mathbf{R}^{r \times 1}$ 为采样信号的低维表示, r 是子空间的维数. 本文的目的就是构造合适的神经网络权矩阵迭代更新方程, 使神经网络的权矩阵最终能够收敛到采样信号的主子空间.

基于上述神经网络模型, 给定域 $\Omega = \{W | 0 < W^T R W < \infty, W^T W \neq 0\}$, 提出如下信息准则:

$$W^* = \arg \max_{W \in \Omega} J(W)$$

$$J(W) = \frac{1}{2} \text{tr} [(W^T R W)(W^T W)^{-1}] + \frac{1}{2} \text{tr} [\ln(W^T W) - W^T W] \quad (2)$$

其中, $R = E[\mathbf{x}\mathbf{x}^T]$ 为采样信号的自相关矩阵. 根据矩阵理论可得: 矩阵 R 是一个对称的正定矩阵, 且特征值均是非负的. 对矩阵 R 特征值分解得:

$$R = U \Lambda U^T \quad (3)$$

其中, $U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]$ 是由矩阵 R 的特征向量构成的矩阵, $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 是由矩阵 R 的特征值组成. 为了后续使用方便, 这里将特征值按降序进行排列, 即特征值满足如下方程:

$$\lambda_1 > \lambda_2 > \dots > \lambda_r > \dots > \lambda_n > 0 \quad (4)$$

根据主子空间的定义可知特征值 $\lambda_1, \lambda_2, \dots, \lambda_r$ 所对应的特征向量张成的子空间称为输入信号的主子空间. 从式 (2) 可得 $J(W)$ 是无下界的, 且当 W 趋于无穷大时, $J(W)$ 将趋于负无穷大, 因此研究 $J(W)$ 的极小值是没有意义的. 实际上, 我们关

心的是 $J(W)$ 的极大点. 具体说来, 我们关心以下几个问题:

1) $J(W)$ 有没有全局极大点?

2) $J(W)$ 的全局极大点与信号的主子空间之间的关系是什么?

3) $J(W)$ 有没有其他局部极值?

上述三个问题的回答将由下在一节中完成.

2 信息准则的全局最优值分析

信息准则 (2) 的全局最优值分析通过定理 1 和定理 2 来完成.

定理 1. 在域 $\Omega = \{W | 0 < W^T R W < \infty, W^T W \neq 0\}$ 内, 当且仅当 $W = U'_r Q$ 时, 权矩阵 W 是信息准则 $J(W)$ 的一个平稳点, 其中 $U'_r = [\mathbf{u}_{j1}, \mathbf{u}_{j2}, \dots, \mathbf{u}_{jr}]$ 是由自相关矩阵 R 的任意 r 个特征向量构成的矩阵, Q 是任意一个 $r \times r$ 维正交矩阵.

将 R 的特征向量作为空间 $\mathbf{R}^{n \times n}$ 的一组正交基, 则权矩阵 W 可以表示为: $W = U^T \tilde{W}$, 其中 $\tilde{W} \in \mathbf{R}^{n \times r}$ 称为系数矩阵. 将这一结果代入式 (2) 可得:

$$\tilde{W}^* = \arg \max_{\tilde{W} \in \Omega} E(\tilde{W})$$

$$E(\tilde{W}) = \frac{1}{2} \text{tr}[(\tilde{W}^T \Lambda \tilde{W})(\tilde{W}^T \tilde{W})^{-1}] + \frac{1}{2} \text{tr}[\ln(\tilde{W}^T \tilde{W}) - \tilde{W}^T \tilde{W}] \quad (5)$$

显然式 (2) 和式 (5) 是等价的, 即定理 1 的证明是可以通过对下述推论的证明来完成.

推论 1. 在域 $\tilde{\Omega} = \{\tilde{W} | 0 < \tilde{W}^T \Lambda \tilde{W} < \infty, \tilde{W}^T \tilde{W} \neq 0\}$ 中, 当且仅当 $\tilde{W} = P_r Q$ 时, \tilde{W} 是 $E(\tilde{W})$ 的一个平稳点, 其中 $P_r \in \mathbf{R}^{n \times r}$ 是任意一个 $n \times r$ 维置换矩阵.

证明参见附录 A. \square

定理 2. 在域 Ω 内, 当且仅当 $W = \tilde{U}_r Q$ 时, 其中 $\tilde{U}_r = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r]$ 是由自相关矩阵 R 的前 r 个特征向量构成的矩阵且 Q 是任意一个 $r \times r$ 维正交矩阵, 信息准则 $J(W)$ 达到全局极大. $J(W)$ 没有其他局部极值, 在全局最大点处有:

$$J(W) = \frac{1}{2} \sum_{i=1}^r \lambda_i - \frac{r}{2} \quad (6)$$

同定理 1 的证明过程一样, 这里将通过推论 2 的证明来完成定理 2 的证明.

推论 2. 在域 $\tilde{\Omega}$ 中, 当且仅当 $\tilde{W} = \tilde{P} Q$, 信息准则 $E(\tilde{W})$ 达到全局最大点, 其中 $\tilde{P} = (\tilde{P} \ 0)^T \in \mathbf{R}^{n \times r}$ 一个 $n \times r$ 维矩阵, \tilde{P} 是一个置换矩阵, 而

$E(\tilde{W})$ 所有其他的平稳点都是鞍点.

证明参见附录 B. \square

通过定理 1 和定理 2 可知, 当神经网络权矩阵 W 刚好收敛到自相关矩阵 R 的主子空间的一组正交基时, $J(W)$ 取得全局极大值, 从而建立起神经网络与信号主子空间之间的关系. 由于信息准则 $J(W)$ 只有一个全局最大点, 而没有局部极值, 因此采用非线性规划算法 (如梯度法、共轭梯度法、牛顿法等) 来求解该优化问题. 梯度算法是采用 $J(W)$ 的梯度作为迭代步长; 共轭梯度法需要不断修正算法的共轭方向, 所导出的算法通常具有较高的计算复杂度; 牛顿法则用到的 Hessian 矩阵, 而当 W 是一个矩阵时该 Hessian 矩阵是非常难以获得的. 相比共轭梯度法和牛顿算法, 梯度法具有算法结构更为简单, 计算复杂度低, 因此下一节将采用梯度算法导出新型的主子空间跟踪算法.

3 主子空间跟踪算法

假定 $\mathbf{x}(k), k = 0, 1, 2, \dots$ 是一个平稳的随机过程, 这里将其作为神经网络模型的输入. 根据随机学习理论, 权矩阵 W 的变化规律与输入向量 $\mathbf{x}(k)$ 并不相关. 取式 (2) 作为最优化函数, 则可以得出所提信息准则的梯度流为:

$$\frac{dW(t)}{dt} = \left[\frac{W(t)}{W^T(t)W(t)} - W(t) \right] + \left[RW(t) - \frac{W(t)W^T(t)RW(t)}{W^T(t)W(t)} \right] \times \{W^T(t)W(t)\}^{-1} \quad (7)$$

应用随机近似理论可得:

$$\frac{dW(t)}{dt} = \left[W(t)\{W^T(t)W(t)\}^{-1} - W(t) \right] - \left[W(t)W^T(t)\mathbf{x}(t)\mathbf{x}^T(t)W(t)\{W^T(t)W(t)\}^{-1} - \mathbf{x}(t)\mathbf{x}^T(t)W(t) \right] \{W^T(t)W(t)\}^{-1} \quad (8)$$

对式 (8) 进行离散化操作后获得如下方程:

$$W(k+1) = W(k) - \eta \left[W(k)(W^T(k)W(k))^{-1}\mathbf{y}(k)\mathbf{y}^T(k) - \mathbf{x}(k)\mathbf{y}^T(k) \right] (W^T(k)W(k))^{-1} - \eta W(k) + \eta W(k)(W^T(k)W(k))^{-1} \quad (9)$$

其中, η 是神经网络的学习因子, 且满足 $0 < \eta < 1$. 式 (9) 所描述的算法在每一步迭代过程中计算复杂度为: $nr^2 + 4nr^3/3$, 这点与 UIC 算法^[12] 是相同的,

要少于 NIC 算法^[11] 中的 $2n^2r + O(nr^2)$ 的计算量和共轭梯度算法^[13] $12n^2r + O(nr^2)$ 的计算量.

4 仿真实验

本节通过两个仿真实例来对所提算法的性能进行验证. 第一个实验考察所提算法提取多维主子空间的能力并将仿真结果与其他同类型算法进行对比; 第二个是应用所提算法解决图像重构问题.

4.1 主子空间跟踪实验

在本实验中, 所提算法将与 UIC 算法和 SDPM 算法进行对比. 为了衡量算法的收敛性能, 这里采用如下两个评价函数, 第一个是第 k 次迭代时的权矩阵模值

$$p(W_k) = \|W_k^T W_k\|_F \quad (10)$$

第二个是指标参数

$$\text{dist}(W_k) = \left\| W_k^T W_k \text{diag}\{W_k^T W_k\}^{-1} - I_r \right\|_F \quad (11)$$

指标参数代表着权矩阵正交化的偏离程度. 显然, 如果 $\text{dist}(W_k)$ 收敛到 0, 则意味着权矩阵收敛到了主子空间的一个正交基.

本实验中信号产生方法与文献 [12] 相同, 即输入信号 $X_k = B\mathbf{z}_k$, 其中 $B = \text{randn}(31, 31)/31$ 是一个随机产生的矩阵, $\mathbf{z}_k \in \mathbf{R}^{31 \times 1}$ 是高斯的、瞬时的、随机产生的向量. 在本实验中, 分别采用 UIC 算法、SDPM 算法和所提算法对信号的 12 维主子空间进行提取跟踪. 三种算法采用相同的学习因子 $\eta = 0.1$, 初始权矩阵是随机产生的 (即矩阵的每一个元素均服从均值为零方差为 1 的高斯分布). 通常情况下, 取不同的初始化权矩阵时, 算法具有不同的收敛速度. 为了更全面地衡量算法性能, 通常取多次实验的平均值来描述算法收敛过程. 在图 1 和图 2 分别给出了三种算法在迭代过程中的权矩阵模值曲线和指标参数曲线, 该结果曲线是 100 次独立实验结果平均得到的. 图中实线代表着所提算法, 虚线代表 UIC 算法, 点划线代表 SDPM 算法.

图 1 和图 2 中可以发现, 所提算法的权矩阵模值曲线收敛到了一个常数而且指标参数收敛到了零. 这就表明所提算法具备跟踪信号主子空间的能力. 从图 1 中我们还可以发现, 所提算法的权矩阵模值在 200 步时就已经收敛, 而 UIC 算法则需要 300 步, SDPM 算法需要 800 步, 即所提算法的权矩阵模值曲线具有最快的收敛速度. 同理, 从图 2 中可以发现所提算法指标参数的收敛速度要优于其他两个算法. 综合两图可以得出结论: 三种算法中, 所提算法具有最快的收敛速度.

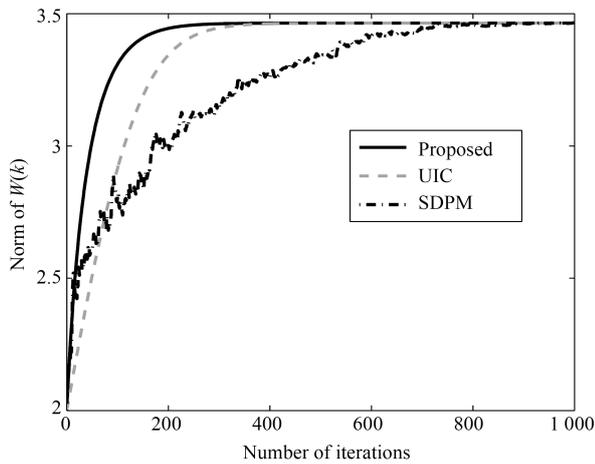


图 1 权矩阵模值曲线

Fig. 1 Curves of the weight matrix norm

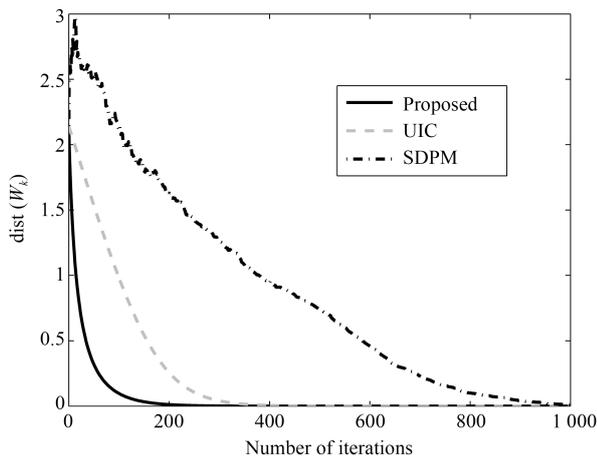


图 2 指标参数曲线

Fig. 2 Curves of the parameter index

4.2 图像压缩重构离实验

数据压缩是主子空间算法的一个很重要的应用. 本实验将利用所提算法对著名的 Lena 图像进行压缩重构. 如图 3 所示, 原始 Lena 图像的像素为 512×512 . 在本实验中, 将 Lena 图像分解成若干个 8×8 不重叠的小块. 将每一个小块中的数据按照从左至右从上到下的顺序排列起来, 构成一个 64 维向量. 在去掉均值和标准化后, 将这些图像数据构成一个输入序列. 然后采用 SDPM 算法、UIC 算法以及所提算法对该图像进行压缩重构, 这里重构维数为 5. 与实验 1 相同, 本实验同样采用权矩阵模值和指标参数两个评价函数.

本实验中三种算法的初始化参数设置方法与实验 1 相类似, 具体参数如下: 学习因子 $\eta = 0.2$, 初始化权矩阵是随机产生的. 图 4 是经过所提算法压缩后重构出来的 Lena 图像, 图 5 是三种算法的权矩阵模值曲线, 图 6 是三种算法的指标参数曲线, 该

结果曲线都是 100 次独立实验的平均值. 对比图 3 和图 4 可以发现, 重构的 Lena 图像是很清晰的, 即所提算法能够有效解决图像压缩重构问题. 通过图 5 和图 6 可以发现: 不论是权矩阵模值曲线还是指标参数曲线, 所提算法的收敛速度均要快于 UIC 算法和 SDPM 算法. 这进一步证实了所提算法在收敛速度方面的优势.



图 3 原始的 Lena 图像

Fig. 3 The original image of Lena



图 4 所提算法压缩重构的 Lena 图像

Fig. 4 Reconstructed image by the proposed algorithm

5 结论

主子空间跟踪算法在现代信息科学各个领域均有着很重要的应用. 基于 Hebbian 神经网络的主子空间跟踪是近些年来国际上的一个研究热点. 然而目前大多数主子空间跟踪神经网络算法是基于启发式推理而提出来的, 能够提供信息准则的算法并不多见. 针对这一问题, 本文提出了一种新型的信息准则, 并对所提信息准则的前景进行了严格的证明. 通

过梯度上升法导出了一个主子空间跟踪算法. 仿真实验表明: 相比一些现有主子空间跟踪算法, 所提算法具有更快的收敛速度.

进一步的研究方向是寻找新型的主子空间信息准则, 创新信息准则的平稳点分析方法, 将主子空间算法应用于更广泛的领域.

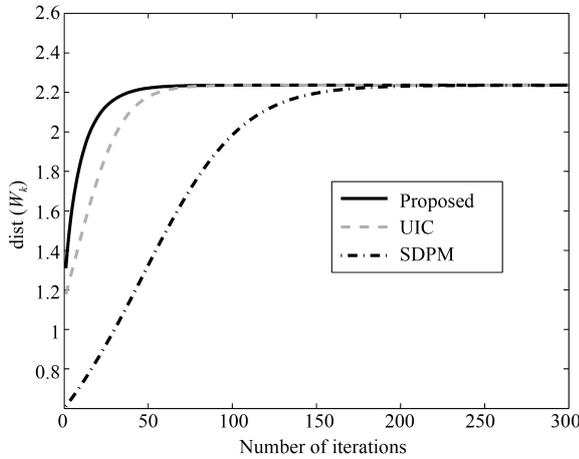


图 5 三种算法权重矩阵模值曲线

Fig. 5 Norm curves of the three algorithms

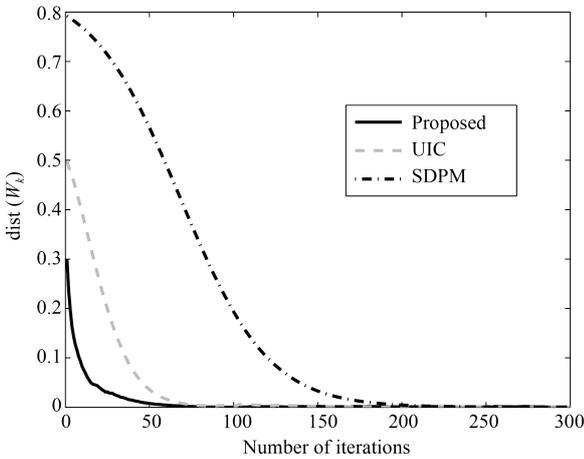


图 6 三种算法的指标参数曲线

Fig. 6 Parameter index curves of the three algorithms

附录 A. 推论 1 证明

证明. 在域 $\tilde{\Omega}$ 内, $E(\tilde{W})$ 对于矩阵 \tilde{W} 的一阶微分存在, 且有

$$\nabla E(\tilde{W}) = \Lambda \tilde{W} (\tilde{W}^T \tilde{W})^{-1} - \tilde{W} - \tilde{W} (\tilde{W}^T \tilde{W})^{-2} \tilde{W}^T \Lambda \tilde{W} + \tilde{W} (\tilde{W}^T \tilde{W})^{-1} \quad (\text{A1})$$

定义一个矩阵集 $\{\tilde{W} | \tilde{W} = P_r Q\}$, 则在该集合内的任

意一点均有:

$$\begin{aligned} \nabla E(\tilde{W})|_{\tilde{W}=P_r Q} &= \\ \Lambda P_r Q (Q^T P_r^T P_r Q)^{-1} + P_r Q (Q^T P_r^T P_r Q)^{-1} - \\ P_r Q (Q^T P_r^T P_r Q)^{-2} Q^T P_r^T \Lambda P_r Q - P_r Q &= \\ \Lambda P_r Q + P_r Q Q^T P_r^T \Lambda P_r Q &= 0 \end{aligned} \quad (\text{A2})$$

反之, 根据定义可得, 在 $E(\tilde{W})$ 平稳点处有 $\nabla E(\tilde{W}) = 0$ 成立, 即

$$\begin{aligned} \Lambda \tilde{W} (\tilde{W}^T \tilde{W})^{-1} + \tilde{W} (\tilde{W}^T \tilde{W})^{-1} - \\ \tilde{W} (\tilde{W}^T \tilde{W})^{-2} \tilde{W}^T \Lambda \tilde{W} - \tilde{W} &= 0 \end{aligned} \quad (\text{A3})$$

将上式左右两边各乘以 \tilde{W}^T 可得:

$$\begin{aligned} \tilde{W}^T \Lambda \tilde{W} (\tilde{W}^T \tilde{W})^{-1} + \tilde{W}^T \tilde{W} (\tilde{W}^T \tilde{W})^{-1} - \\ \tilde{W}^T \tilde{W} (\tilde{W}^T \tilde{W})^{-2} \tilde{W}^T \Lambda \tilde{W} - \tilde{W}^T \tilde{W} &= \\ \tilde{W}^T \Lambda \tilde{W} (\tilde{W}^T \tilde{W})^{-1} - \tilde{W}^T \tilde{W} - \\ (\tilde{W}^T \tilde{W})^{-1} \tilde{W}^T \Lambda \tilde{W} + I &= \\ I - \tilde{W}^T \tilde{W} &= 0 \end{aligned} \quad (\text{A4})$$

根据上式可得, 在 $E(\tilde{W})$ 平稳点处有

$$\tilde{W}^T \tilde{W} = I \quad (\text{A5})$$

上式表明在 $E(\tilde{W})$ 平稳点处矩阵 \tilde{W} 的各个列向量之间是相互正交的. 将式 (A5) 代入式 (A3) 可得:

$$\Lambda \tilde{W} = \tilde{W} \tilde{W}^T \Lambda \tilde{W} \quad (\text{A6})$$

令矩阵 \tilde{W} 的行向量为 \tilde{u}_i ($i = 1, 2, \dots, n$), 即有 $\tilde{W} = [\tilde{u}_1^T, \tilde{u}_2^T, \dots, \tilde{u}_n^T]^T$, 同时定义矩阵 $B = \tilde{W}^T \Lambda \tilde{W}$, 则根据式 (11) 可得:

$$\sigma_i \tilde{u}_i = \tilde{u}_i B, \quad i = 1, 2, \dots, n \quad (\text{A7})$$

显然, 上式可以看作是矩阵 B 的特征值分解. 由于 B 是一个 $r \times r$ 维对称正定矩阵, 只有 r 个相互正交的左行特征向量, 即矩阵 \tilde{W} 只有 r 个相互正交的行向量. 更进一步, 矩阵 \tilde{W} 的这 r 个非零行向量正好构成了一个正交矩阵, 也就是说此时矩阵 \tilde{W} 可以通过 $\tilde{W} = P_r Q$ 来表示. \square

附录 B. 推论 2 证明

证明. 定义一个置换矩阵 $P_r \neq \bar{P}$, 则矩阵 P_r 中的第 $r+1$ 到 n 个行向量之中必有一个非零的行向量. 由于 P_r 和 \bar{P} 同为两个置换矩阵, 则必定存在两个对角矩阵 $\bar{\Lambda}$ 和 $\hat{\Lambda}$ 使得下式成立:

$$\bar{P}^T \Lambda \bar{P} = \bar{\Lambda}, P_r^T \Lambda P_r = \hat{\Lambda} \quad (\text{B1})$$

根据上式可得:

$$\begin{cases} \text{tr}(\bar{P}^T \Lambda \bar{P}) = \sum_{i=1}^r \lambda_i \\ \text{tr}(P_r^T \Lambda P_r) = \sum_{i=1}^r \lambda_{j_i} \end{cases} \quad (\text{B2})$$

将特征值 λ_{j_i} ($i = 1, 2, \dots, r$) 按照降序排列, 即有 $\hat{\lambda}_{j_1} > \hat{\lambda}_{j_2} > \dots > \hat{\lambda}_{j_r}$, 则对于 $P_r \neq \bar{P}$, 则必定存在有

$\lambda_i > \hat{\lambda}_{j_i}$ ($i = 1, 2, \dots, r$) 成立, 即有:

$$\text{tr}(\bar{P}^T \Lambda \bar{P}) > \text{tr}(P_r^T \Lambda P_r) \quad (\text{B3})$$

由于

$$\begin{aligned} E(P_r, Q) &= \frac{1}{2} \text{tr}[(Q^T P_r^T \Lambda P_r Q)(Q^T P_r^T P_r Q)^{-1}] + \\ &\frac{1}{2} \text{tr}[\ln(Q^T P_r^T P_r Q) - Q^T P_r^T P_r Q] = \\ &\frac{1}{2} \text{tr}[(Q^T P_r^T \Lambda P_r Q)] - \frac{r}{2} = \\ &\frac{1}{2} \text{tr}[(P_r^T \Lambda P_r)] - \frac{r}{2} \end{aligned} \quad (\text{B4})$$

$$\begin{aligned} E(\bar{P}^T Q) &= \frac{1}{2} \text{tr}[(Q^T \bar{P}^T \Lambda \bar{P} Q)(Q^T \bar{P}^T \bar{P} Q)^{-1}] + \\ &\frac{1}{2} \text{tr}[\ln(Q^T \bar{P}^T \bar{P} Q) - Q^T \bar{P}^T \bar{P} Q] = \\ &\frac{1}{2} \text{tr}[(Q^T \bar{P}^T \Lambda \bar{P} Q)] - \frac{r}{2} = \\ &\frac{1}{2} \text{tr}[(\bar{P}^T \Lambda \bar{P})] - \frac{r}{2} \end{aligned} \quad (\text{B5})$$

根据式 (B3) 有:

$$E(P_r, Q) < E(\bar{P}^T Q) \quad (\text{B6})$$

即集合 $\{P_r, Q | Q^T \Lambda Q > 0 \& P_r \neq \bar{P}\}$ 的点并不是全局极大点.

由于 $P_r \neq \bar{P}$, 则矩阵 \bar{P} 中必定存在一列向量 \bar{p}_i ($1 \leq i \leq r$), 使得:

$$\bar{p}_i^T P_r = 0 \quad (\text{B7})$$

同理, 矩阵 P_r 中也存在一列向量 $p_{r,j}$ ($1 \leq j \leq r$) 使得

$$p_{r,j} \bar{P} = 0 \quad (\text{B8})$$

令 \bar{p}_i 的非零元素为 \bar{j}_i 行, $p_{r,j}$ 的非零元素为 \hat{j}_j 行, 则有 $\bar{j}_i > \hat{j}_j$ 和 $\lambda_{\bar{j}_i} > \lambda_{\hat{j}_j}$. 定义矩阵:

$$B = \begin{bmatrix} p_{r,1}, \dots, \frac{p_{r,i} + \varepsilon \bar{p}_i}{\sqrt{1 + \varepsilon^2}}, \dots, p_{r,r} \end{bmatrix} \quad (\text{B9})$$

其中, ε 是正任意小数. 由 \bar{p}_i 和 $p_{r,j}$ 有且仅有一个非零元素得

$$\Lambda B = \text{diag} \left\{ \lambda_{\hat{j}_1} p_{r,1}, \dots, \frac{\lambda_{\bar{j}_j} p_{r,i} + \varepsilon \lambda_{\bar{j}_i} \bar{p}_i}{\sqrt{1 + \varepsilon^2}}, \dots, \lambda_{\hat{j}_r} p_{r,r} \right\} \quad (\text{B10})$$

更进一步有:

$$B^T \Lambda B = \text{diag} \left\{ \lambda_{\hat{j}_1}, \dots, \frac{\lambda_{\bar{j}_j} + \varepsilon \lambda_{\bar{j}_i}}{1 + \varepsilon^2}, \dots, \lambda_{\hat{j}_r} \right\} \quad (\text{B11})$$

通过式 (A8) 可得:

$$\begin{aligned} B^T \Lambda B - P_r^T \Lambda P_r &= \\ &\text{diag} \left\{ \lambda_{\hat{j}_1}, \dots, \frac{\lambda_{\bar{j}_j} + \varepsilon \lambda_{\bar{j}_i}}{1 + \varepsilon^2}, \dots, \lambda_{\hat{j}_r} \right\} - \\ &\text{diag} \{ \lambda_{\hat{j}_1}, \dots, \lambda_{\hat{j}_j}, \dots, \lambda_{\hat{j}_r} \} = \\ &\text{diag} \left\{ 0, \dots, \frac{(-\lambda_{\bar{j}_j} + \lambda_{\bar{j}_i}) \varepsilon^2}{1 + \varepsilon^2}, \dots, 0 \right\} \end{aligned} \quad (\text{B12})$$

由于 $\lambda_{\bar{j}_j} > \lambda_{\bar{j}_i}$, 所以 $B^T \Lambda B - P_r^T \Lambda P_r$ 是一个负定矩阵, 因此有:

$$E(BQ) = \frac{1}{2} \text{tr}[B^T \Lambda B] - \frac{r}{2} < E(P_r, Q) \quad (\text{B13})$$

即集合 $\{P_r, Q | Q^T \Lambda Q > 0 \& P_r \neq \bar{P}\}$ 中所有平稳点都是不稳定的鞍点.

接下来将证明: $J(W)$ 没有其他局部极值. 令 $\dot{U}_r = \ddot{U}'_r + \varepsilon M_1$, 其中 $M_1 = [0, \dots, \mathbf{u}_k, \dots, 0]$, $1 \leq k \leq r$. 即 \dot{U}_r 是 \ddot{U}'_r 沿 \mathbf{u}_k 的方向增长而成. 由于 $\ddot{U}'_r \neq \ddot{U}_r$, 则必定有 $\lambda_k > \lambda_{j_k}$.

当 $W = \dot{U}_r Q$ 时, 有:

$$\begin{aligned} J(W)|_{W=\dot{U}_r Q} - J(W)|_{W=\ddot{U}'_r Q} &= \\ &\frac{1}{2} (\lambda_k - \lambda_{j_k}) \varepsilon^2 + o(\varepsilon^2) \end{aligned} \quad (\text{B14})$$

令 $\ddot{U}_r = \ddot{U}'_r + \varepsilon M_2$, 其中 $M_2 = [0, \dots, \mathbf{u}_{j_k}, \dots, 0]$, $1 \leq k \leq r$. 即 \ddot{U}_r 是 \ddot{U}'_r 沿 \mathbf{u}_{j_k} 的方向增长而成. 当 $W = \ddot{U}_r Q$ 时, 有:

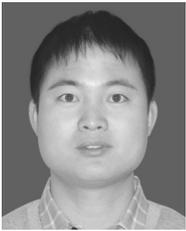
$$J(W)|_{W=\ddot{U}_r Q} - J(W)|_{W=\ddot{U}'_r Q} = -2\varepsilon^2 + o(\varepsilon^2) \quad (\text{B15})$$

从式 (B14) ~ (B15) 得当 $W = U'_r Q$ 且 $W \neq \ddot{U}_r Q$ 时, $J(W)$ 沿 \mathbf{u}_k 方向是增的, 而沿 \mathbf{u}_{j_k} 方向是减的, 所以 $J(W)$ 在该平稳点处不可能取得局部极值. \square

References

- Helwani K. Fundamentals of adaptive filter theory. *Adaptive Identification of Acoustic Multichannel Systems Using Sparse Representations*. Cham: Springer International Publishing, 2015.
- Haddadi F. Steady-state statistical performance analysis of subspace tracking methods. *IEEE Transactions on Signal Processing*, 2016, **64**(18): 4781–4791
- Kong Ling-Zhi, Gao Ying-Bin, Li Hong-Zeng, Zhang Hua-Peng. A fast algorithm that extracts multiple principle components in parallel. *Acta Automatica Sinica*, 2017, **43**(5): 835–842
(孔令智, 高迎彬, 李红增, 张华鹏. 一种快速的多个主成分并行提取算法. *自动化学报*, 2017, **43**(5): 835–842)
- Nguyen V D, Abed-Meraim K, Linh-Trung N, Weber R. Generalized minimum noise subspace for array processing. *IEEE Transactions on Signal Processing*, 2017, **65**(14): 3789–3802
- Du K L, Swamy M N S. Principal component analysis. *Neural Networks and Statistical Learning*. London: Springer, 2014.
- Doukopoulos X G, Moustakides G V. Fast and stable subspace tracking. *IEEE Transactions on Signal Processing*, 2008, **56**(4): 1452–1465
- Golub G H, Van Loan C F. Matrix computations. *Mathematical Gazette*, 1996, **47**(5 Series II): 392–396
- Wang R, Yao M L, Zhang D M, Zou H X. A novel orthonormalization matrix based fast and stable DPM algorithm for principal and minor subspace tracking. *IEEE Transactions on Signal Processing*, 2012, **60**(1): 466–472

- 9 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
- 10 Yang B. Projection approximation subspace tracking. *IEEE Transactions on Signal Processing*, 1995, **43**(1): 95–107
- 11 Miao Y, Hua Y. Fast subspace tracking and neural network learning by a novel information criterion. *IEEE Transactions on Signal Processing*, 1998, **46**(7): 1967–1979
- 12 Kong X Y, Hu C H, Han C Z. A dual purpose principal and minor subspace gradient flow. *IEEE Transactions on Signal Processing*, 2012, **60**(1): 197–210
- 13 Kong X Y, Hu C H, Duan Z S. *Principal Component Analysis Networks and Algorithms*. Singapore: Springer, 2017.



高迎彬 中国电子科技集团公司第五十四研究所工程师。2016 年获得火箭军工程大学博士学位。主要研究方向为自适应信号处理, 神经网络。

E-mail: welcome8793@sina.com

(**GAO Ying-Bin** Engineer at the 54th Research Institute of China Electronics Technology Group Corporation.

He received his bachelor degree from Rocket Force University of Engineering in 2016. His research interest covers adaptive signal processing and neural networks.)



孔祥玉 火箭军工程大学教授。主要研究方向为故障诊断, 自适应信号处理, 神经网络。本文通信作者。

E-mail: xiangyukong01@163.com

(**KONG Xiang-Yu** Professor at Rocket Force University of Engineering.

His research interest covers fault diagnosis, adaptive signal processing, and neural networks. Corresponding author of this paper.)



崔巧花 中国电子科技集团公司第五十四研究所工程师。主要研究方向为散射通信。E-mail: gybcqh@sina.com

(**CUI Qiao-Hua** Engineer at the 54th Research Institute of China Electronics Technology Group Corporation. Her main research interest is scatter communication.)



侯立安 中国工程院院士, 火箭军工程大学教授。主要研究方向为大气污染治理, 故障诊断, 信号处理。

E-mail: houlian678@hotmail.com

(**HOU Li-An** Academician of the Chinese Academy of Engineering, professor at Rocket Force University of Engineering. His research interest covers

atmospheric pollution treatment, fault diagnosis, and signal processing.)