



单位代码 10635

学 号 112015333002140

# 西南大学

## 硕士学位论文

基于混合核及最大相关熵的自适应滤波算法

论文作者：孙启堂

指导教师：王世元 副教授

学科专业：信号与信息处理

研究方向：自适应信号处理

提交论文日期：2018 年 4 月 07 日

论文答辩日期：2018 年 6 月 02 日

学位授予单位：西南大学

中 国 • 重 庆

2018 年 6 月

# **Adaptive Filtering Algorithms under Mixed Kernels and Maximum Correntropy**



A Thesis Submitted to Southwest University in  
Partial Fulfillment of the Requirement for the  
Master's Degree of Engineering

**By**  
**Sun Qitang**

**Supervised by Associate Prof. Wang Shiyuan**

**Specialty: Signal and Information Processing**

College of Electronic and Information Engineering of  
Southwest University, Chongqing, China

June, 2018

# 独创性声明

学位论文题目： 基于混合核及最大相关熵的自适应滤波算法

本人提交的学位论文是在导师指导下进行的研究工作及取得的研究成果。论文中引用他人已经发表或出版过的研究成果，文中已加了特别标注。对本研究及学位论文撰写曾做出贡献的老师、朋友、同仁在文中作了明确说明并表示衷心感谢。

学位论文作者： 孙启堂 签字日期： 2018 年 6 月 6 日

## 学位论文版权使用授权书

本学位论文作者完全了解西南大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权西南大学研究生院（筹）可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后适用本授权书，本论文：☒不保密，☐保密期限至      年      月止）。

学位论文作者签名： 孙启堂 导师签名： 王世

签字日期： 2018 年 6 月 6 日      签字日期： 2018 年 6 月 6 日

## 目 录

摘 要.....	I
Abstract.....	III
第一章 引言.....	1
1.1 选题背景及其意义.....	1
1.2 国内外研究现状.....	3
1.3 论文结构.....	5
第二章 自适应滤波算法原理.....	7
2.1 经典的线性自适应滤波算法.....	7
2.1.1 最小均方算法.....	7
2.1.2 仿射投影算法.....	7
2.1.3 递归最小二乘法.....	8
2.2 核方法与核自适应滤波器.....	9
2.3 经典的核自适应滤波算法.....	11
2.3.1 核最小均方算法.....	11
2.3.2 核仿射投影算法.....	12
2.3.3 核递归最小二乘法.....	13
2.4 本章小结.....	15
第三章 混合核最小均方算法.....	17
3.1 核函数及混合核函数.....	17
3.2 混合核最小均方算法.....	18
3.3 仿真实验.....	20
3.3.1 MG 时间序列预测模型及仿真.....	21
3.3.2 非线性回归模型及仿真.....	22
3.4 本章小结.....	24
第四章 基于最大相关熵的仿射投影算法.....	25
4.1 最大相关熵.....	25
4.1.1 最大相关熵准则.....	25
4.1.2 广义最大相关熵准则.....	26
4.1.3 广义最大相关熵的基本性质.....	28
4.1.4 广义最大相关熵算法.....	28
4.2 基于最大相关熵的仿射投影算法.....	29
4.2.1 基于广义最大相关熵的非核仿射投影算法.....	29
4.2.2 基于广义最大相关熵的核仿射投影算法.....	31

4.3 仿真实验 .....	33
4.4 本章小结 .....	39
第五章 总结与展望 .....	41
5.1 总结 .....	41
5.2 展望 .....	41
参考文献 .....	41
致 谢 .....	47
攻读硕士期间已发表的学术论文 .....	49
攻读硕士期间参加的科研项目 .....	49

# 基于混合核及最大相关熵的自适应滤波算法

学科专业：信号与信息处理

研究方向：自适应信号处理

指导教师：王世元 副教授

作者：孙启堂

## 摘 要

自适应滤波器 (AF, adaptive filter) 作为一种十分有效的随机信号处理工具, 已被应用于信息科学的各个领域。自适应滤波算法的研究是信号处理领域较为活跃的研究课题。自适应滤波算法可以归结为两大类, 即线性自适应滤波算法和非线性自适应滤波算法。其中经典的自适应滤波算法有: 最小均方算法 (LMS, least mean square algorithm)、核最小均方算法 (KLMS, kernel least mean square algorithm)、核仿射投影算法(KAPA, kernel affine projection algorithm)等。在核自适应滤波算法中, 核函数的选择是极其重要的。然而可供选择的核函数有很多种, 比如: 高斯核(GK, Gaussian kernel)函数、拉普拉斯核(LK, Laplace kernel)函数、指数核函数、多项式核函数, 其中最常用的是高斯核函数。在本论文中, 我们提出一种混合核函数, 并基于此混合核函数提出了混合核最小均方算法 (KLMS-MK, kernel least mean square algorithm with mixed kernel)。

大多数现有的自适应滤波算法都是在均方误差 (MSE, mean squares error) 准则下研究的, 这是高斯噪声下理想的最优化准则。然而, 这个假设未能模拟在实践中发现的非高斯噪声的行为。作为核空间中一种鲁棒的非线性相似性度量, 相关熵在机器学习和信号处理领域受到越来越多的关注。尤其是最大相关熵准则 (MCC, maximum correntropy criterion) 已被成功地应用于随机信号处理中。相关熵中默认的核函数是高斯核函数, 当然, 这并不是最好的选择。广义最大相关熵准则 (GMCC, generalized maximum correntropy criterion) 作为 MCC 的推广已被应用于自适应滤波算法中。本论文主要从自适应滤波算法的代价函数角度进行研究, 其工作内容主要为以下两个方面:

(1) 提出一种混合核函数, 用于替代 KLMS 中的高斯核函数。本文运用该混合核函数, 在 KLMS 中进行理论推导, 从而提出了 KLMS-MK。同时将凸组合的方法应用到 KLMS 的核函数上, 提出的 KLMS-MK 具有高斯核和拉普拉斯核两者的优点。在 KLMS-MK 中, 凸组合的混合参数是由梯度下降法自动调整的。同时, 我们还证明了混合参数的收敛性。因此, 在 KLMS-MK 中, 将传统的单核函数拓展到多核函数; 进一步改善了收敛速度和 MSE。

(2) 提出了基于最大相关熵(MC, maximum correntropy)的仿射投影算法。首先, 介绍最大相关熵的相关理论, 包括: 定义、最大相关熵的相关性质以及与其相关的自适应滤波算法。其次, 研究了基于最大相关熵的仿射投影算法, 推导得到了其权重更新方式。验证了提出算法的有效性。

关键词: 自适应算法; KLMS-MK; 最大相关熵; 广义最大相关熵; 基于广义最大相关熵的核仿射投影算法

---

# Adaptive Filtering Algorithms under Mixed Kernels and Maximum Correntropy

---

Major: Signal and Information Processing

Direction: Adaptive Signal Processing

Supervisor: Associate Prof. Wang Shiyuan

Author: Sun Qitang

## Abstract

Adaptive filter (AF), as an effective random signal processing tool, has been used in various fields of information society. The research of adaptive filter algorithm is an active topic in the field of signal processing. Adaptive filtering algorithms can be categorized into two major categories, including linear adaptive filtering algorithms and nonlinear adaptive filtering algorithms. The classical adaptive filter algorithms include least mean square algorithm (LMS), kernel least mean square algorithm (KLMS), kernel affine projection algorithm (KAPA), etc. Kernel adaptive filter (KAF) is a typical representative of nonlinear AF. In kernel adaptive filtering algorithms, the choice of kernel function is extremely important. There are many kinds of available kernel functions, such as Gaussian kernel (GK) function, Laplace kernel (LK) function, exponential kernel function and polynomial kernel function. The most widely used kernel function is Gaussian kernel function. In this paper, we propose a mixed kernel function and a mixed kernel least mean square algorithm (KLMS-MK) based on the mixed kernel function.

Most existing adaptive filtering algorithms are derived under the mean square error criterion (MSE), which is optimal in the presence of Gaussian noise. However, this assumption failed to simulate the behavior of non-Gaussian noise in practice. As a robust nonlinear similarity measure in kernel space, the correntropy has received more and more attention in the field of machine learning and signal processing. In particular, the maximum correntropy criterion (MCC) has been successfully applied to signal processing. The default kernel function in the correntropy is the Gaussian kernel function. Of course, this is not the best choice. The generalized maximum correntropy criterion (GMCC) has been applied to adaptive filter algorithms as a generalization of MCC. This paper focuses on the cost function of adaptive filtering algorithm, and the



main work mainly includes the following two aspects:

(1) We propose a mixed kernel function to replace the Gaussian kernel function in KLMS. We use the mixed kernel function to perform theoretical derivation in KLMS and propose KLMS-MK. The convex combination method is applied to the KLMS kernel function. KLMS-MK has the advantages of both Gaussian kernel and Laplace kernel. In KLMS-MK, the convex combination of mixing parameters is automatically adjusted by the stochastic gradient descent method. At the same time, we also prove the convergence of the mixed parameter. In KLMS-MK, KLMS-MK extends traditional single-kernel functions to multi-kernel functions, which leads to the improvement of convergence rate and MSE.

(2) We propose affine projection algorithms based on maximum correntropy. First of all, the related theory of maximum correlation entropy is introduced, including: definition, correlation properties of maximum correlation entropy and related adaptive filtering algorithms. Secondly, we mainly study APA and KAPA based on the maximum correntropy, and derive their weight updating methods. Simulation results validate the efficiency of the proposed algorithm.

**Keywords: Adaptive Algorithm; KLMS-MK; MC; GMC; KAPA based on GMC**

## 第一章 引言

### 1.1 选题背景及其意义

当今社会是信息化社会,信息科学技术<sup>[1,2]</sup>已经渗入到社会生活的各个方面,并深刻地影响着全球人们的工作和生活。信息的表现形式是多种多样的,其中数据是信息重要的表现形式。数据已经渗入社会的各个方面,化身为重要的生产因素。但社会中的数据是复杂多样的,并且新的数据在不断地产生。数据在社会中有着重要的地位,因此对于数据的研究是十分有必要的。数据研究的目的是:获得数据内在规律,通过获得的内在规律指导人们的工作和生活。然而,由于数据种类的多样性以及同种数据的复杂程度,人们自身研究数据的难度非常巨大。随着科学技术的发展和人们认识程度的不断提高,人们对于数据的研究在不断的深入。计算机技术<sup>[3]</sup>的发展促使机器学习<sup>[4]</sup>的概念被提出;机器学习的目标是:为数据建立一种衡量其内在规律的模型。在机器学习逐渐被人所知的同时,许多的机器学习算法<sup>[5]</sup>被提出。随着电信号的发展,人们对于信号处理的研究日益深入,形成了较为完整的信号处理过程。AF<sup>[6,7]</sup>是信号处理的重要环节,它与机器学习有着密不可分的联系。AF能够根据所处环境的不同,使用自适应算法来改变滤波器的参数和结构。通常,自适应滤波算法不改变AF的结构;而AF的系数是通过自适应滤波算法来更新的时变系数。AF可以归结为两大类,即线性AF和非线性AF。线性AF结构简单,但是其滤波能力有限。非线性AF是在线性AF的基础上发展起来的;相比于线性AF,非线性AF滤波性能得到了显著的提高。AF是一种自设计的滤波器,它能够根据自动地调整自身的参数。

线性自适应滤波算法<sup>[8]</sup>与非线性自适应算法<sup>[9]</sup>存在着很多的联系,研究者们不断地改善自适应滤波算法。一般来说,系统的输入、输出之间满足一定的关系,但是人们无法得出它们之间准确的关系。因此,人们对输入、输出关系进行一定的假设,希望通过测试输入与输出得到一种一定误差范围内的函数关系。简单的线性AF是基于某种线性关系假设进行研究的。简单线性自适应滤波算法中典型的算法包括:LMS、APA和RLS<sup>[10-12]</sup>。当然还存在着其他稍微复杂一些的线性自适应滤波算法。然而,在现实中许多问题都是非线性的,超出了线性假设的研究范围。因此,我们需要输入、输出之间的非线性假设模型来描述这些非线性问题,这促使了非线性AF的产生。早期的非线性AF主要为:Hammerstein<sup>[13]</sup>和Wiener模型<sup>[14]</sup>,它们是通过线性AF的优化所获得。优化的主要方法就是在线性AF的基础上串联一个非线性部分,从而获得非线性。这种模型具有很强的局限性,建模能力也十分有限。随着研究的进行,出现了Volterra模型<sup>[15]</sup>。这种模

型相比于早期的非线性 AF 模型具有较好的建模能力, 但是其计算复杂度增加了很多。Volterra 模型中存在着线性组合器, 这限制了它处理非线性问题的能力。针对这一问题, 研究者们用径向基函数网络<sup>[16]</sup>、循环神经网络<sup>[17]</sup>等替换线性组合器, 以期获得较好的处理非线性问题的能力。替换线性组合器的方法提高了解决非线性问题的能力。但是这些方法应用范围有限, 这归咎于它们是非凸优化<sup>[18]</sup>的。

由于 KAF<sup>[19]</sup>在实际中的有效应用, 研究者们对 KAF 的研究愈加深入。KAF 是一种凸优化模型<sup>[20]</sup>, 避免了非凸优化的弊端, 不会出现局部极小值。KAF 通过运用通用逼近能力好的核函数, 解决了系统输入、输出非线性问题, 同时具有很好的滤波性能。核自适应滤波算法<sup>[21]</sup>作为非线性自适应滤波算法中的典型代表, 它使得非线性 AF 理论发展到了一种新的高度。研究者们坚持不懈的追求收敛速度快, 计算复杂度低, 稳定性好的自适应滤波算法。

在 AF 应用于越来越多的领域同时, 研究者们想尽各种各样的方法不断的改善 AF 的性能, 拓展 AF 的内容, 这使得 AF 理论更加的完善。因为熵<sup>[22]</sup>成功的应用于很多领域中, 从而使人们对熵的研究不断的深入。最大熵原理<sup>[23]</sup>是一种信息论方法, 在很多的领域中已被成功应用。在机器学习和信号处理领域中相关熵已受到越来越多的关注。现在, 广义最大相关熵已被应用于自适应滤波算法中, 研究者们不断的对其进行改进和完善。

选择合适的代价函数(通常是误差信号的统计测量)是自适应滤波理论和应用中的关键问题。MSE 被广泛用作代价函数, 因为它具有吸引人的特性, 如平滑性, 凸性, 数学易处理性, 低计算负担和高斯假设下的最优性。在此准则下开发了众所周知的 LMS 及其变体, 如归一化 LMS (NLMS) 和变步长 LMS (VSSLMS)。如果信号是高斯分布的, 则 MSE 是可取的。然而, 在非高斯情况下, 其性能可能会大大降低, 并且在这些情况下, 非平方成本通常会优于 MSE。

一般而言, 有两种类型的非高斯分布: 轻尾(例如统一, 二进制等)和重尾(例如拉普拉斯)分布。当期望的信号受轻尾非高斯噪声干扰时, 通常更希望误差的高阶统计量度量。一个典型的例子是最小均值四(LMF, least mean fourth)族算法, 它使用误差的平均偶次幂作为代价函数。与 LMS 算法相比, LMF 可以实现更快的收敛速度和更低的稳态均方差(MSD, steady-state mean square deviation), 尤其是轻尾噪声。然而, LMF 算法的一个缺点是不能保证稳定性, 这取决于输入和噪声功率以及权值的初始值。更普遍的一类算法是最小均值 p-功率(LMP, leastmean p-power)族算法, 其采用误差的 p 阶绝对矩作为自适应代价。

当期望信号受到重尾脉冲噪声(可能导致大离群值)的干扰时, 误差的低阶统计测量通常更稳健(即对脉冲干扰较不敏感)。例如, 使用误差的平均绝对值作

为代价函数的符号算法 (SA, sign algorithm) 对于大噪声的存在是相当稳健的。然而, SA 算法的收敛速度和稳态性能通常不是很好。特别是近年来, MCC 已成功用于鲁棒自适应滤波, 其中滤波器权重被调整为使得期望信号和滤波器输出之间的相关性最大化。相关熵是一种非线性和局部相似性度量, 与两个随机变量在由核心带宽控制的关节空间邻域内的相似程度直接相关, 这也是 Renyi's 熵的根源。由于相关熵对异常点不敏感, 特别是在内核带宽较小的情况下, 在重尾脉冲噪声的情况下, 它自然是一个强大的适应成本。

相关熵中的核函数通常是一个高斯核函数, 由于其平滑性和严格的正定性, 这是合乎需要的。对于高斯核, 相对于核宽度, 相关性相对较小时, 相关性会引起称为相关诱导度量的非线性度量, 其行为如同  $L_2$  范数, 数据相对较小, 数据变大时为  $L_1$  范数,  $L_1$  当数据远离原点时是标准的。但是, 高斯内核当然并不总是最好的选择。因此, 使用了广义高斯密度 (GGD, generalized Gaussian density) 函数作为相关函数的核函数, 而新的相关熵被称为广义相关熵。

类似于具有高斯核的原始相关性, 广义相关性也可以用作与估计相关的问题的优化成本。可以证明, 实质上, GMCC 的估计是平滑的最大后验概率 (MAP, maximum a posteriori probability) 估计, 包括 MAP 和 LMP 估计作为极端情况。除了自相关矩阵和互相关向量由误差非线性加权外, GMCC 滤波的最优解与形式类似于著名的维纳解。如果涉及的信号是零均值高斯, 则最优解将等于维纳解。在 GMCC 准则下, GMCC 算法已被提出并且应用。

在本课题中, 我们研究了基于最大相关熵的 APA 和 KAPA, 尤其注重对基于广义最大相关熵<sup>[24]</sup>的 APA 和 KAPA 的研究。一方面, 它完善了对于最大相关熵的自适应滤波算法理论, 另一方面, 为下面的研究提供了理论基础。在核自适应滤波算法中, 大多数的核算法应用一种核函数, 但核函数多种多样, 也具有其独特的优点。在本课题中, 我们还研究了一种混合核最小均方算法, 这为核自适应滤波算法的完善和改进提供了新的方向。

## 1.2 国内外研究现状

21 世纪是信息化时代, 信息通信<sup>[25]</sup>事业得到超速发展, 自适应信号处理技术已被广泛应用。20 世纪 40 年代, 维纳滤波理论<sup>[26]</sup>被提出, 它是基于平稳随机信号建立的。到了 60 年代初, 由于多位空间理论和空间技术<sup>[27]</sup>的发展, 卡尔曼滤波 (KF, Kalman filter) 理论<sup>[28, 29]</sup>被提出。卡尔曼滤波利用状态变量模型对非平稳、多输入多输出的随机序列作最优估计。1967 年, 自适应滤波理论<sup>[30]</sup>被 Widrow.B 等人提出, 这促进了自适应滤波器的发展。自适应滤波器能够自动地调整自适应滤波系统的参数, 以期望于获得最佳滤波。在自适应滤波器设计过程中, 我们

对信号与噪声的先验统计知识的需求极低。早期的自适应滤波器主要为线性 AF，因为具有结构简单、滤波性能较好等特点，它发展很快。在线性 AF 发展的同时，许多的线性自适应滤波算法被提出，主要的线性自适应滤波算法为：LMS、APA 和 RLS。

1950 年，Aronzajn 发表了题为“Theory of Reproducing Kernels”的论文<sup>[31, 32]</sup>，在这篇论文中研究了再生核 Hilbert 空间(RKHS, reproducing kernel Hilbert spaces)<sup>[33]</sup>理论，并且介绍了多种再生核的构造方法。1964 年，Aizerman 等通过对位势函数法<sup>[34]</sup>的讨论，提出使用核函数替代特征空间内积的方法。1990 年，Wahba 等人研究了再生核在核正则化理论中的应用。同年，Poggio 和 Girosi 将再生核应用于正则网络<sup>[35]</sup>研究。1992 年，Boser 等人研究了核方法理论和大边缘分类器<sup>[36]</sup>理论。通过研究，他们提出了支持向量机<sup>[37]</sup>，进而使核方法广泛应用于机器学习领域中。

KAF 是核方法的一类典型应用，它在 RKHS 中实现了线性 AF 的非线性化。1999 年，Fried 和 Harrison 在 RKHS 中运用梯度下降法<sup>[38]</sup>推导线性自适应滤波算法，得到了核自适应滤波算法。2004 年，Engel 等人运用上述方法推导获得了 KRLS<sup>[11]</sup>。几年后，刘伟峰等人基于 LMS 算法核 RKHS，推导出了著名的 KLMS<sup>[39]</sup>。随后，刘伟峰、Richard 等人从不同角度对 KAPA<sup>[40]</sup>进行了研究，一般可以把 KAPA 看成是一种介于 KLMS 和 KRLS 之间的方法。而今，KAF 作为非线性 AF 的典型代表，在自适应信号处理中有着十分重要的地位。

正是由于 KAF 表现出的优越性能，使研究者们不断的对其进行探索。一般 KAF 中的核函数<sup>[41]</sup>多采用一般的单核函数，其中 GK<sup>[42]</sup>与 LK<sup>[43]</sup>居多。但是对于混合核函数的研究一直在进行着。混合核函数不仅保留了单核核函数的优点，而且体现出了新的优点，它具有更好的抗干扰能力，同时混合核函数的 KLMS 能表现出更好的性能。当然混合核函数的 KAF 也尤其缺点存在，比如其复杂度会增加。对于混合核的 KAF 研究虽然在不断的进行着，但是还没有形成完善的理论体系。最优化的混合核函数方式依然没被找到。

熵是一个十分重要的概念，它被用来描述客观事件的无序性，在许多领域中被应用。对于熵的研究开始的比较早，1865 年，R.Clausius 在熵概念的基础上提出了热力学熵<sup>[44]</sup>的物理概念；后来，R.Clausius 提出了统计热力学熵。1929 年 Lszilard 使信息科学引用熵的概念成为可能，他首次提出了熵与信息不确定性的关系。1948 年，C.Shannon 创立的信息论<sup>[45]</sup>，使熵的概念有了信息熵（香农熵）<sup>[45]</sup>这个全新的解释。信息熵描述的是随机事件的不确定性程度，但是它对事件中概率分配问题没有进行说明。1957 年 E.T.Jaynes 在“信息论与统计力学”一文中<sup>[46]</sup>，提出了最大熵原理。最大熵原理为我们提供了一个选择标准，使我们从诸多相容

分布中挑选“最佳”、“最合理”的分布成为可能。众所周知，当熵最大的时候，随机变量最不确定。虽然最大熵原理有着它主观的一面，但是对于当时来说它是最“客观”的。我国学者钟义信先生 1981 年针对香农信息论的熵公式只能度量概率信息的缺陷，提出了一种广义信息函数。1985 年他又把该公式推广到语义信息和语用信息的度量，得到了语法，语义和语用信息的综合测度公式。1993 年鲁晨光提出广义信息论，指出香农信息论的局限性具体表现在：不偏于度量语义信息、感觉信息、信源信道可变时的信息以及单个信号信息的广义信息测度；不分明集这些不确定性方面做出了许多的研究。斯托尼尔在 60 年代初提出统一的信息理论，他把信息看成是宇宙的一种基本属性，其目标是在信息物理学的基础上，并综合现有的控制论、符号学、语言学、认知心理学等学科，构建一个所谓一般信息理论的框架。

信息熵和最大熵原理在信息科学领域中被广泛的应用，尤其是信号处理方面。最大熵原理已被应用于自适应滤波算法中，很多基于最大相关熵的自适应滤波算法被提出。陈霸东等人提出了有约束的最大相关熵准则（CMCC, constrained maximum correntropy criterion）<sup>[47]</sup>，特别在非高斯环境中 CMCC 算法易于实现，计算复杂度较低。在收敛性和稳定性方面由于基于 MSE 的受约束的自适应算法。紧接着 LMS-MCC 和 RLS-MCC<sup>[48]</sup>算法被提出，都表现了良好的性能。MCC 作为 GMCC 的一个特例，然后陈霸东等人又提出了 GMCC 算法，介绍了 GMCC 的相关性质，并通过实验仿真验证了算法优越的滤波性能；在此基础上被研究的算法包括：LMS-GMCC、RLS-GMCC、KLMS-GMCC 和 KLMS-GMCC<sup>[49, 50]</sup>等。本论文主要从代价函数角度对自适应滤波算法进行改进和完善，从而完善自适应滤波算法理论。

### 1.3 论文结构

AF 作为数据处理的一个重要工具，它能够对未知的系统进行学习，从而估计系统内部的内在规律；而自适应滤波算法是 AF 的核心环节。本论文中我们为完善自适应滤波算法理论做了一些努力。一方面；从核函数角度出发，研究了混合核的 KLMS-MK。另一方面：研究了基于最大相关熵的 APA 和 KAPA。

第一章，阐述了 AF 和最大相关熵及其它们产生的背景及研究意义，纵览了 AF 和最大相关熵的研究过程与现状，体现出 AF 和最大相关熵的学术研究意义。这些也都是本课题考虑改善的问题。

第二章，首先，介绍经典的自适应滤波算法：LMS、APA 和 RLS。其次，介绍核方法和核自适应滤波器。最后，介绍经典的核自适应滤波算法：KLMS、KAPA 和 KRLS。

第三章，介绍了几种核函数，提出了一种混合核函数，并在这种混合核函数的基础上，提出了 KLMS-MK。通过测试 KLMS-MK 在 MG 时间序列预测和非线性回归的表现，证明了其优越的性能。

第四章，介绍了最大相关熵相关理论，比如：MCC，GMC，GMCC。研究了基于最大相关熵的 APA 和 KAPA，在 MG 时间序列预测模型中进行了仿真实验。

第五章，总结本论文主要的研究内容，同时根据本文研究内容遇到的疑惑提出下面研究工作的方向。

## 第二章 自适应滤波算法基础

在本章中，我们介绍了一些经典的自适应滤波算法。首先，介绍经典的线性自适应滤波算法，其主要包括：LMS、APA 和 RLS。其次，介绍核方法和核自适应滤波器的有关的理论。核自适应滤波器是非线性自适应滤波器的一个典型代表，通过介绍核自适应滤波器，能够更加充分的认识非线性自适应滤波器。最后，介绍了经典的核自适应滤波算法（KLMS、KAPA 和 KRLS）。在本章中目的是：为接下来介绍本文自己工作的成果，提供一定的理论支撑。

### 2.1 经典的线性自适应滤波算法

#### 2.1.1 最小均方算法

LMS 是最经典的线性自适应滤波算法，由于它比较简单、计算复杂度低和滤波性能较好，从而被广泛应用。LMS 的目标是最小化(2-1)中的代价函数。

$$J(n) = 1/2 e^2(n) \quad (2-1)$$

LMS 的滤波输出为  $y(n) = \omega(n-1)^T \mathbf{x}(n)$ ， $d(n)$  为期望信号，从而我们可得到误差信号为：

$$e(n) = d(n) - y(n) = d(n) - \omega(n-1)^T \mathbf{x}(n) \quad (2-2)$$

根据最陡梯度下降法，我们可以得到 LMS 的权矢量更新形式：

$$\omega(n) = \omega(n-1) - \mu \Delta J_{\omega}(n-1) = \omega(n-1) + \mu e(n) \mathbf{x}(n) \quad (2-3)$$

其中  $\Delta J_{\omega}(n-1)$  表示代价函数  $J(n)$  对  $\omega(n-1)$  的梯度，显然上述的 LMS 构建的是一种线性模型。LMS 对自适应滤波器结构要求低，具有计算复杂度低，在平稳信号环境中收敛性好的优点，使其成为应用最广的线性自适应滤波算法。

#### 2.1.2 仿射投影算法

设横向自适应滤波器的阶数为  $L$ ，滤波器的系数为  $\omega_i(n) (i=0,1,\dots,L-1)$ 。APA 主要通过重复利用过去的  $K$  个输入信号，其收敛速度比 LMS 快，它的误差为：

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{y}(n) = \mathbf{d}(n) - \mathbf{X}^T(n) \mathbf{W}(n) \quad (2-4)$$



其中输入向量为:

$$\mathbf{X}(n) = [\mathbf{x}(n-K+1), \dots, \mathbf{x}(n)]_{L \times K} \quad (2-5)$$

滤波器的权重向量为:

$$\mathbf{W}(n) = [\omega_0(n), \omega_1(n), \dots, \omega_{L-1}(n)]^T \quad (2-6)$$

期望信号向量为:

$$\mathbf{d}(n) = [d(n-K+1), \dots, d(n)]^T \quad (2-7)$$

APA 的更新形式为:

$$\boldsymbol{\omega}(n) = \boldsymbol{\omega}(n-1) + \mu \mathbf{X}(n) \mathbf{e}(n) \quad (2-8)$$

APA 是可以有效的提高线性自适应滤波算法的收敛性能, 但是其计算复杂度较高。APA 可以在最终失调和收敛速度中取得平衡。APA 是基于标准最小均方算法 (NLMS, normalized least mean square) [51] 的自适应滤波算法。快速仿射投影算法 (FAP, fast affine projection algorithm) 已被提出, 用于降低 APA 的计算复杂度。FAP[52] 获得了较快的收敛速度的同时, 降低了计算的复杂度, 其复杂度与 LMS 相当。APA 类算法是线性自适应滤波算法的重要组成部分。

### 2.1.3 递归最小二乘算法

RLS 的代价函数如下所示:

$$J = \sum_{l=1}^n |d(l) - \mathbf{x}(l)^T \boldsymbol{\omega}|^2 \quad (2-9)$$

令  $\mathbf{X}(n) = [\mathbf{x}(1), \dots, \mathbf{x}(n)]_{L \times n}$ ,  $\mathbf{d}(n) = [d(1), \dots, d(n)]^T$ , 可得在  $n$  时刻, 方程(2-9) 的解为

$$\boldsymbol{\omega}(n) = (\mathbf{X}(n) \mathbf{X}(n)^T)^{-1} \mathbf{X}(n) \mathbf{d}(n) \quad (2-10)$$

显然有  $\mathbf{X}(n) = [\mathbf{X}(n-1), \mathbf{x}(n)]$ ,  $\mathbf{d}(n) = [d(n-1), d(n)]$ ,  $\mathbf{P}(n) = (\mathbf{X}(n) \mathbf{X}(n)^T)^{-1}$ , 从而存在

$$\mathbf{P}(n)^{-1} = \mathbf{P}(n-1)^{-1} + \mathbf{x}(n) \mathbf{x}(n)^T \quad (2-11)$$

借助于矩阵求逆引理公式:

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{C}^{-1} - \mathbf{DA}^{-1} \mathbf{B})^{-1} \mathbf{DA} \quad (2-12)$$

这里令  $\mathbf{P}(n)^{-1} = \mathbf{A}$ ,  $\mathbf{x}(n) = \mathbf{B}$ ,  $\mathbf{C} = 1$ ,  $\mathbf{D} = \mathbf{x}(n)^T$ , 从而有

$$\mathbf{P}(n) = \mathbf{P}(n-1) - \frac{\mathbf{P}(n-1)\mathbf{x}(n)\mathbf{x}(n)^T\mathbf{P}(n-1)}{1 + \mathbf{x}(n)^T\mathbf{P}(n-1)\mathbf{x}(n)} \quad (2-13)$$

由以上可得 RLS 的权重矢量  $\boldsymbol{\omega}(n)$  有如下的递归更新形式:

$$\begin{aligned} \boldsymbol{\omega}(n) &= \mathbf{P}(n)\mathbf{X}(n)d(n) \\ &= \left[ \mathbf{P}(n-1) - \frac{\mathbf{P}(n-1)\mathbf{x}(n)\mathbf{x}(n)^T\mathbf{P}(n-1)}{1 + \mathbf{x}(n)^T\mathbf{P}(n-1)\mathbf{x}(n)} \right] [\mathbf{X}(n-1)d(n-1) + \mathbf{x}(n)d(n)] \\ &= \boldsymbol{\omega}(n-1) + \frac{\mathbf{P}(n-1)\mathbf{x}(n)}{1 + \mathbf{x}(n)^T\mathbf{P}(n-1)\mathbf{x}(n)} [d(n) - \mathbf{x}(n)^T\boldsymbol{\omega}(n-1)] \end{aligned} \quad (2-14)$$

RLS 具有比 APA 更高的计算复杂度, 但是它的收敛速度最快, 同时具有比 LMS 和 APA 更好的滤波性能。

## 2.2 核方法与核自适应滤波器

在随机信号处理中, 我们面临的问题大多是非线性的, 因此对于非线性滤波器的研究是十分有必要的。在这一部分, 我们主要介绍核自适应滤波器。核自适应滤波器, 是从核方法发展起来的。核方法首先将原始线性空间中的数据映射到高维特征空间中, 然后利用线性分析和处理方法分析这个高维空间中的数据, 从而解决非线性问题的分析。Mercer 核<sup>[53]</sup>实际上就是一个函数  $\kappa(\cdot, \cdot): \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$ , 输入域表示为:  $\mathbb{X} \subset \mathbb{R}^N$ 。  $\kappa(\cdot, \cdot)$  函数必须同时满足连续性、对称性以及正定性。在众多 Mercer 核中, 最常用的是以下两种:

$$\text{高斯核: } \kappa(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2h^2}\right) \quad (2-15)$$

$$\text{拉普拉斯核: } \kappa(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|}{2h^2}\right) \quad (2-16)$$

其中,  $h$  表示核函数的核宽度。在给定核函数一个自由参数的条件下, 我们有  $\varphi(\mathbf{x}) = \kappa(\mathbf{x}, \cdot)$ ,  $\varphi$  将原始输入空间  $\mathbb{X}$  映射到高维特征空间<sup>[54]</sup>  $\mathbb{F}$  上, 它表示一种非线性映射关系。通常认为这个特征空间  $\mathbb{F}$  是 RKHS。

Miccheli 等人证明了高斯核具有通用的逼近能力, 即对于一个任意的连续映射  $f: \mathbb{X} \rightarrow \mathbb{R}$  以及任意小的常数  $\xi > 0$ , 存在一组输入数据点  $\{\mathbf{x}_l\}_{l=1}^L \in \mathbb{X}$  满足

$$\left\| f(\cdot) - \sum_{l=1}^L c_l \kappa(\mathbf{x}_l, \cdot) \right\| \leq \xi \quad (2-17)$$

其中  $\{c_l\}_{l=1}^L$  是一组实数集合。除了通用的近似能力外，高斯核还表现出期望的平滑性以及数值稳定性等，因而实际应用中往往都会选用高斯核。

用  $\mathbb{H}$  来表示一个向量空间，它是由所有与  $\mathbf{x}$  相关的实值函数构成的，并且它的每个元素都可以通过  $\kappa(\mathbf{x}, \bullet)$  导出，从中任取两个函数  $h(\bullet)$  和  $g(\bullet)$ ，假设他们有如下的形式：

$$h(\bullet) = \sum_{l=1}^L a_l \kappa(\mathbf{x}_l, \bullet) \quad (2-18)$$

$$g(\bullet) = \sum_{j=1}^{L'} b_j \kappa(\mathbf{x}'_j, \bullet) \quad (2-19)$$

其中， $\mathbf{x}_l$  和  $\mathbf{x}'_j$  属于输入空间  $\mathbb{X}$ ； $a_l$  和  $b_j$  表示相应的扩展系数。

进一步，定义  $h(\bullet)$  和  $g(\bullet)$  的双线性形式为：

$$\langle h, g \rangle = \sum_{l=1}^L \sum_{j=1}^{L'} a_l \kappa(\mathbf{x}_l, \mathbf{x}'_j) b_j \quad (2-20)$$

其满足如下性质：

(1) 对称性

$$\langle h, g \rangle = \langle g, h \rangle \quad (2-21)$$

(2) 按比例缩放性以及分配性

$$\langle \langle cf + dg \rangle, h \rangle = c \langle f, h \rangle + d \langle g, h \rangle \quad (2-22)$$

(3) 范数平方非负性

$$\|f\|^2 = \langle f, f \rangle \geq 0 \quad (2-23)$$

如果令  $g(\bullet) = \kappa(\mathbf{x}', \bullet)$ ，则容易得到：

$$\langle h, \kappa(\mathbf{x}^*, \bullet) \rangle = \sum_{l=1}^L a_l \kappa(\mathbf{x}_l, \mathbf{x}^*) = h(\mathbf{x}^*) \quad (2-24)$$

这就是所谓的再生性。上式中， $\kappa(\mathbf{x}_l, \mathbf{x}^*)$  是两个矢量的函数，如果他们具备以下两个特征：(1) 对于任意的  $\mathbf{x} \in \mathbb{X}$ ，矢量  $\mathbf{x}^*$  的函数  $\kappa(\mathbf{x}_l, \mathbf{x}^*)$  都属于空间  $\mathbb{H}$ ；(2) 满足再生性，则就可称它为  $\mathbb{H}$  空间中的一个再生核函数，显然 Mercer 核具备上述两个特征，因而也可以称之为再生核。

如果  $\mathbb{H}$  是一个向量空间，并且它是一个 RKHS。 $\mathbb{H}$  中的每个元素与  $\mathbf{x}$  有关，

它是通过  $\kappa(\mathbf{x}, \cdot)$  导出的。RKHS 中的重要概念之一是核戏法，即

$$\varphi(\mathbf{x})^T \varphi(\mathbf{x}') = \kappa(\mathbf{x}, \mathbf{x}') \quad (2-25)$$

RKHS 中的元素  $\varphi(\mathbf{x})$  具有很高的维度，甚至是无限维的。因此直接通过 RKHS 中的元素计算是十分困难的。核方法避免了高维特征空间中的直接计算，这一点是非常重要的。

对于输入输出数据  $\{x(n), d(n)\}$ ，我们需要一种映射关系  $f$  构建其相应的输出  $f(\mathbf{x}(n)) = \langle f(\cdot), \kappa(\cdot, \mathbf{x}(n)) \rangle$ ，这个输出值是  $d(n)$  的估计值。一般来说， $d(n) = f^*(\mathbf{x}(n)) + \xi(n)$ ，这里  $f^*(\mathbf{x}(n))$  是  $d(n)$  的最优估计值。KAF 的目标是通过学习一个连续地输入、输出映射关系  $f: \mathbb{X} \rightarrow \mathbb{R}$ ，找到它们之间的内在规律。根据表示理论， $f(\cdot)$  可以表示成如下的线性形式：

$$f(\cdot) = \boldsymbol{\Omega}^T \varphi(\cdot) \quad (2-26)$$

其中  $\boldsymbol{\Omega}$  是 RKHS 中的权重矢量， $\boldsymbol{\Omega}$  可以表示为线性组合的形式，其表现形式如下式：

$$\boldsymbol{\Omega} = \sum_{l=1}^n a_l \varphi(\mathbf{x}(l)) \quad (2-27)$$

将(2-26)带入(2-27)，并结合核技巧，可以进一步得到

$$f(\cdot) = \sum_{l=1}^n a_l \kappa(\cdot, \mathbf{x}(l)) \quad (2-28)$$

其中  $\kappa(\cdot, \mathbf{x}(l))$  是以  $\mathbf{x}(l)$  为中心的核函数。自适应滤波算法通过采用线性逼近的方式，使  $f$  尽可能的逼近  $f^*$ 。

## 2.3 经典的核自适应滤波算法

### 2.3.1 核最小均方算法

在随机信号处理中，大多数问题的输入、输出关系是非线性的。前面介绍的 LMS 构建的是一种线性模型，其主要解决输入、输出的线性问题。虽然 LMS 比较简单，但是它的滤波性能具有一定的局限性。为了得到更好的滤波效果，需要采用非线性自适应滤波算法，在 RKHS 中重新推导 LMS 得到 KLMS。 $\mathbf{x}(n)$  经过映射在 RKHS 中表示为  $\varphi(\mathbf{x}(n))$ ，同时用  $\boldsymbol{\Omega}(n)$  表示映射后的权重矢量，则有

$$\boldsymbol{\Omega}(n) = \boldsymbol{\Omega}(n) + \mu e(n) \varphi(\mathbf{x}(n)) \quad (2-29)$$

令  $\mathbf{\Omega}(0) = 0$ ，经过反复迭代，可以得到

$$\begin{aligned}
 \mathbf{\Omega}(n) &= \mathbf{\Omega}(n-1) + \mu e(n) \varphi(\mathbf{x}(n)) \\
 &= [\mathbf{\Omega}(n-2) + \mu e(n-1) \varphi(\mathbf{x}(n-1))] + \mu e(n) \varphi(\mathbf{x}(n)) \\
 &\dots \\
 &= \mathbf{\Omega}(0) + \mu \sum_{l=1}^n e(l) \varphi(\mathbf{x}(l)) \\
 &= \mu \sum_{l=1}^n e(l) \varphi(\mathbf{x}(l))
 \end{aligned} \tag{2-30}$$

当一个新的输入数据  $\mathbf{x}(n)$  到来的时候，相应的 KLMS 的输出为：

$$\begin{aligned}
 y(n) &= \mathbf{\Omega}(n-1)^T \varphi(\mathbf{x}(n)) \\
 &= \mu \sum_{l=1}^{n-1} e(l) \varphi(\mathbf{x}(l))^T \varphi(\mathbf{x}(n)) \\
 &= \mu \sum_{l=1}^n \mathbf{a}_l(n-1) \kappa(\mathbf{x}(l), \mathbf{x}(n))
 \end{aligned} \tag{2-31}$$

其中  $\mathbf{a}_l(n-1) = \mu e(l)$  表示  $n-1$  时刻第  $l$  个“字典”中心对应的系数。如果用  $f_n$  表示  $n$  时刻的输入、输出非线性映射关系估计，KLMS 可以总结为：

$$\begin{cases} f_0 = 0 \\ e(n) = d(n) - f_{n-1}(\mathbf{x}(n)) \\ f_n = f_{n-1} + \mu e(n) \kappa(\mathbf{x}(n), \cdot) \end{cases} \tag{2-32}$$

由此可知，KLMS 拓扑结构类似一个 RBF 网络，主要有三个不同点：首先，权重矢量是每个样本的预测误差；其次，这是一个不断增长的网络，每个新单元都放置在每个新输入上；第三， $\kappa$  不限于径向基函数，可以是任何 Mercer 核。因此 KLMS 是一种简单的算法，这里需要关注以下三个方面：（1）如何选择核函数。

（2）如何选择步长参数  $\eta$ 。（3）如何应对线性增长的网络结构。

### 2.3.2 核仿射投影算法

类似于 KLMS，运用核方法公式(2-8)可以重新表述为：

$$\mathbf{\Omega}(n) = \mathbf{\Omega}(n-1) + \mu \Phi(n) [\mathbf{d}(n) - \Phi(n)^T \mathbf{\Omega}(n-1)] \tag{2-33}$$

其中  $\Phi(n) = [\varphi(\mathbf{x}(n-K+1)), \dots, \varphi(\mathbf{x}(n))]$ 。令  $\mathbf{\Omega}(0) = 0$ ，通过迭代可得到：

$$\mathbf{\Omega}(n) = \sum_{l=1}^n \mathbf{a}_l(n-1) \varphi(\mathbf{x}(n)) \tag{2-34}$$

扩展系数  $\mathbf{a}(n)$  的更新分几种不同的情况，具体如下：

$$a_l(n) = \begin{cases} \mu \left( d(n) - \sum_{j=1}^{n-1} a_j(n-1) \kappa(x(n), x(j)) \right), & l = n \\ a_l(n-1) + \mu \left( d(l) - \sum_{j=1}^{n-1} a_j(n-1) \kappa(x(j), x(l)) \right), & n-K+1 \leq l \leq n-1 \\ a_l(n-1), & 1 \leq l \leq n-K+1 \end{cases} \quad (2-35)$$

对于给定输入  $\mathbf{x}(n)$ ，容易得到 KAPA 的输出为：

$$y(n) = \sum_{l=1}^{n-1} \mathbf{a}_l(n-1) \kappa(\mathbf{x}(l), \mathbf{x}(n)) \quad (2-36)$$

类似于前面的 KLMS，KAPA 的学习规则可以总结如下：

$$\begin{cases} f_0 = 0 \\ \mathbf{e}_l(n) = \mathbf{d}_l(n) - \mathbf{y}_l(n), & n-K+1 \leq l \leq n-1 \\ f_n = f_{n-1} + \mu \sum_{l=n-K+1}^{n-1} \mathbf{e}_l(n) \kappa(\mathbf{x}(l), \cdot) \end{cases} \quad (2-37)$$

### 2.3.3 核递归最小二乘算法

通过将  $\mathbf{x}(n)$  映射到 RKHS 中，即  $\varphi(\mathbf{x}(n))$ ，在 RLS 基础上进一步推导 KRLS。KRLS 的代价函数表示为：

$$J = \min_{\mathbf{Q}} \sum_{l=1}^n \left| d(l) - \mathbf{Q}^T \varphi(\mathbf{x}(l)) \right|^2 + \zeta \|\mathbf{Q}\|^2 \quad (2-38)$$

解方程(2-38)，可以能够得到下式：

$$\mathbf{Q}(n) = \left[ \zeta \mathbf{I} + \mathbf{\Phi}(n)^T \mathbf{\Phi}(n) \right]^{-1} \mathbf{\Phi}(n) \mathbf{d}(n) \quad (2-39)$$

其中  $\mathbf{\Phi}(n) = [\varphi(x(1)), \dots, \varphi(x(n))]$ ， $\mathbf{I}$  为单位矩阵， $\zeta$  为正则化参数。根据矩阵求逆引理，我们可以将(2-39)重新表示为：

$$\begin{aligned} \mathbf{Q}(n) &= \mathbf{\Phi}(n) \left[ \zeta \mathbf{I} + \mathbf{\Phi}(n)^T \mathbf{\Phi}(n) \right]^{-1} \mathbf{d}(n) \\ &= \mathbf{\Phi}(n) \mathbf{a}(n) \end{aligned} \quad (2-40)$$

其中  $\mathbf{a}(n) = [\zeta \mathbf{I} + \mathbf{\Phi}(n)^T \mathbf{\Phi}(n)]^{-1} \mathbf{d}(n)$ 。如若表示  $\mathbf{Q}(n) = [\zeta \mathbf{I} + \mathbf{\Phi}(n)^T \mathbf{\Phi}(n)]^{-1}$ ， $\mathbf{h}(n) = \mathbf{\Phi}(n-1)^T \varphi(\mathbf{x}(n))$ ，则容易得到

$$\mathbf{Q}(n)^{-1} = \begin{bmatrix} \mathbf{Q}(n-1)^{-1} & \mathbf{h}(n) \\ \mathbf{h}(n)^T & \zeta + \varphi(\mathbf{x}(n))^T \varphi(\mathbf{x}(n)) \end{bmatrix} \quad (2-41)$$

为了进一步推导，这里引入块矩阵求逆引理公式，即

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -\mathbf{A}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix} \quad (2-42)$$

其中  $\mathbf{A}$  和  $\mathbf{D}$  为任意方阵。令  $\mathbf{z}(n) = \mathbf{Q}(n-1)\mathbf{h}(n)$ ， $r(n) = \zeta + \varphi(\mathbf{x}(n))^T \varphi(\mathbf{x}(n)) - \mathbf{z}(n)^T \mathbf{h}(n)$ ，可得到  $\mathbf{Q}(n)$  有如下的递归更新表达式：

$$\mathbf{Q}(n) = r(n)^{-1} \begin{bmatrix} \mathbf{Q}(n-1) + \mathbf{z}(n)\mathbf{z}(n)^T & -\mathbf{z}(n) \\ -\mathbf{z}(n)^T & 1 \end{bmatrix} \quad (2-43)$$

从而得到

$$\begin{aligned} \mathbf{a}(n) &= \mathbf{Q}(n)\mathbf{d}(n) \\ &= \begin{bmatrix} \mathbf{a}(n-1) - \mathbf{z}(n)r(n)^{-1}e(n) \\ r(n)^{-1}e(n) \end{bmatrix} \end{aligned} \quad (2-44)$$

对于输入  $\mathbf{x}(n)$ ，相应的预测输出为：

$$\begin{aligned} y(n) &= \mathbf{h}(n)^T \mathbf{a}(n) \\ &= \sum_{l=1}^{n-1} \mathbf{a}_l(n-1) \kappa(\mathbf{x}(l), \mathbf{x}(n)) \end{aligned} \quad (2-45)$$

显然，KRLS 有如下顺序学习规则：

$$\begin{cases} f_0 = 0 \\ e(n) = d(n) - f_{n-1}(\mathbf{x}(n)) \\ f_n = f_{n-1} + r(n)^{-1} \left( \kappa(\mathbf{x}(n), \cdot) - \sum_{l=1}^{n-1} \mathbf{z}_l(n) \kappa(\mathbf{x}(n), \cdot) \right) e(n) \end{cases} \quad (2-46)$$

通过使用核方法将 RLS 映射到高维特征空间。然后，通过利用称为矩阵求逆引理，推导出了 KRLS 算法。该算法的一个重要特点是其收敛速度通常比 KLMS 快一个数量级。然而，这种性能的改善是以增加计算复杂度为代价来实现的。但是现在有很多稀疏化的方法在保证其收敛速度的同时，大大减小了其计算复杂度。因此，KRLS 在随机信号处理中应用较多。

## 2.4 本章小结

本章首先介绍了经典的线性自适应滤波算法：LMS、APA 和 RLS，推导了它们的权重更新式，并对算法进行了简单的介绍。其次，介绍了核方法的相关理论与核自适应滤波器。最后，在经典线性自适应算法的基础上，运用核方法相关理论，将线性空间中的数据映射到高维特征空间中，从而推导出相应的经典的核自适应滤波算法。本文接下来将在本章理论的基础上，主要从代价函数角度，运用新的方法，对自适应滤波算法进行研究。





### 第三章 混合核最小均方算法

本章中介绍了提出的一种混合核函数，并基于此混合核函数得到了 KLMS-MK，它提高了 KLMS 的滤波性能。通过把凸组合的方法应用于 KLMS 的核函数，KLMS-MK 同时具有高斯核和拉普拉斯核的优点。在 KLMS-MK 中，凸组合的混合参数用梯度下降法进行更新。通过仿真实验，可以看出，KLMS-MK 的稳态均方误差 (MSE, steady-state mean square error)<sup>[56]</sup> 和收敛速度都有所提高。

#### 3.1 核函数及混合核函数

在上一章介绍的经典核自适应滤波算法中，KLMS 的广泛应用，归功于它的简单性。然而，核自适应滤波算法面临的主要挑战之一是：随着训练数据的增加，可再生核函数网络结构尺寸将呈线性增长。这将增加核自适应滤波算法的计算复杂度和存储容量。为了解决这个问题，因此一个有效的稀疏化的方法需要实现一个紧凑的网络结构。一般来说，现有的稀疏化准则<sup>[57]</sup>包括：近似线性依赖性<sup>[12]</sup> (ALD, approximate linear dependency)，新颖性准则<sup>[58]</sup> (NC, novelty criterion)，相关性准则 (CC, coherence criterion) 和量化准则等。

众所周知，在核自适应滤波算法中使用的核函数对自适应滤波性能具有重要的影响。径向基函数 (RBF, radial basis function) 是最常用的核函数，它为训练数据中的隐藏关系提供了一个很好的近似。经典核函数包括高斯核和拉普拉斯核。基于  $L_2$  范数的高斯核函数是一种具有良好的抗干扰能力强大的 RBF 核函数，但对核宽度非常敏感。基于  $L_2$  范数平方根的拉普拉斯核可以看作是高斯核的一个变种。此外，拉普拉斯核降低了性能对核宽度的依赖性。它们的表达式如下：

$$\text{高斯核: } \kappa(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2h^2}\right) \quad (3-1)$$

$$\text{拉普拉斯核: } \kappa(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|}{2h^2}\right) \quad (3-2)$$

其中， $h$  表示核函数的核宽度。通常，在核自适应滤波算法中，仅有单个核函数被应用于 RBF 网络。但是，当数据比较复杂时，只有一个核函数不能完全覆盖数据的分布特征。因此，需要具有多核的核自适应滤波算法呈现数字特征。

下面，我们将介绍我们提出的一种混合核函数。我们通过  $\gamma_1(n) = \|\mathbf{x} - \mathbf{x}'\|^2 / 2h^2$  表示  $L_2$  范数，它的平方根为  $\gamma_2(n) = \|\mathbf{x} - \mathbf{x}'\| / 2h^2$ 。混合核因此被定义为：

$$\begin{aligned}\kappa(\mathbf{x}, \mathbf{x}') &= \exp\left[-\left(\lambda\gamma_1(n) + (1-\lambda)\gamma_2(n)\right)\right], \\ &= \exp\left[-\left(\lambda\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2h^2} + (1-\lambda)\frac{\|\mathbf{x}-\mathbf{x}'\|}{2h^2}\right)\right]\end{aligned}\quad (3-3)$$

这里  $\lambda(n) \in [0,1]$  是混合参数。构造的混合核函数的一些特征给出如下：

1、混合核的结构仍然是 RBF 核的一种形式，因此保留了简单内核的优点。当混合参数接近零时，混合核函数降低为高斯核函数。当  $\lambda(n)=1$  时，混合核为拉普拉斯核。混合参数作为组合的一个参数，决定了两个核函数在混合核参数中的比例。因此，自适应策略用于  $\lambda(n)$  的选择，这将在下面给出。

2、高斯和拉普拉斯核的结合仍处于希尔伯特空间，避免了不匹配的问题。

### 3.2 混合核最小均方算法

利用(3-3)公式设计的混合核函数，KLMS-MK 的输出估计为：

$$y(n) = f_n(\mathbf{x}(n)) = \mu \sum_{i=1}^{n-1} e(i) \kappa(\mathbf{x}(i), \mathbf{x}(n)) \quad (3-4)$$

期望的输出如下式所表示：

$$d(n) = f^*(\mathbf{x}(n)) \quad (3-5)$$

其中  $f^*$  表示被估计的学习函数。系统的误差表示为：

$$\begin{aligned}e(n) &= d(n) - y(n) + \varepsilon(n) \\ &= f^*(\mathbf{x}(n)) - f_{n-1}(\mathbf{x}(n)) + \varepsilon(n) \\ &= \tilde{f}_{n-1}(\mathbf{x}(n)) + \varepsilon(n)\end{aligned}\quad (3-6)$$

$$\tilde{f}_{n-1} = f^* - f_{n-1} \quad (3-7)$$

表示在  $n-1$  此迭代的额外映射。

通过 KLMS 的更新公式(2-24)，可以得到：

$$f_n(\mathbf{x}(n)) = f_{n-1}(\mathbf{x}(n)) + \mu e(n) \kappa(\mathbf{x}(n-1), \mathbf{x}(n)) \quad (3-8)$$

结合公式(3-7)，我们重写公式(3-8)，可得到：

$$\tilde{f}_n(\mathbf{x}(n)) = \tilde{f}_{n-1}(\mathbf{x}(n)) - \mu e(n) \kappa(\mathbf{x}(n-1), \mathbf{x}(n)) \quad (3-9)$$

不合适的混合参数的选择可能会导致 KLMS-MK 性能的下降。为了规避这个问题，混合参数  $\lambda$  应该自动调整。因此，下面我们用  $\lambda(n)$  代替  $\lambda$ 。通过梯度下降法更新每次迭代的混合参数  $\lambda(n)$ ，具体如下：

$$\begin{aligned}
 \lambda(n) &= \lambda(n-1) - \frac{\eta}{2} \frac{\partial}{\partial \lambda(n-1)} [e^2(n)] \\
 &= \lambda(n-1) - \eta e(n) \frac{\partial}{\partial \lambda(n-1)} [\tilde{f}_{k-1}(\mathbf{x}(n)) + \varepsilon(n)] \\
 &= \lambda(n-1) - \eta e(n) \frac{\partial}{\partial \lambda(n-1)} \left[ \begin{array}{c} \tilde{f}_{k-2}(\mathbf{x}(n)) + \varepsilon(n) \\ -\mu e(n-1) \kappa(\mathbf{x}(n-1), \mathbf{x}(n)) \end{array} \right] \\
 &= \lambda(n-1) + \eta \mu e(n) e(n-1) (\gamma_1(n) + \gamma_2(n)) \kappa(\mathbf{x}(n-1), \mathbf{x}(n))
 \end{aligned} \tag{3-10}$$

其中  $\tilde{f}_{n-1}(\mathbf{x}(n)) = \tilde{f}_{n-2}(\mathbf{x}(n)) - \mu e(n-1) \kappa(\mathbf{x}(n-1), \mathbf{x}(n))$ 。

从(3-10)中可以看出，混合参数是通过自适应切换两类核函数来实现的。

此外，KLMS-MK 中的混合参数的收敛性被推导。公式(3-10)两边同时计算期望：

$$E[\lambda(n)] = E[\lambda(n-1)] + \Delta_\lambda(n) \tag{3-11}$$

这里  $\Delta_\lambda(n)$  的表示如下：

$$\begin{aligned}
 \Delta_\lambda(n) &= E[\lambda(n)] - E[\lambda(n-1)] \\
 &= E[e(n) e(n-1) (\gamma_1 - \gamma_2) \kappa(\mathbf{x}(n-1), \mathbf{x}(n))] \\
 &\stackrel{(a_1)}{=} \mu E[e(n) e(n-1)] E[(\gamma_1 - \gamma_2) \kappa(\mathbf{x}(n-1), \mathbf{x}(n))] \\
 &\stackrel{(a_2)}{=} \mu E[e_a(n) e_a(n-1)] E[(\gamma_1 - \gamma_2) \kappa(\mathbf{x}(n-1), \mathbf{x}(n))] \\
 &= 0
 \end{aligned} \tag{3-12}$$

其中  $(a_1)$  和  $(a_2)$  分别由输入和误差独立和输入和噪声独立得到。因此，从(3-12)可得到  $\Delta_\lambda(n)$  收敛的。最终，KLMS-MK 总结在**算法 1**。

### 算法 1: 混合核最小均方算法

**输入:**

输入—输出对为  $\{\mathbf{x}(n), d(n)\}, n=1, 2, \dots$

**初始化:**

让学习率:  $\mu > 0$  和  $\eta > 0$ ,

核宽度:  $h > 0$ ,

初始化系数矢量:  $a(1) = \mu d(1)$ 。

当  $\{\mathbf{x}(n), d(n)\} (n > 1)$  可用时, 开始循环

**计算:**

1) 计算核函数:

$$\kappa(\mathbf{x}, \mathbf{x}') = \exp \left[ - \left( \lambda(n) \frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2h^2} + (1 - \lambda(n)) \frac{\|\mathbf{x} - \mathbf{x}'\|}{2h^2} \right) \right],$$

这里,  $\lambda(n)$  通过公式(3-10)计算

2) 计算输出:

$$y(n) = f_n(\mathbf{x}(n)) = \mu \sum_{i=1}^{n-1} a_i(n-1) \kappa(\mathbf{x}(i), \mathbf{x}(n))$$

3) 计算误差:  $e(n) = d(n) - y(n)$

**更新:**

$$a_n(n) = \mu e(n)$$

停止循环

### 3.3 仿真实验

在这一部分中, 混沌时间序列预测<sup>[59]</sup>和非线性回归<sup>[60]</sup>被用来验证所提出的 KLMS-MK 在高斯和非高斯噪声下的效率。在下面的仿真中, 比较的算法包括: LMS、高斯核的 KLMS(KLMS-G)、拉普拉斯核的 KLMS(KLMS-L)和 KLMS-MK, 同时取 50 次独立的蒙特卡洛<sup>[61]</sup>。在非高斯噪声下, 考虑  $\alpha$  稳定分布, 其特征函数为:

$$f(t) = \exp \left\{ j\delta t - \gamma |t|^\sigma \left[ 1 + j\beta \operatorname{sgn}(t) S(t, \sigma) \right] \right\} \quad (3-13)$$

对  $\alpha$  稳定噪声相关的参数是由  $V = (\sigma, \beta, \gamma, \delta)$  给出, 其中特征因子  $\sigma \in (0, 2]$ ,

$\beta \in (-1, 1)$  衡量不对称性, 分布参数的  $\gamma > 0$ , 位置参数  $\delta \in (-\infty, \infty)$  和

$$S(t, \sigma) = \begin{cases} \tan \frac{\sigma\pi}{2}, & \sigma \neq 1 \\ \frac{2}{\pi} \log|t|, & \sigma = 1 \end{cases} \quad (3-14)$$

为了评估 KLMS-MK 的性能, 定义测试 MSE 为:

$$\text{MSE} = \frac{1}{N} \sum_{n=1}^N (d(n) - f(\mathbf{x}(n)))^2 \quad (3-15)$$

这里,  $N$  是测试数据的长度。

### 3.3.1 MG 时间序列预测模型及仿真

MG 混沌时间序列表现出混沌动力学、可预测性和周期性等特性, 由下式产生,

$$\frac{dx(t)}{dt} = -0.1x(t) + \frac{0.2x(t-\tau)}{1+x(t-30)^{10}} \quad (3-16)$$

为了获得离散序列  $x(n)$ , 取采样周期为 6。采用前十个输入为:  $\mathbf{u}(n) = [x(n-10), x(n-9), \dots, x(n-1)]^T$ , 来估计所需的输出  $x(n)$ 。训练和测试数据的长度分别为 1000 和 50。

1、高斯环境: 在这种情况下, 训练数据受到加性高斯噪声的干扰, 白噪声的均值为 0, 方差为 0.1。LMS、KLMS-G、KLMS-L 和 KLMS-MK 的步长分别设置为: 0.3, 0.6, 0.6, 和 0.6; 它们的核宽度设置为 1。其中, 在步长参数的设置中, KLMS-G、KLMS-L 和 KLMS-MK 的步长参数设置是一样的, LMS 选择了适合的步长参数。图 3.1, 展示了在高斯噪声下四种算法的测试 MSEs。从图 3.1, 在高斯环境中, 我们看到 KLMS-MK 相比于其他三种算法达到了更好的滤波性能。因此, 高斯核和拉普拉斯核的组合核函数相比于单核核函数有更小的误差。

2、非高斯环境: 在非高斯情况下, 使用与图 3.1 相同的学习速率和核宽度。对  $\alpha$  稳定分布噪声的参数设置为  $V = [1.4, 0, 0.02, 0]$ 。图 3.2 给出了测试  $\alpha$  稳定噪声下的测试 MSEs。从图 3.2 可以看出, 在  $\alpha$  稳定分布噪声下 KLMS-MK 相比于其他比较的算法也达到最佳的滤波性能。这也意味着 KLMS-MK 对于非高斯噪声具有更好的鲁棒性比单核核函数的算法。在 KLMS-MK 鲁棒性改善的原因是, 高斯和拉普拉斯核能根据混合参数自适应的调整。

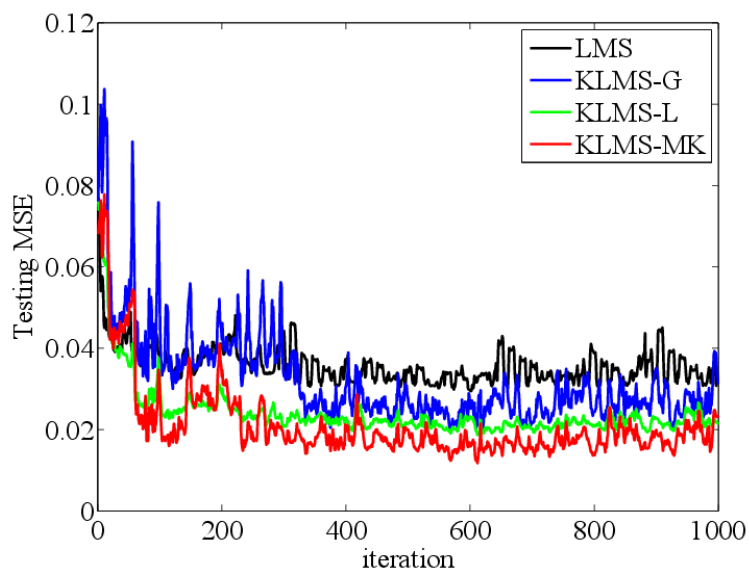


图 3.1 高斯噪声下不同算法在 MG 时间时间序列预测模型的性能曲线

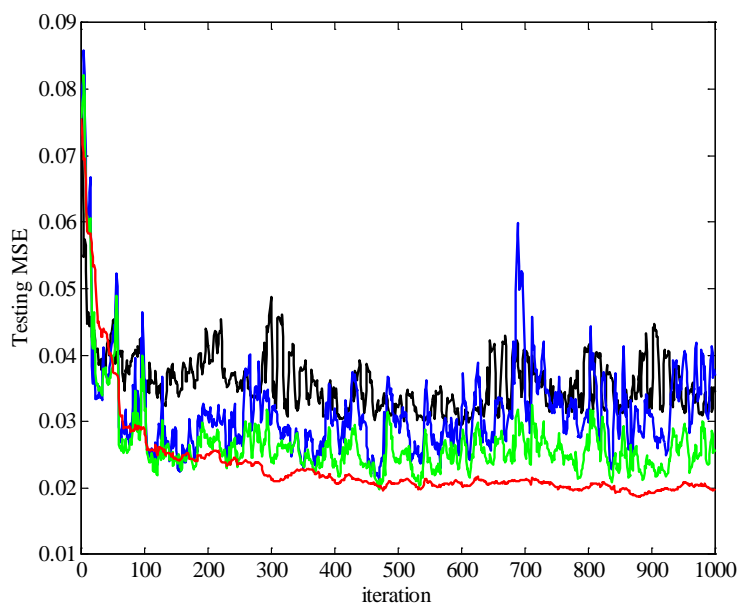


图 3.2 非高斯噪声下不同算法在 MG 时间时间序列预测模型的性能曲线

### 3.3.2 非线性回归模型及仿真

考虑如下所示的非线性系统：

$$x(n) = x(n-1)(0.8 - 0.5e^{-x(n-1)^2}) - (0.3 + 0.9x(n-2)e^{-x(n-1)}) + 0.1\sin(x(n-1)\pi) \quad (3-17)$$

它的初始值为：  $x(-1) = 0.1$  和  $x(-2) = 0.1$ 。我们选择  $\mathbf{u}(n) = [x(n-1), x(n-2)]^T$  去预测当前输出  $x(n)$ 。选取 2500 个数据段作为训练数据，100 个数据作为测试数据。

1、高斯环境：在这个例子中，LMS、KLMS-G、KLMS-L 和 KLMS-MK 的学习率分别设置为：0.1，1，0.1，和 0.85；它们的核宽度分别设置为：1，2，和 2。上述参数的设置，是在算法最优性能时设置的。图 3.3 给出了高斯噪声下非线性回归中的测试 MSE 曲线。从这个仿真图可以看出，可以得到与图 1 相同的结论。

2、非高斯环境：在非高斯情况下，配置了与图 3.3 相同的学习速率和核宽度。对  $\alpha$  稳定分布噪声的参数设置为  $V = [1.2, 0, 0.6, 0]$ 。图 3.4 显示了在  $\alpha$  稳定噪声下，非线性回归模型中测试 MSEs。从图 3.4 中我们看到，在非线性回归模型中，与其他算法相比，本文提出的 KLMS-MK 由于混合核也对非高斯噪声有一个较好的鲁棒性。

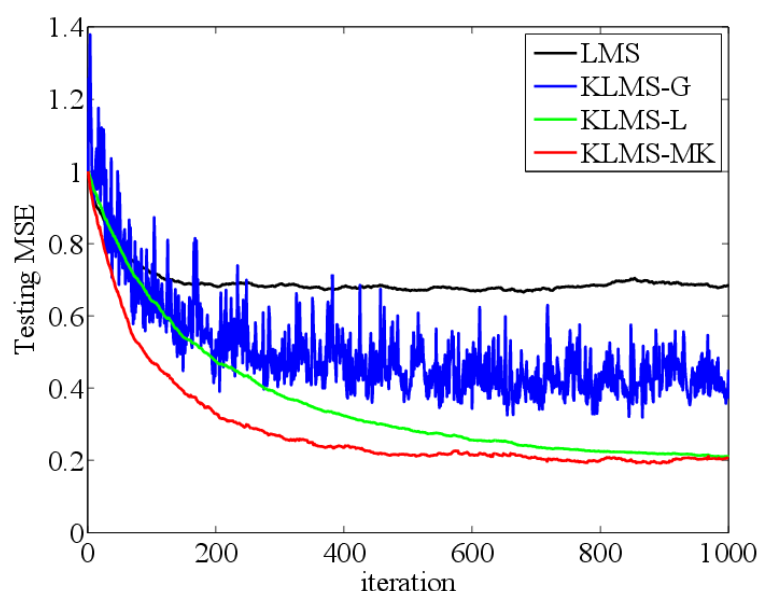


图 3.3 高斯噪声下不同算法在非线性回归模型的性能曲线



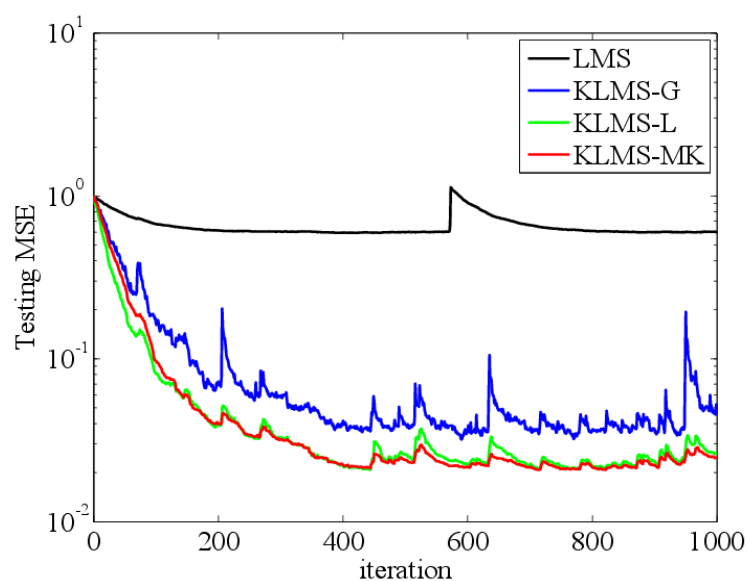


图 3.4 高斯噪声下不同算法在非线形回归模型的性能曲线

### 3.4 本章小结

为了解决 KLMS 中高斯核函数和拉普拉斯核函数的权衡问题, KLMS-MK 在本章中被提出。相比较于 KLMS, KLMS-MK 通过自适应混合参数结合了高斯核和拉普拉斯核, 提供了更快的收敛速度和更高的估计精度将高斯和拉普拉斯内核的自适应混合参数。同时, 证明了混合参数的收敛性。仿真结果表明, KLMS-MK 从收敛速度和估计精度角度实现了性能优势。

## 第四章 基于最大相关熵的仿射投影算法

前面提到的经典的线性自适应滤波算法和经典的核自适应滤波算法，都是基于 MSE 准则的。但是在某些非高斯噪声的情况下，它们的滤波性能不太理想；其原因在于：它们仅仅利用误差的二阶统计信息。基于最大相关熵的自适应滤波算法在面对较大的异常值或脉冲噪声时候表现了良好的鲁棒性，具有较好的滤波性能，因此 MCC、GMCC 都是可供选择的误差准则。在本文中，我们将最大相关熵分为：普通最大相关熵和广义最大相关熵，它们分别基于 MCC 和 GMCC。此外，我们将仿射投影算法分为：非核的仿射投影算法和核仿射投影算法。我们将系统的研究基于最大相关熵的仿射投影算法。APA 可以看成是 LMS 和 RLS 两者的折中，它能够均衡算法的计算复杂度和收敛速度。同时，仿射投影算法的更新形式更加的多样化，为接下来更新形式的研究提供理论基础。

### 4.1 最大相关熵

#### 4.1.1 最大相关熵准则

作为两个随机变量  $X$  和  $Y$  之间的相似性度量<sup>[62]</sup>，相关熵的定义如下：

$$\begin{aligned} V(X, Y) &= E[\kappa_h(X, Y)] \\ &= \int \kappa_h(x, y) dF_{X,Y}(x, y) \end{aligned} \quad (4-1)$$

其中： $\kappa_h(\cdot, \cdot)$  表示一个核宽度大小为  $h$  控制的 Mercer 核， $F_{X,Y}(x, y)$  表示  $X$  和  $Y$  的联合分布函数。它利用了将输入空间非线性映射到更高维特征空间的核技巧。在目前的工作中，除非另有说明，相关熵中的核函数为高斯核，表示为：

$$\begin{aligned} \kappa_h(x, y) &= \kappa_h(x - y) = G_h(e) \\ &= \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{e^2}{2h^2}\right) \\ &= \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{(x-y)^2}{2h^2}\right) \\ &= \frac{1}{\sqrt{2\pi}h} \exp(-\lambda e^2) \end{aligned} \quad (4-2)$$

这里， $e = x - y, \lambda = 1/2h^2$ 。在大多数的实际情况中，联合分布函数  $F_{X,Y}(x, y)$  通常是未知的，只有有限的数据  $\{(x(n), y(n))\}_{n=1}^N$  是可用的。在这些情况下，我们可以用下面的公式估计相关熵，

$$\hat{V}_{N,h} = \frac{1}{N} \sum_{n=1}^N \kappa_h(x(n) - y(n)) \quad (4-3)$$

在最大相关准则（MCC）下，自适应滤波器将通过最大化期望响应和滤波器输出之间的相关性来训练，公式如下：

$$\max_{\omega} J_{MCC} = \frac{1}{N} \sum_{n=1}^N \kappa_h(e(n)) \quad (4-4)$$

#### 4.1.2 广义最大相关熵准则

相关熵  $V(X, Y)$  也可以表示为下式：

$$V(X, Y) = E[\varphi(X)^T \varphi(Y)] = \text{Tr}(\mathbf{R}_{\varphi(X)\varphi(Y)}) \quad (4-5)$$

这里， $\varphi(\cdot)$  为由  $\kappa$  而来的一个非线性映射，同时满足： $\varphi(X)^T \varphi(Y) = \kappa(X, Y)$ 。因此，相关熵本质上是映射特征空间数据的二阶统计量。

广义高斯密度（GCD, generalized Gaussian density）<sup>[63]</sup> 函数为众人所知，具有零均值的广义高斯密度（GGD）函数由下式给出：

$$G_{\alpha,\beta}(e) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left|\frac{e}{\beta}\right|^\alpha\right) = \gamma_{\alpha,\beta} \exp(-\lambda|e|^\alpha) \quad (4-6)$$

其中， $\Gamma(\cdot)$  是伽玛函数， $\alpha > 0$  是形状参数， $\beta > 0$  是带宽参数， $\lambda = 1/\beta^\alpha$  是核参数， $\gamma_{\alpha,\beta} = \alpha/2\beta\Gamma(1/\alpha)$  为归一化常数。该参数对称分布族包括作为特例的高斯（ $\alpha = 2$ ）和拉普拉斯（ $\alpha = 1$ ）分布。当  $\alpha \rightarrow \infty$  时，GGD 密度在  $(-\beta, \beta)$  上逐点收敛为均匀密度。

在本文中，我们用 GCD 函数作为相关熵的核函数，表示如下：

$$V_{\alpha,\beta}(X, Y) = E[G_{\alpha,\beta}(X - Y)] \quad (4-7)$$

为了区分具有高斯核的相关熵，我们称上式为广义相关熵。显然，具有高斯核的相关熵对应于  $\alpha = 2$  的广义相关熵。

值得注意的是，在广义相关熵中核函数不一定满足 Mercer 的条件。实际上当且仅当  $0 < \alpha \leq 2$  时，核函数  $\kappa(x, y) = G_{\alpha,\beta}(x - y)$  是正定的。

广义相关熵的样本平均估计量为：

$$\hat{V}_{\alpha,\beta}(X, Y) = \frac{1}{N} \sum_{i=1}^N G_{\alpha,\beta}(x_i - y_i) \quad (4-8)$$

## 4.1.3 广义最大相关熵的基本性质

这部分，主要介绍广义相关熵的几个基本性质。通过这些基本性质，使我们对于广义相关熵有了更充分的认识，也可以从侧面说明在信号处理中对熵研究的意义。

性质 1: 同时  $V_{\alpha,\beta}(X,Y)$  具有对称性, 即  $V_{\alpha,\beta}(X,Y) = V_{\alpha,\beta}(Y,X)$ 。

性质 2:  $V_{\alpha,\beta}(X,Y)$  是正定的并且是有界的,  $0 < V_{\alpha,\beta}(X,Y) \leq G_{\alpha,\beta}(0) = \gamma_{\alpha,\beta}$ , 当且仅当  $X = Y$  时, 它达到最大值。

性质:3: 广义相关熵包含误差变量  $E_{X-Y} = X - Y$  的高阶绝对矩,

$$V_{\alpha,\beta}(X,Y) = \gamma_{\alpha,\beta} \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{n!} E[|X-Y|^{\alpha n}] \quad (4-9)$$

当核参数  $\lambda$  足够小时, 能够得到  $V_{\alpha,\beta}(X,Y) \approx \gamma_{\alpha,\beta} (1 - \lambda E[|X-Y|^{\alpha}])$ 。在这种情况下, 广义相关熵可近似为误差  $E_{X-Y}$   $\alpha$  阶绝对时刻的仿射线性函数。

性质 4: 假设样本  $\{(x_i, y_i)\}_{i=1}^N$  从联合概率密度函数  $p_{XY}(x, y)$  得出。假设  $\hat{p}_E(e)$  是来自样本的误差 PDF 的 Parzen 估计, 其中 GGD 密度函数  $G_{\alpha,\beta}$  作为 Parzen 窗口核函数。能够得到:

$$\hat{V}_{\alpha,\beta}(X,Y) = \hat{p}_E(0) \quad (4-10)$$

$$\text{这里, } \hat{p}_E(e) = \frac{1}{N} \sum_{i=1}^N G_{\alpha,\beta}(e - e_i)。$$

性质 5: 对于  $0 < \alpha \leq 2$  的情况, 广义相关熵是映射空间数据的二阶统计量。

性质 6: 令  $\mathbf{X} = [x_1, \dots, x_N]^T, \mathbf{Y} = [y_1, \dots, y_N]^T$ , 广义相关诱导度量 (GCIM, generalized correntropy induced metric),  $GCIM(\mathbf{X}, \mathbf{Y}) = \sqrt{\hat{J}_{GC-loss}(X,Y)}$ 。当  $0 < \alpha \leq 2$  时, GCIM 是在  $N$  维样本向量空间中定义的一个度量。

$$\text{性质 7: 当 } \lambda \rightarrow 0_+ \text{ (或者 } x_i = 1, \dots, N), \text{ 函数 } L_{\alpha,\beta}(\mathbf{X}) = \left( \frac{N}{\lambda \gamma_{\alpha,\beta}} \hat{J}_{GC-loss}(X, 0) \right)$$

趋近于  $\bar{X}$  的  $L_{\alpha}$  范数, 即

$$L_{\alpha,\beta}(\mathbf{X}) \approx \|\mathbf{X}\|_{\alpha} = \left( \sum_{i=1}^N |x_i|^{\alpha} \right)^{1/\alpha}, \lambda \rightarrow 0 \quad (4-11)$$

性质 8: 假设  $|x_i| > \delta, \forall i: x \neq 0$ , 这里  $\delta$  是一个小的正整数, 能够得到:

$$\min_{\mathbf{X} \in \Omega} L_{\alpha, \beta}(\mathbf{X}) \sim \min_{\mathbf{X} \in \Omega} L_{\alpha, \beta} \|\mathbf{X}\|_0, \lambda \rightarrow \infty \quad (4-12)$$

这里,  $\Omega$  表示  $\mathbf{X}$  的可行集合。

性质 9: 令  $\mathbf{e} = [e_1, \dots, e_N]^T$ , 下面的语句能够成立。

(1) 如果  $0 < \alpha \leq 1$ , 对于所有的  $e_i \neq 0 (i = 1, \dots, N)$ , GC-loss  $\hat{J}_{GC-loss}$  是凹的。

(2) 如果  $\alpha > 1$ , 对于所有的  $0 < |e_i| \leq [(\alpha - 1) / \alpha \lambda]^{1/\alpha} (i = 1, \dots, N)$ , GC-loss  $\hat{J}_{GC-loss}$  是凸的。

#### 4.1.4 广义最大相关熵算法

通常, 基于 GMC 的自适应滤波算法都有着如下的代价函数:

$$J_{GMC} = E \left[ \exp(-\lambda |e|^\alpha) \right] \quad (4-13)$$

这里, 误差为:

$$e(n) = d(n) - y(n) = d(n) - \mathbf{W}^T \mathbf{X}(n) \quad (4-14)$$

其中,  $\mathbf{W} = [\omega_1, \omega_2, \dots, \omega_m]^T$ ,  $\mathbf{X}(n) = [x(n), x(n-1), \dots, x(n-m+1)]$ 。

基于代价函数(4-13), 运用随机梯度法, 可以简单地推导 GMC 算法的权重更新为:

$$\begin{aligned} \mathbf{W}(n+1) &= \mathbf{W}(n) + \mu \frac{\partial}{\partial \mathbf{W}(n)} \exp(-\lambda |e|^\alpha) \\ &= \mathbf{W}(n) - \mu \lambda \alpha \exp(-\lambda |e|^\alpha) |e(n)|^{\alpha-1} \text{sign}(e(n)) \frac{\partial e(n)}{\partial \mathbf{W}(n)} \\ &= \mathbf{W}(n) + \eta \exp(-\lambda |e|^\alpha) |e(n)|^{\alpha-1} \text{sign}(e(n)) \mathbf{X}(n) \end{aligned} \quad (4-15)$$

这里  $\eta = \mu \lambda \alpha$  为步长参数。

通过理论推导, 我们发现以下几点:

1、当  $\alpha = 2$ , GMC 算法权重更新变为 MC 算法的权重更新, 即:

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \eta \exp(-\lambda e^2(n)) e(n) \mathbf{X}(n) \quad (4-16)$$

2、权重更新公式(4-17), 也可以写为:

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \eta(n) |e(n)|^{\alpha-1} \text{sign}(e(n)) \mathbf{X}(n) \quad (4-17)$$

这里  $\eta(n) = \eta \exp(-\lambda |e|^\alpha)$ 。因此, 可以将 GMCC 算法看作是具有  $p = \alpha$  和可变步长  $\eta(n)$  的 LMP 算法。LMP 算法是基于最小平均 p 次方 (LMP, least mean P-power) [64] 准则推导获得的。

3、当  $\eta(n) \rightarrow 0+$ ，得到  $\eta(n) \rightarrow \eta$ 。在这种情况下，GMCC 算法简化为传统的 LMP 算法，其中  $p = \alpha$ ：

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \eta |e(n)|^{\alpha-1} \text{sign}(e(n)) \mathbf{X}(n) \quad (4-18)$$

特别的是，当  $\alpha = 2$  时，上式成为众所周知的 LMS 算法。当  $|e(n)| \rightarrow \infty$  时，我们有  $\eta(n) \rightarrow 0$ 。因此，大的误差对滤波器权重影响不大。这意味着 GMCC 算法对于大的异常值（或脉冲噪声）的处理是十分有效的。

运用核方法在 RKHS 中重新推导 GMC，能够得到 KGMC。KGMC 有如下的更新形式：

$$\begin{cases} \boldsymbol{\Omega}(0) = 0 \\ e(n) = d(n) - \boldsymbol{\Omega}(n-1)^T \boldsymbol{\varphi}(\mathbf{x}(n)) \\ \boldsymbol{\Omega}(n) = \boldsymbol{\Omega}(n-1) + \eta(n) |e(n)|^{\alpha-1} \text{sign}(e(n)) \boldsymbol{\varphi}(\mathbf{x}(n)) \end{cases} \quad (4-19)$$

与前面核算法推到类似，上式也可以表述为：

$$\begin{cases} f_0 = 0 \\ e(n) = d(n) - f_{n-1}(\mathbf{x}(n)) \\ f_n = f_{n-1} + \eta(n) |e(n)|^{\alpha-1} \text{sign}(e(n)) \kappa_{h_2}(\mathbf{x}(n), \cdot) \end{cases} \quad (4-20)$$

## 4.2 基于最大相关熵的仿射投影算法

由上面的介绍可知，普通最大相关熵是当  $\alpha = 2$  的广义最大相关熵的一个特例。因此，在这一部分我们将最大相关熵准则运用到仿射投影算法中，对其进行理论推导，从而获得基于最大相关熵的仿射投影算法。

### 4.2.1 基于广义最大相关熵的非核仿射投影算法

设横向自适应滤波器的阶数为  $L$ ，滤波器的系数为  $\omega_i(n) (i = 0, 1, \dots, L-1)$ ，APA 的误差为：

$$e(n) = d(n) - y(n) = d(n) - \mathbf{X}^T(n) \mathbf{W}(n) \quad (4-21)$$

其中输入向量为：

$$\mathbf{X}(n) = [\mathbf{x}(n-K+1), \dots, \mathbf{x}(n)]_{L \times K} \quad (4-22)$$

滤波器的权系数向量为：

$$\mathbf{W}(n) = [\omega_0(n), \omega_1(n), \dots, \omega_{L-1}(n)] \quad (4-23)$$

基于广义最大相关熵的非核仿射投影算法，我们将其表示为：APA-GMC。它的代价函数表示如下：

$$\mathbf{J}_{APA-GMC} = \exp \left( - \left( -\lambda \left| \mathbf{e}(n) \right|^\alpha \right) \right) \quad (4-24)$$

则：

$$\frac{\partial \mathbf{J}}{\partial w} = \alpha \lambda U_i \exp \left( -\lambda \left| \mathbf{e}(n) \right|^\alpha \right) \left| \mathbf{e}(n) \right|^{\alpha-1} \text{sign}(\mathbf{e}(n)) \quad (4-25)$$

类似于 APA 的推导方法，可推导 APA-GMC 的权重更新方程为：

$$\mathbf{W}(n) = \mathbf{W}(n-1) + \mu \alpha \lambda \mathbf{X}(n) \mathbf{G}_1(n) \text{sign}(\mathbf{e}(n)) \quad (4-26)$$

$$\text{其中：} \mathbf{G}_1(n) = \begin{pmatrix} \exp(-\lambda \left| \mathbf{e}(n) \right|^\alpha) \left| \mathbf{e}(n) \right|^{\alpha-1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \exp(-\lambda \left| \mathbf{e}(n) \right|^\alpha) \left| \mathbf{e}(n) \right|^{\alpha-1} \end{pmatrix}。$$

APA-GMC 是运用最大相关熵准则，由 APA 推导而来。它不涉及高维空间，所以算法推导比较容易，这里简单介绍了一下。APA-GMC 具有 GMC 的优点，对于大的异常值（或脉冲噪声）的滤波作用是非常好的。接下来将介绍基于广义最大相关熵的核仿射投影算法，我们把该算法简称为 APA-GMC。最终，APA-GMC 总结在**算法 4.1**。

**算法 4.1: 基于最大相关熵的非核仿射投影算法****输入:**输入—输出对为  $\{\mathbf{x}(n), d(n)\}, n=1, 2, \dots$ **初始化:**让学习率:  $\mu > 0$  和  $\lambda > 0$ ,核宽度参数:  $\beta > 0$ ,**计算:**1) 计算参数矩阵  $\mathbf{G}_1(n)$ :

$$\mathbf{G}_1(n) = \begin{pmatrix} \exp(-\lambda |e(n)|^\alpha) |e(n)|^{\alpha-1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \exp(-\lambda |e(n)|^\alpha) |e(n)|^{\alpha-1} \end{pmatrix}$$

2) 权重更新

$$\mathbf{W}(n) = \mathbf{W}(n-1) + \mu \alpha \lambda \mathbf{X}(n) \mathbf{G}_1(n) \text{sign}(e(n))$$

3) 计算输出:

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

**4.2.2 基于广义最大相关熵的核仿射投影算法**

KAPA-GMC 是 APA-GMC 在 RKHS 中推导得来。在 KAPA-GMC 的推导过程中, 我们需要一个非线性映射  $\varphi(\cdot)$ , 通过它我们可以将输入数据  $\mathbf{x}(n)$  映射到 RKHS 中, 重复利用过去的  $K$  个输入信号。我们用  $\varphi(\mathbf{x}(n))$  表示映射后的输入, 用  $\boldsymbol{\Omega}(n)$  表示相应的  $n$  时刻的权重矢量。

原始的输入向量为:

$$\mathbf{X}(n) = [\mathbf{x}(n-K+1), \dots, \mathbf{x}(n)]_{L \times K} \quad (4-27)$$

经过映射后我们表示输入向量为:

$$\boldsymbol{\Phi}(n) = [\varphi(\mathbf{x}(n-K+1)), \dots, \varphi(\mathbf{x}(n))] \quad (4-28)$$

滤波器的权系数向量为:

$$\mathbf{W}(n) = [\omega_0(n), \omega_1(n), \dots, \omega_{L-1}(n)] \quad (4-29)$$



KAPA-GMC 的代价函数为:

$$J = \sum_{j=n-k+1}^n \exp\left(-\lambda |e(j)|^\alpha\right) \quad (4-30)$$

其中,  $e(j) = d(j) - \mathbf{w}^T(n-1)\varphi(\mathbf{x}(j))$ 。

由公式(4-30)可以求得:

$$\frac{J(n)}{w(n-1)} = -\lambda \alpha \mathbf{G}_2(n) \text{sign}(\mathbf{e}(n)) \boldsymbol{\Phi}(n) \quad (4-31)$$

$$\text{其中 } \mathbf{G}_2(n) = \begin{pmatrix} \exp(-\lambda |e(n)|^\alpha) |e(n)|^{\alpha-1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \exp(-\lambda |e(n)|^\alpha) |e(n)|^{\alpha-1} \end{pmatrix}。$$

从而, KAPA-GMC 的权重更新表达式表示如下:

$$\begin{aligned} \boldsymbol{\Omega}(n) &= \boldsymbol{\Omega}(n-1) + \eta \boldsymbol{\Phi}(n) \mathbf{G}_2(n) \text{sign}(\mathbf{e}(n)) \\ &= \sum_{j=1}^n \alpha_j(n) \varphi(\mathbf{x}(j)) \end{aligned} \quad (4-32)$$

这里,  $\alpha_j(n)$  具体的表达形式如下:

$$\alpha_j(n) = \begin{cases} \eta \sum_{j=n-k+1}^n \exp\left(-\lambda |e(j)|^\alpha\right) (e(j))^{\alpha-1} \text{sign}(\mathbf{e}(n)) & j=i \\ \alpha_j(n) + \eta \mathbf{G}_2(n) \text{sign}(\mathbf{e}(n)) & i-K+1 \leq j \leq i-1 \\ \alpha_j(n) & 1 \leq j < i-K+1 \end{cases} \quad (4-33)$$

实际的输出为:  $\mathbf{y}(n) = \boldsymbol{\Phi}^T(n) \mathbf{w}(n)$ 。最终, KAPA-GMC 总结在 **算法 4.2**。

**算法 4.2: 基于最大相关熵的核仿射投影算法****输入:**输入—输出对为  $\{\mathbf{x}(n), d(n)\}, n=1, 2, \dots$ **初始化:**让学习率:  $\mu > 0$  和  $\eta > 0$ ,核宽度参数:  $h > 0, \beta > 0$ 初始化系数矢量:  $\mathbf{a}(1) = \mu d(1)$ 。当  $\{\mathbf{x}(n), d(n)\} (n > 1)$  可用时, 开始循环**计算:**

1) 计算参数矩阵:

$$\mathbf{G}_2(n) = \begin{pmatrix} \exp(-\lambda |\mathbf{e}(n)|^\alpha) |\mathbf{e}(n)|^{\alpha-1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \exp(-\lambda |\mathbf{e}(n)|^\alpha) |\mathbf{e}(n)|^{\alpha-1} \end{pmatrix}$$

2) 参数更新:

$$\alpha_j(n) = \begin{cases} \eta \sum_{j=n-k+1}^n \exp(-\lambda |e(j)|^\alpha) (e(j))^{\alpha-1} \text{sign}(e(n)) & j=i \\ \alpha_j(n) + \eta \mathbf{G}_2(n) \text{sign}(e(n)) & i-K+1 \leq j \leq i-1 \\ \alpha_j(n) & 1 \leq j < i-K+1 \end{cases}$$

3) 权重更新:

$$\begin{aligned} \boldsymbol{\Omega}(n) &= \boldsymbol{\Omega}(n-1) + \eta \boldsymbol{\Phi}(n) \mathbf{G}_2(n) \text{sign}(e(n)) \\ &= \sum_{j=1}^n \alpha_j(n) \boldsymbol{\phi}(\mathbf{x}(j)) \end{aligned}$$

4) 计算误差:  $e(n) = d(n) - y(n)$ 5) 计算输出:  $\mathbf{y}(n) = \boldsymbol{\Phi}^T(n) \mathbf{w}(n)$ **4.3 仿真实验**

在本章中, 主要研究了基于最大相关熵的仿射投影算法。将最大相关熵分为: 普通最大相关熵和广义最大相关熵; 将仿射投影算法分为: 非核仿射投影算法和核仿射投影算法。研究了基于最大相关熵的非核仿射投影算法: APA-GMC 和 APA-MCC, 还研究了基于最大相关熵的核仿射投影算法: KAPA-GMC 和 KAPA-MCC。此外, 我们还研究了 GMC 和 KGMC。同时, 推导了它们的权重更

新形式。在这一部分，我们将借助 MG 时间序列预测模型分别在非高斯环境中对 APA、KAPA、GMC、KGMC、APA-GMC 和 KAPA-GMC 进行仿真，并对得到的仿真图进行一定的介绍。同时，还仿真了不同  $\alpha$  值的条件下，APA 和 KAPA 的学习曲线；通过对所得学习曲线的分析，得到了一些认识和启示，这对我们下面的研究提供了一定的思路。

在本章的仿真实验中，在 MG 时间序列预测模型中对高斯环境下和非高斯环境下进行仿真实验。MG 混沌时间序列表现出混沌动力学、可预测性和周期性，由下式产生：

$$\frac{dx(t)}{dt} = -0.1x(t) + \frac{0.2x(t-\tau)}{1+x(t-30)^{10}} \quad (4-34)$$

这里为了获得离散序列  $x(n)$ ，取采样周期为 6，前八个输入为： $u(n)=[x(n-8), x(n-7), \dots, x(n-1)]^T$ ，来估计所需的输出  $x(n)$ 。训练和测试数据的长度分别为 3000 和 200。希望选择较为合适的算法参数，期望于获得理想的性能。在本次仿真实验中，设置的算法参数具体如下。设置相同的参数为：训练数据为 3000，测试数据为 50，所有算法学习率为 0.01，核函数的核宽度为 1，蒙特卡洛次数为 150， $\alpha = 2, \lambda = 0.01$ 。在非高斯环境中，选取  $\alpha$  稳定分布的噪声，其参数设置为  $V = [1.4, 0, 0.01, 0]$ 。训练数据为 3000，测试数据为 200，所有算法学习率为 0.01，核函数的核宽度为 1，蒙特卡洛次数为 100， $\alpha = 2, \lambda = 0.01$ 。在高斯环境中，训练数据受到加性高斯噪声的干扰，白噪声的均值为 0，方差为 0.0001。

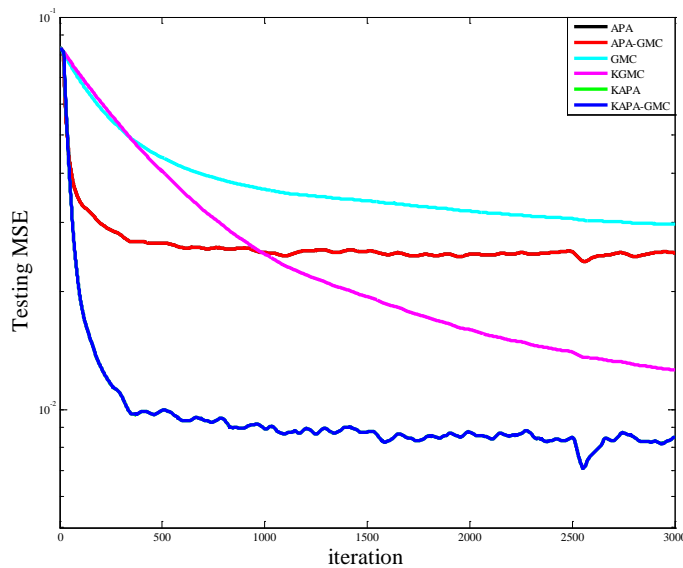
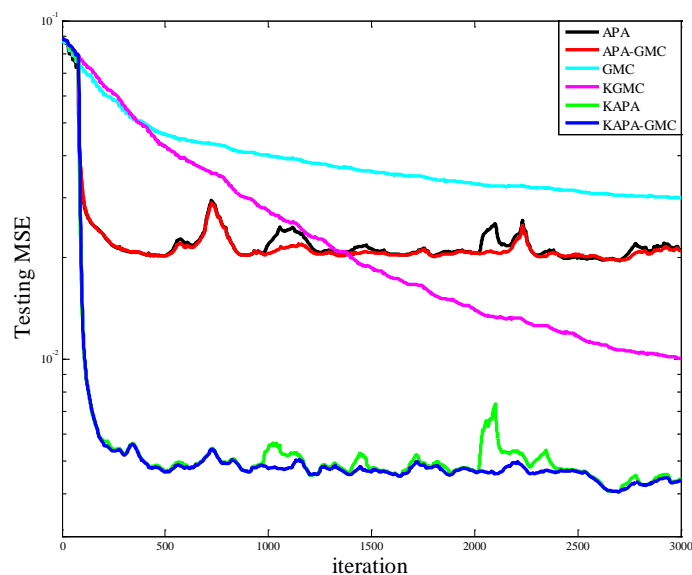
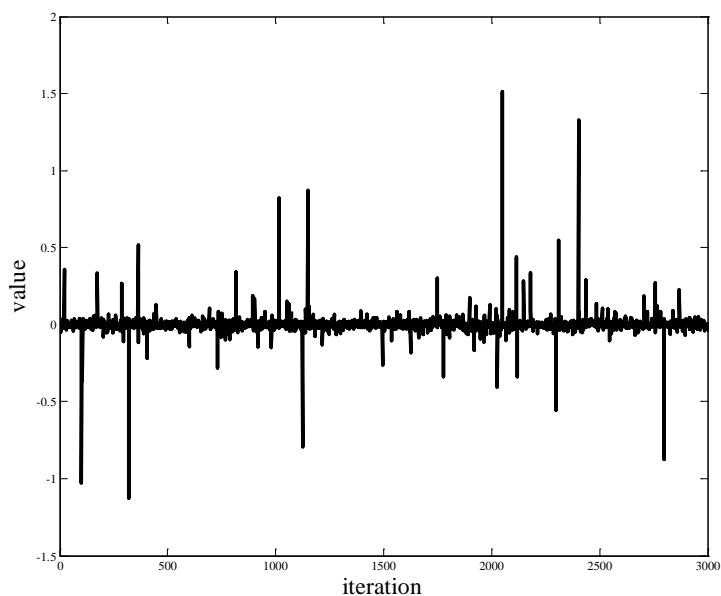


图 4.1 MG 时间序列预测中，不同算法在高斯噪声环境下的学习曲线

从图 4.1 可以在高斯环境中看出以下几点：（1）核自适应算法比其对应的非核自适应算法具有更好的滤波性能。（2）基于最大相关熵的仿射投影算法相比于基于最大相关熵的最小均方算法具有更快的收敛速度和更小的均方误差。（3）基于最大相关熵的仿射投影算法与其对应的仿射投影算法的滤波性能差别不大



(a)

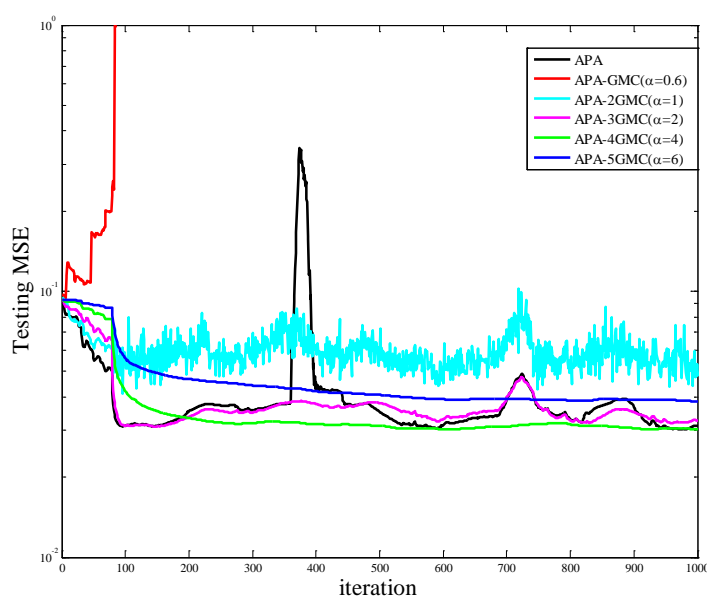


(b)

图 4.2 MG 时间序列预测中的仿真曲线：(a) 不同算法的学习曲线；(b) 与(a)对应的  $\alpha$  稳定分布的噪声

从图 4.2 可以看出  $\alpha$  稳定分布的噪声下：（1）核自适应算法比其对应的非核自适应算法具有更好的滤波性能。（2）基于最大相关熵的仿射投影算法相比于基于最大相关熵的最小均方算法具有更快的收敛速度和更小的均方误差。（3）APA 受较大异常值影响较大，而 APA-MCC 受影响较小，稳定条件下 APA 与 APA-MCC 的均方误差十分接近。（4）KAPA 受较大异常值影响较大，而 KAPA-MCC 受影响较小，稳定条件下 KAPA-MCC 具有相似的滤波性能。

此外，通过图 4.1 和图 4.2 还可以看出：（1）在高斯环境和非高斯环境下，非核仿射投影算法与基于最大相关熵的非核仿射投影算法在稳定条件下的均方误差十分接近。但基于最大相关熵的非核自适应算法抗噪声能力较强，表现了更好的滤波性能。（2）基于最大相关熵的核仿射投影算法在高斯和非高斯的环境下，都表现了良好的滤波性能。尤其是在非高斯环境中，KAPA-MCC 具有更好的滤波性能。（3）基于最大相关熵的仿射投影算法也具有较好的收敛速度。同时，也可以明显的看出基于最大相关熵的自适应算法对于较大异常值和脉冲噪声具有更好的滤波作用。



(a)

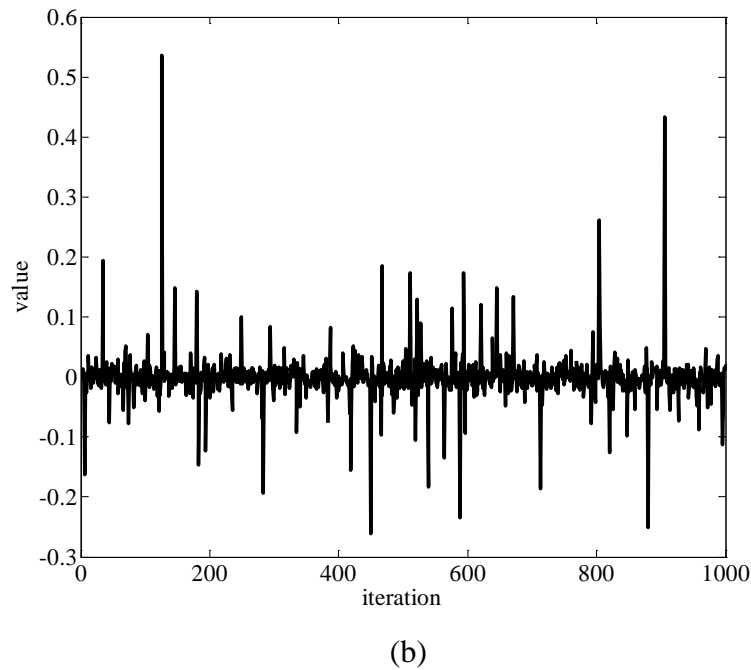


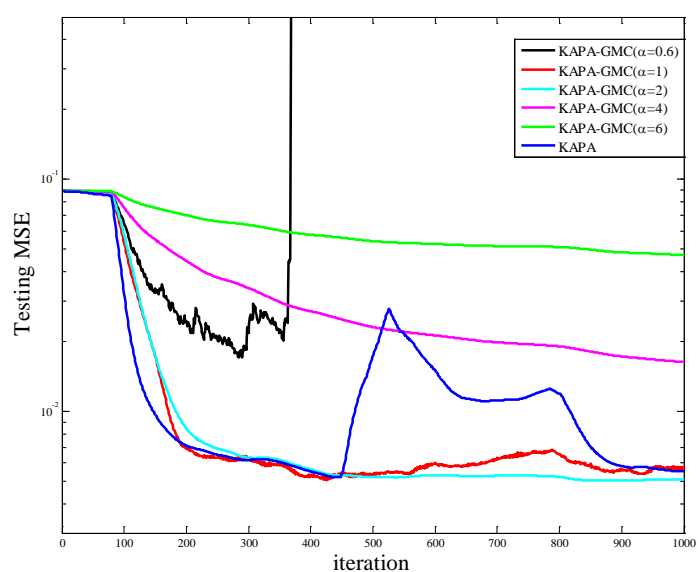
图 4.3 MG 时间序列预测中的仿真曲线：(a) APA-GMC 不同 $\alpha$ 值的学习曲线；(b) 与(a)对应的 $\alpha$ 稳定分布的噪声

表 4.3 不同 $\alpha$ 下 APA-GMC 的稳态 MSE

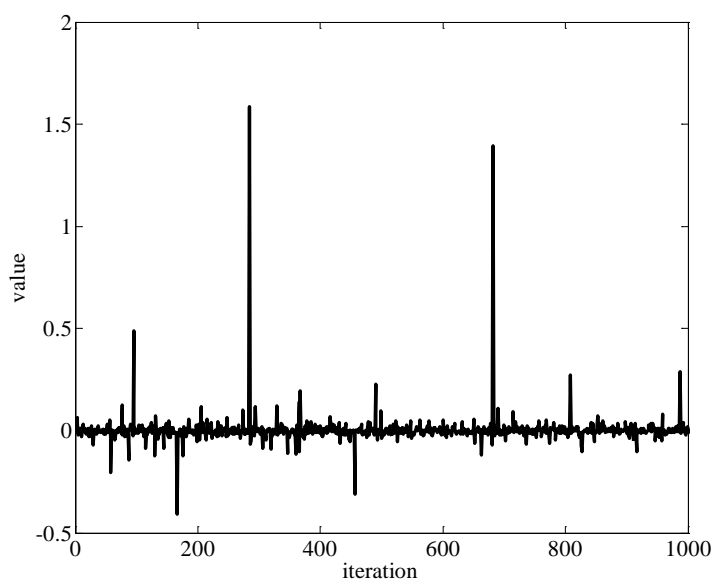
	$\alpha$	稳态MSE
APA	无	0.031165
APA-GMC	0.6	发散
APA-GMC	1	0.051266
APA-GMC	2	0.032797
APA-GMC	4	0.030312
APA-GMC	6	0.038528

在本仿真实验中，训练数据为 1000，测试数据为 200，算法的学习率为 0.05， $\lambda = 5$  其余参数与上面介绍的参数一致。从图 4.3 发现 $\alpha$  稳定分布的噪声下：（1）APA-GMC 在 $\alpha = [2, 4, 6]$ 时比 $\alpha = [0.6, 1]$ 时具有更好滤波性能。当 $\alpha = 0.6$ 时，算法发散。当 $\alpha = 4$ 时，算法性能最好。（2）APA 受噪声影响较大，会出现较大的误

差，而 APA-GMC 受影响较小，具有相对稳定的滤波性能。



(a)



(b)

图 4.4 MG 时间序列预测中的仿真曲线：(a) KAPA-GMC 不同  $\alpha$  值的学习曲线；  
(b)与(a)对应的  $\alpha$  稳定分布的噪声

表 4.4 不同  $\alpha$  下 KAPA-GMC 的稳态 MSE

	$\alpha$	稳态MSE
KAPA	无	0.005534
KAPA-GMC	0.6	发散
KAPA-GMC	1	0.005704
KAPA-GMC	2	0.005086
KAPA-GMC	4	0.016363
KAPA-GMC	6	0.047260

在本仿真实验中，训练数据为 1000，测试数据为 200，算法的学习率为 0.007， $\lambda = 5$ ，核函数的核宽度为 8，其余参数与上面介绍的参数一致。从图 4.4 可以看出：（1）KAPA-GMC 在  $\alpha = [1, 2, 4]$  时比  $\alpha = [0.6, 6]$  时具有更好滤波性能， $\alpha = 2$  时性能最好， $\alpha = 0.6$  时性能最差。（2）KAPA-GMC 的滤波性能比 KAPA 的滤波性能稳定，KAPA-GMC 具有更好的鲁棒性。

#### 4.4 本章小结

在本章中系统化的介绍了最大相关熵的相关知识，并研究了基于最大相关熵的仿射投影算法。主要的研究了基于广义最大相关熵的仿射投影算法，并推导了它们的权重更新形式。在此基础上，我们 MG 时间序列预测模型中仿真了在高斯和非高斯这两种情况下，不同算法的学习曲线。同时我们还仿真了  $\alpha$  值不同的情况下 APA-GMC 和 KAPA-GMC 的学习曲线。可以看出，基于最大相关熵的仿射投影算法在所给数据中有大的异常或被脉冲噪声干扰时，表现出了优越的滤波性能。基于最大相关熵的仿射投影算法理论，在自适应信号处理中起着积极的作用。





## 第五章 总结与展望

### 5.1 总结

自适应滤波器经过研究者的完善与发展,已经发展的比较成熟。但是,由于科技的不断发展和人们认识程度的不断提高,自适应滤波器得到了快速发展。自适应滤波算法在自适应滤波器中具有核心地位,人们对于它的研究一直在进行。人们想尽各种方法,不断的提高算法的滤波性能。同时,算法的滤波性能在不同的情况下也有着极大的差异。如何在某种特定的环境中,选择合适的自适应滤波算法,也是人们需要思考的问题。为了获得收敛速度快,计算复杂度低,滤波性能好以及所受外界环境影响小的自适应滤波算法,研究者们通常对算法的代价函数和更新形式进行一定的改进与完善。在本论文中我们主要研究了以下两个方面:

(1) 提出了一种混合核函数,并在此核函数的基础上,提出了一种混合核最小均方算法,该算法改善了 **KLMS** 的滤波性能。**KLMS-MK** 把凸组合的方法应用到 **KLMS** 的核函数上。同时我们还从理论上证明了混合参数的收敛性。在 **KLMS-MK** 中,收敛速度和稳态均方误差得到了进一步的改善。由于 **KLMS-MK** 具有高斯核和拉普拉斯核两者的优点,它的鲁棒性能更好。最后,我们还进行了仿真实验,验证了 **KLMS-MK** 的优越性。

(2) 系统的研究了基于最大相关熵的仿射投影算法。在高斯和非高斯环境中,分别进行了基于最大相关熵算法的仿真实验。同时我们还在非高斯环境中仿真了不同 $\alpha$ 值时 **APA-GMC** 和 **KAPA-GMC** 的学习曲线。通过所获得的学习曲线图,我们获得了一些认识,同时我们也获得了一些启示,这为我们接下来的研究提供了一定的思路。本论文主要从代价函数改善的角度进行研究仿真,在此基础上我们还可以从更新方式上进行优化、研究,以期获得更好的自适应滤波算法。

### 5.2 展望

尽管本论文中,对上面的两个地方进行了一定的改善,但还存在一些需要思考的问题:

(1) 在本论文中,我们提出了一种混合核函数,由通用的单核函数变为多核函数。我们知道核函数的种类是较多的,那么在多核函数的组成和构造方式上,是否有最优的选择?在多核函数的构造方面也是我们需要考虑的一个问题

(2) 最大相关熵和广义的最大相关熵对自适应滤波算法的代价函数进行了一定的改进。尤其是,广义最大相关熵对于代价函数的改进比较大。同时也增添了几个人为可控的变量。由此很容易的想到,可以在此基础上对自适应算法的更新方式进行一定的改进。变量之间是否存在着一一定的联系,结合现有的更新方式,

是否可以提出一些新的更新方式等等一系列的问题也是需要我们考虑的。

(3) 在广义最大相关熵算法中,  $\alpha$  值到底是怎么影响算法的。由于所处噪声环境的复杂, 对于这一方面的研究十分困难, 目前还没有定论。

---

## 参考文献

- [1] Wiley. Journal of the association for information science and technology. American Journal of Comparative Law, 2014, (4): 704-710.
- [2] 孙小礼, 冯国瑞. 信息科学技术与当代社会. 北京: 高等教育出版社, 2001.
- [3] 蔡芝蔚. 计算机技术发展研究. 电脑与电信, 2008, (2): 54-55.
- [4] Mitchell T M. Machine learning. China: China Machine Press : McGraw-Hill Education (Asia), 2003.
- [5] 陈小燕. 机器学习算法在数据挖掘中的应用. 陕西: 现代电子技术, 2015, (20): 11-14.
- [6] Haykin S. Adaptive filter theory (2nd ed.) . Prentice Hall, 2002, 4(96): 469-490.
- [7] Sayed A H. Adaptive filters. Wiley : Wiley-IEEE Press, 2008.
- [8] 蔡卫菊. 线性自适应滤波算法综述. 科技资讯, 2011, (36): 233-233.
- [9] Erdogmus D, Principe J C. An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems. Signal Processing IEEE Transactions on, 2002, 50(7): 1780-1786.
- [10] Gay S L. Affine projection algorithms. American : John Wiley & Sons, Inc, 2003.
- [11] Liu W, Park I, Wang Y. Extended kernel recursive least squares algorithm . IEEE Transactions on Signal Processing, 2009, 57(10): 3801-3814.
- [12] Liu W, Pokharel P P, Principe J C. The kernel least-mean-square Algorithm . IEEE Transactions on Signal Processing, 2008, 56(2): 543-554.
- [13] Ding F, Chen T. Identification of Hammerstein nonlinear Armax systems. British: Pergamon Press, Inc, 2005.
- [14] Norquay S J, Palazoglu A, Romagnoli J. Model predictive control based on Wiener models . Chemical Engineering Science, 1998, 53(1): 75-84.
- [15] Figueiredo R J P D. The Volterra and Wiener theories of nonlinear systems . Proceedings of the IEEE, 2005, 70(3): 316-317.
- [16] Billings S A, Zheng G L. Radial basis function network configuration using genetic algorithms. Elsevier Science Ltd, 1995.
- [17] Williams R J, Zipser D. A learning algorithm for continually running fully recurrent neural networks. American: MIT Press, 1989.
- [18] Panagopoulos H. Design of PI controllers based on non-convex optimization . British: Pergamon Press, Inc, 1998.

- 
- [19] Greer N P. Kernel adaptive filter . Patho Publishing, 2012.
- [20] Chandrasekaran V, Parrilo P A, Willsky A S. Rejoinder: Latent variable graphical model selection via convex optimization . Annals of Statistics, 2010, 40(4): 1610 - 1613.
- [21] 陈乾. 核自适应滤波算法研究: [硕士学位论文]. 湖北: 华中师范大学, 2014.
- [22] 冯端, 冯步云. 熵. 北京: 科学出版社, 1992.
- [23] Xiao X Y, Ma C, Yang H G, et al. Stochastic estimation of equipment sensitivity to voltage sag based on voltage sag severity index and maximum entropy principle . Proceedings of the Csee, 2009, 29(31): 115-121.
- [24] He Y, Wang F, Yang J, et al. Kernel adaptive filtering under generalized maximum correntropy criterion. Vancouver: The International Joint Conference on Neural Networks, 2016.
- [25] Buhalis D, O'Commor P. Information communication technology revolutionizing tourism. Tourism Recreation Research, 2006, 30(3): 196-209.
- [26] Robbins H. An extension of Wiener filter theory to partly sampled systems. circuit theory Ire transactions on, 1959, 6(4): 362-370.
- [27] King-Hele D G. Space Technology. Nature, 1962, 196(4854): 506-507.
- [28] Welch G, Bishop G. An untroduction to the Kalman filter. American: University of North Carolina at Chapel Hill, 2001.
- [29] 秦永元, 张洪钺, 汪叔华. 卡尔曼滤波与组合导航原理. 陕西: 西北工业大学出版社, 2015.
- [30] Davisson L D. A theory of adaptive filtering. IEEE Transactions on Information Theory, 1966, 12(2): 97-102.
- [31] Saitoh S. Theory of reproducing kernels. Transactions of the American Mathematical Society, 2003, 68(3): 337-404.
- [32] Liu W K, Jun S, Zhang Y F. Reproducing kernel particle methods . International Journal for Numerical Methods in Fluids, 1995, 20(10): 1655-1679.
- [33] Wegman E J. Reproducing kernel Hilbert spaces. John Wiley & Sons, Inc, 2004.
- [34] Ajzerman M A, Braverman E M, Rozonoehr L I. Theoretical foundations of the potential function method in pattern recognition learning. Automation & Remote Control, 1964, 25(6): Doc2.
- [35] Arden B W, Lee N H. A regular network for multicomputer systems. IEEE Transactions on Computers, 2006, C-31(1): 60-69.
- [36] Wang Q, Garrity G M, Tiedje J M, et al. Naive Bayesian classifier for rapid

- assignment of rRNA sequences into the new bacterial taxonomy. *Applied & Environmental Microbiology*, 2007, 73(16): 52-61.
- [37] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers. netherlands: Kluwer Academic Publishers, 1999.
- [38] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 2002, 5(2): 157-166.
- [39] Liu W, Principe J C, Haykin S. 2. Kernel least-mean-square algorithm. John Wiley & Sons, Inc, 2010.
- [40] Liu W, Principe J C. Kernel affine projection algorithms. *Eurasip Journal on Advances in Signal Processing*, 2008(1): 1-12.
- [41] Bergman S. The kernel function and conformal mapping. *Mathematical Gazette*, 1950, 5(35): 257.
- [42] Babaud J, Witkin A P, Baudin M, et al. Uniqueness of the gaussian kernel for scale-space filtering. *IEEE Trans Pattern Anal Mach Intell*, 1986, 8(1): 26-33.
- [43] Paclik P, Novovicova J, Pudil P, et al. Road sign classification using Laplace kernel classifier. *Pattern Recognition Letters*, 2000, 21(13 - 14): 1165-1173.
- [44] Entropy P T. Information processing and Thermodynamic entropy. *Stanford Encyclopedia of Philosophy*, 2009, (8), 45-49.
- [45] Rimoidl B, Urbanke R. Information theory. *The Communications Handbook - Second edition*, 2002, 133(3462).
- [46] Jaynes E T. Information theory and statistical mechanics . *Physical Review*, 1957, 106(4): 620-630.
- [47] Peng S, Chen B, Sun L, et al. Constrained maximum correntropy adaptive filtering. *Signal Processing*, 2016, 140(TENCON): 116-126.
- [48] Zhao S, Chen B, Principe J C. Kernel adaptive filtering with maximum correntropy criterion. *American: The International Joint Conference on Neural Networks*, 2011.
- [49] Zhao J, Zhang H. Kernel recursive generalized maximum correntropy. *IEEE Signal Processing Letters*, 2017, PP(99): 1-2.
- [50] Liu C, Qi Y, Ding W. The data-reusing MCC-based algorithm and its performance analysis. *Chinese Journal of Electronics*, 2016, 25(4): 719-725.
- [51] Benesty J, Rey H, Vega L R, et al. A nonparametric VSS nlms algorithm. *IEEE Signal Processing Letters*, 2006, 13(10): 581-584.
- [52] Gay S L, Tavathia S. Fast affine projection algorithm. *The International*

- Conference on Acoustics, Speech, and Signal Processing, 1995.
- [53] 刘华富. 支持向量机 Mercer 核的若干性质. 北京联合大学学报:自然科学版, 2005, 19(1): 45-46.
- [54] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Analysis & Machine Intelligence, 2002, 24(5): 603-619.
- [55] Roth M W. Survey of neural network technology for automatic target recognition . IEEE Transactions on Neural Networks, 1990, 1(1): 28-43.
- [56] Lin B, He R, Wang X, et al. The steady-state mean-square error analysis for least mean p -order algorithm. IEEE Signal Processing Letters, 2009, 16(3): 176-179.
- [57] Saad Y. Iterative method for sparse linear systems. 2003.
- [58] Platt J. A resource-allocating network for function interpolation. Neural Computation, 2008, 3(2): 213-225.
- [59] 吕金虎. 混沌时间序列分析及其应用. 湖北: 武汉大学出版社, 2002.
- [60] Bate S, Watts. 非线性回归分析及其应用. 北京: 中国统计出版社, 1998.
- [61] 方再根. 计算机模拟和蒙特卡洛方法. 北京: 北京工业学院出版社, 1988.
- [62] 刘宝生, 闫莉萍, 周东华. 几种经典相似性度量的比较研究. 计算机应用研究, 2006, 23(11): 1-3.
- [63] Do M N, Vetterli M. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. IEEE Trans Image Process, 2002, 11(2): 146-158.
- [64] Pei S C, Tseng C C. Adaptive IIR notch filter based on least mean p-power error criterion. IEEE Transactions on Circuits & Systems II Analog & Digital Signal Processing, 1993, 40(8): 525-528.

## 致 谢

从 2015 年第一次踏进西南大学校门，到 2018 年我即将离开我的母校，回顾这三年，我经历了丰富多彩的研究生生涯，同时我收获颇多。一方面：我收获了很多的宝贵知识和经验，同时认识了很多新的朋友。另一方面：由于自身的原因，走了一些歪路，经历了很多曲折。通过这段不太好的道路，收获了很多的教训，也深深地体会到老师、同学对我的关心与照顾。尤其是王老师对我学习、生活等各方面的指导与关心。在这期间，我给您添了很多麻烦，真的对不起您。总体来说，我的研究生生涯是成功的。通过这几年的学习、生活，我取得了很多的进步；在未来的日子里我会努力完善自己，使自己成为有用的人才。

首先要感谢导师王世元副教授，在我的研究生生涯中，王老师扮演着别人无法替代的重要角色。在学习中，王老师点拨迷津，同时为我们提供了科学的学习计划。在生活中，他对我们关怀备至，同时为我们指明了正确的人生道路。在我的研究生生涯中，能够师从王老师，我感到十分庆幸。在这里，向您表达我诚挚的敬意和感谢。同时，感谢电子信息工程学院所有教导过我、关心过我的老师。你们为我的研究生学业倾注了大量的心血，严谨治学的态度令我敬佩。

非常感谢一直支持与关心我的同学和朋友们，尤其是王万里、尹超和党路娟对我的帮助和鼓励。还要感谢我的师兄师姐郑云飞、冯亚丽和张文杰以及所有的朋友们。三年来，我们共同进步，非常感谢你们的关心与帮助。

在此要特别感谢我的母校——西南大学，母校为我们提供了宽阔的学习平台和良好的生活环境。在母校，我们不断的吸取新知识，获得新技能，为我们将来的生活、工作打下了坚实的基础。

需要特别感谢的是我的父母，他们是我求学道路上的坚实后盾。当我面临人生选择的时候，为我排忧解难。他们无私的爱与照顾是使我不断前进的源动力，父母之的养育之恩无以为报。

2018 年 4 月 7 日





## 攻读硕士期间已发表的学术论文

1. **Sun Qitang**, Dang Lujuan. Zheng Yunfei. Kernel Least Mean Square Algorithm With Mixed Kernel. 10th International Conference on Advanced Computational Intelligence, 2018.
2. Dang Lujuan, Zheng Yunfei. **Sun Qitang**. A novel least Cauchy error algorithm and its kernel extension. 9th IEEE International Congress on Image and Signal Processing, Biomedical Engineering and Informatics, 2016. (EI: 20171303496857)
3. Dang Lujuan, Wang Wanli. **Sun Qitang**. Hammerstein Adaptive Filter with Single Feedback under Minimum Mean Square Error. 20th International Conference on Information Fusion, 2017. (EI: 20173804176103)

## 攻读硕士期间参加的科研项目

1. 中央高校基本科研业务费专项——学生项目，XDJK2017D178，基于核方法的非线性自适应滤波器设计，2017/03-2018/03，1 万，已结题，参与。
2. 中央高校基本科研业务费专项——学生双创项目，XDJK2016E029，量化核自适应滤波器新方法的研究，2016/05-2016/12，1 万，已结题，参与。