孔祥玉　胡昌华　韩崇昭　著

系统特征信息提取神经网络与算法

科学出版社

系统特征信息提取神经网络与算法

孔祥玉　胡昌华　韩崇昭　著

科学出版社

北　京

内 容 简 介

本书主要讨论了矩阵理论相关知识、特征值与奇异值分析、主成分分析及神经网络分析方法、次成分分析及神经网络分析方法、子空间跟踪及神经网络分析方法、总体最小二乘方法、特征提取方法应用等。全书内容新颖，不但包含信息特征提取与优化的若干方法，而且对这些迭代方法的神经网络算法的性能分析方法也进行了较为详细的分析，反映了国内外信息处理和神经网络领域在该方向上研究和应用的最新进展。

本书适合作为电子、通信、自动控制、计算机、系统工程、模式识别和信号处理等信息科学与技术学科高年级本科生和研究生教材，也可供相关专业研究人员和工程技术人员参考。

图书在版编目(CIP)数据

系统特征信息提取神经网络与算法/孔祥玉，胡昌华，韩崇昭著. —北京：科学出版社，2012

ISBN 978-7-03-033141-0

Ⅰ. 系… Ⅱ. ①孔… ②胡… ③韩 Ⅲ. 人工神经网络 Ⅳ. TP183

中国版本图书馆 CIP 数据核字(2011)第 273774 号

责任编辑：魏英杰 杨向萍 / 责任校对：刘小梅
责任印制：张 倩 / 封面设计：陈 敬

科学出版社出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com
骏杰印刷厂印刷
科学出版社发行 各地新华书店经销
*
2012 年 5 月第 一 版 开本：B5 (720×1000)
2012 年 5 月第一次印刷 印张：17 1/2
字数：339 000

定价：60.00 元

(如有印装质量问题，我社负责调换)

前　言

随着人们认识世界的深入，改造世界的拓展，出现了一些大型复杂系统，这些系统规模大、造价高，一旦出现故障，其后果往往是灾难性的，因此要求这种系统具有极高的安全性和可靠性。对系统进行安全性和可靠性分析的前提是快速准确地掌握系统的大量信息，尤其是反映系统本质属性的特征信息。哪些信息是系统的特征信息？如何描述这些特征信息？如何及时发现，并从大量复杂的数据信息中提取有用的特征信息？如何应用这些特征信息？这些问题都是现代信息科学与技术各学科，如电子、通信、自动控制、计算机、系统工程、模式识别、信号处理等面临的共性问题。

我们一直从事信号处理、自动控制、神经网络和模式识别的教学、科研和工程应用等工作，深刻感受到特征信息提取和神经网络方法在信息科学与技术各学科科学研究中所起的重要作用，感到有必要对过去的工作加以总结，为相关的学习研究提供一些参考，因此总结了十余年来在国内外权威和著名杂志上发表的学术研究成果，写成此书。我们相信，本书的出版对我国信息科学与技术领域特征信息提取和神经网络方法的研究将起到积极的推动作用。

全书共八章，其主要内容可概括如下：

第一章绪论，概述了复杂系统特征信息提取领域的国内外研究现状和发展趋势，以及本书主要概况和章节安排。第二章阐述了矩阵理论中特征分析和奇异值分析理论。第三章阐述了主成分分析神经网络与算法，分类介绍了现有的主要主成分分析方法。第四章研究了次成分分析神经网络及算法性能分析方法，重点对我们提出的一种次成分分析神经网络算法及算法收敛性能进行了系统分析。第五章研究了特征信息提取神经网络算法的性能分析方法，对近年来发展起来的确定性离散时间系统方法进行了系统研究，重点对一种次成分神经网络方法的确定性离散时间系统进行了深入分析。第六章研究了双目的主/次子空间跟踪神经网络方法，力图在一个算法中通过简单的符号变换实现主成分分析或次成分分析或者主/次子空间跟踪，重点阐述了提出的一种双目的主/次子空间跟踪算法及性能分析方法。第七章研究了输入和输出数据中均含有噪声情况下实现最优估计的总体最小二乘方法，该方法是特征信息次成分分析的重要应用，重点研究了一种新的总体最小二乘迭代分析方法及其性能分析。第八章简要地分析了特征信息提取神经网络方法在信号处理、系统辨识、自适应滤波和频谱分析等方面的应用。

本书作者之一孔祥玉同志 1997 年师从第二炮兵工程大学黄先祥院士、郭晓松

教授，攻读机电专业硕士学位；2001 年师从西安交通大学韩崇昭教授攻读自动控制专业博士学位；在第二炮兵工程大学胡昌华教授指导下进行了博士后研究，为本书的写作打下了坚实的理论基础。在本书出版之际，谨向他们表示衷心的感谢！在本书的写作过程中，参考了大量的国内外有关矩阵理论、特征信息提取、总体最小二乘等方向的论文和著作，特别参考了冯大政教授、张贤达教授、陈天平教授、张毅教授、徐雷教授、欧阳膳教授等发表在 *IEEE Trans. Signal Processing*，*IEEE Trans. Neural Network* 等杂志上的相关研究论文，也参考了国际上 Oja、Cirrincione、Zufiria、Davila、Markovsky、Huffel、Moller 等学者的文章。为了知识的系统性和完整性，在部分章节中也将上述多位学者的研究成果编入书中，在此向几位学者表示衷心的感谢！

在本书的出版过程中，得到了第二炮兵工程大学马红光教授、西安交通大学曹建福教授、第二炮兵工程大学三系邵军勇主任、第二炮兵工程大学政治部王虎副处长、山西师范大学安秋生教授、铁道部科学研究院吕晓军博士等热情的推荐和帮助，在此深表感谢！

衷心感谢国家自然科学基金面上项目(61074072)、国家杰出青年基金项目(61025014)、国家自然科学基金面上项目(61174207)、国家自然科学基金重点项目(60736026)、中国博士后基金特别资助项目(200801480)等课题的支持，及第二炮兵工程大学导航制导与控制国家重点学科专项建设经费的支持。感谢科学出版社魏英杰等同志给予的支持和帮助！

由于水平有限，书中难免存在不妥之处，敬请广大读者批评指正。

作　者

2011 年 7 月于西安

目录

第一章　绪　　论

1.1　特征信息提取

复杂系统输出信号中包含着丰富的反映系统本质属性的特征信息，如何描述并提取这些特征信息，进一步如何应用这些特征信息？这些问题引起了信息领域学者的广泛关注，新技术、新算法不断出现，形成了信号处理、数据分析和神经网络领域非常活跃的研究热点。经过这些年的发展，特征信息提取技术形成了主成分分析、次成分分析、子空间跟踪和独立成分分析等不同的研究方向，这几个研究方向之间既有联系，又有相对的独立性。

1.1.1　主/次成分分析与子空间跟踪

主成分(principal component，PC)是指信号有最大偏差的方向，只利用数据向量的 K 个主分量进行的数据或者信号分析称为主成分分析(principal component analysis，PCA)。次成分(minor component，MC)是指信号有最小偏差的方向，基于次成分的信号分析、系统分析或者模式分析则统称为次成分分析(minor component analysis，MCA)。主成分分析在数据或图像压缩、多重信号分类、波达方向估计、通信技术等领域得到广泛应用，而次成分分析也已经应用在总体最小二乘(total least square，TLS)、运动目标识别、曲线与曲面匹配、数字成形束、频域估计和故障诊断等领域。通常主/次成分分析都是单维的，而实际中主成分或次成分以多维为主。与数据向量的自相关矩阵 r 个最小特征值对应的特征向量被称为次成分。与数据向量的自相关矩阵的 r 个最大特征值对应的特征向量被称为主成分，这里 r 是主成分或次成分的个数。在一些实际应用中，有时并非要得到多个主成分或者次成分，而只要求跟踪由特征成分张成的子空间。这里将主成分张成的子空间称为主子空间(principal subspace，PS)，而将由次成分张成的子空间称为次子空间(minor subspace，MS)。一个对称矩阵的主成分和次成分分析器可以分别收敛到主成分和次成分。类似地，一个对称矩阵的主子空间和次子空间分析器可以分别收敛到一个主子空间和次子空间。主子空间是由一个高维向量序列的自相关矩阵的主特征值相关的所有特征向量张成的一个子空间，而与该高维向量序列的自相关矩阵的次特征值相关的所有特征向量所张成而形成的子空间被称为次子空间。主子空间有时也称为信号子空间，而次子空间也称为噪声子空间。主子空间

分析(principal subspace analysis,PSA)为许多信息处理领域,如特征提取和数据压缩等提供了有效的方法。在许多实时信号处理领域,如自适应方向波达方向估计、自适应信号处理中的总体最小二乘的自适应解、曲线与曲面匹配等应用中,次子空间分析(minor subspace analysis,MSA)是一个主要的需求。

通过数学分析,可以得出结论:所谓数据的主成分就是数据向量自相关矩阵的最大特征值所对应的特征向量,而数据的次成分是数据向量自相关矩阵的最小特征值所对应的特征向量。这样通过数学上相关矩阵特征值处理或数据矩阵奇异值处理可以得到主成分或次成分。相关矩阵特征值或数据矩阵奇异值处理的方法是基于数据的集中处理,本质上是一种批处理算法,无法实时应用,而且对于维数大的数据来说,其计算复杂度是相当大的,也容易出现数据不稳定的情况。这样寻求可以实时处理、迭代运算、数值稳定、算法简单的主成分分析、次成分分析方法或者子空间跟踪是近 20 年来国际上自动控制、信号处理和神经网络领域的一个研究热点,受到广泛关注。

1.1.2 主/次成分神经网络分析算法

为了实现特征信息的在线迭代与自适应提取,大量的迭代及自适应算法被提出,主/次成分分析或者主/次子空间跟踪迭代求取算法包括逆迭代、常规的和逆 Chebyshev 迭代、Rayleigh 商迭代、神经网络等方法。其中神经网络方法是一种有效的迭代求取算法,尤其是求取主/次成分的单层神经网络及其 Hebbian 型算法由于其算法的简单性和有效性受到人们的高度重视,得到迅速发展,成为自适应主/次成分分析的主流算法。在该类神经网络算法的研究中,新算法不断出现,算法的性能也得到深入透彻的研究。芬兰学者 Oja[1] 和华裔学者 Xu[2] 等是该领域的开创者,他们的工作为该领域的发展奠定了良好的基础。

早在 1979 年,Thompson[3] 就提出了估计与样本协方差矩阵最小特征值对应的特征向量的最小均方(LMS)型自适应方法,并结合 Pisarenko 谱估计子提供了角度/频率的自适应跟踪算法。其后,许多学者开展了特征向量及其子空间跟踪算法的研究[4],更多的是跟踪信号子空间算法的研究,有的同时更新特征值和特征向量,有的是采用矩阵理论计算如经典的特征值分解/奇异值分解的批处理方法,有的采用优化理论来解决子空间跟踪问题[5]。

在主子空间神经网络跟踪研究领域,基于启发式推理的算法,如 Oja 算法[6]、对称误差修正算法[7] 和对称后向传播算法[8] 等相继被提出。分析表明,这几个算法本质上是相同的,被通称为 Oja 算法。后来,最小均方误差重构算法(LMSER)被提出[9],在该算法中著名的梯度搜索概念用来最小化一个均方误差。不像 Oja 算法,该算法是全局收敛的,因为只有主子空间实现均方误差的全局最小而均方误差的其他平稳点都是鞍点。基于该均方误差,投影近似子空间跟踪算法

(PAST)[10]、共轭梯度算法[11]、高斯牛顿方法[12]等算法被提出。近来，一个新颖的信息准则(NIC)被提出，基于该准则，一种新颖的梯度算法和递归类主子空间跟踪算法被提出[13]。后来，基于 NIC 准则和加权矩阵，一个快速提取多个主成分的梯度算法和递归类算法被提出[14]。

20 世纪 90 年代以来，基于反馈神经网络模型进行次子空间跟踪受到高度关注，相继有多个次子空间神经网络跟踪算法被提出来[15~22]。使用膨胀方法，Luo 等[15]提出了一个次子空间分析算法，该算法在运行过程中不需要任何标准化操作；Douglas 等[16]提出了一个自稳定的次子空间分析算法，该算法不需要周期性的标准化操作，也没有矩阵的逆运算；Chiang 等[17]显示出一个学习算法采用合适的初始化而不是膨胀方法，就可以并行抽取多个次成分。基于一个信息准则，Ouyang 等[4]提出了一个自适应次成分跟踪器，该算法可以自动发现次子空间而不需要采用膨胀方法。近年来，Feng 等[19]提出了一个 OJAm 算法，将该算法推广到用来跟踪多个次成分或次子空间，使相应的状态矩阵收敛到次子空间的列正交基。最近，性能更为优良的次成分及子空间跟踪算法[23~25]被提出来，该领域新算法仍然在不断发展中。

1.1.3 该领域目前的研究热点

1. 神经网络主/次成分分析算法的收敛性与稳定性分析

对迭代或神经网络算法进行收敛性和稳定性分析是主/次成分分析领域十多年来的研究热点。算法收敛性与稳定性的直接研究和分析是一个非常难的课题，传统上这些算法的收敛性是通过某种确定性的连续时间系统(deterministic continuous-time，DCT)来间接分析的。由随机系统描述的特征提取神经网络算法可以由相应的确定性连续时间系统来表示，这种表示需要许多假设性条件，其中之一是要求学习因子收敛到零，这在很多实际应用是一个强加的不合理的要求。通过 DCT 系统证明已经收敛的算法，是否存在发散或不稳定的可能？2002 年意大利学者 Cirrincione 等对一些次成分分析神经网络学习算法进行了研究[20]，首次根据黎曼度量来分类次成分分析线性核，并通过误差损失退化的分析证明了在接近收敛的时候算法的不同行为。同时，对算法进行了直接的随机离散时间系统(stochastic discrete-time，SDT)分析，发现了突然发散、动态发散和数值发散，这一发现推动了该领域的研究。然而，DCT 和 SDT 虽然可以分析得出算法是否收敛与稳定，却不能求出具体收敛与稳定的充分条件或边界条件。西班牙学者 Zufiria[26]提出一种确定性的离散时间系统(deterministic discrete-time，DDT)来间接解释由随机离散时间系统描述的神经网络学习算法的动力学系统，DDT 刻画的是核节点的平均进化行为，保持了原网络的离散时间形式，要求学习因子保持常

数，得到的是该类学习算法的更真实的动态行为。在此基础上，近年来我国学者 Yi 等[27]研究团队对 DDT 方法进行了深入研究和推广，研究了几乎所有现有的主/次成分分析神经网络学习算法，推导了一系列算法各自收敛和稳定的成分条件及边界条件，大大推进了次成分分析神经网络学习算法性能的研究，形成了从 2005～2009 年国际上神经网络领域的一个学术研究热点。

2. 神经网络主/次成分分析自稳定算法

一个神经网络主/次成分分析方法以及主/次子空间跟踪算法如果在算法迭代更新过程中，神经网络的权向量或权矩阵的模值随着时间的进行发散到无穷，则不利于算法的实际应用。解决的途径有两个，一是在迭代更新过程中每步或者定期将权向量或权矩阵的模值实行规范化处理，使其模值长度等于 1；另一种探索权向量模值自稳定的算法[16,19,21]，无论初始权向量模值大小，使算法在更新过程中权向量或权矩阵的模值自动收敛于某一固定值或者为 1。为了克服神经网络算法在迭代过程中权向量模值发散问题，寻求权向量模值自稳定的主/次成分神经网络算法是该领域一个研究热点，在这些自稳定的学习算法中，神经网络核的权向量保证收敛到一个规范化的主/次成分或者神经网络核的权矩阵保证收敛到一个规范化的主/次子空间。当前，自稳定特性已经成为神经网络主/次成分分析方法及主/次子空间跟踪算法的一个必备的特性。

3. 统一或双目的主/次成分分析自稳定算法

最初的主/次成分分析算法以及主/次子空间分析是各自独立发展的，大量的算法被提出来，并得到广泛的应用。主成分与次成分算法之间存在怎样的关系？一个自然的想法是在一个主成分分析(或者主子空间跟踪)算法中，通过改变相关矩阵的符号或者取原矩阵的逆矩阵，或者仅仅改变学习因子的符号便可以实现次成分分析(或者次子空间跟踪)，反过来也一样。实践证明这样的变换常常不成立，要么不能实现另一种成分分析或者子空间跟踪，要么虽然可以实现预期的功能但是算法更新过程中，神经网络的权向量或者权矩阵由收敛变成发散。在文献[28]，[29]中，Chen 等提出了主成分分析/主子空间分析和次成分分析/次子空间分析之间的一种转换机制，分析显示通过这种转换机制，每一个主成分分析算法都配有一个次成分分析算法，反过来也一样。这样基本解决了上述问题，通过这一转换机制，导出的双目算法具有不同的算法结构形式。那么，是否有一个统一的神经网络算法，该算法仅仅通过改变同一学习规则中的符号就能够进行主成分分析与次成分分析或者主/次子空间跟踪，无疑这样的算法更具有现实意义，可以减少硬件设施的复杂性和成本。近十多年来，寻求统一或双目的主/次成分分析(或者主/次子空间分析)算法是该领域的一个研究热点[29～31]。

1.2 特征提取与子空间跟踪基础

由次成分张成的子空间称为次子空间，而由主成分张成的子空间称为主子空间，单维主成分分析或单维次成分分析可以认为是主子空间跟踪或次子空间跟踪的特殊形式。在主成分或次成分神经网络领域，芬兰学者 Oja 等作出了开创性的工作。为了对次成分、主成分、次子空间以及主子空间及其应用有一个清晰的理解，下面从子空间的角度对这些概念及其数学描述与物理意义进行介绍[5]。

1.2.1 子空间概念

子空间定义：若 $\boldsymbol{S}=\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m\}$ 是向量空间 $\boldsymbol{V}$ 的向量子集合，则 $\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m$ 的所有线性组合的集合 $\boldsymbol{W}$ 称为由 $\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m$ 张成的子空间，即

$$\boldsymbol{W}=\mathrm{Span}\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m\}=\{\boldsymbol{u}:\boldsymbol{u}=a_1\boldsymbol{u}_1+a_2\boldsymbol{u}_2+\cdots+a_m\boldsymbol{u}_m\} \tag{1.1}$$

其中，$\boldsymbol{W}$ 的每个向量称为 $\boldsymbol{W}$ 的生成元；所有生成元组成的集合 $\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m\}$ 称为子空间的张成集（spanning set）；若 $\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m\}$ 是一组线性无关的集合，则称为 $\boldsymbol{W}$ 的一组基。

如果 $\boldsymbol{A}=[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n]\in\boldsymbol{C}^{m\times n}$ 为复矩阵，那么列向量的所有线性组合的集合构成一个子空间，称为矩阵 $\boldsymbol{A}$ 的列空间（column space），用符号 Col($\boldsymbol{A}$)表示，即

$$\mathrm{Col}(\boldsymbol{A})=\mathrm{Span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}=\left\{\boldsymbol{y}\in\boldsymbol{C}^m:\boldsymbol{y}=\sum_{j=1}^{n}\alpha_j\boldsymbol{a}_j:\alpha_j\in\boldsymbol{C}\right\} \tag{1.2}$$

行空间可以类似地定义。

如前所述，矩阵 $\boldsymbol{A}_{m\times n}$ 的列空间和行空间分别由 $\boldsymbol{A}$ 的 n 个列向量和 m 个行向量张成。如果矩阵的秩 $r=\mathrm{rank}(\boldsymbol{A})$ 时，则只需要矩阵 $\boldsymbol{A}$ 的 r 个线性无关列向量或行向量(基)，即可分别生成列空间 Span($\boldsymbol{A}$) 和行空间 Span($\boldsymbol{A}^{\mathrm{H}}$)。显然使用基向量是一种经济的、更好的子空间表示方法，可以采用初等变换法进行子空间的基构造，也可以采用奇异值分解法进行基本空间的标准正交基构造。

设数据矩阵 $\boldsymbol{A}$ 不可避免地存在观测误差或噪声，令观测数据矩阵为

$$\boldsymbol{X}=\boldsymbol{A}+\boldsymbol{W}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\in\boldsymbol{C}^{m\times n} \tag{1.3}$$

其中 $\boldsymbol{x}_i\in\boldsymbol{C}^{m\times 1}$。

在信号处理和系统科学等领域中，观测数据矩阵的列空间

$$\mathrm{Span}(\boldsymbol{X})=\mathrm{Span}\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\} \tag{1.4}$$

称为观测数据矩阵。

定义相关矩阵为

$$\boldsymbol{R}_X=E\{\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\}=E\{(\boldsymbol{A}+\boldsymbol{W})^{\mathrm{H}}(\boldsymbol{A}+\boldsymbol{W})\} \tag{1.5}$$

假设误差矩阵 $\boldsymbol{W}=[\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_n]$ 与真实数据矩阵 $\boldsymbol{A}$ 统计不相关，则

$$\boldsymbol{R}_X = E\{\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\} = E\{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\} + E\{\boldsymbol{W}^{\mathrm{H}}\boldsymbol{W}\} \tag{1.6}$$

令 $\boldsymbol{R}=E\{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\}$ 和 $E\{\boldsymbol{W}^{\mathrm{H}}\boldsymbol{W}\}=\sigma_w^2\boldsymbol{I}$，即各观测噪声相互统计不相关，并且具有相同的方差 σ_w^2，则

$$\boldsymbol{R}_X = \boldsymbol{R} + \sigma_w^2\boldsymbol{I} \tag{1.7}$$

令 $\mathrm{rank}(\boldsymbol{A})=r$，则矩阵 $\boldsymbol{R}=E\{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\}$ 的特征值分解

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}} + \sigma_w^2\boldsymbol{I} = \boldsymbol{U}(\boldsymbol{\Lambda}+\sigma_w^2\boldsymbol{I})\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{U}\boldsymbol{\Pi}\boldsymbol{U}^{\mathrm{H}}$$

其中，$\boldsymbol{\Pi}=\boldsymbol{\Sigma}+\sigma_w^2\boldsymbol{I}=\mathrm{diag}(\sigma_1^2+\sigma_w^2,\cdots,\sigma_r^2+\sigma_w^2,\sigma_w^2,\cdots,\sigma_w^2)\boldsymbol{\Sigma}=\mathrm{diag}(\sigma_1^2,\cdots,\sigma_r^2,0,\cdots,0)$，且 $\sigma_1^2\geqslant\sigma_2^2\geqslant\cdots\geqslant\sigma_r^2$ 为真实自相关矩阵 $\boldsymbol{R}=E\{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\}$ 的非零特征值。

显然，如果信噪比足够大，即 σ_r^2 比 σ_w^2 明显大，则将含噪声的自相关矩阵 $\boldsymbol{R}_X$ 的前 r 个大特征值，即

$$\lambda_1=\sigma_1^2+\sigma_w^2,\lambda_2=\sigma_2^2+\sigma_w^2,\cdots,\lambda_r=\sigma_r^2+\sigma_w^2$$

称为主特征值(principal eigenvalue)。

$$\lambda_{r+1}=\sigma_w^2,\lambda_{r+2}=\sigma_w^2,\cdots,\lambda_n=\sigma_w^2$$

称为次特征值(minor eigenvalue)。

这样，自相关矩阵 $\boldsymbol{R}_X$ 的特征值分解即可写为

$$\boldsymbol{R}_X = [\boldsymbol{U}_S,\boldsymbol{U}_n]\begin{bmatrix}\boldsymbol{\Sigma}_S & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{\Sigma}_n\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_S^{\mathrm{H}}\\ \boldsymbol{U}_n^{\mathrm{H}}\end{bmatrix} = \boldsymbol{S}\boldsymbol{\Sigma}_S\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{G}\boldsymbol{\Sigma}_n\boldsymbol{G}^{\mathrm{H}} \tag{1.8}$$

其中，$\boldsymbol{S}\overset{\mathrm{def}}{=}[\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_r]=[\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_r]$；$\boldsymbol{G}\overset{\mathrm{def}}{=}[\boldsymbol{g}_1,\boldsymbol{g}_2,\cdots,\boldsymbol{g}_{n-r}]=[\boldsymbol{u}_{r+1},\boldsymbol{u}_{r+2},\cdots,\boldsymbol{u}_n]$；$\boldsymbol{\Sigma}_S=\mathrm{diag}(\sigma_1^2+\sigma_w^2,\sigma_2^2+\sigma_w^2,\cdots,\sigma_r^2+\sigma_w^2)$；$\boldsymbol{\Sigma}_n=\mathrm{diag}(\sigma_w^2,\sigma_w^2,\cdots,\sigma_w^2)$。
因此，$m\times r$ 酉矩阵 $\boldsymbol{S}$ 和 $m\times(n-r)$ 酉矩阵 $\boldsymbol{G}$ 分别是与 r 个主特征值和 $n-r$ 个次特征值对应的特征向量构成的矩阵。

令 $\boldsymbol{S}$ 是与观测数据的自相关矩阵的 r 个大特征值 $\lambda_1,\lambda_2,\cdots,\lambda_r$ 对应的特征向量矩阵，其列空间 $\mathrm{Span}(\boldsymbol{S})=\mathrm{Span}\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_r\}$ 称为观测数据空间 $\mathrm{Span}(\boldsymbol{X})$ 的信号子空间，而与另外 $n-r$ 个次特征值对应的特征向量矩阵 $\boldsymbol{G}$ 的列空间 $\mathrm{Span}(\boldsymbol{G})=\mathrm{Span}\{\boldsymbol{u}_{r+1},\boldsymbol{u}_{r+2},\cdots,\boldsymbol{u}_n\}$ 称为观测数据空间的噪声子空间。

由子空间的构造方法及酉矩阵的特点，信号子空间与噪声子空间正交，即

$$\mathrm{Span}\{\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_r\} \perp \mathrm{Span}\{\boldsymbol{g}_1,\boldsymbol{g}_2,\cdots,\boldsymbol{g}_{n-r}\} \tag{1.9}$$

由于 $\boldsymbol{U}$ 是酉矩阵，故

$$\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}} = [\boldsymbol{S},\boldsymbol{G}]\begin{bmatrix}\boldsymbol{S}^{\mathrm{H}}\\ \boldsymbol{G}^{\mathrm{H}}\end{bmatrix} = \boldsymbol{S}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{G}\boldsymbol{G}^{\mathrm{H}} = \boldsymbol{I}$$

即

$$\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}} = \boldsymbol{I} - \boldsymbol{S}\boldsymbol{S}^{\mathrm{H}} \tag{1.10}$$

定义信号子空间上的投影矩阵为

$$P_S \stackrel{\text{def}}{=} S\langle S,S\rangle^{-1}S^{\mathrm{H}} = SS^{\mathrm{H}} \tag{1.11}$$

其中,矩阵内积 $\langle S,S\rangle = S^{\mathrm{H}}S = I$。

于是,$P_S x$ 可视为向量 x 在信号子空间上的投影,而 $(I - P_S)x$ 则代表向量 x 在信号子空间上的正交投影。由 $\langle G,G\rangle = G^{\mathrm{H}}G = I$,可得噪声子空间上的投影矩阵 $P_n = G\langle G,G\rangle^{-1}G^{\mathrm{H}} = GG^{\mathrm{H}}$。因此,常将

$$GG^{\mathrm{H}} = I - SS^{\mathrm{H}} = I - P_S \tag{1.12}$$

称为信号子空间的正交投影矩阵。

子空间应用具有以下几个特点[32]:

① 只需要少数几个奇异向量或者特征向量。由于矩阵 $A_{m\times n}$ 的大奇异值(或者特征值)个数比小奇异值(或者特征值)个数少,所以使用维数比较小的信号子空间比噪声子空间更有效。

② 在很多应用中,并不需要奇异值或者特征值,而只需知道矩阵的秩及奇异向量或者特征向量即可。

③ 多数情况下,并不需要准确知道奇异向量或者特征向量,而只需知道张成信号子空间或者噪声子空间的基向量即可。

④ 子空间方法可以应用于多重信号分类(multiple signal classification,MUSIC)[33~35]、波达方向估计、盲信号辨识和子空间白化[36,37]等应用。

1.2.2 子空间跟踪方法

对一个极端(最大或最小)特征对(特征值与特征向量)进行迭代计算最早可以追溯到 1966 年[38]。Thompson 于 1980 年提出了估计与样本协方差矩阵最小特征值对应的特征向量的最小均方型自适应方法,并结合 Pisarenko 谱估计子提供了角度/频率的自适应跟踪算法[39]。Sarkar 等[40]应用共轭梯度法,对与慢时变信号协方差矩阵的最小特征值对应的极端特征向量的变化进行跟踪,并证明该方法比 Thompson 的最小均方方法收敛快得多。这些方法只跟踪单个极端特征值和特征向量,虽然应用有限,但是后来被推广为特征子空间的跟踪和更新方法。1990 年,Comon 与 Golub[41] 提出了跟踪奇异值和奇异向量的极端对的 Lanczos 方法,Lanczos 方法原是解决大的、稀疏的对称特征问题 $Ax = \lambda x$ 的一种常用方法[42]。

最早的特征值和特征向量更新方法是由 Golub[43] 于 1973 年提出的。后来,Golub 的更新思想被 Bunch 等[44,45]加以扩展。推广方法的基本思想就是在每次秩 1 修正后更新协方差矩阵的特征值分解,然后利用交织定理(interlacing theorem)将矩阵的特征根定位,用迭代求根方法更新特征根的位置,从而更新特征向量。Schereiber[46]引入一种变换,将大部分的复数算术运算变成实数运算,并使用 Karasalo 的子空间平均方法[47]进一步减少运算量。DeGroat 与 Roberts[48] 发