

主奇异子空间跟踪算法与性能分析

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: 主奇异子空间分析是一种自适应的神经网络信号处理技术, 广泛应用于现代信号处理中. 本文提出一种新的主奇异子空间跟踪信息准则, 并以此为基础推导出一种在线的梯度流神经网络算法. 理论分析表明, 信息准则具有唯一的全局最小值, 且最小值对应的状态矩阵能够恰好张成输入信号的主奇异子空间. 该算法具有良好的收敛能力, 强大的自稳性能, 且当输入信号呈现出奇异互相关特性时, 仍呈现出良好的跟踪效果. 分别采用李雅普诺夫函数方法和常微分方程方法分析算法的收敛性能和自稳性能. MATLAB 仿真算例验证了算法的性能.

关键词: 主奇异子空间; 收敛性分析; 自稳性能分析; 神经网络

引用格式: 杜柏阳, 孔祥玉, 冯晓伟. 主奇异子空间跟踪算法与性能分析. 控制理论与应用, 2020, 37(7): 1491 – 1500

DOI: 10.7641/CTA.2020.90498

Algorithm and its performance analysis of principal singular subspace tracking

DU Bo-yang^{1†}, KONG Xiang-yu¹, FENG Xiao-wei²

(1. College of Missile Engineering, Rocket Force University of Engineering, Xi'an Shaanxi 710025, China;
2. College of Nuclear Engineering, Rocket Force University of Engineering, Xi'an Shaanxi 710025, China)

Abstract: Principal singular subspace analysis is an adaptive neural network signal processing technique which has been widely applied in modern signal processing. In this paper, a novel information criterion for principal singular subspace tracking is proposed and based on the information criterion an online gradient flow neural network algorithm is derived. Theoretical analysis shows that the information criterion exhibits a unique global minimum point where the state matrices corresponding to the minimum point can exactly span the principal singular subspace of the input signals. The proposed algorithm has a good performance in convergence and an excellent self-stabilizing property. What is more, even if the input signals present a singular cross-correlation characteristic, the proposed algorithm can still track the principal singular subspace of the input signals efficiently. The convergence and self-stability are analyzed via the Lyapunov function approach and ordinary differential equation approach, respectively. MATLAB simulation results verify the effectiveness of the proposed algorithm.

Key words: principal singular subspace (PSS); convergence analysis; self-stabilizing property; neural networks

Citation: DU Boyang, KONG Xiangyu, FENG Xiaowei. Algorithm and its performance analysis of principal singular subspace tracking. *Control Theory & Applications*, 2020, 37(7): 1491 – 1500

1 引言

在现代信号处理中, 许多问题可以转化协相关矩阵的分解问题, 从而通过奇异值分解的方法最终有效地解决^[1]. 奇异特征空间提取方法广泛应用于信息产业等领域, 例如低秩优化^[2]、人脸识别^[3]、功耗分析^[4]、故障诊断^[5]、波达角估计^[6]等. 通常, 主奇异特征向量 (principal singular component, PSC) 指的是在由两路输入信号得来的协相关矩阵中, 与几个主要奇

异值对应的特征向量, 该向量是成对存在的, 区分为左和右两个向量. 由这些特征向量张成的空间称为主奇异特征子空间 (principal singular subspace, PSS).

实现奇异特征提取的最初方法主要依赖矩阵分解理论^[7-8], 基于矩阵分解的处理算法具有良好的鲁棒性, 但他们是批处理算法, 只能处理离线数据, 因而该方法计算效率很受限, 而且对硬件设备要求比较高. 神经网络算法具有在线学习的特点, 能够实现奇

收稿日期: 2019-06-29; 录用日期: 2020-03-14.

[†]通信作者. E-mail: duboyange@163.com; Tel.: +86 15002967483.

本文责任编辑: 孙长银.

国家自然科学基金项目 (61374120, 61673387, 61903375, 61833016) 资助.

Supported by the National Natural Science Foundation of China (61374120, 61673387, 61903375, 61833016).

异特征的在线提取,因而被广泛开展研究,例如Hebbian规则算法^[9-10].早期的神经网络算法需要将协相关矩阵 R 改造成 $R^T R$ 或者 RR^T .这种做法确保了被分解矩阵的对称性,但是还存在秩缺失情况下的算法稳定性问题^[11],这在实际应用中是不可避免的.针对这个问题,有学者提出基于梯度的神经网络算法^[12-14],降低了对协相关矩阵的要求.Kung和Diamantaras^[13]提出了互相关神经网络(cross-correlation neural network, CNN)模型,该模型可以直接从高维信息流中提取互相关特征.然而,这一类算法能提取信号特征中与最大的奇异值相对应的一对奇异成分,还没有考虑在信号中提取奇异特征子空间的问题.目前,针对奇异特征子空间提取方法的研究亟待进行.

通常来讲,诸如Hebbian类的神经网络算法大多是通过优化信息准则的方式获得的^[15].信息准则包含算法的优化目标,由梯度下降法、牛顿法等不同方法进行推导,可得到相应的特征提取算法.优化方式不同,对应的算法也不尽相同.因而信息准则对于优化算法的搜索方向,提高算法的自稳定性,以及扩展算法的形式等具有十分重要的意义.Feng等^[16]在Kung等^[13]的CNN模型基础上,设计了新的优化框架,并给出了一个信息准则

$$\begin{aligned} \min J_{\text{PSS}}(U, V), \\ J_{\text{PSS}}(U, V) = \text{tr}[U^T R V], \end{aligned} \quad (1)$$

其中: $U \in \mathbb{R}^{M \times r}$ 和 $V \in \mathbb{R}^{N \times r}$ 分别为左右奇异特征向量张成的奇异子空间,当 $r = 1$ 时, U 和 V 退化为单个主奇异特征向量.该信息准则表达了算法优化目标的核心内容,即寻找能够使矩阵 $U^T R V$ 的秩最小的左右奇异子空间.该信息准则的提出对于后来学者^[17-18]的研究工作具有一定的启发意义.一方面,该信息准则本身体现了从信息角度对奇异特征信息提取过程的物理意义;另一方面,该信息准则为后续信息准则在数学和物理两种意义上的延伸提供了借鉴和参考的依据,可以丰富算法的功能,提高算法在工程中的应用价值.

众所周知,在实际工程应用中,收敛性和自稳定性是神经网络算法十分重要的性质.收敛性是算法最基本的性能,这里不做赘述.自稳定性是指算法不局限于特定的初始设定值,在信息变化较大的情况下,算法依然保持收敛的性质.Kong等^[17]分析Kung等^[13]的算法表明,该算法在某些情况下不具有自稳定性,特别是当协相关矩阵接近奇异状态的时候.这在一定程度上降低了Kung算法在实际应用中的价值.

在本文中,针对输入信号奇异特征的在线提取问题,作者提出一种新的信息准则,采用梯度下降法推导出对应的奇异子空间提取算法.通过李雅普诺夫函数方法对提出的算法进行收敛性分析,求解算法实现

收敛的基本条件.通过常微分方程(ordinary differential equation, ODE)方法对提出的算法进行自稳定分析,求解算法实现自稳定的基本条件.最后通过MATLAB仿真验证算法的理论分析结果,检验算法的性能,为该算法在实际生产生活中的应用奠定了良好的基础.

2 信息准则和算法的提出

对于一个 N 维数据序列 $\{\mathbf{y}(k)\}$ 和 M 维数据序列 $\{\mathbf{x}(k)\}$, 定义其协相关矩阵为 $R_{xy} = E[\mathbf{y}(k)\mathbf{x}^T(k)]$. 根据该定义, 获取协相关矩阵必定要求遍历全部输入信号, 而这对在线的信号处理过程是无法实现的. 为实现算法在线提取信号特征, 需要通过在线估计的方式求得信号的协相关矩阵^[19]. 估计的方式大体分为两种情况. 如果两个数据序列同为平稳序列, 那么它们的协相关矩阵可以估计为

$$R(k+1) = \frac{k}{k+1}R(k) + \frac{1}{k+1}\mathbf{x}(k)\mathbf{y}^T(k), \quad (2)$$

其中 $R(k)$ 为第 k 次迭代中协相关矩阵的估计. 当两个数据序列任一个是缓变序列时, 它们的协相关矩阵被估计为

$$R(k+1) = \alpha R(k) + (1-\alpha)\mathbf{x}(k)\mathbf{y}^T(k), \quad (3)$$

其中 α 为遗忘因子, 通过遗忘因子可以保证过去输入的权重总是小于最近输入的权重. 一般来说, $\alpha \in (0, 1)$ 其具体取值需要依赖于信号的性质, 缓变序列的变化越缓慢的系统 α 的取值越接近 1, 反之则越接近 0. 相比而言, 第 2 种估计方式相比于第 1 种估计方式, 更加重视邻近迭代中的信号信息.

对一个矩阵特征值分解, 可以分为两个部分^[1]:

$$R = L_1 S_1 R_1^T + L_2 S_2 R_2^T, \quad (4)$$

其中: $L_1 = [l_1 \ l_2 \ \cdots \ l_r] \in \mathbb{R}^{M \times r}$ 和 $R_1 = [r_1 \ r_2 \ \cdots \ r_r] \in \mathbb{R}^{N \times r}$ 分别为左右主奇异特征向量张成的 PSS, 奇异特征值则由矩阵 $S_1 = \text{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_r\} \in \mathbb{R}^{r \times r}$ 对角线上各个元素表示. 本文仅讨论主奇异值为正的情况. $L_1 S_1 R_1^T$ 是最小均方误差意义下对 R 阵的 r 秩估计, 是 R 阵的主要成分, r 小于等于 M 和 N 中的较小者. $L_2 S_2 R_2^T$ 表示 R 矩阵 $SV D$ 的次要成分, 其中 $L_2 = [l_{r+1} \ l_{r+2} \ \cdots \ l_M] \in \mathbb{R}^{M \times (M-r)}$, $R_2 = [r_{r+1} \ r_{r+2} \ \cdots \ r_N] \in \mathbb{R}^{N \times (N-r)}$. 另外, $L_1^T L_1 = R_1^T R_1 = I_r$, $L_2^T L_2 = R_2^T R_2 = I_{\min\{M, N-r\}}$ 都是正交矩阵.

构造一个新型的信息准则如下:

$$\begin{aligned} \min \{J(U, V)\}, \\ J(U, V) = \\ \text{tr}(-4U^T R V + U^T R V U^T U + U^T R V V^T V), \end{aligned} \quad (5)$$

对该信息准则细节分析通过两个定理^[1]给出.

定理 1 L_1 和 R_1 为 R 的左右主奇异特征矩阵, A

为奇异值矩阵为 $S_1 = L_1^T R R_1$, 则存在旋转矩阵 Q_U 和 Q_V , 使 $J(U, V)$ 处于平衡状态, 其中: $Q_U = L_1^T U$, $Q_V = R_1^T V$.

证 $J(U, V)$ 存在平衡状态的充分必要条件是存在 U 和 V , 使 $J(U, V)$ 对 U 的偏导数 $\frac{\partial J(U, V)}{\partial U} = 0$ 和对 V 的偏导数 $\frac{\partial J(U, V)}{\partial V} = 0$ 同时成立, 于是有

$$4RV - RVU^T U - UU^T RV - UV^T R^T U - RVV^T V = 0, \quad (6)$$

$$4R^T U - R^T UVV^T V - VU^T RV - VV^T R^T U - R^T UU^T U = 0. \quad (7)$$

对式(6)等号两边同时左乘 U^T , 对式(7)等号两边同时左乘 V^T , 可得

$$4U^T RV = U^T RVU^T U + U^T UU^T RV + U^T UV^T R^T U + U^T RVV^T V \quad (8)$$

和

$$4V^T R^T U = V^T R^T UVV^T V + V^T VU^T RV + V^T VV^T R^T U + V^T R^T UU^T U. \quad (9)$$

式(8)与式(9)相加, 并化简可得

$$4(U^T RV + V^T R^T U) = (U^T RV + V^T R^T U)(U^T U + V^T V) + (U^T U + V^T V)(U^T RV + V^T R^T U). \quad (10)$$

再由 $U^T RV = (V^T R^T U)^T$, 将式(9)等式两边同时转置, 再与式(8)相减可得

$$U^T UV^T R^T U + U^T RVU^T U = V^T VU^T RV + V^T R^T UVV^T V. \quad (11)$$

考虑到式(10)–(11)对任意 $J(U, V)$ 和 R 都成立, 则设 $U = [u_1 \ u_2 \ \dots \ u_r]$, $V = [v_1 \ v_2 \ \dots \ v_r]$, 代入式(10)–(11)中, 可得方程组

$$4(v_i^T R^T u_j + u_i^T R v_j) = \sum_{p=1}^r (v_i^T R^T u_p + u_i^T R v_p)(u_p^T u_j + v_p^T v_j) + (u_i^T u_p + v_i^T v_p)(v_p^T R^T u_j + u_p^T R v_j), \quad (12)$$

$$\sum_{p=1}^r u_i^T u_p v_p^T R^T u_j + \sum_{p=1}^r u_i^T R v_p u_p^T u_j = \sum_{p=1}^r v_i^T v_p u_p^T R v_j + \sum_{p=1}^r v_i^T R^T u_p v_p^T v_j, \quad (13)$$

其中: $i, j \in \{1, 2, \dots, r\}$, 且标量 $u_i^T R v_j = v_j^T R^T u_i$. 对式(12)分析可得

$$u_i^T u_i + v_i^T v_i = 2, \quad (14)$$

$$u_i^T u_j + v_i^T v_j = 0 \ (i \neq j). \quad (15)$$

对式(13)分析可得

$$u_i^T u_j = v_i^T v_j, \quad (16)$$

$$u_i^T u_i = v_j^T v_j \ (i \neq j). \quad (17)$$

联立式(14)–(17), 可得 U 和 V 均为单位正交矩阵. 再结合式(6)–(7)可得

$$U^T R V = V^T R^T U. \quad (18)$$

已知 L_1 和 R_1 为左右奇异矩阵, $S_1 = L_1 R R_1^T$ 为奇异值矩阵, 则 U 和 V 在 L_1 和 R_1 空间的投影分别为 $Q_U = L_1^T U$ 和 $Q_V = R_1^T V$. 由

$$Q_U^T Q_U = (L_1^T U)^T (L_1^T U) = I_r, \quad (19)$$

$$Q_V^T Q_V = (R_1^T V)^T (R_1^T V) = I_r, \quad (20)$$

可得 Q_U 和 Q_V 为旋转矩阵.

下面进行充分性证明, 已知 U 和 V 均为单位正交矩阵, $U^T R V = V^T R^T U$, 代入式 $\frac{\partial J(U, V)}{\partial U} = 0$ 和 $\frac{\partial J(U, V)}{\partial V} = 0$, 并化简可得

$$\frac{\partial J(U, V)}{\partial U} = 2R R_1 Q_V - 2L_1 S_1 Q_V, \quad (21)$$

$$\frac{\partial J(U, V)}{\partial V} = 2R^T L_1 Q_U - 2R_1 S_1 Q_U. \quad (22)$$

对式(21)和式(22)的等式两边分别同时左乘 L_1^T 和 R_1^T 可得

$$L_1^T \frac{\partial J(U, V)}{\partial U} = 0, \quad (23)$$

$$R_1^T \frac{\partial J(U, V)}{\partial V} = 0. \quad (24)$$

由 $\text{rank}(L_1) = \text{rank}(R_1) = r$, $J(U, V)$ 仅与前 r 个成分相关, 可知式(23)–(24) 的解空间维度均为 0, 因而 $\frac{\partial J(U, V)}{\partial U} = 0$ 和 $\frac{\partial J(U, V)}{\partial V} = 0$ 成立. 证毕.

注 1 本文算法主要适用于实数域的信号, 或者经过转化最终在实数域对信号进行处理, 因而 $J(U, V)$ 规定 Q_U 和 Q_V 的取值空间为 $\{(Q_U, Q_V) | Q_U^T \Lambda Q_V > 0\}$.

由定理 1 可知, $J(U, V)$ 在 $U = L_1 Q_U, V = R_1 Q_V$ 时能够达到平衡状态. 但是使 $J(U, V)$ 达到平衡状态的点有极值点和鞍点. 极值点是信息准则的稳定收敛点, 除掉该点的邻域就是该点的吸引域, 吸引域内的方向导数均指向该点. 鞍点是不稳定收敛点, 除掉该点的邻域只有部分或者没有该点的吸引域, 在该邻域中方向导数的指向是不唯一的. 信息准则在该平衡点处取得极值才能用于算法推导. 为进一步探讨信息准则在平衡点出的性质, 给出如下定理.

定理 2 当且仅当 $U = L_1 Q_U, V = R_1 Q_V, Q_U$ 和 Q_V 为旋转矩阵时, $J(U, V)$ 全局极小, 极小值为 $-2\text{tr}(S_1)$. Q_U 和 Q_V 为非旋转矩阵时的平衡点均为 J

(U, V) 的鞍点.

证 由式(5)容易看出,当 $\text{tr}(Q_U^T Q_U) \rightarrow \infty$ 或者 $\text{tr}(Q_V^T Q_V) \rightarrow \infty$ 时, $J(U, V)$ 也是趋向于无穷大的. 因而 $J(U, V)$ 是一个有下界而无上界的函数. 根据文献[21], 函数的极值可通过计算Hessian矩阵判断.

$$H_U J(U, V) = \frac{\partial}{\partial(\text{vec}(U))^T} \left(\frac{\partial J(U, V)}{\partial(\text{vec}(U))^T} \right)^T, \quad (25)$$

$$H_V J(U, V) = \frac{\partial}{\partial(\text{vec}(V))^T} \left(\frac{\partial J(U, V)}{\partial(\text{vec}(V))^T} \right)^T. \quad (26)$$

由于计算过程复杂, 计算步骤相似, 本文主要说明式(25)的推导过程. 由 $U = L_1 Q_U, V = R_1 Q_V$ 可知

$$\begin{aligned} dJ(U, V) = & \text{tr}(-4d(Q_U^T S_1 Q_V) + d(Q_U^T S_1 Q_V Q_U^T Q_U) + \\ & d(Q_U^T S_1 Q_V Q_V^T Q_V)) = \\ & \text{tr}(-4dQ_U^T S_1 Q_V + dQ_U^T S_1 Q_V Q_U^T dQ_U + \\ & Q_U^T S_1 Q_V dQ_U^T Q_U + Q_U^T S_1 Q_V Q_U^T dQ_U + \\ & dQ_U^T S_1 Q_V Q_V^T dQ_V), \end{aligned} \quad (27)$$

然后, 二阶微分为

$$\begin{aligned} d^2 J(U, V) = & \text{tr}(-4d(Q_U^T S_1 Q_V) + \\ & d(Q_U^T S_1 Q_V Q_U^T Q_U) + d(Q_U^T S_1 Q_V Q_V^T Q_V)) = \\ & \text{tr}(dQ_U^T S_1 Q_V dQ_U^T Q_U + dQ_U^T S_1 Q_V Q_U^T dQ_U + \\ & dQ_U^T S_1 Q_V dQ_U^T Q_U + Q_U^T S_1 Q_V dQ_U^T dQ_U + \\ & dQ_U^T S_1 Q_V Q_U^T dQ_U + Q_U^T S_1 Q_V dQ_U^T dQ_U) = \\ & \text{tr}(2dQ_U^T S_1 Q_V dQ_U^T Q_U + 2dQ_U^T S_1 Q_V Q_U^T dQ_U + \\ & 2Q_U^T S_1 Q_V dQ_U^T dQ_U), \end{aligned} \quad (28)$$

那么, 可得

$$\begin{aligned} \hat{H}_U J(U, V) = & K_{rM}(Q_U^T \otimes S_1 Q_V + Q_V^T S_1 \otimes Q_U) + \\ & I_r \otimes (S_1 Q_V Q_U^T + Q_U Q_V^T S_1) + \\ & (Q_U^T S_1 Q_V + Q_V^T S_1 Q_U) \otimes I_r = \\ & (Q_U^T \otimes I_r + I_r \otimes Q_U) K_{rr} \cdot \\ & (I_r \otimes S_1 + S_1 \otimes I_r) \cdot \\ & (I_r \otimes Q_V + Q_V^T \otimes I_r) + \\ & I_r \otimes (S_1 Q_V Q_U^T + Q_U Q_V^T S_1) + \\ & (Q_U^T S_1 Q_V + Q_V^T S_1 Q_U) \otimes I_r, \end{aligned} \quad (29)$$

其中: \otimes 表示直积运算; K_{rr} 表示一个 $r^2 \times r^2$ 的置换矩阵, $K_{rr} = K_{rr}^T = K_{rr}^{-1}$; 式(29)是通过如下公式计算而来的:

$$d^2 \phi(\mathbf{X}) = \text{tr}(B(d\mathbf{X})^T C d\mathbf{X}) \Leftrightarrow$$

$$H\phi(\mathbf{X}) = \frac{1}{2}(B^T \otimes C + B \otimes C^T), \quad (30)$$

$$d^2 \phi(\mathbf{X}) = \text{tr}(B(d\mathbf{X})C d\mathbf{X}) \Leftrightarrow$$

$$H\phi(\mathbf{X}) = \frac{1}{2}K_{pq}(B^T \otimes C + C^T \otimes B), \quad (31)$$

其中: $\phi(\mathbf{X})$ 为被二阶微分的函数, $\mathbf{X} \in \mathbb{R}^{p \times q}$. 析式(29)可知等式右边的3项式子均为对角线矩阵, 由 Q_U 和 Q_V 为旋转矩阵可知对角线矩阵分别为 $I_r \otimes S_1 + S_1 \otimes I, 2I_r \otimes S_1$ 和 $2S_1 \otimes I_r$. $\hat{H}_U J(U, V)$ 矩阵的对角线元素均大于零, 那么 $\hat{H}_U J(U, V)$ 为正定函数. 由此得证, $U = L_1 Q_U, V = R_1 Q_V$ 就是 $J(U, V)$ 全局极小点. 将 $U = L_1 Q_U, V = R_1 Q_V$ 代入式(5)可得该处的值为 $-2\text{tr}(S_1)$.

下面证明其他平衡点均为鞍点. 判断一个平衡点为鞍点的标准^[14]是在平衡点无限小的邻域内存在一点 (U', V') , 使 $J(U', V') < J(L_1 Q_U, R_1 Q_V)$. 假设存在某种不同于 Q_U 和 Q_V 的旋转矩阵 $Q_{\bar{U}}$ 和 $Q_{\bar{V}}$, 有

$$Q_{\bar{U}}^T S_1 Q_{\bar{V}} = \hat{S}_1, \quad (32)$$

$$Q_{\bar{U}}^T S_1 Q_{\bar{V}} = \bar{S}_1, \quad (33)$$

其中: $\hat{S}_1 = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_r\}, \bar{S}_1 = \text{diag}\{\sigma_{j_1}, \sigma_{j_2}, \dots, \sigma_{j_r}\}$, 并且 $\hat{S}_1 - \bar{S}_1$ 至少存在一个正的对角元素. 则可知

$$\begin{aligned} J(L_1 Q_U, R_1 Q_V) = \text{tr}(-2Q_U^T S_1 Q_V) < \\ \text{tr}(-2Q_{\bar{U}}^T S_1 Q_{\bar{V}}) = J(L_1 Q_{\bar{U}}, R_1 Q_{\bar{V}}). \end{aligned} \quad (34)$$

这表明 $\{(L_1 Q_{\bar{U}}, R_1 Q_{\bar{V}}) \mid Q_{\bar{U}}^T S_1 Q_{\bar{V}} \succ 0, Q_{\bar{U}} \neq Q_U, Q_{\bar{V}} \neq Q_V\}$ 不存在全局极小点. 由于 $Q_{\bar{U}} \neq Q_U$ 和 $Q_{\bar{V}} \neq Q_V$, 必然有 $Q_{\bar{U}}$ 中存在 $i \in (1, r)$ 列 $q_{\bar{U}i}$ 与 Q_U 各列均正交, 或者 Q_U 中存在 $j \in (1, r)$ 列 q_{Uj} 与 $Q_{\bar{U}}$ 各列均正交. 令 $q_{\bar{U}i}$ 中第 j_i 行的元素和 q_{Uj} 中第 j_i 行的元素为对应列的唯一非零元素, 则可以定义一个标准正交矩阵 $B_{\bar{U}} = [q_{U1} \ \dots \ (q_{Uj} + \varepsilon q_{\bar{U}i})/\sqrt{1 + \varepsilon^2} \ \dots \ q_{Ur}]$ 和 $B_{\bar{V}} = [q_{V1} \ \dots \ (q_{Vi} + \varepsilon q_{\bar{V}i})/\sqrt{1 + \varepsilon^2} \ \dots \ q_{Vr}]$, 其中 ε 是一个正的无穷小值. 将 $(L_1 B_{\bar{U}}, R_1 B_{\bar{V}})$ 代入式(33)可得

$$B_{\bar{U}}^T S_1 B_{\bar{V}} = \text{diag}\{\sigma_{j_1}, \dots, \frac{(\sigma_{j_j} + \varepsilon \sigma_{j_i})}{\sqrt{1 + \varepsilon^2}}, \dots, \sigma_{j_r}\}. \quad (35)$$

此时

$$\begin{aligned} B_{\bar{U}}^T S_1 B_{\bar{V}} - \bar{S}_1 = \\ \text{diag}\{0, \dots, 0, \frac{\sigma_{j_j}(1 - \sqrt{1 + \varepsilon^2}) + \varepsilon \sigma_{j_i}}{\sqrt{1 + \varepsilon^2}}, 0, \dots, 0\}. \end{aligned} \quad (36)$$

由于 $\sigma_{j_j}(1 + \varepsilon - \sqrt{1 + \varepsilon^2}) > 0$ 和 $\varepsilon(\sigma_{j_i} - \sigma_{j_j}) > 0$ 同时成立, 则 $B_{\bar{U}}^T S_1 B_{\bar{V}} - \bar{S}_1$ 为非零的正半定矩阵, 也即

$$J(L_1 B_{\bar{U}}, R_1 B_{\bar{V}}) = \text{tr}(-2B_{\bar{U}} S_1 B_{\bar{V}}) <$$

$$\text{tr}(-2Q_U^T \bar{S}_1 Q_V) = J(L_1 Q_U, R_1 Q_V), \quad (37)$$

这说明 $\{(L_1 Q_U, R_1 Q_V) | Q_U^T S_1 Q_V > 0, Q_U \neq Q_U, Q_V \neq Q_V\}$ 是一个鞍点的点集. 证毕.

注2 通过分析定理2可知, 信息准则只有全局极值点而不存在局部极小点, 由此可得, 该信息准则的梯度下降搜索算法总能够全局收敛到期望的主奇异子空间.

下面采用梯度下降法求解单层神经网络算法.

分别对 U 和 V 对信息准则 $J(U, V)$ 求梯度可得

$$\nabla_U J(U, V) = -4RV + RVU^T U + UU^T RV + UV^T R^T U + RVV^T V, \quad (38)$$

$$\nabla_V J(U, V) = -4R^T U + R^T UV^T V + VU^T RV + VV^T R^T U + R^T UU^T U. \quad (39)$$

该梯度函数可以理解为信息准则在 (U, V) 的动力学系统, 按照梯度下降法^[22]的基本思路, 认为此时算法的梯度即为 U 和 V 的更新方向, 即为 $\nabla_U J(U, V) = -\Delta U$ 和 $\nabla_V J(U, V) = -\Delta V$, 于是可知自适应算法如下:

$$\begin{aligned} U(k+1) &= U(k) + \eta[4R(k)V(k) - R(k)V(k)U^T(k)U(k) - U(k)U^T(k)R(k)V(k) - U(k)V^T(k)R^T(k)U(k) - R(k)V(k)V^T(k)V(k)], \end{aligned} \quad (40)$$

$$\begin{aligned} V(k+1) &= V(k) + \eta[4R^T(k)U(k) - R^T(k)U(k)V^T(k)V(k) - V(k)U^T(k)R(k)V(k) - V(k)V^T(k)R^T(k)U(k) - R^T(k)U(k)U^T(k)U(k)], \end{aligned} \quad (41)$$

其中学习因子 $\eta \in (0, 1)$, $U(k)$ 和 $V(k)$ 为 k 时刻的状态向量. 特别注意的是, 当 $r = 1$ 时, 矩阵 U 和 V 就退化为两个状态向量 $\mathbf{u} \in \mathbb{R}^{M \times 1}$ 和 $\mathbf{v} \in \mathbb{R}^{N \times 1}$, 算法可以实现单个主奇异特征向量的提取,

$$\begin{aligned} \mathbf{u}(k+1) &= \mathbf{u}(k) + \eta[4R(k)\mathbf{v}(k) - R(k)\mathbf{v}(k)\mathbf{u}^T(k)\mathbf{u}(k) - \mathbf{u}(k)\mathbf{u}^T(k)R(k)\mathbf{v}(k) - \mathbf{u}(k)\mathbf{v}^T(k)R^T(k)\mathbf{u}(k) - R(k)\mathbf{v}(k)\mathbf{v}^T(k)\mathbf{v}(k)], \end{aligned} \quad (42)$$

$$\begin{aligned} \mathbf{v}(k+1) &= \mathbf{v}(k) + \eta[4R^T(k)\mathbf{u}(k) - R^T(k)\mathbf{u}(k)\mathbf{v}^T(k)\mathbf{v}(k) - \mathbf{v}(k)\mathbf{u}^T(k)R(k)\mathbf{v}(k) - \mathbf{v}(k)\mathbf{v}^T(k)R^T(k)\mathbf{u}(k) - R^T(k)\mathbf{u}(k)\mathbf{u}^T(k)\mathbf{u}(k)]. \end{aligned} \quad (43)$$

这个算法可以适用于一些应用中仅需要提取最大的奇异特征向量的情况.

3 算法的性能分析

本文考察算法的性能主要从收敛性和自稳定性两

个角度分别展开. 收敛性是算法能够完成基本提取任务所必须具备的性能, 收敛性能的优劣对算法质量评价具有决定性意义. 本文主要采用李雅普诺夫方法对收敛性展开分析, 从而得到理论结果. 自稳定性也是算法非常重要的性质. 算法如果仅在特定初始条件以及限定的干扰条件下具备良好的收敛性能, 那么它的应用范围会受到很大的限制. 具备良好的自稳定性的算法相比不具有该性能的算法拥有更高的应用价值. 本文通过ODE方法, 开展对算法自稳定性的分析.

3.1 算法收敛性分析

对式(40)–(41)进行分析可知, 在学习因子足够小的情况下, 该式可以通过一个连续时间的常微分方程来近似^[23]:

$$\begin{aligned} \frac{dU}{dt} &= 4R(t)V(t) - R(t)V(t)U^T(t)U(t) - U(t)U^T(t)R(t)V(t) - U(t)V^T(t)R^T(t)U(t) - R(t)V(t)V^T(t)V(t), \end{aligned} \quad (44)$$

$$\begin{aligned} \frac{dV}{dt} &= 4R^T(t)U(t) - R^T(t)U(t)V^T(t)V(t) - V(t)U^T(t)R(t)V(t) - V(t)V^T(t)R^T(t)U(t) - R^T(t)U(t)U^T(t)U(t). \end{aligned} \quad (45)$$

算法(40)–(41)的收敛性条件等价于式(44)–(45)的全局收敛条件, 因此考虑采用李雅普诺夫理论分析式(44)–(45)的全局收敛性. 显然, 信息准则 $J(U, V)$ 是有下界的, 连续而且还是连续可微的, 满足作为常微分公式李雅普诺夫函数的条件. 具体条件通过定理3分析可得.

定理3 考虑初始状态 (U_0, V_0) , 如果学习因子足够小, 并且状态向量 $(U(t), V(t))$ 满足ODE约束, 那么当时间 $t \rightarrow \infty$ 时, 状态向量以概率1收敛到固定点 $(L_1 Q_U, R_1 Q_V)$.

证 根据式(5)可设计一个能量函数

$$\begin{aligned} E(U, V) &= -\text{tr}(4U^T RV) + \text{tr}(U^T RVU^T U) + \text{tr}(U^T RVV^T V). \end{aligned} \quad (46)$$

研究该式可以发现, 当 $t \rightarrow \infty$ 时, $\|U\|_F$ 和 $\|V\|_F$ 有上界, 此时 $E(U, V)$ 存在上界, 那么对应的自变量 (U, V) 就存在一个有界的定义域. 容易知道, $E(U, V)$ 是连续函数, 而且是一阶微分连续函数. 那么, 对 $E(U, V)$ 求梯度可得

$$\begin{aligned} \nabla_U E(U, V) &= -4R(k)V(k) + R(k)V(k)U^T(k)U(k) + U(k)U^T(k)R(k)V(k) + U(k)V^T(k)R^T(k)U(k) + \end{aligned}$$

$$R(k)V(k)V^T(k)V(k), \tag{47}$$

$$\nabla_V E(U, V) = -4R^T(k)U(k) + R^T(k)U(k)V^T(k)V(k) + V(k)U^T(k)R(k)V(k) + V(k)V^T(k)R^T(k)U(k) + R^T(k)U(k)U^T(k)U(k). \tag{48}$$

由此, 根据微分矩阵的链式规则^[24], 可以求解 $E(U, V)$ 的微分

$$\frac{dE(U, V)}{dt} = \text{tr}\left(\frac{dU^T}{dt} \cdot \nabla_U E(U, V) + \frac{dV^T}{dt} \cdot \nabla_V E(U, V)\right) - \text{tr}\left(\frac{dU^T}{dt} \cdot \frac{dU}{dt} + \frac{dV^T}{dt} \cdot \frac{dV}{dt}\right), \tag{49}$$

这表明 $\frac{dE(U, V)}{dt}$ 是一个非正定的矩阵函数, 即为当 $t \rightarrow \infty$ 的时候, $\frac{dU(t)}{dt} \rightarrow 0$ 且 $\frac{dV(t)}{dt} \rightarrow 0$, 也就是 $\nabla_U E(U, V) \rightarrow 0$ 和 $\nabla_V E(U, V) \rightarrow 0$. 综合上述分析可知, 状态向量 $(U(t), V(t))$ 满足李雅普诺夫收敛性条件, 因此, $(U(t), V(t))$ 全局收敛到 $E(U, V)$ 的稳定点. 结合定理1和定理2的结论可知, $E(U, V)$ 只具有一个全局最小点 (L_1Q_U, R_1Q_V) , 其他平衡点均为鞍点或者不稳定点. 由此可以得出结论, $(U(t), V(t))$ 随着 $t \rightarrow \infty$ 以概率1收敛到 (L_1Q_U, R_1Q_V) .

证毕.

注3 收敛性分析的内容结合了前期对信息准则的分析结论, 这表明李雅普诺夫意义下的能量函数, 在一定程度上同表达信息分布的信息准则在信息论分析的角度具有相关性.

3.2 算法自稳定性分析

自稳定性是指算法迭代过程中, 以任意初始值开始, 状态向量总是能够移动到一个特定的点, 其具体表现是状态向量的模值总是收敛到一个固定值. 有学者^[25]通过研究多个算法的稳定发现, 缺乏自稳定性能的算法是潜在发散的, 因此对自稳定性能的分析十分必要. 根据范数理论状态向量 U 的自稳定性按照 F 范数意义下的稳定性展开分析.

定理4 在学习因子足够小, 输入的向量有界, 并且初始状态模量 $\|U(0)\|_F^2 + \|V(0)\|_F^2$ 存在上界 $2r$ 的情况下, 在算法(40)–(41)中, U 和 V 的 F 范数总能够收敛到常数1.

证 首先求解 $U(k+1)$ 和 $V(k+1)$ 的 F 范数, 根据式(40)–(41)可知

$$U^T(k+1)U(k+1) + V^T(k+1)V(k+1) = U^T(k)U(k) + V^T(k)V(k) + 8\eta[U^T(k)RV(k) + V^T(k)R^T U(k)] -$$

$$2\eta \left\{ \begin{matrix} [U^T(k)RV(k) + V^T(k)R^T U(k)] \\ [U^T(k)U(k) + V^T(k)V(k)] \end{matrix} \right\} - 2\eta \left\{ \begin{matrix} [U^T(k)U(k) + V^T(k)V(k)] \\ [U^T(k)RV(k) + V^T(k)R^T U(k)] \end{matrix} \right\} + o(\eta^2), \tag{50}$$

其中考虑到学习因子足够小, 高阶项 $o(\eta^2)$ 对范数影响在后续的研究中可以被忽略. 由此可知

$$\frac{\|U(k+1)\|_F^2 + \|V(k+1)\|_F^2}{\|U(k)\|_F^2 + \|V(k)\|_F^2} = \frac{\text{tr}(U^T(k+1)U(k+1) + V^T(k+1)V(k+1))}{\text{tr}(U^T(k)U(k) + V^T(k)V(k))} \approx 1 - \frac{4\eta}{\|U(k)\|_F^2 + \|V(k)\|_F^2} \times \text{tr}((U^T(k)RV(k) + V^T(k)R^T U(k)) \times (U^T(k)U(k) + V^T(k)V(k) - 2I_r)). \tag{51}$$

经过式(50)中关于矩阵的迹的一些数学计算和推导, 可得 $\|U(k)\|_F^2 + \|V(k)\|_F^2$ 两步之间模值关系的几种不同的情况:

$$\frac{\|U(k+1)\|_F^2 + \|V(k+1)\|_F^2}{\|U(k)\|_F^2 + \|V(k)\|_F^2} \begin{cases} > 1, & \|U(k)\|_F^2 + \|V(k)\|_F^2 < 2r, \\ = 1, & \|U(k)\|_F^2 + \|V(k)\|_F^2 = 2r, \\ < 1, & \|U(k)\|_F^2 + \|V(k)\|_F^2 > 2r. \end{cases} \tag{52}$$

该式表明, 状态向量的模值比例情况可分为如下3种: 当 $\|U(k)\|_F^2 + \|V(k)\|_F^2 < 2r$ 时, 算法状态向量的模值在本次更新过程中会增大, 而且在 $\|U(k)\|_F^2 + \|V(k)\|_F^2 > 2r$ 时会减小, 最终的趋势是状态向量的模值 $\|U(k)\|_F^2 + \|V(k)\|_F^2$ 能够逐渐稳定在常数 $2r$ 附近. 这表明, 提出的算法的收敛性能不依赖于状态向量之前的数值, 在协方差稳定的情况下总是使得状态向量向着确定的方向收敛. 这表明该算法具有自稳定性. 证毕.

4 仿真实验

仿真实验主要从3个角度展开, 首先从状态矩阵的模值和方向余弦两个方面体现所提出算法的收敛性; 其次验证算法的自稳定性; 最后通过所提出算法在高维波束频率估计的应用, 展示该算法在实际应用算例中的性能. 在算例1和算例2中, 蒙特卡洛仿真次数为40, 实验的结果就是这40次独立运行结果的平均.

4.1 算例1: 算法收敛性验证

随机生成一组五维的高斯白噪声序列 $x(k)$, 并且 $y(k) = Ax(k)$, 其中矩阵 A 通过3个矩阵计算得出:

$$A = [u_0 \ u_1 \ \cdots \ u_4]A_0[v_0 \ u_1 \ \cdots \ u_4], \tag{53}$$

其中: u_i 和 v_i ($i = 0, 1, \dots, 4$)分别为一个七维和五

维的正交离散余弦基函数; A_0 是一个对角矩阵, 从上到下分别为 10, 10, 10, 2, 10^{-2} . 理论上, 通过式(53)计算出来的协方差矩阵, 其奇异值就等于 A_0 矩阵对角线上的各个元素. 设置三重根 10, 并且条件数为 10^4 , 此时协方差矩阵是病态的. 并且所提取的奇异子空间对应信号前 3 个主奇异特征成分所张成的子空间. 另外根据定理约束, 确定学习因子 $\eta = 0.1$, 状态向量的初始值 U_0 和 V_0 都是随机生成的, 其中每一个列向量的模值均为 1. 本文通过对比所提出算法, Kong 算法^[17]和 Kaiser 算法^[20]体现出所提出算法在收敛速度、收敛精度等方面的优良性能.

为清晰地描述算法的收敛速度和精度, 分别从方向和长度两个方面定义状态矩阵的方向余弦和模值长度. 一般而言, 状态矩阵最终收敛到的子空间总是与前 3 个主奇异特征向量张成的子空间存在一个旋转矩阵的相位差, 因此定义 $U_{lss} = [\mathbf{u}_{1lss} \ \mathbf{u}_{2lss} \ \mathbf{u}_{3lss}]$ 在 k 时刻的方向余弦为

$$DC_l(i, k) = \sqrt{\sum_{j=1}^r \left(\frac{\|\mathbf{l}_j^T \mathbf{u}_{ilss}(k)\|}{\|\mathbf{l}_j\| \|\mathbf{u}_{ilss}(k)\|} \right)^2}, \quad (54)$$

$$DC_r(i, k) = \sqrt{\sum_{j=1}^r \left(\frac{\|\mathbf{r}_j^T \mathbf{v}_{ilss}(k)\|}{\|\mathbf{r}_j\| \|\mathbf{v}_{ilss}(k)\|} \right)^2}. \quad (55)$$

其中 \mathbf{l}_j 和 \mathbf{r}_j 分别为协方差矩阵的真实的左右奇异成分. 这样无需预先计算旋转关系矩阵, 就可以验证两个子空间等价. 另外, 子空间的模值长度通过状态向量的 2 范数估计

$$\text{Norm}_l(i, k) = \|\mathbf{u}_{ilss}(k)\|, \quad (56)$$

$$\text{Norm}_r(i, k) = \|\mathbf{v}_{ilss}(k)\|. \quad (57)$$

为了更清晰的展现子空间的空间特性, 考察子空间正交性, 定义向量的正交积为

$$\text{Orth}_l(i, k) = \mathbf{u}_{ilss}^T(k) \mathbf{u}_{jlss}(k), \quad i \neq j, \quad (58)$$

$$\text{Orth}_r(i, k) = \mathbf{v}_{irss}^T(k) \mathbf{v}_{jrss}(k), \quad i \neq j. \quad (59)$$

去掉重复的计算, 分别计第 1 个值为第一、二向量的正交积, 第 2 个值为第二、三向量的正交积, 第 3 个值为第一、三向量的正交积.

如图 1 所示, 在左空间中, 本文所提出算法的方向余弦变化情况用实线表示, Kong 算法的余弦变化情况用虚线表示, Kaiser 算法的余弦变化情况用段线表示, 与前 3 个最大奇异特征值相对应的 3 个特征向量分别用红绿蓝 3 种颜色标示. 容易看出, 3 种算法的 3 条线都是逐渐收敛到 1 的位置, 这表明算法能够在方向上跟踪得到输入信号的左奇异特征子空间. 3 种算法中, 本文提出的算法明显需要最少的迭代步数, 并且算法到达 1 以后, 存在最小的波动变化. 右空间的分析与此相似. 这验证了理论分析的结果.

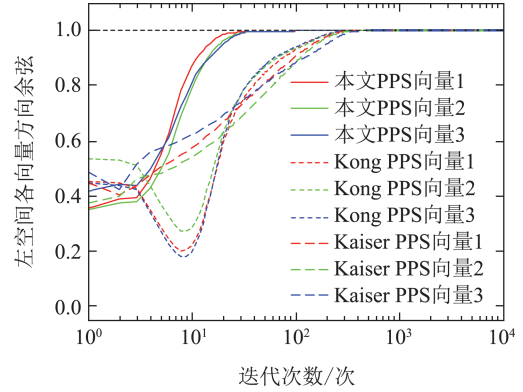


图 1(a) 左奇异子空间的方向余弦收敛过程

Fig. 1(a) Direction cosine convergence of the left principal singular subspace

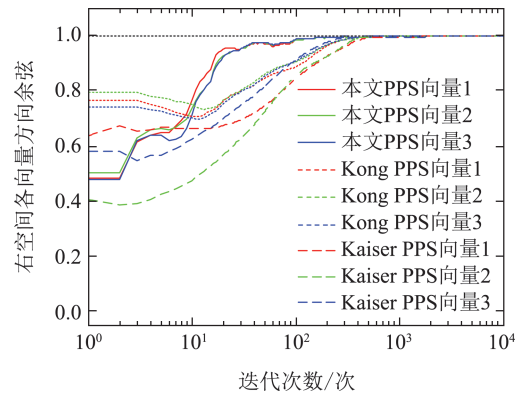


图 1(b) 右奇异子空间的方向余弦收敛过程

Fig. 1(b) Direction cosine convergence of the right principal singular subspace

图 2 展示的是各个算法模值变化的情况, 各种线条对应的算法以及向量与图 1 相同, 以左空间为例, 本文所提出算法的模值变化情况用实线表示, Kong 算法的模值变化情况用虚线表示, Kaiser 算法的余弦变化情况用段线表示, 3 个向量分别用红绿蓝 3 种颜色标示. 图中显示, 3 种算法不同向量的曲线都是逐渐收敛到一个固定的实数值的位置, 其中, 本文提出算法和 Kaiser 算法都是收敛到 1 的, Kong 算法的曲线则收敛到 1.44 附近. 显然, 本文算法的各个曲线都是以最少的迭代次数收敛到稳定状态的, 这体现了本文提出算法在收敛速度方面的优越性. 另外, 本文提出算法的曲线在收敛到 1 以后, 几乎看不到任何波动的迹象, 而 Kong 算法和 Kaiser 算法在收敛到稳定状态的初期存在一定的波动. 这又说明了本文提出算法在收敛精度方面的优势. 对右空间的分析也能够得到相似的结论.

图 3 中, 以左空间为例, 本文所提出算法的正交积的变化情况用实线表示, Kong 算法的正交积的变化情况用虚线表示, Kaiser 算法的正交积变化情况用段线表示, 3 个向量分别用红绿蓝 3 种颜色标示. 3 种算法的各条曲线均不为零则表明 3 条向量存在不相互垂直的部分, 经过短期震荡后, 3 条线均收敛到零值, 此时各

个向量完全正交,并一直保持到最后.右空间的分析与此相似.但是需要指出的是,图3(b)中由于Kaiser算法的曲线变化幅值较大,故将图中 $-0.1 \sim 0.4$ 的部分放大,用来显示本文算法和Kong算法的曲线变化情况.

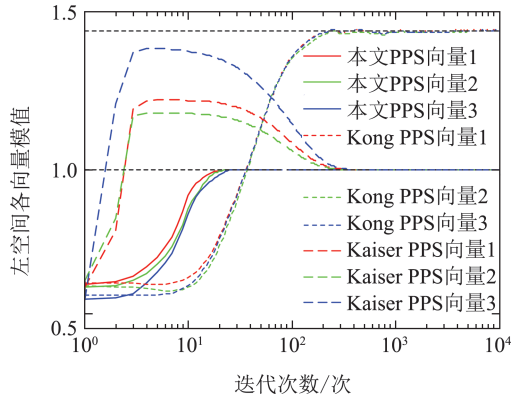


图 2(a) 左奇异子空间的模值长度收敛过程
Fig. 2(a) Norm convergence of the left principal singular subspace

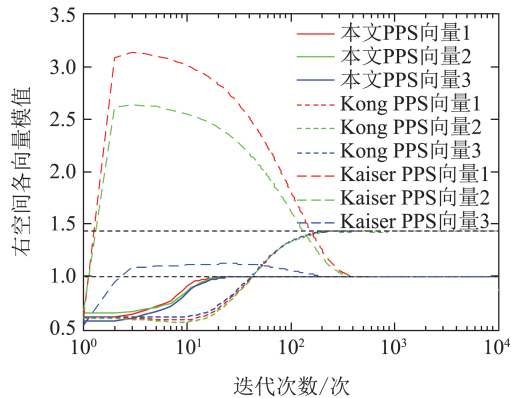


图 2(b) 右奇异子空间的模值长度收敛过程
Fig. 2(b) Norm convergence of the right principal singular subspace

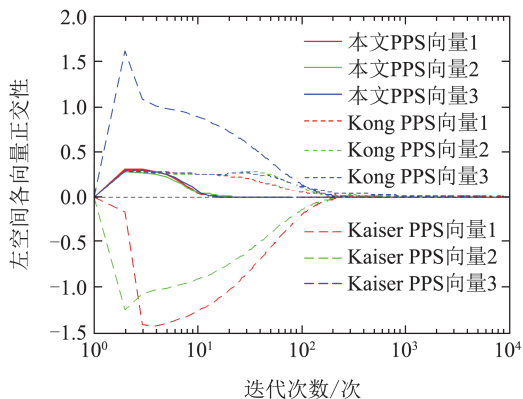


图 3(a) 左奇异子空间的各向量正交积曲
Fig. 3(a) Orthogonality of each vector in the left principal singular subspace

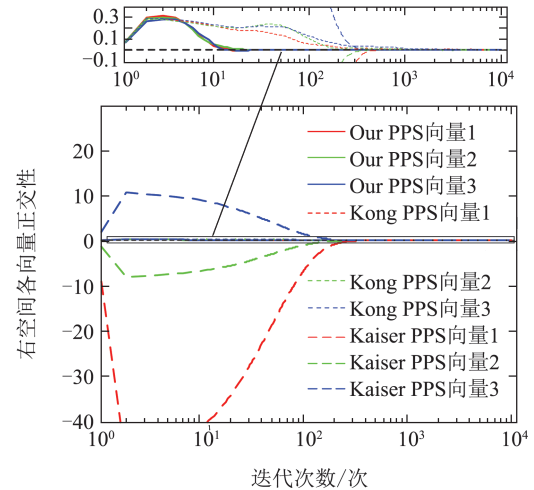


图 3(b) 右奇异子空间的各向量正交积曲线
Fig. 3(b) Orthogonality of each vector in the right principal singular subspace

另外,本文还比较了算法的计算复杂性.本文提出算法的计算复杂度每次更新分别需 $MNr + 5Mr^2 + Nr^2$ 和 $MNr + 5Nr^2 + Mr^2$ 个浮点计算.Kong算法的计算复杂度为每次更新分别需 $MNr + 4Mr^2 + Nr + Mr$ 和 $MNr + 4Nr^2 + Nr + Mr$ 个浮点运算.Kaiser算法的计算复杂度每次更新均需要 $2MNr + Mr^2 + Nr^2$ 个浮点运算.

4.2 算例2: 算法自稳定性

与第4.1节算例1相似,本算例的输入序列 $\mathbf{x}(k)$ 和 $\mathbf{y}(k)$ 均为高斯白噪声序列,并同样将他们的关系设置为 $\mathbf{y}(k) = A\mathbf{x}(k)$,并且 A 值保持不变.不同之处是本算例研究目的着重在初始状态矩阵的模值,算法对收敛性的保持状态.因此,设置3种不同情况展开验证.具体而言,首先随机生成一组向量作为状态向量的初始值 $U(0)$ 和 $V(0)$,然后将该向量的范数规范为2.2, 1.7和1.0,分别对应大于、等于以及小于 $\sqrt{3}$ 这三种情况,最后分别验证这样的设置下算法状态矩阵的 F 范数变化情况.其中,状态矩阵的列数 $r = 3$,根据定理4可知, F 范数最终稳定在 $\sqrt{r} = \sqrt{3}$ 处.另外,根据定理约束,确定学习因子 $\eta = 0.1$.

算法自稳定性的仿真结果如图4所示.图4(a)和4(b)分别表示左和右奇异子空间的变化情况,3种初始条件的变化曲线分别用实线、虚线和点线表示.在左空间中,不同模值的初始状态使3条曲线起始于完全不同的3个位置,但是在无论哪一种初始条件情况下,算法总是最终稳定在 $\sqrt{3}$ 这一同一条曲线上.

由此,可以说明该算法具有良好的自稳定性.另外,状态空间收敛到稳定状态的迭代次数也是十分接近,这一方面呼应了算例1中的算法迭代次数,另一方面也体现出了算法收敛的行为具有相似性.右空间的变化情况与左空间类似,在左空间中相应的结论也适

用于右空间.

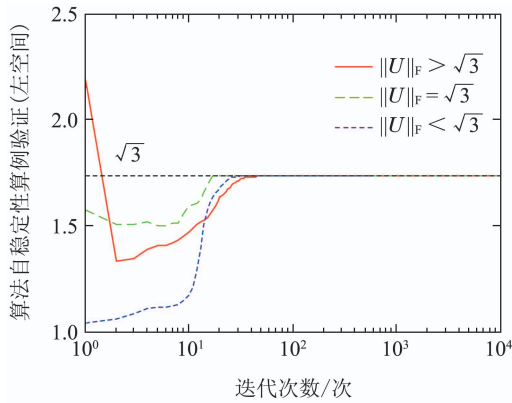


图 4(a) 算法自稳定性数值验证(左空间)

Fig. 4(a) Self-stabilizing property verification in the left principal singular subspace

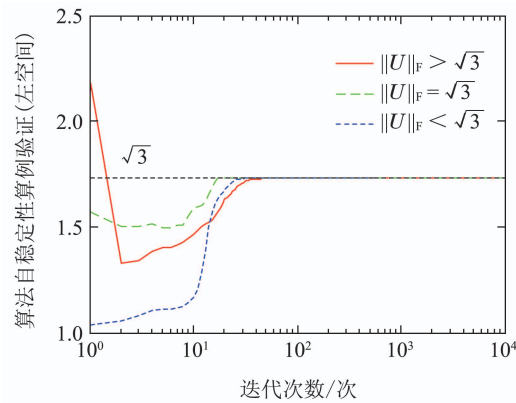


图 4(b) 算法自稳定性数值验证(右空间)

Fig. 4(b) Self-stabilizing property verification in the right principal singular subspace

4.3 算例3: 本文算法在高维波束频率分析中的应用

本算例是验证提出方法在高维波束跟踪问题中的应用效果. 设置高维波束存在3个阶段的变化, 阶段之间的变化是突然的, 每个阶段都存在3个频率, 具体的波束函数如下设置.

当 $1 \leq k < 170$ 时, 有

$$x(k) = \exp(0.1\pi k + 0.513) + \tau n(k) + \exp(0.2\pi k + 0.714) + \exp(0.7\pi k + 0.146); \quad (60)$$

当 $171 \leq k < 340$ 时, 有

$$x(k) = \exp(0.2\pi k + 0.513) + \tau n(k) + \exp(0.3\pi k + 0.714) + \exp(0.7\pi k + 0.146); \quad (61)$$

当 $341 \leq k < 500$ 时, 有

$$x(k) = \exp(0.3\pi k + 0.513) + \tau n(k) + \exp(0.5\pi k + 0.714) + \exp(0.7\pi k + 0.146); \quad (62)$$

其中: $n(k) \in [-0.5, 0.5]$ 表示白噪声, τ 是噪声信号的系数. 有兴趣的读者可通过文献[1]了解关于该波函数的详细知识. 高维波束频率分析问题可以通过提取两个数据流的主奇异成分获得. 这两个数据流可以构造为

$$\mathbf{x}(k) = [x(k) \ x(k+1) \ \cdots \ x(k+M-1)]^T, \quad k = 0, 1, \dots, K-M, \quad (63)$$

$$\mathbf{y}(k) = [x(k) \ x(k+1) \ \cdots \ x(k+N-1)]^T, \quad k = 0, 1, \dots, K-N, \quad (64)$$

其中: K 为样本总数, M 为 \mathbf{x} 数据流的维度数, N 为 \mathbf{y} 数据流的维度数.

图5展示了算法在频率的跟踪速度和估计精度两个方面的性能. 图中蓝色虚曲线表示通过奇异值分解, 红色实曲线表示本文算法跟踪的子空间方法对频率的估计. 从算法曲线在不同阶段间的表现来看, 算法能够在150次迭代中成功跟踪发生激烈变化的频率, 体现了跟踪算法的快速性. 算法在每后半阶段中, 震荡幅度逐渐减小, 最终与奇异值分解方法的曲线高度重合, 这体现了算法具有较高的估计精度. 总之, 本文提出算法在高维波束频率估计问题中有良好的效果.

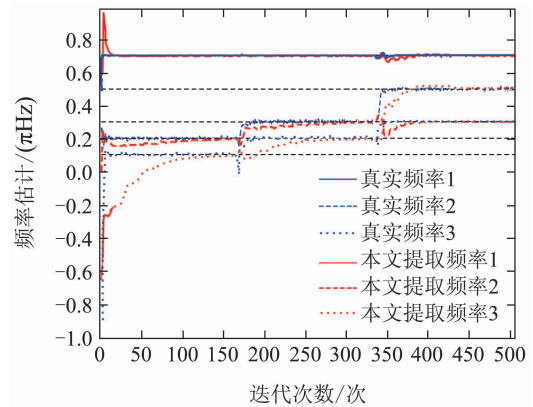


图 5 本文算法在频率估计中的变化曲线

Fig. 5 Evolution curves of the proposed algorithm for estimating the frequencies

5 结论

本文提出了一种新型的主奇异子空间跟踪算法. 首先提出一种新的信息准则, 理论分析了信息准则极值点的存在性和唯一性. 然后基于提出的信息准则, 通过梯度下降法推导出一种神经网络搜索算法, 实现了对输入信号主奇异子空间的跟踪. 算法的收敛性和自稳定性分别通过李雅普诺夫函数方法和ODE方法进行了分析. 最后, 不同形式的MATLAB仿真验证实验表明, 该算法具有较快的收敛速度, 较高的收敛精度以及良好的自稳定性能. 因此, 提出算法具有较高的实际应用价值.

参考文献:

- [1] FENG D Z, ZHANG X D, BAO Z. A neural network learning for adaptively extracting cross-correlation features between two high-dimensional data streams. *IEEE Transaction on Neural Networks*, 2004, 15(6): 1541 – 1554.
- [2] OH T, MATSUSHITA Y, TAI Y, et al. Fast randomized singular value thresholding for low-rank optimization. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2017, 40(2): 376 – 391.
- [3] WAN W, ZHOU Z, ZHAO J, et al. A novel face recognition method: using random weight networks and quasi-singular value decomposition. *Neurocomputing*, 2015, 151: 1180 – 1186.
- [4] ZHOU Xiping, SUN Degang, WANG Zhu, et al. A method based on singular value decomposition for enhancement of differential power analysis. *Acta Electronica Sinica*, 2017, 45(9): 2250 – 2255. (周新平, 孙德刚, 王竹, 等. 一种基于奇异值分解的功耗轨迹筛选方法. *电子学报*, 2017, 45(9): 2250 – 2255.)
- [5] PANG B, TANG G, TIAN T, et al. Enhanced singular spectrum decomposition and its application to rolling bearing fault diagnosis. *IEEE Access*, 2019, 7: 87769 – 87782.
- [6] XU K, NIE W, FENG D, et al. A multi-direction virtual array transformation algorithm for 2D DOA estimation. *Signal Processing*, 2016, 125: 122 – 133.
- [7] BUNCH J R, NIELSEN C P. Updating the singular value decomposition. *Numerical Mathematics*, 1978, 31(2): 111 – 129.
- [8] MOONEN M, Van DOOREN P, VANDEWALLE J. A singular value decomposition updating algorithm for subspace tracking. *Siam Journal on Matrix Analysis and Applications*, 1992, 13(4): 1015 – 1038.
- [9] CHATTERJEE C, KANG Z, ROYCHOWDHURY V P. Algorithms for accelerated convergence of adaptive PCA. *IEEE Transaction on Neural Networks*, 2000, 11(2): 338 – 355.
- [10] YUILLE A, KAMMEN D, COHEN D. Quadrature and the development of orientation selective cortical cells by Hebb rules. *Biological Cybernetics*, 1989, 61(3): 183 – 194.
- [11] GOLUB G H, VAN LOAN C F. *Matrix Computations (3rd edition)*. Baltimore: John Hopkins University Press, 1996.
- [12] SAMARDZIJA N, WATERLAND R L. A neural network for computing eigenvectors and eigenvalues. *Biological Cybernetics*, 1991, 65(4): 211 – 214.
- [13] DIAMANTARAS K I, KUNG S Y. Cross-correlation neural network models. *IEEE Transactions on Signal Processing*, 1994, 42(11): 3218 – 3223.
- [14] FENG D Z, BAO Z, ZHANG X D. A cross-associative neural network for SVD of non-squared data matrix in signal processing. *IEEE Transaction on Neural Networks*, 2001, 12(5): 1215 – 1221.
- [15] MÖLLER R, KÖNIES. Coupled principal component analysis. *IEEE Transaction on Neural Networks*, 1995, 6(5): 1255 – 1263.
- [16] FENG D Z, ZHANG X D, BAO Z. A neural network learning for adaptively extracting cross-correlation features between two high-dimensional data streams. *IEEE Transaction on Neural Networks*, 2004, 15(6): 1541 – 1554.
- [17] KONG X Y, MA H G, AN Q S, et al. An effective neural learning algorithm for extracting cross-correlation feature between two high-dimensional data streams. *Neural Processing Letters*, 2015, 42(2): 459 – 477.
- [18] FENG X W, KONG X Y, XU D H, QIN J Q. A fast and effective principal singular subspace tracking algorithm. *Neural Processing Letters*, 2015, 42(2): 459 – 477.
- [19] HAYKIN S. *Adaptive Filter Theory*. 2nd edition. Englewood Cliffs, NJ: Prentice-Hall, 1991.
- [20] KAISER A, SCHECK W.. Coupled singular value decomposition of a cross-covariance matrix. *International Journal of Neural Systems*, 2010, 20(4): 293 – 318.
- [21] MIAO Y F, HUA Y B. Fast subspace tracking and neural network learning by a novel information criterion. *IEEE Transaction on Signal Processing*, 1998, 46(7): 1967 – 1979.
- [22] KAARE B P, MICHEAL S P. *The matrix cookbook*. Denmark, Finland: Technical University of Denmark, 2012.
- [23] LJUNG L. Analysis of recursive stochastic algorithms. *IEEE Transactions on Automatic Control*, 1977, 22(4): 551 – 575.
- [24] MAGNUS J R, NEUDECKER H. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York, USA: Wiley, 1991.
- [25] CIRRINCIONE G, CIRRINCIONE M, HÉRAULT J, et al. The MCA EXIN neuron for the minor component analysis. *IEEE Transaction on Neural Networks*, 2002, 13(1): 160 – 187.

作者简介:

杜柏阳 博士研究生, 主要研究方向为信号特征提取、故障检测与诊断, E-mail: duboyangepegc@163.com;

孔祥玉 教授, 博士生导师, 主要研究方向为复杂系统建模、大型工业过程状态监测与故障诊断, E-mail: xiangyukong01@163.com;

冯晓伟 讲师, 主要研究方向为大型工业过程状态监测与故障诊断、信号特征提取, E-mail: xiaowei121416@163.com.