

基于 Rayleigh 商的次子空间准则跟踪函数

徐中英, 高迎彬, 孔祥玉, 杜伯阳

(火箭军工程大学控制工程系, 陕西 西安 710025)

摘要: 针对次子空间信息准则函数缺乏的问题, 通过对 Rayleigh 商函数添加惩罚项提出了新型的准则函数。平稳点分析表明, 当且仅当神经网络权矩阵收敛到次子空间的一组基时, 准则函数达到全局极大值。采用梯度上升法推导出一个新型的次子空间跟踪算法, 并分析了算法的全局收敛性。仿真实验和实例应用验证了准则函数和导出算法的正确性。

关键词: Rayleigh 商; 准则函数; 次成分分析; 李雅普诺夫函数

中图分类号: TP391

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2019216

Novel criterion function for minor subspace tracking based on Rayleigh quotient

XU Zhongying, GAO Yingbin, KONG Xiangyu, DU Boyang

Department of Control Engineering, Rocket Force University of Engineering, Xi'an 710025, China

Abstract: Aiming at the lack of information criterion functions of sub subspace, a novel criterion function was proposed by adding a penalty term to the Rayleigh quotient. Through analyzing the properties of all the stable points, it was proven that the criterion function exhibited the global maximum attained if and only if the weight matrices span the minor subspace. A minor subspace tracking algorithm was derived by gradient ascending method and its global convergence analysis was also accomplished. Numerical simulations and real application verifies the correctness of the criterion function and derived algorithm.

Key words: Rayleigh quotient, criterion function, minor component analysis, Lyapunov function

1 引言

在现代信号处理和数据分析领域, 次成分分析是一种常用的分析方法。次成分代表了信号或数据具有最小偏差的方向, 由多个次成分张成的空间称为次子空间。次成分分析是指利用次成分或次子空间进行信号处理的方法^[1]。次成分分析方法可以应用在很多领域, 如滤波器设计^[2]、Pisarenko 频率估计^[3]、阵列天线设计^[4]、总体最小二乘估计^[5]、波达方向 (DoA, direction of arrival) 估计^[6]等。

早期的次成分分析是通过特征值分解 (EVD,

eigenvalue decomposition) 或者奇异值分解 (SVD, singular value decomposition) 来实现的, 这类算法本质上是采用批处理方式, 存在计算量大和无法实时运算的问题^[7]。因此, 有研究者提出基于神经网络的算法, 该类算法的优点有: 1) 避免了自相关矩阵的直接计算, 2) 能够实现非平稳信号的跟踪, 3) 能够处理高维数据^[7]。基于神经网络的算法自提出以来迅速成为国内外研究的热点。

目前, 学者们提出了许多次成分分析算法, 如 Möller 算法^[8]、SDPM (stable data projection method)^[9]、PAST (projection approximation subspace tracking) 算法^[10]、Douglas 算法^[11]等, 这些算法都是基于启

收稿日期: 2019-07-04; 修回日期: 2019-09-09

通信作者: 高迎彬, welcome8793@sina.com

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

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

发式推理得到的, 缺乏对应的准则函数。信息准则规定了次成分分析方法的搜索方式, 利用准则函数可以快速确定算法的全局收敛域, 因此发展准则函数具有十分重要的意义。目前, 已经提出的准则函数有 AMEX (adaptive minor component extraction) 函数^[12]、Hasan 函数^[13]、OJAm 函数^[3]等。相比丰富的次成分分析算法而言, 准则函数并不多见, 需要进一步丰富和发展。

本文的主要工作是, 通过对 Rayleigh 商函数添加适当的惩罚项, 提出了一个新型的次子空间准则函数; 通过对所提准则函数的平稳点进行分析, 证明了准则函数的最优解是次子空间的一组基, 构建了准则函数与次子空间之间的联系; 通过梯度上升法导出了一个次子空间跟踪算法; 通过李雅普诺夫函数法确定了导出算法的全局收敛域。

2 新型准则函数的提出

2.1 预备知识

假设输入信号向量序列 $\{\mathbf{x}(k) | \mathbf{x}(k) \in \mathbb{R}^{n \times 1}, k = 0, 1, 2, \dots\}$ 是一个 0 均值的平稳随机过程, 其中 n 是信号向量的维数。 $\mathbf{R} = E[\mathbf{x}(k)\mathbf{x}^T(k)]$ 是输入信号序列的自相关矩阵。根据矩阵理论, \mathbf{R} 是一个正定对称矩阵。假设 λ_i 和 \mathbf{v}_i 分别是矩阵 \mathbf{R} 的特征值和其对应的特征向量, 其中 $i = 1, 2, \dots, n$, 则 \mathbf{R} 的特征值分解可以表示为

$$\mathbf{R} = \mathbf{U}\mathbf{A}\mathbf{U}^T = \mathbf{U}_r\mathbf{A}_r\mathbf{U}_r^T + \mathbf{U}_{n-r}\mathbf{A}_{n-r}\mathbf{U}_{n-r}^T \quad (1)$$

其中, $\mathbf{U} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$ 是由 \mathbf{R} 的全部 n 个特征向量构成的矩阵, $\mathbf{A} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ 是由 \mathbf{R} 所有特征值构成对角线元素的对角矩阵, \mathbf{U}_r 和 \mathbf{U}_{n-r} 分别是由任意 r 个特征向量和剩余的特征向量构成的矩阵, \mathbf{A}_r 和 \mathbf{A}_{n-r} 分别是对应的特征值为对角线元素构成的对角矩阵, r 是子空间的维数。

将矩阵 \mathbf{R} 的特征值按照从小到大的顺序排列, 即

$$0 < \lambda_1 \leq \dots \leq \lambda_r \leq \lambda_{r+1} \leq \dots \leq \lambda_n \quad (2)$$

特征值对应的特征向量也相应排列, 可以得到 \mathbf{R} 的另外一种特征值分解表示, 如式(3)所示。

$$\mathbf{R} = \mathbf{U}\mathbf{A}\mathbf{U}^T = \mathbf{U}_1\mathbf{A}_1\mathbf{U}_1^T + \mathbf{U}_2\mathbf{A}_2\mathbf{U}_2^T \quad (3)$$

其中, \mathbf{U}_1 是由最小的 r 个特征值对应的特征向量构成的矩阵, \mathbf{U}_2 则是由剩余的 $n-r$ 个特征值对应的特征向量构成的矩阵。根据信号处理理论^[14]可知,

自相关矩阵 \mathbf{R} 的最小的 r 个特征值对应的特征向量张成的空间等于信号的次子空间。

2.2 新型准则函数

为了跟踪信号次子空间, 在给定域 $\Omega = \{\mathbf{W} | \mathbf{W}^T\mathbf{R}\mathbf{W} > 0, \mathbf{W}^T\mathbf{W} \neq 0\}$ 内, 提出如式(4)所示的准则函数。

$$\mathbf{W}^* = \underset{\mathbf{W} \in \Omega}{\text{argmax}} J(\mathbf{W})$$

$$J(\mathbf{W}) = -\frac{1}{2} \text{tr} \left(\frac{\mathbf{W}^T\mathbf{R}\mathbf{W}}{\mathbf{W}^T\mathbf{W}} \right) + \frac{1}{2} \text{tr} [\ln(\mathbf{W}^T\mathbf{R}\mathbf{W}) - \mathbf{W}^T\mathbf{R}\mathbf{W}] \quad (4)$$

其中, \mathbf{W} 是神经网络的权矩阵, $\text{tr}(\cdot)$ 是矩阵的迹。 $J(\mathbf{W})$ 是一个无约束优化函数, 它由两部分构成, 第一部分是 Rayleigh 商函数, 作用是使算法能够收敛到需要的次子空间; 第二部分是一个惩罚函数, 作用是对权矩阵模值施加一个隐形约束, 保证算法迭代过程中模值的收敛性。

3 准则函数的平稳点分析

对于准则函数式(4)而言, 是否存在全局极大值、是否可以建立与次子空间的联系、是否存在其他局部极值是 3 个非常重要的问题, 本文以定理 1 和定理 2 对上述问题进行回答。

定理 1 在域 $\Omega = \{\mathbf{W} | \mathbf{W}^T\mathbf{R}\mathbf{W} > 0, \mathbf{W}^T\mathbf{W} \neq 0\}$ 内, 当且仅当 $\mathbf{W} = \mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q}$ 时, \mathbf{W} 是准则函数 $J(\mathbf{W})$ 的平稳点, 其中 $\mathbf{Q} \in \mathbb{R}^{r \times r}$ 是一个正交矩阵。

证明 由于 $\mathbf{W}^T\mathbf{R}\mathbf{W} > 0$ 和 $\mathbf{W}^T\mathbf{W} \neq 0$ 均是对称正定矩阵, 因此都是可逆矩阵。由式(4)可得, $J(\mathbf{W})$ 对矩阵 \mathbf{W} 的一阶微分存在, 且有

$$\begin{aligned} \nabla J(\mathbf{W}) = & -\left[\mathbf{R}\mathbf{W} - \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{R}\mathbf{W} \right] \cdot \\ & (\mathbf{W}^T\mathbf{W})^{-1} + \left[\mathbf{R}\mathbf{W}(\mathbf{W}^T\mathbf{R}\mathbf{W})^{-1} - \mathbf{R}\mathbf{W} \right] \end{aligned} \quad (5)$$

显然, 当 $\mathbf{W} = \mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q}$ 时, 有

$$\begin{aligned} \nabla J(\mathbf{W}) \Big|_{\mathbf{W}=\mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q}} = & -\left[\mathbf{R}\mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q} - \mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q}\mathbf{A}_r\mathbf{Q}^T\mathbf{A}_r^{-\frac{1}{2}}\mathbf{U}_r^T\mathbf{R}\mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q} \right] \mathbf{A}_r + \\ & \left[\mathbf{R}\mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q}(\mathbf{Q}^T\mathbf{A}_r^{-\frac{1}{2}}\mathbf{U}_r^T\mathbf{R}\mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q})^{-1} - \mathbf{R}\mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q} \right] = \\ & -\left[\mathbf{U}_r\mathbf{A}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q} - \mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q}\mathbf{A}_r\mathbf{Q}^T\mathbf{A}_r^{-\frac{1}{2}}\mathbf{U}_r^T\mathbf{R}\mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q} \right] \mathbf{A}_r + \\ & \left[\mathbf{U}_r\mathbf{A}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q}(\mathbf{Q}^T\mathbf{A}_r^{-\frac{1}{2}}\mathbf{U}_r^T\mathbf{R}\mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q})^{-1} - \mathbf{U}_r\mathbf{A}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q} \right] = \\ & -\left[\mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q} - \mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q} \right] \mathbf{A}_r + \left[\mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q} - \mathbf{U}_r\mathbf{A}_r^{-\frac{1}{2}}\mathbf{Q} \right] = \mathbf{0} \end{aligned} \quad (6)$$

反之，由平稳点定义可得，在 $J(W)$ 的平稳点有 $\nabla J(W) = \mathbf{0}$ ，即

$$\begin{aligned} & -[RW - W(W^T W)^{-1} W^T R W](W^T W)^{-1} + \\ & [RW(W^T R W)^{-1} - RW] = \mathbf{0} \end{aligned} \quad (7)$$

式(7)同时左乘 W^T 并化简得

$$W^T R W = I \quad (8)$$

令 $P = A_r^{\frac{1}{2}} U_r^T W$ ，并将其代入式(8)可得， $P^T P = I$ ，即 P 是一个正交矩阵。也就是说 $W = U_r \cdot A_r^{\frac{1}{2}} P$ 是 $J(W)$ 的平稳点，且 P 是一个正交矩阵。

证毕。

定理 2 在域 $\Omega = \{W | W^T R W > 0, W^T W \neq 0\}$ 内，当且仅当 $W = U_1 A_1^{-\frac{1}{2}} Q$ 时，准则函数 $J(W)$ 达到全局极大值。在全局极大值处， $J(W) = -\frac{1}{2} \sum_{i=1}^r (\lambda_i - 1)$ ，其他所有的平稳点 $W = U_r A_r^{\frac{1}{2}} Q (U_r \neq U_1)$ 都是 $J(W)$ 的鞍点，其中 $Q \in \mathbb{R}^{r \times r}$ 任意一个正交矩阵。

证明 根据定理 1 可知，任意的 $W = U_r A_r^{\frac{1}{2}} Q$ 均是准则函数 $J(W)$ 的平稳点，其中 U_r 是由 R 的任意 r 个特征向量构成的矩阵。令 U_r 中特征向量的位置序号构成的集合为 S_1 ，即 $S_1 = \{i_1, i_2, \dots, i_r\}$ 。同理， $S_2 = \{1, 2, \dots, r\}$ 为 U_1 中特征向量的位置序号构成的集合。

对于任意的 $S_1 (S_1 \neq S_2)$ 而言，该集合中必定存在某一个元素 j ，满足

$$j \in S_1 \text{ 且 } j \notin S_2 \quad (9)$$

将矩阵 U_r 中的第 j 列向量替换为 $\mathbf{v}_j + \varepsilon \mathbf{v}_k$ ($k \in S_2$ 且 $k \notin S_1$)，则形成新矩阵 U'_r 。显然，有 $\lambda_j > \lambda_k$ 。令 $M_1 = [\mathbf{0}, \dots, \mathbf{0}, \mathbf{v}_k, \mathbf{0}, \dots, \mathbf{0}]$ ，即矩阵 M_1 中只有第 j 列为 \mathbf{v}_k ，而其他列都是 $\mathbf{0}$ ，则

$$U'_r = U_r + \varepsilon M_1 \quad (10)$$

令 $W_1 = U_r A_r^{\frac{1}{2}} Q$ 和 $W_2 = U'_r A_r^{\frac{1}{2}} Q$ ，则

$$\begin{aligned} & W_2^T R W_2 = \\ & Q^T A_r^{\frac{1}{2}} (U_r + \varepsilon M_1)^T R (U_r + \varepsilon M_1) A_r^{\frac{1}{2}} Q = \\ & I + \varepsilon^2 Q^T C_1 Q \end{aligned} \quad (11)$$

其中， $C_1 = \text{diag}(0, \dots, 0, \frac{\lambda_k}{\lambda_j}, 0, \dots, 0)$ 是一个对角矩阵。

进一步有

$$(W_2^T W_2)^{-1} =$$

$$Q^T A_r^{\frac{1}{2}} (I + \varepsilon^2 M_1^T M_1)^{-1} A_r^{\frac{1}{2}} Q = Q^T A_r C_2 Q \quad (12)$$

其中， $C_2 = \text{diag}(1, \dots, \frac{1}{1 + \varepsilon^2}, 1, \dots, 1)$ 是一个对角矩阵。

将式(11)和式(12)代入式(4)可得

$$\begin{aligned} & J(W_2) = \\ & -\frac{1}{2} \text{tr} [W_2^T R W_2 (W_2^T W_2)^{-1}] + \\ & \frac{1}{2} \text{tr} [\ln(W_2^T R W_2) - W_2^T R W_2] = \\ & -\frac{1}{2} \text{tr} [(I + \varepsilon^2 Q^T C_1 Q)(Q^T A_r C_2 Q)] + \\ & \frac{1}{2} \text{tr} [\ln(I + \varepsilon^2 Q^T C_1 Q)] - \frac{1}{2} \text{tr} (I + \varepsilon^2 Q^T C_1 Q) = \\ & -\frac{1}{2} \left[\sum_{i=1, i \neq j}^r \lambda_i + \frac{\lambda_j + \varepsilon^2 \lambda_k}{1 + \varepsilon^2} + r \right] + o(\varepsilon^4) \end{aligned} \quad (13)$$

同理，有

$$\begin{aligned} & J(W_1) = \\ & -\frac{1}{2} \text{tr} [W_1^T R W_1 (W_1^T W_1)^{-1}] + \\ & \frac{1}{2} \text{tr} [\ln(W_1^T R W_1) - W_1^T R W_1] = \\ & -\frac{1}{2} \text{tr} [Q^T A_r Q] + \frac{1}{2} \text{tr} [\ln(I) - I] = \\ & -\frac{1}{2} \left(\sum_{i=1}^r \lambda_i + r \right) \end{aligned} \quad (14)$$

将式(13)和式(14)代入式(15)有

$$J(W_2) - J(W_1) = \frac{\varepsilon^2}{2(1 + \varepsilon^2)} (\lambda_j - \lambda_k) + o(\varepsilon^4) \quad (15)$$

由于 $\lambda_j > \lambda_k$ ，则根据式(15)可得，在平稳点 $W = U_r A_r^{\frac{1}{2}} Q (U_r \neq U_1)$ 处，沿着向量 \mathbf{v}_k 的方向， $J(W)$ 是递增的。此外，将矩阵 U_r 中的 \mathbf{v}_j 替换为 $\mathbf{v}_j + \varepsilon \mathbf{v}_j$ ，则可以获得新的矩阵 U''_r 。令 $M_2 = [\mathbf{0}, \dots, \mathbf{0}, \mathbf{v}_j, \mathbf{0}, \dots, \mathbf{0}]$ ，即矩阵 M_2 中只有第 j 列为 \mathbf{v}_j ，而其他列都是 $\mathbf{0}$ ，则

$$U''_r = U_r + \varepsilon M_2 \quad (16)$$

定义矩阵 $W_3 = U''_r A_r^{\frac{1}{2}} Q$ ，则

$$\begin{aligned} & W_3^T R W_3 = \\ & Q^T A_r^{\frac{1}{2}} (U_r + \varepsilon M_2)^T R (U_r + \varepsilon M_2) A_r^{\frac{1}{2}} Q = \end{aligned}$$

$$\begin{aligned} & \mathbf{Q}^T \mathbf{A}_r^{-\frac{1}{2}} (\mathbf{A}_r + 2\varepsilon \lambda_j \mathbf{D}_1 + \varepsilon^2 \mathbf{M}_2^T \mathbf{R} \mathbf{M}_2) \mathbf{A}_r^{-\frac{1}{2}} \mathbf{Q} = \\ & \mathbf{I} + (2\varepsilon + \varepsilon^2) \mathbf{Q}^T \mathbf{D}_1 \mathbf{Q} \end{aligned} \quad (17)$$

其中, $\mathbf{D}_1 = \text{diag}(0, \dots, 0, 1, 0, \dots, 0)$ 是一个对角矩阵。

进一步有

$$\begin{aligned} & (\mathbf{W}_3^T \mathbf{W}_3)^{-1} = \\ & \left[\mathbf{Q}^T \mathbf{A}_r^{-\frac{1}{2}} (\mathbf{U}_r + \varepsilon \mathbf{M}_2)^T (\mathbf{U}_r + \varepsilon \mathbf{M}_2) \mathbf{A}_r^{-\frac{1}{2}} \mathbf{Q} \right]^{-1} = \\ & \mathbf{Q}^T \mathbf{A}_r^{-\frac{1}{2}} (\mathbf{I} + 2\varepsilon \mathbf{D}_1 + \varepsilon^2 \mathbf{D}_1)^{-1} \mathbf{A}_r^{-\frac{1}{2}} \mathbf{Q} = \mathbf{Q}^T \mathbf{A}_r \mathbf{D}_2 \mathbf{Q} \end{aligned} \quad (18)$$

其中, $\mathbf{D}_2 = \text{diag}(1, \dots, 1, \frac{1}{(1+\varepsilon)^2}, 1, \dots, 1)$ 也是一个对角矩阵。利用式(17)和式(18)可得

$$\begin{aligned} & J(\mathbf{W}_3) = \\ & -\frac{1}{2} \text{tr} \left[\mathbf{W}_3^T \mathbf{R} \mathbf{W}_3 (\mathbf{W}_3^T \mathbf{W}_3)^{-1} \right] + \\ & \frac{1}{2} \text{tr} \left[\ln(\mathbf{W}_3^T \mathbf{R} \mathbf{W}_3) - \mathbf{W}_3^T \mathbf{R} \mathbf{W}_3 \right] = \\ & -\frac{1}{2} \text{tr} \left[(\mathbf{I} + (2\varepsilon + \varepsilon^2) \mathbf{Q}^T \mathbf{D}_1 \mathbf{Q}) (\mathbf{Q}^T \mathbf{A}_r \mathbf{D}_2 \mathbf{Q}) \right] + \\ & \frac{1}{2} \text{tr} \left[\ln(\mathbf{I} + (2\varepsilon + \varepsilon^2) \mathbf{Q}^T \mathbf{D}_1 \mathbf{Q}) \right] - \\ & \frac{1}{2} \text{tr} (\mathbf{I} + (2\varepsilon + \varepsilon^2) \mathbf{Q}^T \mathbf{D}_1 \mathbf{Q}) = \\ & -\frac{1}{2} \sum_{i=1}^r \lambda_i - \frac{\varepsilon^3}{3} - \frac{r}{2} + o(\varepsilon^3) \end{aligned} \quad (19)$$

根据式(14)和式(19)可得

$$J(\mathbf{W}_3) - J(\mathbf{W}_1) = -\frac{\varepsilon^3}{3} + o(\varepsilon^3) \quad (20)$$

式(20)表明, 在平稳点 $\mathbf{W} = \mathbf{U}_r \mathbf{A}_r^{-\frac{1}{2}} \mathbf{Q}$ ($\mathbf{U}_r \neq \mathbf{U}_1$) 处, 沿着向量 \mathbf{v}_k 的方向, $J(\mathbf{W})$ 是递减的。综合式(15)和式(20)可得, 平稳点 $\mathbf{W} = \mathbf{U}_r \mathbf{A}_r^{-\frac{1}{2}} \mathbf{Q}$ ($\mathbf{U}_r \neq \mathbf{U}_1$) 是 $J(\mathbf{W})$ 的鞍点。

反之, 可以证明当 $\mathbf{W} = \mathbf{U}_1 \mathbf{A}_1^{-\frac{1}{2}} \mathbf{Q}$ 时, 如果受到 \mathbf{v}_i ($i=1, 2, \dots, n$) 扰动, $J(\mathbf{W})$ 均会增减。因此 $\mathbf{W} = \mathbf{U}_1 \mathbf{A}_1^{-\frac{1}{2}} \mathbf{Q}$ 是 $J(\mathbf{W})$ 的全局极大点, 将其代入式(4)可得

$$J(\mathbf{W}) = -\frac{1}{2} \sum_{i=1}^r (\lambda_i - 1) \quad (21)$$

证毕。

定理 1 和定理 2 表明, $J(\mathbf{W})$ 存在全局极大值, 在全局极大值时, \mathbf{W} 是次子空间的一组基, 且

$\mathbf{W}^T \mathbf{R} \mathbf{W} = \mathbf{I}$ 可以被唯一确定。由于 $J(\mathbf{W})$ 只有全局极大值而没有其他局部极值, 因此通过迭代法(如梯度法)导出的算法可以保证收敛到需要的次子空间。

4 次子空间跟踪算法及其收敛域分析

4.1 子空间跟踪算法

通过定理 1 可得, 式(5)是 $J(\mathbf{W})$ 的一阶微分。假设在第 k 次迭代时, 神经网络的权矩阵为 \mathbf{W}_k , 则通过利用梯度上升法可以给出下一时刻的权矩阵更新式, 如式(22)所示。

$$\begin{aligned} \mathbf{W}_{k+1} = & \mathbf{W}_k + \mu \left[\mathbf{R} \mathbf{W}_k (\mathbf{W}_k^T \mathbf{R} \mathbf{W}_k)^{-1} - \mathbf{R} \mathbf{W}_k \right] - \\ & \mu \left[\mathbf{R} \mathbf{W}_k - \mathbf{W}_k (\mathbf{W}_k^T \mathbf{W}_k)^{-1} \mathbf{W}_k^T \mathbf{R} \mathbf{W}_k \right] (\mathbf{W}_k^T \mathbf{W}_k)^{-1} \end{aligned} \quad (22)$$

其中, μ 是算法的学习因子, 满足 $0 < \mu < 1$ 。经过多次迭代运算后, 神经网络权矩阵将最终收敛到次子空间的一组基。

4.2 算法的收敛域

本节对导出算法式(22)的收敛性进行讨论。当学习因子 μ 足够小时, 离散差分式(22)可以近似为与之相对应的连续微分方程^[15]。

$$\begin{aligned} \frac{d\mathbf{W}(t)}{dt} = & \left[\mathbf{R} \mathbf{W}(t) (\mathbf{W}^T(t) \mathbf{R} \mathbf{W}(t))^{-1} - \mathbf{R} \mathbf{W}(t) \right] - \\ & \left[\mathbf{R} \mathbf{W}(t) - \mathbf{W}(t) (\mathbf{W}^T(t) \mathbf{W}(t))^{-1} \mathbf{W}^T(t) \mathbf{R} \mathbf{W}(t) \right] \cdot \\ & (\mathbf{W}^T(t) \mathbf{W}(t))^{-1} \end{aligned} \quad (23)$$

其中, $t = \mu k$ 。由此, 算法的全局收敛性可以通过对该连续微分方程对应的动态系统分析来完成。根据李雅普诺夫第二定律, 动态系统的收敛性分析主要涉及以下两方面: 1) 动态系统是否能够全局收敛到次子空间; 2) 动态系统可以吸引的范围是多少, 即什么样的初始化条件能够保证系统的全局收敛性。

定理 3 将给出上述 2 个问题的答案。

定理 3 给定常微分方程式(23), 假如初始化全矩阵满足 $\mathbf{W}(0) \in \Omega$, 则当 $t \rightarrow \infty$ 时, $\mathbf{W}(t)$ 必将全局渐进收敛到集合 $\{\mathbf{W} | \mathbf{W} = \mathbf{U}_1 \mathbf{A}_1^{-\frac{1}{2}} \mathbf{Q}\}$ 中的一个点, 其中 $\mathbf{Q} \in \mathbb{R}^{r \times r}$ 任意一个正交矩阵。

证明 令 $L(\mathbf{W}) = -J(\mathbf{W})$ 。根据定理 1 和定理 2, $L(\mathbf{W})$ 与 $J(\mathbf{W})$ 有相同的平稳点, 且在 $\mathbf{W} = \mathbf{U}_1 \mathbf{A}_1^{-\frac{1}{2}} \mathbf{Q}$ 处, $L(\mathbf{W})$ 取得全局极小值。利用求导链式法, 则

$$\frac{dL(\mathbf{W}(t))}{dt} = -\text{tr} \left[\nabla J(\mathbf{W}(t)) \frac{d\mathbf{W}^T(t)}{dt} \right] \quad (24)$$

将式(5)和式(23)代入式(24), 则很容易得到,

对于任意的 $\mathbf{W} \in \Omega$ ，均有 $\frac{dL(\mathbf{W}(t))}{dt} < 0$ 。因此 $L(\mathbf{W})$ 构成了式(23)的李雅普诺夫函数。根据李雅普诺夫第二定律，从任意初始化权矩阵出发的动态系统都收敛到相同的不变集 $P = \{\mathbf{W} | \nabla J(\mathbf{W}(t)) = 0\}$ 。由于 $\mathbf{W}(t) = U_r A_r^{-\frac{1}{2}} \mathbf{Q}$ 是不稳定的鞍点，因此 $\mathbf{W}(t)$ 必将全局渐进收敛到 $\{\mathbf{W}(t) | \mathbf{W}(t) = U_1 A_1^{-\frac{1}{2}} \mathbf{Q}\}$ 中的一个点。

证毕。

5 仿真实验

本节通过 3 个仿真实验来验证所提准则函数和导出算法的正确性。第一个实验提供了一种特殊情况下的准则函数曲线，第二个实验展示了导出算法在次成分分析方面的优越性，第三个实验是算法在 DoA 估计中的应用。

5.1 准则函数曲线

令准则函数 $J(\mathbf{W})$ 中 $\mathbf{R} = \text{diag}(2, 1)$ 。当提取子空间维数为 1，即 $r=1$ 时。如果 $\mathbf{W} = [\mathbf{W}_{11}, \mathbf{W}_{12}]^T$ ，则通过适当简化有

$$J(\mathbf{W}) = -\frac{2\mathbf{W}_{11}^2 + \mathbf{W}_{12}^2}{2(\mathbf{W}_{11}^2 + \mathbf{W}_{12}^2)} + \frac{1}{2} \ln(2\mathbf{W}_{11}^2 + \mathbf{W}_{12}^2) - \frac{1}{2}(2\mathbf{W}_{11}^2 + \mathbf{W}_{12}^2) \quad (25)$$

图 1 是利用 Matlab 绘出上述情况下的 $J(\mathbf{W})$ 曲线。从图 1 中可以看出， $J(\mathbf{W})$ 存在全局极大值，而没有局部极值，这与定理 2 的分析结果相一致。当权矩阵 $\mathbf{W} = [0, \pm 1]^T$ 时，准则函数 $J(\mathbf{W})$ 取得全局极大值，此时 $\mathbf{W} = [0, \pm 1]^T$ 正好是 $\mathbf{R} = \text{diag}(2, 1)$ 的次成分，从而证明了所提准则函数的正确性。

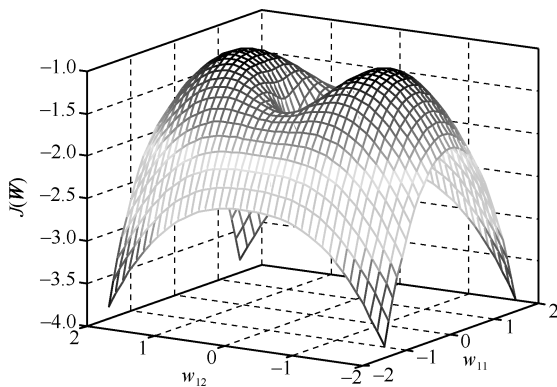


图 1 准则函数的 3D 曲线

5.2 次成分分析实验

5.1 节实验主要通过一个特例来证明所提信息准则的正确性，本节实验将考察准则函数导出算法进行次成分分析的能力。首先，利用文献[10]的方法随机生成一个 5×5 对称正定矩阵，如式(26)所示。

$$\mathbf{R}_1 = \begin{bmatrix} 7.1887 & 1.3265 & -0.8327 & -0.2834 & 0.1340 \\ 1.3265 & 5.1816 & -0.6110 & 0.7910 & 0.1533 \\ -0.8327 & -0.6110 & 1.1608 & 0.3447 & -0.7416 \\ -0.2834 & 0.7910 & 0.3447 & 8.0164 & 0.2479 \\ 0.1340 & 0.1533 & -0.7416 & 0.2479 & 4.4057 \end{bmatrix} \quad (26)$$

该矩阵的最小特征值为 $\lambda_1 = 0.8348$ ，其对应的次成分特征向量为

$$\mathbf{v}_1 = [-0.0968, -0.1107, -0.9674, 0.0616, -0.1968]^T \quad (27)$$

然后，分别利用导出算法、OJAm 算法^[3]和 Douglas 算法^[11]对该矩阵的次成分进行提取。为了定量描述算法的性能，这里引入 2 个评价函数^[10]。

1) 方向余弦

方向余弦 (DC, direction cosine) 实际上是权矩阵与次成分方向之间的夹角，如式(28)所示。如果 $\text{DC} = 1$ ，则表示 \mathbf{W}_k 与 \mathbf{v}_1 方向重合。

$$\text{DC}(k) = \frac{\mathbf{W}_k^T \mathbf{v}_1}{\|\mathbf{W}_k\| \|\mathbf{v}_1\|} \quad (28)$$

2) 权矩阵过程模值

对于 OJAm 算法和 Douglas 算法，已经证明算法收敛时 $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ ，因此过程模值取 $\text{Norm}(k) = \mathbf{W}_k^T \mathbf{W}_k$ ；对于导出算法而言，根据式(8)有 $\mathbf{W}^T \mathbf{R} \mathbf{W} = \mathbf{I}$ ，因此取 $\text{Norm}(k) = \mathbf{W}_k^T \mathbf{R} \mathbf{W}_k$ 。

在迭代过程中，为了保证算法公平比较，3 个算法采用相同的初始化权矩阵和学习因子。本实验中，初始化权矩阵是随机产生的（矩阵的每个元素均是高斯白噪声），学习因子 $\mu = 0.01$ 。3 种算法的仿真结果如图 2 和图 3 所示，该结果曲线是 100 次独立仿真结果的平均值。

从图 2 中可以看出，所提算法的方向余弦曲线最终收敛到了单位 1，即所提算法能够收敛到次成分的方向。从图 3 中可以看出，迭代过程中，权矩阵模值是有界的且最终也收敛到了 1，与定理 1 的

分析结果相吻合。因此，综合图 2 与图 3 结果可知，所提算法具备次成分分析的能力。对比图 2 和图 3 中 3 种算法的表现可知，相比 OJAm 算法和 Douglas 算法，本文所提算法在方向余弦和权矩阵模值方面具有较快的收敛速度。

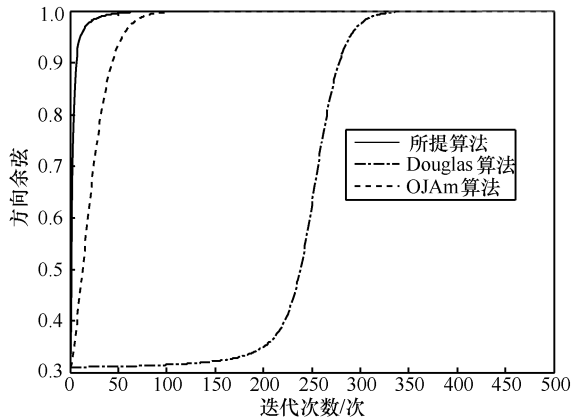


图 2 方向余弦曲线

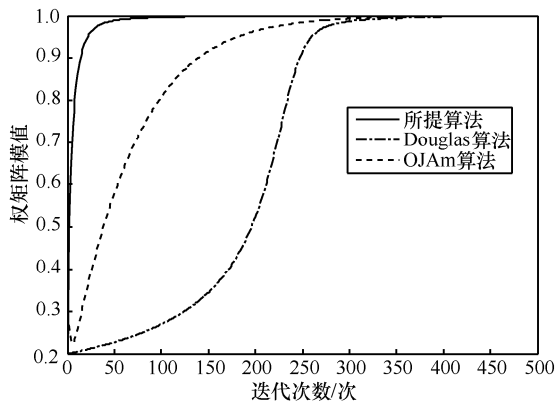


图 3 权矩阵模值曲线

5.3 DoA 估计

假设一个具有 8 个阵子的线性等距天线阵列，阵子的间距为半个波长。3 个远场信号的入射角（即 DoA 角）分别为 12° 、 19° 和 27° 。信号中的噪声为加性高斯白噪声，且信噪比为 20 dB。分别利用所提算法、OJAm 算法^[3]和 Douglas 算法^[11]对信号的次子空间进行估计，通过 MUSIC 法计算 DoA 角。本节实验中，3 种算法的初始化方法与实验 2 相同，初始化权矩阵是随机产生的，学习因子 $\mu = 0.1$ 。为了评价算法对子空间的估计结果，算法迭代过程中分别计算 3 个算法估计子空间与单位矩阵之间的正交误差，如式(29)所示。

$$De(k) = \|A - I_r\|_F \quad (29)$$

其中，对于 Douglas 算法和 OJAm 算法有 $A = W^T W$ ；对于所提算法有 $A = W^T R W$ 。图 4 是 3 个算法的正交误差曲线，图 5 是利用导出算法得到的谱密度曲线。

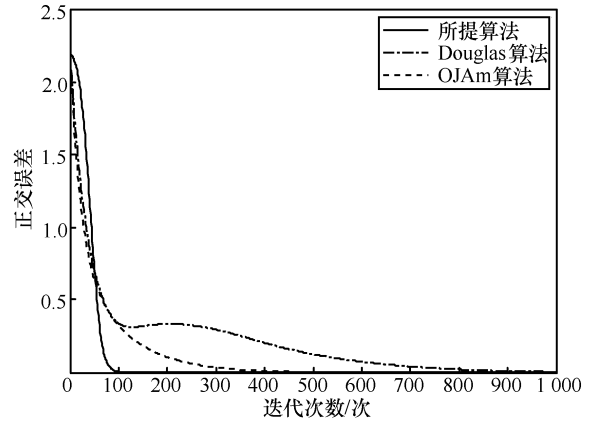


图 4 正交误差曲线

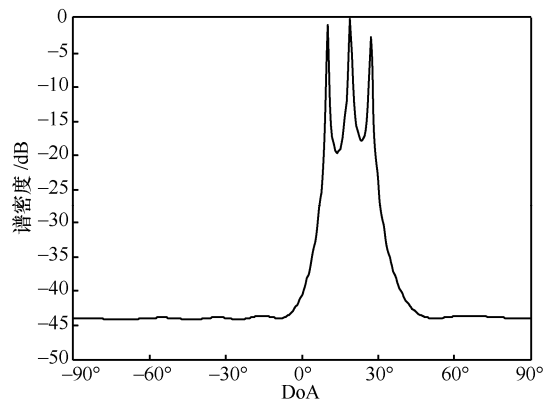


图 5 所提算法的 DoA 估计曲线

从图 4 中可以看出，经过大约 100 次迭代运算后，所提算法的正交误差已经趋于 0，即权矩阵已经收敛到了信号的次子空间。通过与其他 2 个算法对比可以发现，所提算法的收敛速度要快于其他 2 种算法。从图 5 中可以看出，谱密度曲线分别在 3 个 DoA 角处取得极值。因此可以得出结论，所提算法能够有效地解决 DoA 估计问题，而且收敛速度优于同类型其他算法。

6 结束语

次成分分析是信号处理和数据分析领域一门重要的工具，而寻找准则函数在次成分分析算法中具有十分重要的地位。本文首先提出了一个新型的次子空间跟踪准则函数，并通过分析准则函数平稳点建立起了准则函数与次子空间之间的关系；基于

此信息准则构建了一个新的次子空间跟踪算法,并证明了算法的全局收敛性。仿真实验表明,与一些现有算法相比,本文提出的算法具有较快的收敛速度。

参考文献:

- [1] FENG X, KONG X, MA H, et al. Unified and coupled self-stabilizing algorithms for minor and principal eigen-pairs extraction [J]. *Neural Processing Letters*, 2017, 45(1): 197-222.
- [2] KALLA U K. Minor Component Analysis Based Anti-Hebbian Neural Network Scheme of Decoupled Voltage and Frequency Controller (DVFC) for Nanohydro System [C]//Power India International Conference. 2017: 25-27.
- [3] GAO Y, KONG X, ZHANG H, et al. A weighted information criterion for multiple minor components and its adaptive extraction algorithms [J]. *Neural Networks*, 2017, 89(5): 1-10.
- [4] ARJOMANDI-LARI M, KARIMI M. Generalized YAST algorithm for signal subspace tracking [J]. *Signal Processing*, 2015, 117: 82-95.
- [5] NGUYEN V D, ABED-MERAİM K, LINH-TRUNG N, et al. Generalized minimum noise subspace for array processing [J]. *IEEE Transactions on Signal Processing*, 2017, 65(14): 3789-3802.
- [6] 尤国红,邱天爽,夏楠,等.基于均匀圆阵的扩展循环 MUSIC 算法[J]. *通信学报*, 2014, 35(2):
YOU G H, QIU T S, XIA N, et al. Novel extended cyclic MUSIC algorithm based on uniform circular array [J]. *Journal on Communications*, 2014, 35(2): 9-15.
- [7] MÖLLER R. Derivation of coupled PCA and SVD learning rules from a Newton zero-finding framework [R]. Berlin: Computer Engineering, Faculty of Technology, Bielefeld University, (2017-06-05)[2019-07-04].
- [8] GAO Y, KONG X, HU C, et al. Convergence analysis of möller algorithm for estimating minor component [J]. *Neural Processing Letters*, 2015, 42(2): 1-14.
- [9] FARZAN H. Steady-state statistical performance analysis of subspace tracking methods [J]. *IEEE Transactions on Signal Processing*, 2016, 64(18): 4781-4791.
- [10] NGUYEN T D, YAMADA I. Necessary and sufficient conditions for convergence of the DDT systems of the normalized PAST algorithms [J]. *Signal Processing*, 2014, 94(5): 288-299.
- [11] LI H, DU B, KONG X, et al. A generalized minor component extraction algorithm and its analysis[J]. *IEEE Access*, 2018, 6: 36771-36779.
- [12] KONG X, HU C, DUAN Z. Principal component analysis networks and algorithms [M]. Beijing: Springer Singapore, 2017.
- [13] FENG X, KONG X, MA H, et al. Unified and coupled self-stabilizing

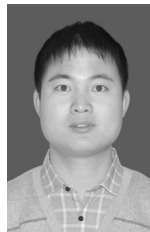
algorithms for minor and principal eigen-pairs extraction [J]. *Neural Processing Letters*, 2016, 1: 1-26.

- [14] DU K L, SWAMY M N S. Principal component analysis [M]. New York: Springer London, 2014.
- [15] LIU L, GE R, MENG J, et al. Dual subspace learning via geodesic search on Stiefel manifold [J]. *International Journal of Machine Learning & Cybernetics*, 2013, 5(5): 753-759.

[作者简介]



徐中英(1978-),男,江苏徐州人,博士,火箭军工程大学讲师,主要研究方向为信号处理。



高迎彬(1986-),男,河北安平人,博士,火箭军工程大学工程师,主要研究方向为自适应信号处理。



孔祥玉(1967-),男,山西洪洞人,博士,火箭军工程大学教授,主要研究方向为信号处理、故障诊断等。



杜伯阳(1990-),男,山东滨州人,火箭军工程大学博士生,主要研究方向为系统特征提取。