**单位代码：11414**

**学 号：2016011647**

****

本科生毕业设计（论文）

|  |  |
| --- | --- |
| **题目** | **奇异谱分解以及时频分析中的应** |
|  | **用研究（题目改为：奇异谱分解及其在信号处理中的应用研究）** |
| **学院名称** | **理学院** |
| **专业名称** | **数学与应用数学** |
| **学生姓名** | **李旺** |
| **指导教师** | **武国宁** |

**起止时间： 2020年2月28日 至 2020年5月24日**

**需要做如下修改：**

1. **修改论文题目；**
2. **重新撰写摘要；**
3. **每张图需要有中英文标题，需要按照学校模版修改；**
4. **第三章改为第二章的一节，因为内容太单薄；**
5. **第四章和第五章合并为一章，分为两小节即可。**
6. **在地震信号处理中尝试截取前1、2，3等主成分合成信号和原始信号比较，分析误差。**

# 摘 要（摘要按照一下格式书写）

奇异谱是一种XXX方法技术。该方法具有什么优势，能够应用于什么领域。

本文研究了奇异谱方法在XXX领域的应用，数值试验表明了：XXX

本论文研究了奇异谱分析在时频分析中的应用。时频分析是一种新兴的信号处理方法，最近受到越来越多的重视。时间——频率分析方法提供了在频域中组合分布信息和时域，清楚地描述随时间变化的信号的频率变化之间的关系。时频分析方法包括：傅里叶变换、短时傅里叶变换、小波变换、S变换、希尔伯特-黄变换和奇异谱分析方法。最经典的是傅里叶变换，而小波变换和短时Fourier变换以及S变换都是基于Fourier变换的方法。并且具有一定的局限性。希尔伯特-黄变换解决了传统变换方法一些局限性，可以处理非线性非平稳的信号，并且具有自适应性的特点。经过发展人们发现奇异谱分析方法处理时频信号数据时在去噪声、提取趋势、稳定性以及预测方面有着更好的效果。

SSA是由Colebrook在1978年首次提出，是一种新颖而强大的的时间序列分析方法技术，融合了经典时间序列分析、多元几何、多元统计量、动力学系统和信号处理。SSA的可能通过应用技术领域是多种形式多样的：从数学和物理学到经济学和金融发展数学，从计量学和海洋学到社会主义科学和市场经济研究。在近年来，余斌、杨少敏等人研究了通过奇异谱方法的北斗恒星日滤波算法，通过采用Cao算法确定嵌入维度，降低了参数选择主观性，而且还提高了奇异谱方法分计算效率，林佩如提出了一种根据经验模式分解来确定奇异谱分析分组的理论，周舟利用多道奇异谱分析方法处理数据重建，赵佳佳在奇异谱分析在地形变数据处理中提出了周期图法和G-P关联维数算法相结合的方法，取得了广泛的应用并且做些了少许改进。

本文着重介绍了奇异谱分析方法的基本原理、奇异值分解和参数选择，而且还介绍了奇异值分解的一些算法，并且采用python将SSA应用于地震信号处理的实例。

**关键词：奇异谱分析；时频分析；python；地震信号**

# Application Research Based on Singular Spectrum Analysis in Time-frequency Analysis

# ABSTRACT

This paper studies the application of singular spectrum analysis in time-frequency analysis. Time-frequency analysis is an emerging signal processing method that has received more and more attention recently. Time-frequency analysis method provides the combination of distribution information and time domain in the frequency domain, clearly describing the relationship between the frequency change of the signal that changes with time. Time-frequency analysis methods include: Fourier transform, short-time Fourier transform, wavelet transform, S transform, Hilbert-Huang transform and singular spectrum analysis methods. The most classic is Fourier transform, and wavelet transform, short-time Fourier transform and S transform are based on Fourier transform. And it has certain limitations. Hilbert-Huang transform solves some of the limitations of traditional transform methods, can deal with nonlinear and non-stationary signals, and has the characteristics of adaptability. After development, it was found that the singular spectrum analysis method has better results in noise removal, trend extraction, stability and prediction when processing time-frequency signal data.

SSA was first proposed by Colebrook in 1978. It is a novel and powerful time series analysis method technology, which combines classic time series analysis, multivariate geometry, multivariate statistics, dynamic systems and signal processing. The possible areas of SSA through applied technology are many and varied: from mathematics and physics to economics and financial development mathematics, from metrology and oceanography to socialist science and market economy research. In recent years, Yu Bin, Yang Shaomin and others have studied the Beidou stellar day filtering algorithm through the singular spectrum method. By using the Cao algorithm to determine the embedding dimension, the subjectivity of parameter selection is reduced, and the efficiency of the singular spectrum method is also improved A theory based on empirical mode decomposition to determine the grouping of singular spectrum analysis is proposed. Zhou Zhou uses multi-channel singular spectrum analysis method to

process data reconstruction. Zhao Jiajia proposed periodic graph method and GP correlation dimension in the processing of terrain variable data in singular spectrum analysis. The method of combining numerical algorithms has been widely used and made some improvements.

This article focuses on the basic principles of singular spectrum analysis, singular value decomposition and parameter selection, and also introduces some algorithms of singular value decomposition, and uses Python to apply SSA to seismic signal processing.

**Keyword：Singular spectrum analysis；Time-frequency analysis；python ；Seismic signal**

# 第1章 绪论

## 1.1 研究背景和意义

奇异频谱分析（singular spectrum analysis, SSA）技术是一种新颖而强大的的时间序列分析方法技术，融合了经典时间序列分析、多元几何、多元统计量、动力学系统和信号处理。SSA的可能通过应用技术领域是多种形式多样的：从数学和物理学到经济学和金融发展数学，从计量学和海洋学到社会主义科学和市场经济研究。

## 1.2 国内外研究现状

### 1.2.1 奇异谱分析方法研究现状

1978年，Colebrook首次提出SSA方法，该方法最早开始于K-L正交展开，它的是一种主成分分析方法（Principal Component Analysis,PCA）在一维时间序列分析中的延伸。SSA不仅具有PCA的优点，可以进行一维数据和图像资料的压缩，但也有自身的优点，比如具有去噪、探测信号、插值、滤波和建立预测模型和稳定的识别周期信号等功能。

国内外研究人员结合SSA及其各自研究领域的优势，利用SSA开展了大量的研究工作。SSA最初在气候学领域被广泛使用。在趋势发展预测和周期特征提取技术方面，Ghil 利用SSA分析了全球变暖趋势和大气环境温度年代际波动的时间管理序列[3]；徐建军等人使用奇异频谱分析方法分析了1873-1990年。近100年来，对东亚季风周期变化进行了一些理论研究，发现准东亚季风周期具有不同的特征。东亚冬季和夏季风的两年年度振荡周期[4]；王宁等，用SSA分析了石羊河流域气象条件的设计周期和趋势，表明SSA变化明显，时间序列趋势和周期值的判断具有较好的合理性和优越性[5]；汪芸等针对当今社会中长期径流的预测和控制精度差的问题，采用1951年以前的奇异谱分析方法，对隔河岩水库入库径流量数据进行图像预处理后，到2009年，建立了混沌支持向量机模型和自回归方程模型。分别用于模拟和预测原始财务数据以及处理后的重建顺序。结果表明，采用SSA分析方法可以对调查数据采集数据进行预处理，可以大大提高全国中长期径流的预报精度[6]；黄嘉佑等利用SSA方法分析了宜昌地区降水和汛期的时间序列，并将奇异谱分析的预测结果与谱分析的结果进行了比较。在短期内，天气预报测试方法SSA的实际效果要优于预测模型[7]。

在去噪方面，Vautard 于1992年将SSA用于混沌信号降噪，证明了SSA具有完善的降噪功能，并开发了相应的软件包；刘元峰等基于SSA的方法在分析基本工作原理的基础上，提出了一种可以降低混沌信号噪声的算法和一种主要基于环境奇异谱特征的最优重构阶数选择方法，并通过这个数值模拟仿真实验验证了此算法的有效性[8]；黄建平提出了一种基于SSA的算法。通过对合成地震数据记录和真实生活数据进行联合降噪和正则化处理，分析了联合降噪和正则化方法，表明联合方法不仅可以完成地震通道的建立，还可以有效地抑制噪声[9]。

在有缺乏数据的应用插值方面，贾永娜等用于地震数据插值的问题，极大提高了计算的速度[10]；根据SSA迭代插补的原理王辉赞等，对于这个方法确定主成分个数K和嵌入维数M上效率不高、精度有缺陷，提出的参数优化算法 - 四分法部，使得迭代内插精度和计算效率显著提高[11]，毛向东等学者为了进一步提高工作经验发展模式分解对微弱信号的提取效果，解决了经验模态设计方法分解获得的几种固有模式函数。奇异频谱分析可以有效地提取小信号弱信号。信噪比[12]。

### 1.2.2 时频分析研究现状

时频分析（JTEA）是时间-频率联合域分析（Joint Time-Frequency Analysis）的一个简称，是分析时变非平稳信号的主要工具，已经成为现代各种信号处理研究的一个主要方法，时频分析是一种新兴的信号处理方法，最近受到越来越多的重视。时间 - 频率分析方法提供了在频域中组合分布信息和时域，清楚地描述随时间变化的信号的频率变化之间的关系。当信号进行时频分析以及频率分析时，提前对信号进行自适应分解和转换是非常重要的。

接下来介绍几种传统时频分析方法：

1. 短时傅里叶变换

短时傅里叶变化是基于传统傅里叶变化（Fourier Transform）的存在的一些不足改进而来，传统傅里叶变化分析方法已经在许多领域产生了巨大的影响。而且在1965之后，快速傅里叶变化（FFt）算法的实现以及改进使得离散傅里叶变换（DFT）实现了更快效率的实现，而且给数字信号处理技术应用提供了条件，促进了离散时间信号和系统分析技术的发展。就像詹姆斯凯赛说过：“最多使用的信号处理工具是FFT，同时也是最多被滥用的”。到了20世纪80年代以后，数字信号处理技术在联合时频分析方法方面有了很大的发展，而且取得了非常多的研究和应用。还形成了一套体系。STFT就是其中一种最简单的分析方法。

鉴于传统傅里叶方法不适用非平稳信号，没有局限性以及时间域和频率域分开的缺陷，提出了联合时频方法。这个概念是尤金维格纳在1932年提出，找到一个二维函数，把信号的时间域和频率域分析结合起来，所以该方法即反映了信号的频率，也反映了频率随着时间变化的规律。此方法可以大体分为两类：线性和非线性联合时频分析方法，线性主要包括短时傅里叶变化、Gabor展开以及小波变换。非线性的包括Wigher-Ville分布和双线性时频分布等。

短时傅里叶变换的基本思想是将数字信号加窗，在进行傅里叶变化。加窗处理可以使得变化结果是时间t附近的小时间段上的局部谱，这个窗函数可以依据时间t在整个时间段上平移，所以利用窗函数就可以使得任意时间t普金的图谱实现局部化，所以就构建成了信号的二维频谱。如果信号是非平稳的，但是经过加窗处理之后，可以假设每一段信号是平稳的，故此方法也能用于非平稳信号或者时变信号的处理。

短时傅里叶变换属于线性联合时频方法，没有造成高次型非平稳分析中出现的交叉干扰项，适合多种分量信号分析，此外，短时傅里叶变换的窗函数方法计算相对来说简单，所以STFT克服了传统的缺陷也便于是实现。

当然STFT也具有一定缺陷，比如：不适合用于信号编码，鉴于测不准原理的影响时频的聚集性有限并且不具有自适应性。可通过下图来说明

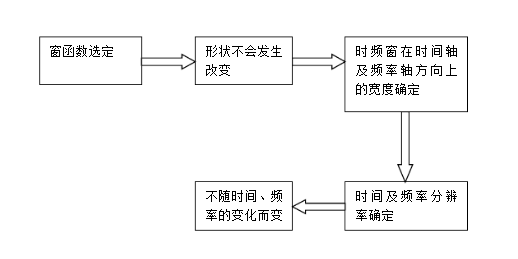


图1.1

1. 小波变换

小波变换也是在STFT的思想上发展起来的，虽然它的窗口大小不变，但是形状会发生改变。它是有法国科学家MORLET1980年提出进行地震数据分析是效率非常好的时频分析工具。已经成功应用于图像处理、信号处理、模式识别等。它的一个重要性质是它在时频和频域都有着很好的局部化特征。它能很好的表现出每个子段的频率信息，这对于信号的分类是费城有用的。小波分类的标准是根据支撑长度（当频率趋于无穷大，长度越小，对奇异点的区分效果越好）、对称性（与信号的重构精度有关）、正则性（平滑效果）来分类的。

1. S变换

S变换也是一种加窗傅里叶变换，是对Gabor和小波变换的延伸。该方法用宽度可变的窗函数，并且宽度与频率成反比，在低频率的窗口宽，可以获得比较高的频率分辨率，在高频率的窗口窄，就可以获得较高的时间分布率，就能观察到一些细微的部分，克服了FFT和小波变换的一些缺陷，又和S变换、傅里叶变化保持联系。在国内众多领域获得应用，比如地震学、医学、电力等领域的信号处理。

1. 希尔伯特─黄转换

美籍华人N.E Huang等人深入研究了瞬时频率的概念，提出了希尔伯特-黄变换（Hilbert-Huang transform，HHT）的新分析方法，并且提出了固有模态信号以及经验模态分解法，用固有模态信号为视频信号的新时频分析方法体系，给瞬时频率定义、物理意义以及求法。这个方法体系脱离了Fourier方法变换理论的束缚。

首先要假设：每一个信号都是由若干固有模态函数和模态信号组成，并且单个信号允许有许多固有模态信号，HHT的主要思想是将时间序列数据经过经验模态分解，分解成多个固有模态函数，然后再进行变换就得到了序列的瞬时频率和振幅，通过上述方法最后得到希尔伯特谱。

经验模态分解通俗称之为“筛选”过程。这个过程依据自适应信号的特点把复杂的信号分解成IMF。满足以下两个条件：（1）信号零点数和极大值点的数量一致或者相差一，（2）信号由极小值定义的和极大值定义的局部均值为零。

EMD筛选过程：

1. 根据输入信号求极大值点和极小值点；
2. 用三次样条函数插值构造信号基于极大值和极小值点上下包络，并且计算上、下包络的均值函数；
3. 需要考察是不是满足IMF的条件，如果不满足，对h1进行前两步操作，求的m，再然后依次下去，直到满足条件；若满足则直接进行下一步；
4. 从而得出第一个残留，对其做上面的三部操作，得到c2以此类推；
5. 直到只有一个极点或成为单调信号为止。

分解成的本征模态函数对应的瞬时频率有其物理意义，所以对每一个本征模态函数求HT变化都能得到相对应的瞬时频率。相对于传统信号和数据处理技术，希尔伯特黄变换有几个优点：

1. 希尔伯特黄变换能分析非线性非平稳的信号，解决了傅里叶变换和小波变换只能处理单一问题的缺陷。
2. 它还有完全自适应性，能自适应产生基是HHT的一大特点，基也就是有经验模态分解产生的IMF。而傅里叶的基三角函数和小波变换的满足可容性条件都是要预先选定的，而在现实情况中预先选定合适的基并不是很容易。而HHT的自适应性就解决了这一问题。
3. 希尔伯特黄变换也不受测不准原理的影响，可以处理突变信号，而传统方法比如傅里叶变换、窗口傅里叶变换和小波变换都受此因素影响，这样就要在牺牲频率精度的情况下提高时间精度，所以不能同时具有很高的精度，这就对于信号处理来说不太方便。而HHT处理之后可以在两者都达到很高的精度。
4. 希尔伯特黄变换采用求导得到瞬时频率，在上述三种传统方法中已经提到它们都是提前选择基函数。但是HHT和这些方法都不一样，它用希尔伯特变换求出相位函数，再求导产生瞬时频率，这样就具有局部性。而Fourier变换产生的频率是全局性的，小波变化是区域性的。

### 1.3 本论文研究内容

奇异值分解是矩阵分析的重要方法之一，该方法在特征表示，聚类等领域有着重要的应用。近年来提出的奇异谱分解是一种非线性时间序列分析方法，该方法根据时间序列构建轨迹矩阵，并对轨迹矩阵进行分解，重构，从而提出代表原时间序列不同成分的信号，如长期趋势信号，周期信号，噪声信号等。从而对时间序列可以进行结构分析，预测，去噪声等研究。本论文主要研究：

（1） 奇异谱分解的原理；

（2） 奇异谱分解的实现算法；

（3） 基于奇异谱分解的时频表示及其在地震信号处理中的应用。

# 第2章 奇异谱分析方法原理

## 2.1 简介

考虑具有足够长度T的实值非零时间序列YT =（y1，...，yT）。SSA的主要目的是将原始序列分解为序列的总和，从而可以使该总和中的每个分量为标识为趋势，周期性或准周期性成分（可能是幅度调制的）噪声。接下来是重建原始系列SSA技术由两个互补阶段组成：分解和重建，并且都包括两个单独的步骤。在第一阶段我们分解序列，在第二阶段我们重建原始序列，并使用重建的序列（无噪声）进行预测新数据点。

## 2.2 奇异谱分析理论基础

### 2.2.1 分解阶段

分解阶段分为嵌入过程和奇异值分解两个步骤：

1. 嵌入过程

嵌入可视为将一维时间序列转化为其自变量满足的多维时间序列，其中K=T-L+1。向量被称为L滞后向量（或简称为“滞后向量”）。嵌入的单个参数是窗口长度L，即的整数。该步骤的结果是轨迹矩阵。



矩阵的每一列是通过移动窗口长度L对原时间序列进行截取的部分序列片段，轨迹矩阵X是汉克尔矩阵，这意味着沿对角线i+j=const的所有元素都是相等的。嵌入是时间序列分析中的标准过程。进行嵌入后未来的分析取决于调查的目的。

1. 奇异值分解

SVD步骤，对轨迹矩阵进行奇异值分解，并将其表示为秩为一的双正交基本矩阵的和。用表示的特征值（），用表示正交系统（即（）=0当ij和对应于矩阵的特征向量的=1（单位范数）时）。（）是向量和的内积，是向量的范数。设：



令，轨迹矩阵的SVD可以表示为：

 （2.1）

其中。矩阵的秩为1，因此他们都是基础矩阵（在SSA文献中被称为“经验因子或正交函数”或简称为EOF）和（通常被称为“主成分”）代表轨迹矩阵的左右特征向量。集合被称为矩阵X的第i本征三元组，（i=1,...,d）为矩阵X的奇异值，集合被称为矩阵X的谱。如果所有特征值都是一重的，则式子（2.1）是唯一定义的。

在所有秩r<d的所有矩阵中，SVD（2.1）是最佳的。矩阵对轨迹方程X的最佳相似，因此是最小的。注意和（i=1,...,d）。因此对于式子（2.1）我们可以将比率视为矩阵贡献的特征。也就是说即第一个r比率的和，是秩为r的矩阵对轨迹矩阵的最佳逼近的特征。

### 2.2.2 重构阶段

重构阶段分为分组和对角平均化两个阶段：

1. 分组

分组是将基本矩阵分成剧组并且将每组内的矩阵求和。令，与相对应的矩阵定义为，一组指标J=1,...,d的溢出进入不相交的子集对应表示为：

 （2.2）

称为本征三元组。在式子（1）中对于给定的组分量的贡献通过对应的特征值的份额来确定：。

1. 对角平均化

对角平均化将每个矩阵转化为时间序列，即最初序列的加法成分。如果代表矩阵的元素，那么通过对取平均值来获取结果序列的第k个项且i,j满足i+j=k+2.这个过程叫做对角平均化或者矩阵z的汉克化，矩阵汉克化的结果：是汉克尔矩阵，是与作为对角平均化的结果而获得的序列相对应的轨迹矩阵。请注意，在相应大小的所有汉克矩阵中，矩阵最接近（相对于矩阵范数）的意义上，汉克化是一种最佳过程。

汉克矩阵将对角线的值与序列中的值相关联来唯一定义序列。通过对（2.2）的所有矩阵分量应用汉克化程序，我们得到另一个式子：

 （2.3）

其中。这等效于将初始序列分解为m个序列的和：

 （2.4）

对于矩阵等效于。

## 2.3 补充信息

我们描述一些信息，这对于识别原始序列的轨迹矩阵的SVD的本征三元组非常有帮助。补充信息可帮助我们建立适当的组以推断趋势，谐波分量和噪声。因此，补充信息可以被视为分解和重建步骤之间的桥梁：

分解→补充信息→重建

下面，我们简要说明一些方法，这些方法可用于将信号分量与噪声分离。辅助信息，在许多实际情况下，辅助信息的可用性提高了构建适当模型的能力。 当然，有关初始序列的辅助信息可以使情况更加清晰，并有助于选择模型的参数。 这些信息不仅可以帮助我们选择合适的组，而且对于基于SSA技术的预测和更改点检测也很有用。

（1）奇异值

通常，每个具有不同频率的谐波分量都会产生两个具有近似奇异值的本征三倍频（除了频率0.5的，后者是一个具有三阶锯齿形奇异矢量的本征三倍频）。 如果N，L和K足够大，它将更加清楚。通过检查特征值谱的断裂，可以提供另一种有用的见解。 通常，纯噪声序列会产生一个递减的奇异值序列。

（2）窗口长度的选择

窗口长度L是分解阶段的唯一参数。 选择合适的窗口长度取决于手头的问题以及有关时间序列的初步信息。理论上L应该足够大但不大于T/2.此外，如果我们知道时间序列的周期分量可能具有整数周期（例如，如果该分量是季节性分量）那么为了获得更好的周期分量的可分离性，建议将窗口长度分离到那个时期，使用这些建议。注意，轨迹矩阵X的行和列是原始时间序列的子序列。 因此，左特征向量和主分量（右特征向量）也具有时间结构，因此也可以被视为时间序列。

1. 趋势识别

趋势是时间序列中变化缓慢的组成部分，其中不包含振荡成分。假设时间序列本身就是这样一个组成部分。实践表明，在这种情况下，一个或多个前导特征向量也将缓慢变化。我们知道，特征向量具有（通常）与初始时间序列的相应分量相同的形式。因此，我们应该找到缓慢变化的特征向量。可以通过考虑特征向量的一维图来完成。

1. 周期图分析

原始序列和特征向量的周期图分析可能对我们进行正确的分组很有帮助。 它告诉我们必须考虑哪个频率。 然后，我们必须寻找频率与原始序列频率一致的本征三元组。如果特征向量的周期图在某些频率附近有尖锐的火花，则必须将相应的特征三元组视为与信号分量有关的特征。

1. 可分离性

研究SSA属性的主要概念是“可分离性”，它表征了不同组件之间的分离程度。系列的SSA分解只有在序列中所得的添加成分彼此大致可分离的情况下才能成功。

以下数量（称为加权相关或w-相关）是自然两个系列和之间的依赖关系的自然度量。



其中，（），。

分组时w相关的绝对值的矩阵，对应于完全分解（在此分解中，每个组仅于SVD的一种矩阵分量）。如果w相关的绝对值较小，则相应的序列几乎是w正交的，但是，如何w相关性的绝对值较大，则这两个序列远不是正交的，因此很难分离。因此，如果两个重构分量的w相关系数为0，这意味着两个分量是可分离的。重建的分量之间相关系数越大，表明这些分量可能应该被收集到一个组中，并且与SSA分解中的同一分量相对应。

## 2.4 本章小结

本章详细介绍了奇异谱分析基础理论主要分为两个阶段分解和重构，并且分解阶段又分为嵌入过程和奇异值分解，而奇异值分解是奇异谱分析中的重点，重构阶段又分为分组和对角平均。接下来又介绍了一些辅助信息奇异值、窗口长度、趋势识别、周期图分析、可分离性。

# 第3章 奇异值方法算法种类简介以及改进（第三章作为第二章的一节即可，内容太少。）

## 3.1 奇异值算法种类

传统QR迭代算法：首先通过正交变换将矩阵变换为双对角矩阵，然后通过QR变换算法对变换后的双对角矩阵进行迭代。Kahan早先证明了双对角矩阵的奇异值可以精确计算，并且精度与原始矩阵相同。零位移QR算法：在1990年，Kahan和demmel提出了一种零偏移QR算法。 该算法不仅通过计算双对角矩阵的奇异矢量具有很高的精度，而且具有很高的奇异值精度。新QD算法：1994年，Fernando和Parlett提出了一种将QD算法应用于奇异值计算的新思路，与第二种算法相比，具有精度更高，计算速度更快的优点。Jacobi算法：虽然与其他算法相比，奇异值和奇异向量的计算精度更高，但其速度比DK算法慢。Bisection算法：同样，该算法也可以求解双对角矩阵，并以相对准确的方式找到所有奇异值。

## 3.2 奇异谱方法改进

2019年，余斌、杨少敏[13]等人研究了通过奇异谱方法的北斗恒星日滤波算法，考虑到此算法计算量过大，计算效率还是有待提高，提出了通过采用Cao算法确定嵌入维度，降低了参数选择主观性，而且还提高了奇异谱方法分计算效率，相比较基于奇异谱方法的北斗恒星日滤波算法更加能够高效的消除多路径误差的影响，明显的提升了北斗动态变形监测的精度。

2018年林佩如[14]提出了一种根据经验模式分解来确定奇异谱分析分组的理论。此理论是应用经验模式分解处理得到的本质模态函数数目来确定分组数目然后通过匹配追踪算法分组把奇异谱分量和相对应的最大相关系数的本质模态函数分为一组。并且和其他分组理论相比较有三大优势：

1. 经过此理论处理得到的组无论是在时域上和频域上都更有利于进行分析，并且组还有着两种方法的分析特性。
2. 此分组理论方法可以自动的确定分组数目，没有人为主观性的干扰，因此具有较强的稳定性。
3. 它还具有计算量小和分组原则容易理解的优点。

该方法与传统的奇异谱分析方法、小波阈值降噪方法、基于主成分分析的方法和EMD降噪方法相比较，有着更好的鲁棒性和降噪效果。这个方法还是有些不完善的，比如通过匹配追踪算法和计算相关系数进行分组，会使得两种方法不够紧密结合，还需要优化模型解决这一问题；降噪效果还稍微偏弱，有待加强的空间。

2014年周舟[17]利用多道奇异谱分析方法处理数据重建的问题，而此方法前提要满足拟线性条件或线性条件，作者将单步预测滤波插值思想引用，并且用已知数据在低频处的Lanczos向量作为约束降秩来插值缺失的高频数据成分，从而实现基于多道奇异谱分析方法的地震数据重建，证明了方法的有效性。

2017年赵佳佳[15]在奇异谱分析在地形变数据处理中提出了周期图法和G-P关联维数算法相结合的方法，用于参数嵌入维度L的选择，不仅避免了主观性选择的缺点，而且还节约了计算成本，在重构阶次P的选择中将方差贡献率和拐点法进行结合应用。

# 第4章 模拟实验分析（第四章，第五章合并为一章，分为第一节模拟试验，第二节地震信号处理）

## 4.1 模拟实验奇异谱方法预测分析

在使用SSA进行模拟实验之前，选取一组含趋势的时间序列数据来验证它的识别趋势，季节性和噪声功能。SSA的一般程序步骤如下：

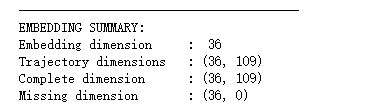
（1）嵌入时间序列通过形成滞后窗口（长度K）向量的Hankel矩阵

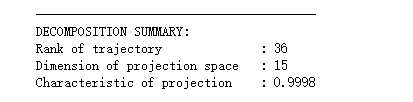
（2）分解基于奇异值分解的嵌入时间序列

（3）特征波纹分组是将特征值-特征向量对识别为趋势、季节性和噪声

（4）重建特征值-特征向量对应时间序列趋势和季节性。这是一个通过对角线平均的办法。

模拟实验奇异谱方法预测分析，选取嵌入维度为36处理时间序列：





然后将每个信号（对应于每个奇异值）的贡献涨时如下图（4.1），从（4.1）我们可以看到，大部分的变化都包含在第一个信号中（对应于第一个奇异向量）。（4.1）图像中的比例使得很难看到剩余的奇异值怎么回事。所以y轴上的刻度是没有意义的，在这种情况下被移除，经过处理：现在我们有了信号分量（信号中有15个可加分量，如图4.2），我们可以在这个范围内循环，并查看每一个，现在我们先只看前五个元素如图（4.3）所示，再重建这五个流的信号，把重建的信号和原始信号放在一张图上进行比较如图（4.4、4.5），看起来很合理！之后可以通过剩余的奇异值中添加额外的信号来接近原始序列。使用前十个奇异值处理的结果如图（4.6）所示，结果表明更接近原序列，之后进行预测，将使用的奇异值索引列表传递给forecast\_recurrent方法预测结果如图（4.7）：再进行原始序列长度一倍的预测：

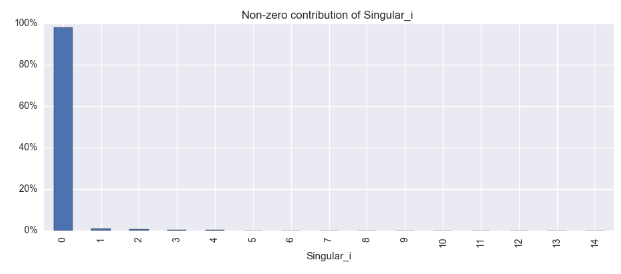


图4.1（每一幅图都应该有中英文标题，见学校论文模版）

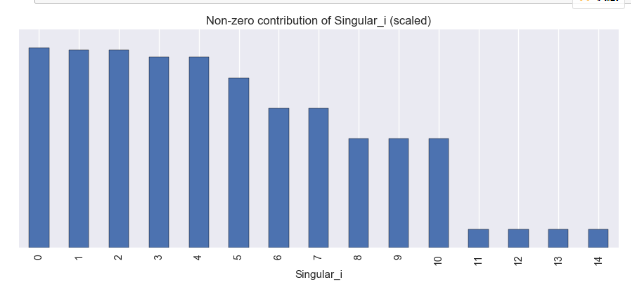


图4.2

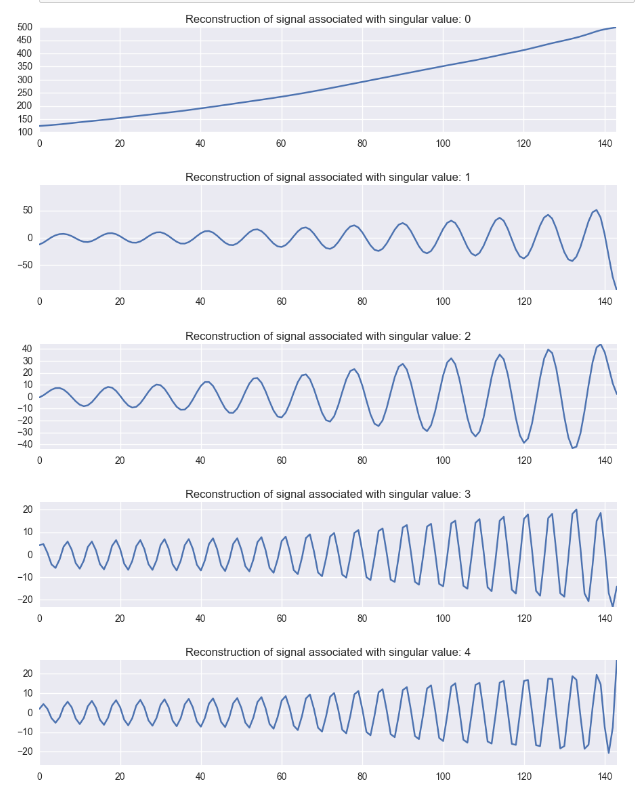


图4.3

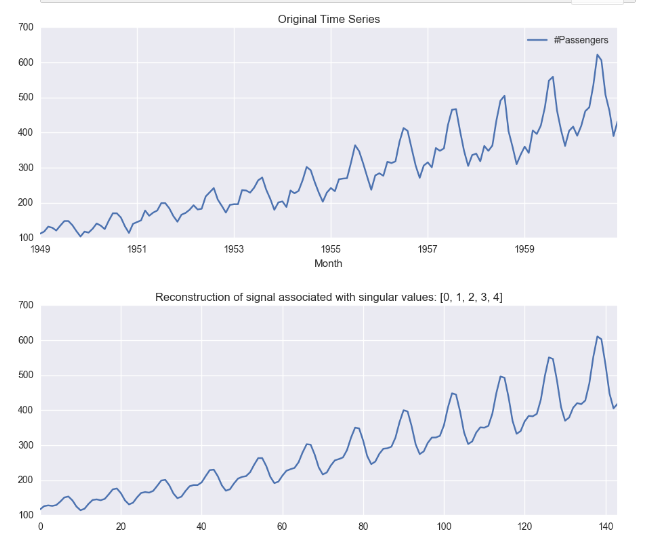


图4.4

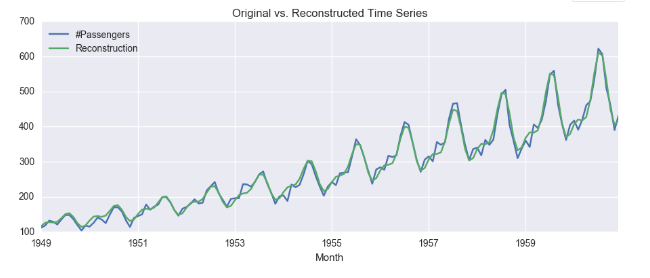


图4.5

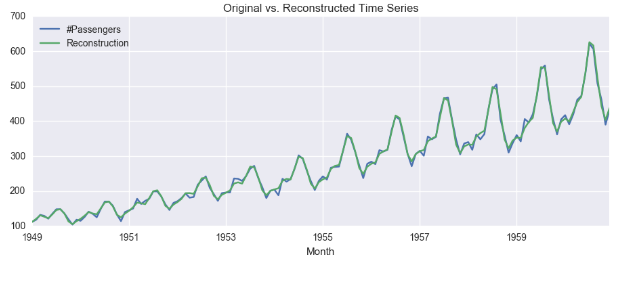


图4.6

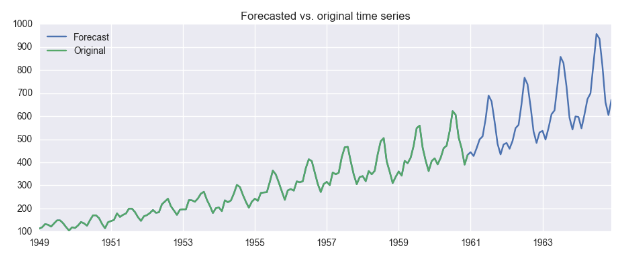


图4.7

同样，我们使用同一组时间序列，不过与上述实验不同的是，我们从中将一段数据截取出来，然后再进行奇异谱分析处理进行预测，以此来证明它的预测效果，选择同样的嵌入维度处理数据，得到的结果如下：

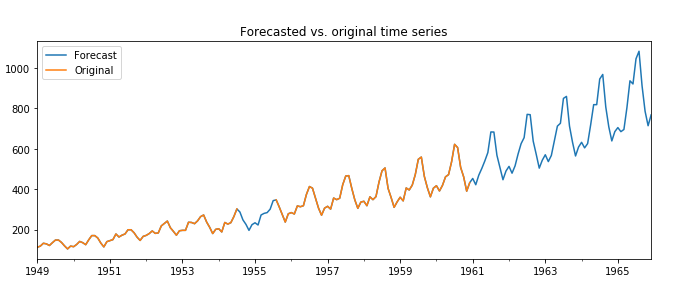


图4.7

从处理结果和原始序列相比较而言，可以明显地看到在截取1955年左右时间段数据之后。经过奇异谱分析方法预测，与原序列几乎一模一样。

## 4.2 本章小结

本章基于奇异谱分析方法用python实现处理成功分离出信号并且利用前五个信号分量和前十个信号进行重构，与原序列进行相比较达到合理的效果，并且

通过对缺失数据的序列的预测结果与原序列结果相比较而言，更加地证明了该方法在预测方面的功能。

# 第5章 地震信号中的应用

## 5.1 地震信号简介

石油和天然气的地震勘探经历了数字革命。勘探过程的大多数阶段都受到影响：数据采集，为准备信号处理而减少这些数据，设计用于检测来自掩埋界面的主要回波（反射）的数字滤波器以及开发从中提取的技术这些检测到的信号提供了有关地下几何形状和物理特性的信息。地震反射通常较弱，必须通过使用信号求和（叠加）程序来增强地震反射。确定目标地平线的深度需要了解地震应力波的传播速度，为此目的已经发展了很多技术。许多勘探工作都发生在近海地区，水层中的混响掩盖了来自下方的反射。预测性去卷积方法在衰减这些混响的能力方面最为有效，这使得检测深度结构的反射成为可能。地震信号处理既不是纯粹的科学也不是纯粹的艺术，并且对两种文化的实践者都构成了持续的挑战。

石油工业会定期记录和处理大量地震数据。在地表网格上进行地震勘测，以便建立一个区域内地下地质的三维图，每个勘测英里包含大约5000万比特的信息，这使得没有高速数字计算机就不可能进行现代数据处理。根据勘探地球物理学家协会1978年发布的数据，在1977年，石油行业在全球范围内获取和处理了大约60万行英里的地震数据，总成本超过10亿美元。

地震信号处理可以分为三类：数据采集，数据处理和数据解释。现代反射地震学方法目前已用于勘探碳氢化合物储量。在地震地震学，核爆探测，地壳研究和建筑工程中也使用了类似的处理技术。总体勘探目标包括地下地质构造图，油气藏的探测以及某个地区总能量储量的估算。大多数储层都与具有向上凸起结构（背斜）和线性位移（断层）的地质构造有关。在全球许多地方，例如在美国墨西哥湾沿岸地区，发现石油和天然气与盐穹顶有关。许多沉积物还与成分的横向变化（地层圈闭）或化石礁（礁圈）有关。反射地震信号之间的时间差映射了结构变形，而反射信号的振幅变化可能表明存在碳氢化合物。

### 5.1.1 数据采集

地震数据采集包括对地震台站连续地震信号的采集和记录。在天文台采集中心进行地震数据自动采集、检测、处理和传播的过程是在三个并行采集系统上进行的：DASP(B.Glavatovic，MSO，黑山)，地震ComP 3(德国波茨坦，GFZ)和SmartGeoHub(Geotech仪器，达拉斯，TX)。模拟信号的采集是在由黑山地震观测站主任Branislav Glavatovic教授和合作者开发的DASP采集系统上进行的。该系统是世界上PC机上地震信号采集时代的开端，已经运行了20多年。  
 数字地震信号采集由两个采集平台(地震ComP 3和SMARTGeoHub服务器)支持，用于全自动地震数据处理和地震信息分析，并同时应用数十个程序。黑山领土及其周边地区发生地震的自动系统反应平均时间为3分钟，通过因特网和短信处理和分发地震数据。地震发生后，利用天文台开发的专用软件进行离线数据处理，生成黑山境内、西巴尔干和世界地震的最终参数。利用计算机技术对地震资料进行数字化处理，从计算机技术在地震学中的应用开始，至今已有30多年的历史。在此期间，天文台开发了原始的地震资料分析和处理方法，综合收集了国产程序。

### 5.1.2 数据处理

地震资料处理[18]主要分为：作图进行时频分析、提高信噪比、提高分辨率以及各种信号预测，所以应用了很多的方法，比如在滤波中应用带通滤波,时变滤波,F一X滤波等，以及一些变换：如K一L变换,r一p变换,小波变换等。虽然办法以及算法较多，但是都基于不同目的来应用，所有的方法都是依据地震信号，地震信号本身存在两个问题：检波器输出信号失真和对地震波的不正确认识。

### 5.1.3 数据解释

简单定义，地震解释是从处理过的地震记录中某个深度推断地质的科学（和艺术）。尽管现代的多通道数据增加了可解释数据的数量和质量，但正确的解释仍然需要解释者从他或她的地质了解中选择数据允许的许多“有效”解释中最有可能的解释。

地震记录包含两个基本要素，以供口译人员研究。首先是来自地质表面的任何反射（或折射）到达的时间。到该表面的实际深度是上覆岩层的厚度和速度的函数。第二个是反射的形状，其中包括信号的强度，信号包含的频率以及频率在脉冲上的分布方式。该信息通常可用于支持有关所评估的地震反射器的岩性和流体含量的结论。

解释过程可以细分为三个相互关联的类别：结构，地层和岩性。结构地震解释的目的是根据观测到的到达时间的三维结构来创建地下结构图。地震层序地层解释将观测到的反射模式与沉积的周期性事件模型联系起来。目的是开发一个周期性的，遗传相关的地层的年代地层学框架。岩石学解释的目的是根据地震数据确定孔隙流体，孔隙率，裂缝强度，岩性等的变化。直接烃指示剂（DHI，HCI，亮点或变暗）是该岩性解释过程中使用的要素。

## 5.2 地震信号的奇异谱分析算法实现

首先我先找了一组长度N=3002的一维时间序列（地震信号）如图5.1所示，基于嵌入维度1<K<L/2,选择窗口长度为300进行处理该序列结果如图5.2和5.3所示

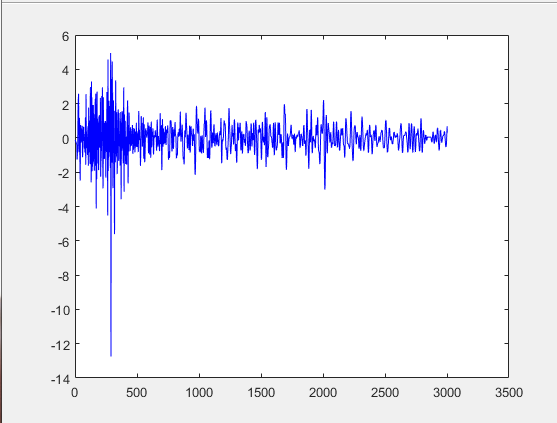


图5.1

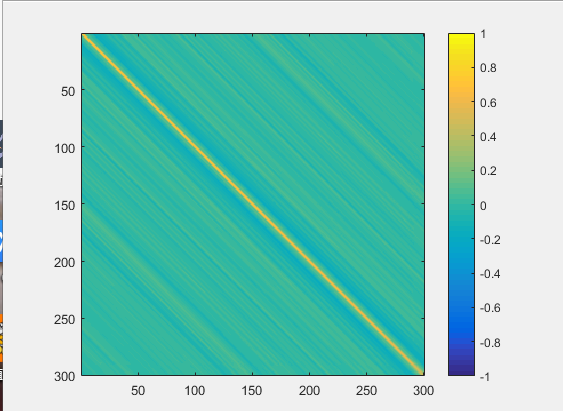


图5.2

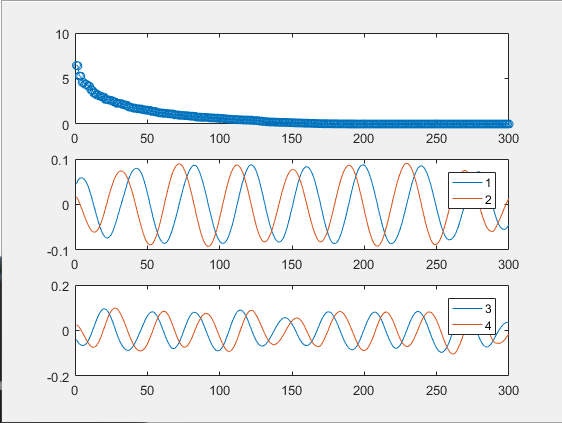


图5.3

之后计算主成分结果得出图5.4所示：

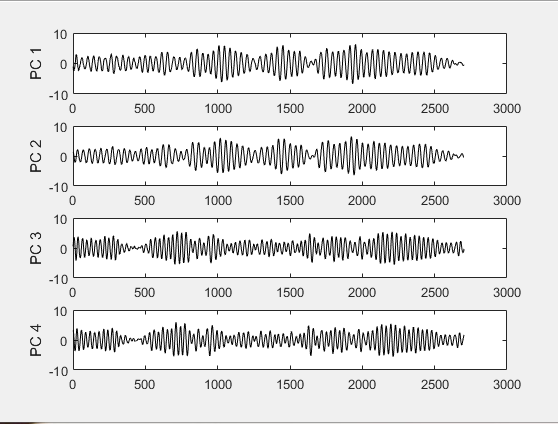


图5.4

之后再进行对角平均化，对角平均就是一种把任意矩阵转化为汉克尔矩阵的方法，转化得到的汉克尔矩阵中的元素对应原矩阵相应对角线上的平均值。这是因为在一个向量中用均值来表示该向量中的元素，得到的向量和原向量的最小均方误差是最小的。结果如图5.5所示：

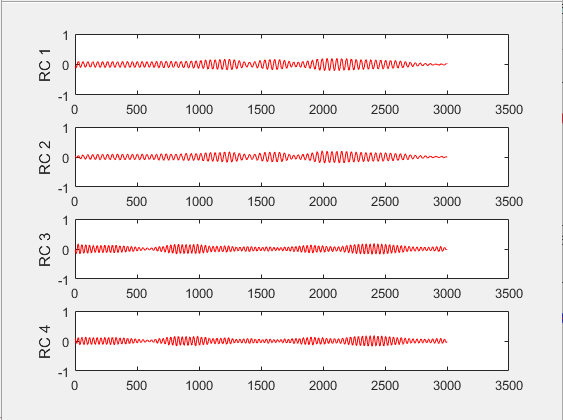


图5.5

最后将重构之后的信号与原信号进行比较，可以看到经过去噪之后的信号如5.6：

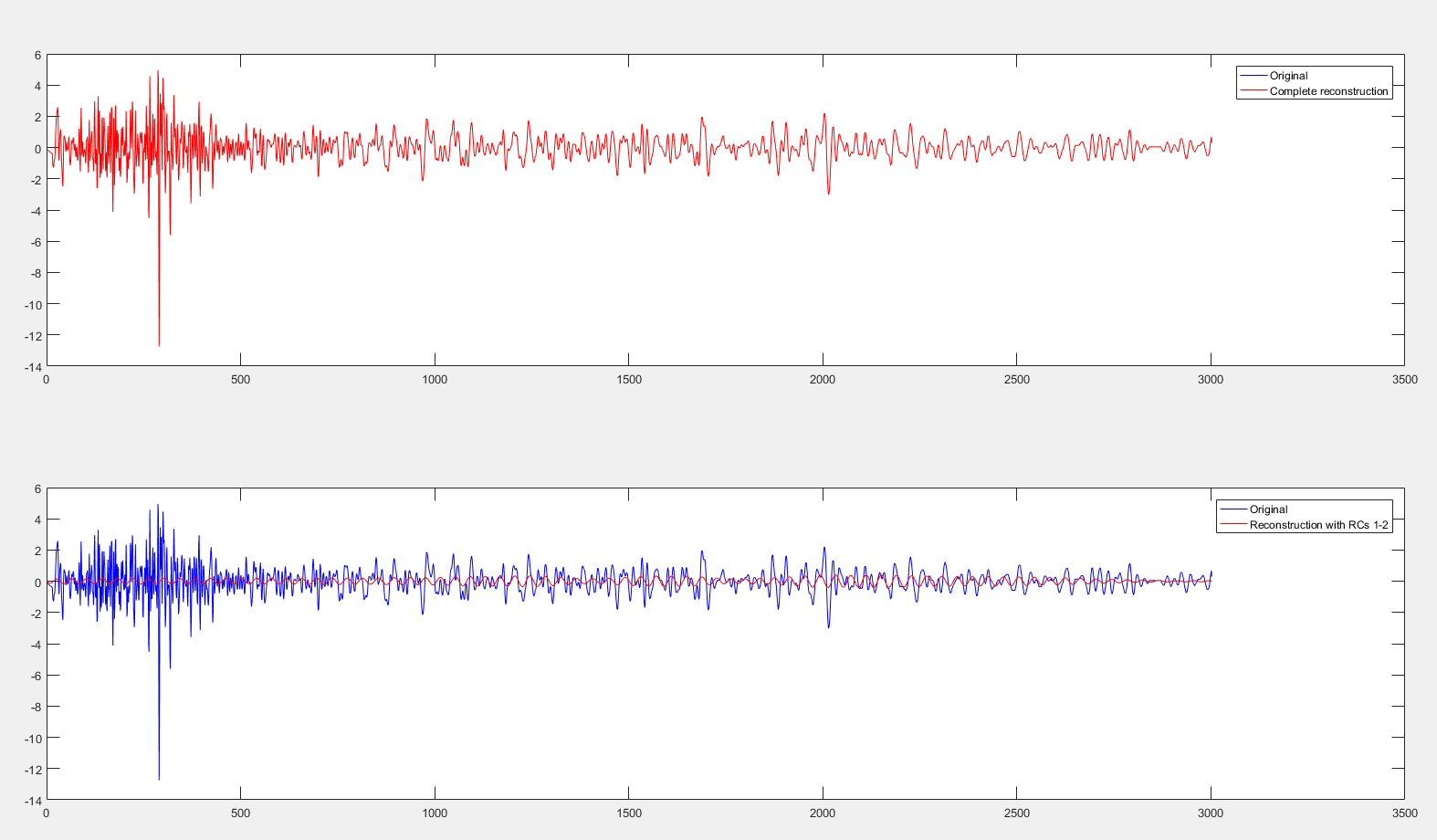


图5.6

# 第六章 结 论

奇异值分解是矩阵分析的重要方法之一，该方法在特征表示，聚类等领域有着重要的应用。近年来提出的奇异谱分解是一种非线性时间序列分析方法，该方法根据时间序列构建轨迹矩阵，并对轨迹矩阵进行分解，重构，从而提出代表原时间序列不同成分的信号，如长期趋势信号，周期信号，噪声信号等。

本论文主要在第一部分主要介绍了奇异谱分析和时频分析的研究现状和背景意义，通过对国内外研究的参考，了解了奇异谱分析和时频分析的发展概况以及方法原理。接下来第二部分详细介绍了SSA方法原理、一些理解原理的难点和实现算法进行数据处理所可能面对的类似于参数选择方面的问题。第三部分简单介绍了奇异谱分析中最重要的奇异值分解的一些算法和近年来一些学者对奇异谱算法的改进。

第四部分进行模拟实验验证了通过与原数据进行对比，验证了SSA方法的预测功能。第五部分先是对地震信号简述，有了一个大致的了解之后，将奇异谱分析方法应用在地震信号的去噪方面，取得了很好的效果。

# 参 考 文 献

[1] 邹红星,周小波,李衍达.时频分析:回溯与前瞻[J].电子学报,2000,28(9):78-84,4.

[2] 陈友良.时频分析在地震信号处理中的研究及应用[D].湖南:中南大学,2011.

[3] Vautard Robert,Yiou Pascal,Ghil Michael. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals[J]. North-Holland,1992,58(1-4):.

[4] 潘裕山,徐建军,张宇,袁帅,朱文清. 基于东亚区域再分析系统模拟2015年西北太平洋热带气旋[J]. 广东海洋大学学报.

[5] 王宁,粟晓玲. 石羊河流域水文气象要素变化的奇异谱分析[J]. 干旱区资源与环境,2013,27(12):180-185.

[6] 汪芸,郭生练,李响. 奇异谱分析在中长期径流预测中的应用研究[J]. 人民长江,2011,42(09):4-7+22.

[7] 黄嘉佑,黄茂怡. 汛期降水的奇异谱分析及预报试验[J]. 应用气象学报,2000,(S1):58-63.

[8] 刘元峰,赵玫. 基于奇异谱分析的混沌序列降噪[J]. 上海交通大学报,2003,(05):778-780.

[9] 黄建平,李闯,李国磊,黄金强,李振春,步长城,腾厚华. 基于奇异谱分析的联合去噪及规则化方法[J]. 地球物理学进展,2014,29(04):1666-1671.

[10] 贾永娜. 基于改进奇异谱分析法的地震数据插值[D].哈尔滨工业大学,2014.

[11] 王辉赞,张韧,刘巍,王桂华,金宝刚. 奇异谱迭代插补的改进算法及其在缺损数据恢复中的应用[J]. 应用数学和力学,2008,(10):1227-1236.

[12] 毛向东,袁惠群,孙华刚. 基于经验模式分解与奇异谱分析的微弱信号提取[J]. 制造业自动化,2014,36(21):61-64.

[13] 余斌,杨少敏.基于改进Cao算法的奇异谱分析法及其在北斗多路径去噪中的应用[J].大地测量与地球动力学,2019,39(01):25-30.

[14] 林佩如. 基于经验模式分解的奇异谱分析分组理论及应用研究[D].广东工业大学,2018.

[15] 赵佳佳. 奇异谱分析在地形变数据处理中的应用研究[D].中国地震局地震研究所,2017.

[16] 赵佳佳,陈志遥,张燕,王嘉沛.奇异谱分析在倾斜应变数据处理中的应用研究[J].大地测量与地球动力学,2017,37(05):541-545.

[17] 周舟. 基于多道奇异谱分析方法的地震数据重建[D].中国地质大学（北京）,2014.

[18]Groth, Andreas,Ghil, Michael. Monte Carlo Singular Spectrum Analysis (SSA) Revisited: Detecting Oscillator Clusters in Multivariate Datasets[J]. Journal of Climate,2015,28(19).

[18] 吴顺和. 地震资料处理和地震信号[C]. 中国地球物理学会.2000年中国地球物理学会年刊——中国地球物理学会第十六届年会论文集.中国地球物理学会:中国地球物理学会,2000:41.

[19]Schoellhamer D H . Singular spectrum analysis for time series with missing data[J]. Geophysical Research Letters, 2001, 28(16):3187-3190.

[20]Hassani H , Zhigljavsky A . SINGULAR SPECTRUM ANALYSIS:METHODOLOGY AND APPLICATION TO ECONOMICS DATA[J]. Journal of Systems Science & Complexity, 2009(03):32-54.

# 附录（程序代码）：

1.

import numpy as np

import pandas as pd

from numpy import matrix as m

from pandas import DataFrame as df

from scipy import linalg

try:

import seaborn

except:

pass

from matplotlib.pylab import rcParams

rcParams['figure.figsize'] = 11, 4

class mySSA(object):

'''Singular Spectrum Analysis object'''

def \_\_init\_\_(self, time\_series):

self.ts = pd.DataFrame(time\_series)

self.ts\_name = self.ts.columns.tolist()[0]

if self.ts\_name==0:

self.ts\_name = 'ts'

self.ts\_v = self.ts.values

self.ts\_N = self.ts.shape[0]

self.freq = self.ts.index.inferred\_freq

@staticmethod

def \_printer(name, \*args):

'''Helper function to print messages neatly'''

print('-'\*40)

print(name+':')

for msg in args:

print(msg)

@staticmethod

def \_dot(x,y):

'''Alternative formulation of dot product to allow missing values in arrays/matrices'''

pass

@staticmethod

def get\_contributions(X=None, s=None, plot=True):

'''Calculate the relative contribution of each of the singular values'''

lambdas = np.power(s,2)

frob\_norm = np.linalg.norm(X)

ret = df(lambdas/(frob\_norm\*\*2), columns=['Contribution'])

ret['Contribution'] = ret.Contribution.round(4)

if plot:

ax = ret[ret.Contribution!=0].plot.bar(legend=False)

ax.set\_xlabel("Lambda\_i")

ax.set\_title('Non-zero contributions of Lambda\_i')

vals = ax.get\_yticks()

ax.set\_yticklabels(['{:3.2f}%'.format(x\*100) for x in vals])

return ax

return ret[ret.Contribution>0]

@staticmethod

def diagonal\_averaging(hankel\_matrix):

'''Performs anti-diagonal averaging from given hankel matrix

Returns: Pandas DataFrame object containing the reconstructed series'''

mat = m(hankel\_matrix)

L, K = mat.shape

L\_star, K\_star = min(L,K), max(L,K)

new = np.zeros((L,K))

if L > K:

mat = mat.T

ret = []

#Diagonal Averaging

for k in range(1-K\_star, L\_star):

mask = np.eye(K\_star, k=k, dtype='bool')[::-1][:L\_star,:]

mask\_n = sum(sum(mask))

ma = np.ma.masked\_array(mat.A, mask=1-mask)

ret+=[ma.sum()/mask\_n]

return df(ret).rename(columns={0:'Reconstruction'})

def view\_time\_series(self):

'''Plot the time series'''

self.ts.plot(title='Original Time Series')

def embed(self, embedding\_dimension=None, suspected\_frequency=None, verbose=False, return\_df=False):

'''Embed the time series with embedding\_dimension window size.

Optional: suspected\_frequency changes embedding\_dimension such that it is divisible by suspected frequency'''

if not embedding\_dimension:

self.embedding\_dimension = self.ts\_N//2

else:

self.embedding\_dimension = embedding\_dimension

if suspected\_frequency:

self.suspected\_frequency = suspected\_frequency

self.embedding\_dimension = (self.embedding\_dimension//self.suspected\_frequency)\*self.suspected\_frequency

self.K = self.ts\_N-self.embedding\_dimension+1

self.X = m(linalg.hankel(self.ts, np.zeros(self.embedding\_dimension))).T[:,:self.K]

self.X\_df = df(self.X)

self.X\_complete = self.X\_df.dropna(axis=1)

self.X\_com = m(self.X\_complete.values)

self.X\_missing = self.X\_df.drop(self.X\_complete.columns, axis=1)

self.X\_miss = m(self.X\_missing.values)

self.trajectory\_dimentions = self.X\_df.shape

self.complete\_dimensions = self.X\_complete.shape

self.missing\_dimensions = self.X\_missing.shape

self.no\_missing = self.missing\_dimensions[1]==0

if verbose:

msg1 = 'Embedding dimension\t: {}\nTrajectory dimensions\t: {}'

msg2 = 'Complete dimension\t: {}\nMissing dimension \t: {}'

msg1 = msg1.format(self.embedding\_dimension, self.trajectory\_dimentions)

msg2 = msg2.format(self.complete\_dimensions, self.missing\_dimensions)

self.\_printer('EMBEDDING SUMMARY', msg1, msg2)

if return\_df:

return self.X\_df

def decompose(self, verbose=False):

'''Perform the Singular Value Decomposition and identify the rank of the embedding subspace

Characteristic of projection: the proportion of variance captured in the subspace'''

X = self.X\_com

self.S = X\*X.T

self.U, self.s, self.V = linalg.svd(self.S)

self.U, self.s, self.V = m(self.U), np.sqrt(self.s), m(self.V)

self.d = np.linalg.matrix\_rank(X)

Vs, Xs, Ys, Zs = {}, {}, {}, {}

for i in range(self.d):

Zs[i] = self.s[i]\*self.V[:,i]

Vs[i] = X.T\*(self.U[:,i]/self.s[i])

Ys[i] = self.s[i]\*self.U[:,i]

Xs[i] = Ys[i]\*(m(Vs[i]).T)

self.Vs, self.Xs = Vs, Xs

self.s\_contributions = self.get\_contributions(X, self.s, False)

self.r = len(self.s\_contributions[self.s\_contributions>0])

self.r\_characteristic = round((self.s[:self.r]\*\*2).sum()/(self.s\*\*2).sum(),4)

self.orthonormal\_base = {i:self.U[:,i] for i in range(self.r)}

if verbose:

msg1 = 'Rank of trajectory\t\t: {}\nDimension of projection space\t: {}'

msg1 = msg1.format(self.d, self.r)

msg2 = 'Characteristic of projection\t: {}'.format(self.r\_characteristic)

self.\_printer('DECOMPOSITION SUMMARY', msg1, msg2)

def view\_s\_contributions(self, adjust\_scale=False, cumulative=False, return\_df=False):

'''View the contribution to variance of each singular value and its corresponding signal'''

contribs = self.s\_contributions.copy()

contribs = contribs[contribs.Contribution!=0]

if cumulative:

contribs['Contribution'] = contribs.Contribution.cumsum()

if adjust\_scale:

contribs = (1/contribs).max()\*1.1-(1/contribs)

ax = contribs.plot.bar(legend=False)

ax.set\_xlabel("Singular\_i")

ax.set\_title('Non-zero{} contribution of Singular\_i {}'.\

format(' cumulative' if cumulative else '', '(scaled)' if adjust\_scale else ''))

if adjust\_scale:

ax.axes.get\_yaxis().set\_visible(False)

vals = ax.get\_yticks()

ax.set\_yticklabels(['{:3.0f}%'.format(x\*100) for x in vals])

if return\_df:

return contribs

@classmethod

def view\_reconstruction(cls, \*hankel, names=None, return\_df=False, plot=True, symmetric\_plots=False):

'''Visualise the reconstruction of the hankel matrix/matrices passed to \*hankel'''

hankel\_mat = None

for han in hankel:

if isinstance(hankel\_mat,m):

hankel\_mat = hankel\_mat + han

else:

hankel\_mat = han.copy()

hankel\_full = cls.diagonal\_averaging(hankel\_mat)

title = 'Reconstruction of signal'

if names or names==0:

title += ' associated with singular value{}: {}'

title = title.format('' if len(str(names))==1 else 's', names)

if plot:

ax = hankel\_full.plot(legend=False, title=title)

if symmetric\_plots:

velocity = hankel\_full.abs().max()[0]

ax.set\_ylim(bottom=-velocity, top=velocity)

if return\_df:

return hankel\_full

def \_forecast\_prep(self, singular\_values=None):

self.X\_com\_hat = np.zeros(self.complete\_dimensions)

self.verticality\_coefficient = 0

self.forecast\_orthonormal\_base = {}

if singular\_values:

try:

for i in singular\_values:

self.forecast\_orthonormal\_base[i] = self.orthonormal\_base[i]

except:

if singular\_values==0:

self.forecast\_orthonormal\_base[0] = self.orthonormal\_base[0]

else:

raise('Please pass in a list/array of singular value indices to use for forecast')

else:

self.forecast\_orthonormal\_base = self.orthonormal\_base

self.R = np.zeros(self.forecast\_orthonormal\_base[0].shape)[:-1]

for Pi in self.forecast\_orthonormal\_base.values():

self.X\_com\_hat += Pi\*Pi.T\*self.X\_com

pi = np.ravel(Pi)[-1]

self.verticality\_coefficient += pi\*\*2

self.R += pi\*Pi[:-1]

self.R = m(self.R/(1-self.verticality\_coefficient))

self.X\_com\_tilde = self.diagonal\_averaging(self.X\_com\_hat)

def forecast\_recurrent(self, steps\_ahead=12, singular\_values=None, plot=False, return\_df=False, \*\*plotargs):

'''Forecast from last point of original time series up to steps\_ahead using recurrent methodology

This method also fills any missing data from the original time series.'''

try:

self.X\_com\_hat

except(AttributeError):

self.\_forecast\_prep(singular\_values)

self.ts\_forecast = np.array(self.ts\_v[0])

for i in range(1, self.ts\_N+steps\_ahead):

try:

if np.isnan(self.ts\_v[i]):

x = self.R.T\*m(self.ts\_forecast[max(0,i-self.R.shape[0]): i]).T

self.ts\_forecast = np.append(self.ts\_forecast,x[0])

else:

self.ts\_forecast = np.append(self.ts\_forecast,self.ts\_v[i])

except(IndexError):

x = self.R.T\*m(self.ts\_forecast[i-self.R.shape[0]: i]).T

self.ts\_forecast = np.append(self.ts\_forecast, x[0])

self.forecast\_N = i+1

new\_index = pd.date\_range(start=self.ts.index.min(),periods=self.forecast\_N, freq=self.freq)

forecast\_df = df(self.ts\_forecast, columns=['Forecast'], index=new\_index)

forecast\_df['Original'] = np.append(self.ts\_v, [np.nan]\*steps\_ahead)

if plot:

forecast\_df.plot(title='Forecasted vs. original time series', \*\*plotargs)

if return\_df:

return forecast\_df

if \_\_name\_\_=='\_\_main\_\_':

from mySSA import mySSA

from pandas import DataFrame as df

import pandas as pd

import numpy as np

from matplotlib.pylab import rcParams

# Construct the data with gaps

ts = pd.read\_csv('AirPassengers.csv', parse\_dates=True, index\_col='Month')

ts\_ = ts.copy()

ts\_.ix[67:79] = np.nan

ts\_ = ts\_.set\_value('1961-12-01','#Passengers', np.nan).asfreq('MS')

ssa = mySSA(ts\_)

# Plot original series for reference

ssa.view\_time\_series()

ssa.embed(embedding\_dimension=36, suspected\_frequency=12, verbose=True)

ssa.decompose(True)

ssa.view\_s\_contributions(adjust\_scale=True)

# Component Signals

components = [i for i in range(13)]

rcParams['figure.figsize'] = 11, 2

for i in range(5):

ssa.view\_reconstruction(ssa.Xs[i], names=i, symmetric\_plots=i!=0)

rcParams['figure.figsize'] = 11, 4

# RECONSTRUCTION

ssa.view\_reconstruction(\*[ssa.Xs[i] for i in components], names=components)

# FORECASTING

ssa.forecast\_recurrent(steps\_ahead=48, plot=True)

2.M=300;

X=csvread('myArray.csv');

N=3002;

t=(1:N)';

figure(1);

set(gcf,'name','Time series X');

clf;

plot(t,X,'b-');

covX = xcorr(X,M-1,'unbiased');

Ctoep=toeplitz(covX(M:end));

figure(2);

set(gcf,'name','Covariance matrix');

clf;

imagesc(Ctoep);

axis square

set(gca,'clim',[-1 1]);

colorbar

Y=zeros(N-M+1,M);

for m=1:M

Y(:,m) = X((1:N-M+1)+m-1);

end;

Cemb=Y'\*Y / (N-M+1);

figure(2);

set(gcf,'name','Covariance matrix');

clf;

imagesc(Cemb);

axis square

set(gca,'clim',[-1 1]);

colorbar

C=Cemb;

[RHO,LAMBDA] = eig(C);

LAMBDA = diag(LAMBDA); % extract the diagonal elements

[LAMBDA,ind]=sort(LAMBDA,'descend'); % sort eigenvalues

RHO = RHO(:,ind); % and eigenvectors

figure(3);

set(gcf,'name','Eigenvectors RHO and eigenvalues LAMBDA')

clf;

subplot(3,1,1);

plot(LAMBDA,'o-');

subplot(3,1,2);

plot(RHO(:,1:2), '-');

legend('1', '2');

subplot(3,1,3);

plot(RHO(:,3:4), '-');

legend('3', '4');

PC = Y\*RHO;

figure(4);

set(gcf,'name','Principal components PCs')

clf;

for m=1:4

subplot(4,1,m);

plot(t(1:N-M+1),PC(:,m),'k-');

ylabel(sprintf('PC %d',m));

ylim([-10 10]);

end;

RC=zeros(N,M);

for m=1:M

buf=PC(:,m)\*RHO(:,m)'; % invert projection

buf=buf(end:-1:1,:);

for n=1:N % anti-diagonal averaging

RC(n,m)=mean( diag(buf,-(N-M+1)+n) );

end

end;

figure(5);

set(gcf,'name','Reconstructed components RCs')

clf;

for m=1:4

subplot(4,1,m);

plot(t,RC(:,m),'r-');

ylabel(sprintf('RC %d',m));

ylim([-1 1]);

end;

figure(6);

set(gcf,'name','Original time series X and reconstruction RC')

clf;

subplot(2,1,1)

plot(t,X,'b-',t,sum(RC(:,:),2),'r-');

legend('Original','Complete reconstruction');

subplot(2,1,2)

plot(t,X,'b','LineWidth',2);

plot(t,X,'b-',t,sum(RC(:,1:2),2),'r-');

legend('Original','Reconstruction with RCs 1-2');figure(1);

set(gcf,'name','Time series X');

clf;

plot(t,X,'b-');

covX = xcorr(X,M-1,'unbiased');

Ctoep=toeplitz(covX(M:end));

figure(2);

set(gcf,'name','Covariance matrix');

clf;

imagesc(Ctoep);

axis square

set(gca,'clim',[-1 1]);

colorbar

Y=zeros(N-M+1,M);

for m=1:M

Y(:,m) = X((1:N-M+1)+m-1);

end;

Cemb=Y'\*Y / (N-M+1);

figure(2);

set(gcf,'name','Covariance matrix');

clf;

imagesc(Cemb);

axis square

set(gca,'clim',[-1 1]);

colorbar

C=Cemb;

[RHO,LAMBDA] = eig(C);

LAMBDA = diag(LAMBDA); % extract the diagonal elements

[LAMBDA,ind]=sort(LAMBDA,'descend'); % sort eigenvalues

RHO = RHO(:,ind); % and eigenvectors

figure(3);

set(gcf,'name','Eigenvectors RHO and eigenvalues LAMBDA')

clf;

subplot(3,1,1);

plot(LAMBDA,'o-');

subplot(3,1,2);

plot(RHO(:,1:2), '-');

legend('1', '2');

subplot(3,1,3);

plot(RHO(:,3:4), '-');

legend('3', '4');

PC = Y\*RHO;

figure(4);

set(gcf,'name','Principal components PCs')

clf;

for m=1:4

subplot(4,1,m);

plot(t(1:N-M+1),PC(:,m),'k-');

ylabel(sprintf('PC %d',m));

ylim([-10 10]);

end;

RC=zeros(N,M);

for m=1:M

buf=PC(:,m)\*RHO(:,m)'; % invert projection

buf=buf(end:-1:1,:);

for n=1:N % anti-diagonal averaging

RC(n,m)=mean( diag(buf,-(N-M+1)+n) );

end

end;

figure(5);

set(gcf,'name','Reconstructed components RCs')

clf;

for m=1:4

subplot(4,1,m);

plot(t,RC(:,m),'r-');

ylabel(sprintf('RC %d',m));

ylim([-1 1]);

end;

figure(6);

set(gcf,'name','Original time series X and reconstruction RC')

clf;

subplot(2,1,1)

plot(t,X,'b-',t,sum(RC(:,:),2),'r-');

legend('Original','Complete reconstruction');

subplot(2,1,2)

plot(t,X,'b','LineWidth',2);

plot(t,X,'b-',t,sum(RC(:,1:2),2),'r-');

legend('Original','Reconstruction with RCs 1-2');

# 致 谢

在本论文的撰写过程中武国宁老师作为我的导师，在一开始为我提供解决本论文的思路，积极指导并且帮助我克服论文中遇到的困难，使得我掌握了其理论依据和研究方法。在整个过程中始终给予我细心的指导。

大学四年时光已经匆匆溜走，转眼已经到了毕业的尾声，回想在大学学习的四年我的心中充满了无限的留恋之情，学校给予我的不仅仅是课本上的理论知识，更重要的是培育了我在阅读、实践中的逻辑思维方式、表达能力和全局视野。很庆幸可以在中国石油大学（北京）遇到这么多和蔼可亲的老师，我十分珍惜学校给与的这种浓厚的学术氛围和同学们之间的互帮互助的娱乐时光，在此我向学校的老师表示衷心的感谢，谢谢老师们的辛勤栽培。

每一个结束也是一个开始，在未来的日子里我会以“厚积薄发，开物成务”这一校训作为人生目标。