

一种基于多尺度核学习的仿射投影滤波算法

李群生^{*①②} 赵 剌^① 寇 磊^③ 王进达^②

^①(北京航空航天大学仪器科学与光电工程学院 北京 100191)

^②(中国空空导弹研究院 洛阳 471009)

^③(中航工业自控所飞行器控制一体化重点实验室 西安 710065)

摘 要: 为了提高强非线性信号的噪声消除和信道均衡能力,在核学习自适应滤波方法的基础上,该文提出一种基于惊奇准则的多尺度核学习仿射投影滤波方法(SC-MKAPA)。在核仿射投影滤波算法的基础上,对核组合函数结构进行改进,将多个不同高斯核带宽作为可变参数,与加权系数共同参与滤波器的更新;利用惊奇准则将计算结果稀疏化,根据仿射投影算法的约束条件对惊奇测度进行改进,简化其方差项,降低了计算的复杂度。将该算法应用于噪声消除、信道均衡以及MG时间序列预测中,与多种自适应滤波算法及核学习自适应滤波算法进行仿真结果的对比分析,验证了该算法的优越性。

关键词: 自适应滤波;核学习;可变核带宽;多核仿射投影;惊奇准则

中图分类号: TN911.7, TP391

文献标识码: A

文章编号: 1009-5896(2020)04-0924-08

DOI: 10.11999/JEIT190023

An Affine Projection Algorithm with Multi-scale Kernels Learning

LI Qunsheng^{①②} ZHAO Yan^① KOU Lei^③ WANG Jinda^②

^①(School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China)

^②(Air-to-Air Missile Research Institute, Luoyang 471009, China)

^③(National Key Laboratory on Flight Vehicle Control Integrated Technology, Xi'an 710065, China)

Abstract: In order to improve the ability of noise elimination and channel equalization of strong non-linear signals, a Multi-scale Kernels learning Affine Projection filtering Algorithm based on Surprise Criterion (SC-MKAPA) is proposed on the basis of kernel learning adaptive filtering method. Based on the kernel affine projection filtering algorithm, the structure of the kernel combination function is improved, and the bandwidths of several different Gaussian kernels are taken as variable parameters to participate in the update of the filter together with the weighted coefficients. The calculation results are sparsed by using the surprise criterion, and the surprise measure is improved according to the constraints of the affine projection algorithm, which simplifies the variance term and reduces the calculation complexity. The algorithm is applied to noise cancellation, channel equalization, and Mackey Glass (MG) time series prediction. The simulation results are compared with the traditional adaptive filtering algorithm and the kernel learning adaptive filtering algorithm, it proves the superiority of the proposed algorithm.

Key words: Adaptive filters; Kernel learning; Variable kernel bandwidth; Affine projection with multi-kernels; Surprise criterion

1 引言

基于核函数学习算法在信号处理领域尤其是核

滤波理论的起步时间较晚^[1],基于核函数思想的自适应滤波算法最早由Frieb和Harrion^[2]于1999年提出,该算法首次以非在线训练方式将Mercer核与Widrow-Hoff最小均方算法结合,应用到函数拟合和非线性系统辨识中去。目前对核函数自适应滤波方法的研究主要集中在对单核^[3,4]和多核的优化^[5]以及核稀疏化^[6]方面。

由于单核滤波器的性能单一,且不同核函数之间的构造差异很大,即使是同一核函数,不同参数

收稿日期: 2019-01-09; 改回日期: 2019-07-30; 网络出版: 2020-01-11

*通信作者: 李群生 570658391@qq.com

基金项目: 国家自然科学基金(61233005), 航空基金(20160812004, 20160112002, 2016ZA12002)

Foundation Items: The National Natural Science Foundation of China (61233005), The Aviation Science Fund (20160812004, 20160112002, 2016ZA12002)

所导致的数据在特征空间的分布也有所不同,针对这种单核映射泛化能力弱的缺点,基于多核组合或混合核学习的方法逐渐成为了新的热点^[7-9]。文献^[10-12]提出了一种基于归一化最小均方误差(LMS)算法的多核自适应滤波方法,在字典设计中引入相干准则(Coherence Criterion, CC)并利用L1范数的正则化对字典进行稀疏化操作,使其在非线性的滤波过程中具有对核参数的不敏感性;Nishikawa和Nakazato^[13]针对KNLMS算法的收敛特性设计了一种混合核结构的自适应滤波器,但这种滤波器本质上还是基于单核的多个滤波器线性组合。而由文献^[14]提出的多核最小均方算法是一种真正的多核混合结构,利用多个核函数的加权组合作为映射机制,权系数的更新依赖系统输出的反馈,从而提高了时间序列的在线预测精度。文献^[15]则在多核自适应滤波算法中采用多字典和正则化策略,每个字典对应一个由多核生成的再生核希尔伯特空间集子集,并采用L2正则化机制避免过匹配现象。文献^[16]主要采用了多个高斯核函数的结构,在迭代中采用归一化方法解决了收敛速度的退化现象。在多核核函数尺度信息的最优利用方面,文献^[17]采用核回归中不同的时间尺度信息与改进的 γ -滤波器相结合,提出一种多核优化的自适应策略,并以递归的形式提高运算速度。文献^[18]则利用每个核函数的输入输出构造一个自适应偏离因子,来最小化不同尺度核函数的带来的负面效益,从而提高多核框架的鲁棒性。

在核稀疏化方面,Liu等人^[19]在核自适应滤波算法的可扩展性方面也做了深入的研究,提出了基于新奇准则的稀疏化核仿射投影算法(Kernel New Criterion-Affine Projection Algorithm, NC-KAPA);Richard等人^[20]将相干准则(CC)与核仿射投影算法相结合,使计算复杂度稀疏化,并在时间序列的在线预测进行了分析;Gao等人^[21]通过增加前馈-反馈环节,对字典进行稀疏正则化,并应用到核最小均方误差方法当中;Takizawa和Yukawa^[22]将基于函数空间和欧氏空间的同构下对核滤波器进行稀疏化,字典可以通过定义一个收缩因子和过去数据再利用来进行自适应改善。

总的看来,进行多核学习可以解决单核学习中对最佳核函数的选择和优化问题,同时能够并行处理数据当中的异构现象。但无论是单核滤波器还是多核滤波器,目前对核函数的优化仅限于对希尔伯特空间形式的变换,并没有涉及对核参数的优化,核参数的选择也仅限于经验或人为的调整。此外,解决核生长结构问题即稀疏化问题

的新奇准则(NC)和相干准则(CC)等也是在直觉层面上调整阈值等关键参数,也需要从数学角度自适应地调整其阈值。

因此,为了最大限度降低人为调整参数的影响,本文在利用多核凸组合的特性对权值系数进行更新的同时,将核函数的尺度因子作为参数进行更新,使得核函数在凸组合方面具有更高的自由度和自适应特性,并根据信息论的惊奇准则对其进行稀疏化选择,提高滤波算法的收敛特性。

2 基于新奇准则的多核仿射投影算法

采用仿射投影算法能够在改进失调和计算复杂度的同时提高稳态的收敛速度。因此将核映射与仿射投影算法相结合可以解决自适应滤波过程中所遇到的非线性问题并能提高收敛特性。

根据再生核理论可知,在线性空间内两个函数 f 和 g 的内积可以定义为

$$\langle f, g \rangle := \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j \kappa(\mathbf{x}_i, \mathbf{x}'_j) \quad (1)$$

其中,函数 f 和 g 可采用Mercer核来表示为

$$f(\cdot) = \sum_{i=1}^m \alpha_i \kappa(\cdot, \mathbf{x}_i) \quad (2)$$

$$g(\cdot) = \sum_{j=1}^n \beta_j \kappa(\cdot, \mathbf{x}'_j) \quad (3)$$

对于非线性映射 $\varphi: \mathbf{u} \rightarrow \phi(\mathbf{u})$,所有的 $f(\cdot)$ 组成由映射 φ 的像张成的线性空间。此时

$$\langle \kappa(\cdot, \mathbf{x}), \kappa(\cdot, \mathbf{x}') \rangle = \kappa(\mathbf{x}, \mathbf{x}') \quad (4)$$

根据再生核的Hilbert空间特性可知, Mercer核是正定的,所以也是再生核,那么再生核的线性组合也符合Hilbert空间特性。

由此设计线性加权组合的多核函数

$$K(\mathbf{u}_i, \mathbf{u}_j) = \sum_{k=1}^N \lambda_k \kappa_k(\mathbf{u}_i, \mathbf{u}_j) \quad (5)$$

其中 $\kappa_k(\mathbf{u}_i, \mathbf{u}_j)$ 为基本核函数。利用式(5)的多核组合核函数可以定义输入数据在特征空间中的内积。

假设在特征空间内参与更新的权向量为 ω ,就有最小化问题

$$\min_{\omega} \|d - \Phi^T \omega\|^2 \quad (6)$$

其中 Φ^T 为 $\mathbf{X}^T$ 在希尔伯特空间的映射矩阵,通过牛顿法可以得到 ω 的更新方程为

$$\begin{aligned} \omega(k) = & \omega(k-1) + \mu \Phi^T(k) [\Phi^T(k) \Phi(k) + \delta \mathbf{I}]^{-1} \\ & \cdot [d(k) - \Phi^T(k) \omega(k-1)] \end{aligned} \quad (7)$$

式中 μ 为步长因子, 由于 $\omega(k)$ 为输入向量在高维空间的隐映射, 因此无法直接求取, 可通过内积运算进行系统的更新。权向量可以按照如式(8)在特征空间内展开

$$\omega(k) = \sum_{i=1}^k \mathbf{a}_i(k) \varphi(k) \quad (8)$$

其中, $\mathbf{a}_i(k)$ 为权向量的扩展系数, 表示为^[12]

$$\mathbf{a}_i(k) = \begin{cases} \mu[\Phi^T(k)\Phi^T(k) + \varepsilon\mathbf{I}]^{-1} \cdot \left(\mathbf{d}(k) - \sum_{j=1}^{k-1} \mathbf{a}_j(k-1) \kappa(\mathbf{u}_j, \mathbf{u}_k) \right), & i = k \\ \mathbf{a}_i(k-1) + \mu[\Phi^T(k)\Phi^T(k) + \varepsilon\mathbf{I}]^{-1} \cdot \left(\mathbf{d}(k) - \sum_{j=1}^{k-1} \mathbf{a}_j(k-1) \kappa(\mathbf{u}_j, \mathbf{u}_k) \right), & k - K + 1 \leq i \leq k - 1 \\ \mathbf{a}_i(k-1), & 1 \leq i < k - K + 1 \end{cases} \quad (9)$$

基于核学习的权值更新往往配合稀疏模型来降低计算量和存储量, 采用新奇准则是一种常用的在线稀疏化准则, 其稀疏化结构是动态创建的, 学习过程中逐步向字典中添加新的样本, 从而使字典能够最大限度体现样本集的特征。通常设置两个阈值 D_1 , D_2 用来判断新样本是否需要加入字典。假设当前字典为

$$\text{Dic}(k) = \{\text{dic}_i\}_{i=1}^{M_k} \quad (10)$$

式中 dic_i 为第 i 个中心, M_k 为集的基数。对于每次更新的新样本对 $\{\mathbf{u}(k+1), \mathbf{d}(k+1)\}$, 首先计算该样本与字典最短距离

$$L = \min_{\text{dic}_i \in \text{Dic}(k)} \|\mathbf{u}(k+1) - \text{dic}_i\| \quad (11)$$

当且仅当 $L > D_1$, $e(k+1) > D_2$ 时, 新样本可以加入到字典中作为新的中心集。

3 多核仿射投影算法改进

3.1 核多尺度化

核函数定义了数据点之间的相似度, 是核方法中最关键的组成部分。多核学习的引入不仅保留了单核函数良好的非线性特性, 同时最大化地利用了不同核函数的优点。但是目前基于多核函数学习的算法仅仅停留在对函数的凸组合优化上, 通常情况下, 核参数的调整对核映射机制的影响更大, 因此本文在建立多核函数组合的基础上, 以高斯核函数为研究对象, 对多核函数的权系数和核参数同时进行优化更新。

高斯核的定义为

$$\kappa(\mathbf{u}, \mathbf{u}') = \exp(-\eta \|\mathbf{u} - \mathbf{u}'\|) \quad (12)$$

其中, η 表示核参数, 其大小决定着输入数据在RKHS空间内的相似性测度, η 过大, 会导致特征空间线性退化, 失去了核学习的非线性映射的基本功能; 而 η 过小, 又会造成RKHS空间内的数据相似, 使得原数据保存在高维空间的特征消失, 从而造成滤波器的性能下降。因此, 核参数的设置和优化对学习过程中输入数据的调整十分必要, 本文在使用多核加权组合函数的基础上, 以高斯核作为基本核函数, 并将核参数作为变量, 学习过程中动态调整以达到参数的最优, 多核函数组合如式(13)所示

$$K(\mathbf{u}_i, \mathbf{u}_j) = \sum_{n=1}^P \lambda_n \kappa_n(\mathbf{u}_i, \mathbf{u}_j, \eta_n) \quad (13)$$

其中, η_n 表示第 n 个核函数的核参数, P 为核的个数。在核仿射投影算法条件下, 考虑最小化均方误差问题

$$F = \|e(k)\|^2 = \left\| \mathbf{d}(k) - \sum_{j=1}^{k-1} \mathbf{a}_j(k-1) K(\mathbf{u}_k, \mathbf{u}_j) \right\|^2 \quad (14)$$

根据式(13)和式(14)对参数 λ_n 和 η_n 进行微分, 可以得到

$$\left. \begin{aligned} \frac{\partial F}{\partial \lambda_n} &= \frac{\partial \|e(k)\|^2}{\partial \lambda_n} = -2 \|e(k)\| \cdot \sum_{j=1}^{k-1} \mathbf{a}_j(k-1) \cdot \kappa_k(\mathbf{u}_k, \mathbf{u}_j, \eta_n(k)) \\ \frac{\partial F}{\partial \eta_n} &= \frac{\partial \|e(k)\|^2}{\partial \eta_n} = 2 \eta_n(k) \|e(k)\| \cdot \sum_{j=1}^{k-1} \mathbf{a}_j(k-1) \cdot \|\mathbf{u}_k - \mathbf{u}_j\|^2 \cdot \kappa_k(\mathbf{u}_k, \mathbf{u}_j, \eta_n(k)) \end{aligned} \right\} \quad (15)$$

根据梯度下降法, 设置合适的迭代步长, 可以得到多核函数中权值和核参数的更新公式为

$$\left. \begin{aligned} \lambda_n(k+1) &= \lambda_n(k) + \frac{\partial F}{\partial \lambda_n} \Delta t \\ \eta_n(k+1) &= \eta_n(k) + \frac{\partial F}{\partial \eta_n} \Delta t \end{aligned} \right\} \quad (16)$$

其中, Δt 为更新步长。为了避免核带宽更新过程中出现发散现象, 通常将更新步长为较小的数, 并对权值作归一化处理如式(17)

$$\eta_n(k+1) = \eta_n(k) / \sum_{i=1}^P \eta_i(k+1) \quad (17)$$

3.2 惊奇准则及其改进

相对于新奇准则和近似线性相关, 惊奇准则是

对这两种方法在严格信息论框架下的统一。由于再生核学习过程独立于观测，且无法知道样本输入的真实联合概率密度，基于样本所含信息测度严重依赖于观测，对于核自适应滤波中在线学习可以通过假设的后验分布来定义一个信息测度，即用给定学习系统的数据分布假设前提下样本的负对数似然作为给定数据的惊奇测度

$$L(\mathbf{u}(t), \mathbf{d}(t)) = -\ln p(\mathbf{u}(t), \mathbf{d}(t) | \text{Stat}(t-1)) \quad (18)$$

其中， $p(\mathbf{u}(t), \mathbf{d}(t) | \text{Stat}(t-1))$ 表示时刻 t 状态下的关于输入数据对 $(\mathbf{u}, \mathbf{d})$ 的后验(条件)概率。同NC准则类似，在学习过程中设置中心集，并假设中心集为数据的当前状态 $\text{Stat}(t)$ ，那么后验概率可以计算得到

$$\begin{aligned} p(\mathbf{u}(t), \mathbf{d}(t) | \text{Stat}(t-1)) &= p(\mathbf{u}(t) | \mathbf{d}(t), \text{Stat}(t-1)) \cdot p(\mathbf{d}(t) | \text{Stat}(t-1)) \\ &= \frac{1}{\sqrt{2\pi}\sigma(t)} \exp\left(-\frac{\|\mathbf{d}(t) - \hat{\mathbf{d}}(t)\|^2}{2\sigma^2(t)}\right) \\ &\quad \cdot p(\mathbf{d}(t) | \text{Stat}(t-1)) \end{aligned} \quad (19)$$

式中， $\hat{\mathbf{d}}(t)$ 表示在 $\text{Stat}(t-1)$ 条件下对 $\mathbf{d}(t)$ 的最大后验估计， $\sigma^2(t)$ 表示 $\text{Stat}(t-1)$ 条件下对 $\mathbf{d}(t)$ 的预测误差方差。因此，根据式(19)可以计算得到 t 时刻的输入数据对所引起的后验对数值为

$$\begin{aligned} L(t) &= -\ln p(\mathbf{u}(t), \mathbf{d}(t) | \text{Stat}(t-1)) \\ &= \ln \sqrt{2\pi} + \ln \sigma(t) + \left(\frac{\|\mathbf{d}(t) - \hat{\mathbf{d}}(t)\|^2}{2\sigma^2(t)}\right) \\ &\quad - \ln p(\mathbf{u}(t) | \text{Stat}(t-1)) \end{aligned} \quad (20)$$

由此可以将式(20)中的 $L(t)$ 作为惊奇测度参数之一引入多核仿射投影自适应算法的稀疏化中去。

稀疏化的程度严重依赖字典的更新机制，根据NC计算样本与字典的最短距离来判断是否需要更新权值更新的原则，SC条件下，仿射投影算法数据具有 K 重使用的特点，而距离 L 可以根据 K 重数据来进行多样化，即在当前字典中选取 K 个中心估计总体距离，并对 K 个中心进行归一化处理，那么SC需要首先考虑式(21)最小化问题

$$\begin{aligned} L_{\text{SC}} &= \min_{\mathbf{u} \in \text{Dic}(k)} \left\| \mathbf{u}(k+1) - \sum_{i=1}^K \beta_i \text{dic}_i \right\| \\ &\quad \cdot \left(\sum_{j=1}^K \beta_j = 1 \right) \end{aligned} \quad (21)$$

采用文献[21]中KRLS算法对加性噪声和正则

化参数的关系公式，可定义多尺度核学习的加性噪声 $\sigma^2(t)$ 与正则化参数 λ 相近，那么将该结论用于多核仿射投影方法中就有式(22)的关系

$$\begin{aligned} r(t) &= \lambda + K(\mathbf{u}(t), \mathbf{u}(t)) - \mathbf{h}_{\text{mkapa}}^T(t) \\ &\quad \cdot [\lambda \mathbf{I} + \mathbf{G}_{\text{sc}}(t)]^{-1} \mathbf{h}_{\text{mkapa}}(t) \end{aligned} \quad (22)$$

其中 $\mathbf{h}_{\text{mkapa}}^T$ ， $\mathbf{G}_{\text{sc}}$ 分别为

$$\begin{aligned} \mathbf{h}_{\text{mkapa}}^T &= [K(\mathbf{u}(i), \mathbf{n}_1), K(\mathbf{u}(i), \mathbf{n}_2), \dots, K(\mathbf{u}(i), \mathbf{n}_K)]^T \\ \mathbf{G}_{\text{sc}} &= \begin{bmatrix} K(\mathbf{n}_1, \mathbf{n}_1) & K(\mathbf{n}_1, \mathbf{n}_2) & \dots & K(\mathbf{n}_1, \mathbf{n}_K) \\ K(\mathbf{n}_2, \mathbf{n}_1) & K(\mathbf{n}_2, \mathbf{n}_2) & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ K(\mathbf{n}_K, \mathbf{n}_1) & K(\mathbf{n}_K, \mathbf{n}_2) & \dots & K(\mathbf{n}_K, \mathbf{n}_K) \end{bmatrix} \end{aligned} \quad (23)$$

式(22)中， $\mathbf{n}_i (i = 1, 2, \dots, K)$ 表示第 i 个最近邻向量。忽略常数项，将式(20)、式(22)代入式(21)惊奇测度公式可以得到基于SC准则的核仿射投影方法中惊奇测度表示为

$$L_{\text{sc}}(t) = \frac{1}{2} \ln r(t) + \frac{e^2(t)}{2r(t)} - \ln p(\mathbf{u}(t), \mathbf{d}(t) | \text{Stat}(t-1)) \quad (24)$$

式(24)的重点在于计算 $r(t)$ ，由于该惊奇测度采用字典中 K 个最近邻数据来估算梯度向量，因此 $r(t)$ 的计算具有较高的复杂度。而其计算核心在于利用这 K 个数据点集所构造的Gram矩阵，该矩阵考虑了 K 个最近邻数据集中两两之间距离所需的所有信息量，并且与原数据点集相比，Gram矩阵还存在一些信息损失，样本点映射到希尔伯特再生核空间内的完整几何形态不仅包括距离信息，还包含相对于原点的方向信息。因此，需要对Gram矩阵进行改进，并简化方差项，使得这种SC准则达到多核仿射投影算法的跟踪性能。

通过Mercer定理可以知道，对于输入向量 $\mathbf{x}$ 和 $\mathbf{z}$ ，核函数在结构上能够以一致收敛序列 $(X \times X)$ 的形式展开

$$\kappa(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{\infty} \phi_i(\mathbf{x}) \phi_i(\mathbf{z}) \quad (25)$$

这种情况下可以将核看成是由函数集的概率分布所得到的协方差函数，那么由此Gram矩阵就是在给定训练集 S 和核函数 $\kappa(*, *)$ 情况下的协方差矩阵。为了提高动态跟踪性能，采用 K 个最近数据点可以最大限度表现当前样本点的相关信息，即对于第 k 个样本点来说，选用 $k-K+1 \sim k$ 这 K 组数据，那么 $\mathbf{h}$ 向量和新型Gram矩阵为

$$\mathbf{h}(t) = [K(\mathbf{u}(t), \mathbf{u}(t-K+1)), K(\mathbf{u}(t), \mathbf{u}(t-K+2)), \dots, K(\mathbf{u}(t), \mathbf{u}(t))]^T$$

$$\mathbf{G}(t) = \begin{bmatrix} K(\mathbf{u}(t-K+1), \mathbf{u}(t-K+1)) & K(\mathbf{u}(t-K), \mathbf{u}(t-K+1)) & \dots & K(\mathbf{u}(t), \mathbf{u}(t-K+1)) \\ K(\mathbf{u}(t-K+1), \mathbf{u}(t-K)) & K(\mathbf{u}(t-K), \mathbf{u}(t-K+1)) & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ K(\mathbf{u}(t-K+1), \mathbf{u}(t)) & K(\mathbf{u}(t-K), \mathbf{u}(t)) & \dots & K(\mathbf{u}(t), \mathbf{u}(t)) \end{bmatrix} \quad (26)$$

由此可以得到简化后的 $r(t)$ 表达式为

$$r(t) = \mathbf{h}^T(t) \mathbf{G}(t) \mathbf{h}(t) \quad (27)$$

3.3 算法迭代步骤

根据3.2节的分析,并结合常规仿射投影算法,可以得到给予惊奇准则的多尺度核仿射投影算法步骤总结如下:

步骤1 初始化,初始化字典 $\mathbf{D}_{ic} = \mathbf{u}(1)$,选择迭代步长 $\mu, \Delta t$,输入数据的重用数 K ,初始化核带宽 $\lambda_i (i=1, 2, \dots)$ 和权系数 $\eta_i (i=1, 2, \dots)$,第1次迭代下扩展系数 $\mathbf{a}(1) = 0$,输出值的估计为 $\hat{\mathbf{d}}(1) = 0$;

步骤2 开始迭代计算,分配新的数组,并使得 $\mathbf{a}_k(k-1) = 0$;

步骤3 计算当前网络输出的估计值 $\hat{d}_k(i)$ 和误差

$$\hat{d}_k(i) = \sum_{j=1}^{k-1} \mathbf{a}_k(j) \kappa(\mathbf{u}(i), \text{Dic}(j)) \quad (28)$$

$$\text{Err}_k(i) = d(i) - \hat{d}_k(i) \quad (29)$$

步骤4 更新核相关参数,按照式(16)和式(17)分别进行核带宽参数更新和归一化处理。

步骤5 根据式(26)构造Gram矩阵,并更新扩展系数,更新公式为

$$\mathbf{a}_{k+1}(k-K+1) = \mathbf{a}_k(k-K+1) + \mu(\xi \mathbf{I} + \mathbf{G})^{-1} \text{Err}_k(i) \quad (30)$$

式中 ξ 为较小的常数,防止Gram矩阵奇异化。

步骤6 字典稀疏化,设定稀疏化阈值

$$\text{thr}_1 = \frac{\text{sum}(L_{sc})}{k-K} \quad (31)$$

$$\text{thr}_2 = \text{thr}_1 + 100 \quad (32)$$

根据式(31)、式(32)判断是否加入字典。

4 数字仿真验证

4.1 噪声消除

将本文提出的算法用于信号噪声的消除,通过已知的参考测量量来消除未知干扰。假定原始输入信号为 $x(t)$,随机噪声 $d(t)$ 通过非线性失真函数得到滤波器的输入 $u(t)$,那么就可以通过多核参数的调整对滤波器进行微调,从而从期望信号 $d(t)$ 中去除噪声。这种噪声去除的滤波器系统结构如图1所示。

噪声源的失真函数为非线性函数,作为确定性模型,假设 $n(t)$ 和 $u(t)$ 之间的关系为

$$u(t) = n(t) - 0.2u(t-1) - u(t-1)n(t-1) + 0.1n(t-1) + 0.4u(t-1) + 0.2e^{-u(t-1)} \quad (33)$$

使用文献[4]中所提供真实数据作为信号的噪声源,其噪声均值为0,方差为0.0026,其波形如图2所示。

该噪声失真模型 $n(t)$ 依赖之前的值,根据自适应滤波的递归特性可以将滤波器输出 $\hat{d}(t)$ 的反馈环节,当由核函数映射到线性条件下时作为输出误差进行递归运算。

假设原始信号 $x(t) = 0$,为了便于计算和分析,本文采用3个高斯基本核作为多核组合函数。为了便于对比,初始化核组合权系数分别为 $\lambda_1 = 1, \lambda_2 = 0, \lambda_3 = 0$,相比于LMS, KLMS, NC-KAPA和NC-MKAPA方法,其中NC-MKAPA的组合权系数初始化为1, 1, 1,其余设置的初始参数如表1所示。

通过100次蒙特卡罗仿真可以得到对数尺度下关于最小均方误差(MSE)的学习曲线如图3所示。对比不同算法下的MSE可以看出SC-MKAPA, NC-MKAPA, NC-KAPA算法相比于LMS和KLMS方

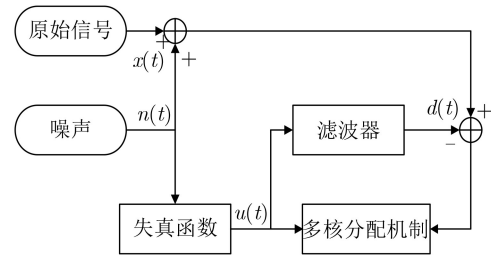


图1 滤波器除噪原理

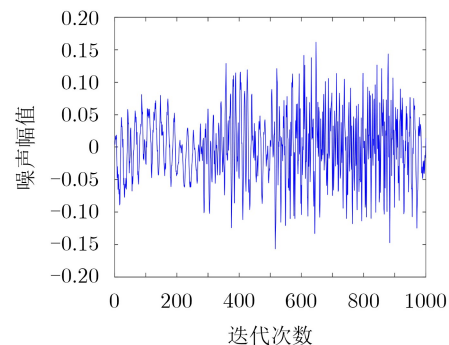


图2 噪声分布

表1 算法参数

算法	核带宽	收敛因子	正则化参数 δ
SC-MKAPA	$\eta_1 = 1.0, \eta_2 = 0.5, \eta_3 = 10$	$\mu = 0.2, \Delta t = 0.01$	5.0×10^{-3}
NC-MKAPA	$\eta_1 = 1.0, \eta_2 = 0.5, \eta_3 = 10$	$\mu = 0.2$	5.0×10^{-3}
NC-KAPA	$\eta_1 = 1.0$	$\mu = 0.2$	5.0×10^{-3}
KLMS	$\eta_1 = 1.0$	$\mu = 0.2$	5.0×10^{-3}
LMS	$\eta_1 = 1.0$	$\mu = 0.2$	5.0×10^{-3}

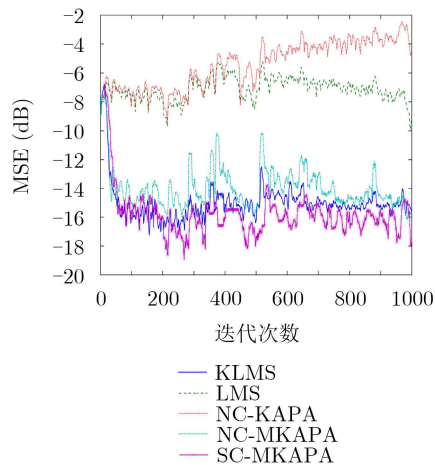


图3 对数条件下MSE的学习曲线

法性能有明显的提升,且SC-MKAPA算法优于NC-KAPA和NC-MKAPA,并且随着迭代次数的增加,SC-MKAPA的动态表现能力更优。

式(31)描述了含噪声的非线性失真函数与滤波器输入之间的关系。更一般地,为了说明本文提出的SC-MKAPA的适应性,在式(33)右端第3项和第6项添加关于 $u(t-1)$ 的高次项来提高系统模型的非线性度,如式(34)所示:

$$u(t) = n(t) - 0.2u(t-1) - u^N(t-1)n(t-1) + 0.1n(t-1) + 0.4u(t-1) + 0.2e^{-u^N(t-1)} \quad (34)$$

横向对比SC-MKAPA方法与其他方法在不同高次项下的对数均方MSE(MMSE)。如表2所示。

从表2中可以看出,SC-MKAPA法在不同非线性度下均优于其他4种方法,但当非线性度过高时,5种方法的MMSE相近。

4.2 非线性信道均衡

核学习自适应算法在信道均衡中有大量的应

表2 不同高次项下5种方法MMSE(dB)

高次项 N	SC-MKAPA	NC-MKAPA	NC-KAPA	KLMS	LMS
2	-71.2	-62.8	-67.2	-32.7	-25.6
3	-62.1	-56.9	-60.6	-24.4	-19.3
6	-33.9	-29.3	-30.2	-21.5	-17.8
7	-18.3	-16.3	-15.2	-13.3	-12.9

用,尤其是对于非线性信道均衡问题。其基本原理是通过估计一个传输函数来补偿由信道引起的非线性失真。这种失真通常具有较高的非线性度和非线性维数。因此,可以根据核自适应特性设计出具有非线性滤波特性的均衡器。

假设非线性信道的输入为二进制信号 $\{u(1), u(2), \dots, u(N), s(N)\}$,受到加性高斯白噪声的干扰,信道接收端的观测值为 $\{d(1), d(2), \dots, d(N)\}$,那么该非线性信道模型的输入 $x(t)$ 和输出 $d(t)$ 分别为

$$x(t) = u(t) + 0.5u(t-1) \quad (35)$$

$$d(t) = x(t) - 0.6x^2(t) + n(t) \quad (36)$$

其中, $n(t)$ 为均值为0,方差为0.1的高斯白噪声。在样本训练过程中引入通道的突变,以观察对信道的跟踪能力。样本训练总数为1500,在500次训练后突变,突变后的输出为

$$d(t) = -x(t) + 0.6x^2(t) + n(t) \quad (37)$$

输入信号维数为3,其他相关参数可参见表1,得到在对数条件下LMS, KLMS, NC-KAPA, NC-MKAPA和SC-MKAPA这5种算法的学习曲线,如图4所示。从图4中可以看出,基于惊奇准则的多核仿射投影投影算法在信道均衡应用中相对于其他方法具有明显的优势。与基于新奇准则的单核KAPA与多核加权组合的KAPA方法对比,可变核参数的多核KAPA方法收敛速度最快,其MSE达到的精度

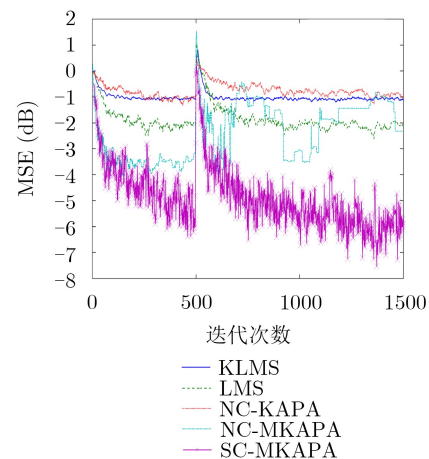


图4 对数条件下MSE的学习曲线

最高,尤其是在第500次迭代发生突变时,能够很快地实现信道跟踪。

4.3 MG非线性时间序列预测

由于MG混沌时间序列具有非线性特性,是最具代表性的时间序列预测问题。因此,本文将SC-MKAPA应用于MG时间序列的预测中,并与KLMS, LMS, NC-KAPA和NC-MKAPA方法进行对比。设置当前时刻数据点 $z(t)$ 前的7个点作为训练集,并取300个样本作为训练数据,100个点作为测试数据。其学习曲线如图5所示。

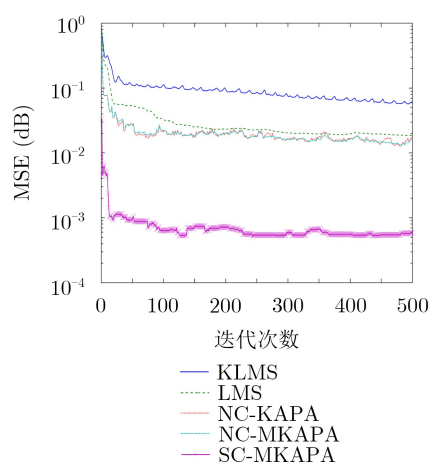


图 5 MG时间序列的预测学习曲线

从图5可以看出,本文提出的SC-MKAPA算法与NC-MKAPA算法在MG时间序列预测上性能相近,但与KLMS, LMS和NC-KAPA算法相比优势明显,原因是在多核尺度因子的自适应调整下,改进的SC准则要比NC准则阈值设定更为灵活,并具有更高的核配置自由度。

5 结论

本文在研究多核自适应滤波算法的基础上,动态调整多个高斯核带宽和组合系数,对在线主动学习的信息论准则进行改进,在改进的惊奇准则中实现自动设定阈值,高效量化了学习过程中样本所含的信息量,最终提高了多核仿射投影滤波算法的灵活性,与其他固定核带宽参数的多核算法相比,在噪声去除、信道均衡和MG时间序列预测的应用当中均能获得较优的效果。

参 考 文 献

- [1] AIZERMAN A, BRAVERMAN E M, and ROZONER L I. Theoretical foundations of the potential function method in pattern recognition learning[J]. *Automation and Remote Control*, 1964, 25(5): 821–837.
- [2] FRIEß T and HARRISON R F. A kernel-based adaline for function approximation[J]. *Intelligent Data Analysis*, 1999, 3(4): 307–313. doi: [10.3233/IDA-1999-3405](https://doi.org/10.3233/IDA-1999-3405).
- [3] 庞业勇, 王少军, 彭宇, 等. 一种在线时间序列预测的核自适应滤波器向量处理器[J]. *电子与信息学报*, 2016, 38(1): 53–62. doi: [10.11999/JEIT150157](https://doi.org/10.11999/JEIT150157).
- [4] PANG Yeyong, WANG Shaojun, PENG Yu, et al. A kernel adaptive filter vector processor for online time series prediction[J]. *Journal of Electronics & Information Technology*, 2016, 38(1): 53–62. doi: [10.11999/JEIT150157](https://doi.org/10.11999/JEIT150157).
- [5] BLANDON J S, VALENCIA C K, ALVAREZ A, et al. Shape classification using Hilbert space embeddings and kernel adaptive filtering[C]. *The 15th International Conference Image Analysis and Recognition*, Portugal, 2018: 245–251. doi: [10.1007/978-3-319-93000-8_28](https://doi.org/10.1007/978-3-319-93000-8_28).
- [6] GAO Wei, YAN Yi, ZHANG Lingling, et al. Convex combinations of multiple kernel adaptive filters[C]. *2017 IEEE International conference on Signal Processing, Communications and Computing*, Xiamen, China, 2017: 1–5. doi: [10.1109/ICSPCC.2017.8242551](https://doi.org/10.1109/ICSPCC.2017.8242551).
- [7] 孙丹华, 孙亮, 王彬, 等. α 稳定分布噪声下基于核方法的非线性信道均衡算法[J]. *信号处理*, 2017, 33(3): 223–228. doi: [10.16798/j.issn.1003-0530.2017.02.013](https://doi.org/10.16798/j.issn.1003-0530.2017.02.013).
- [8] SUN Danhua, SUN Liang, WANG Bin, et al. Nonlinear channel equalization algorithm based on kernel method for α -stable noise[J]. *Journal of Signal Processing*, 2017, 33(3): 223–228. doi: [10.16798/j.issn.1003-0530.2017.02.013](https://doi.org/10.16798/j.issn.1003-0530.2017.02.013).
- [9] SHIN B S, YUKAWA M, CAVALCANTE R L G, et al. Distributed adaptive learning with multiple kernels in diffusion networks[J]. *IEEE Transactions on Signal Processing*, 2018, 66(21): 5505–5519. doi: [10.1109/TSP.2018.2868040](https://doi.org/10.1109/TSP.2018.2868040).
- [10] HAN Yina, YANG Yixin, LI Xuelong, et al. Matrix-regularized multiple kernel learning via (r, p) Norms[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(10): 4997–5007. doi: [10.1109/TNNLS.2017.2785329](https://doi.org/10.1109/TNNLS.2017.2785329).
- [11] LIU Yuqi, SUN Chao, and JIANG Shouda. A reduced gaussian kernel least-mean-square algorithm for nonlinear adaptive signal processing[J]. *Circuits, Systems, and Signal Processing*, 2019, 38(1): 371–394. doi: [10.1007/s00034-018-0862-0](https://doi.org/10.1007/s00034-018-0862-0).
- [12] SHOAB B, QURESHI I M, BUTT S A, et al. Adaptive step size kernel least mean square algorithm for Lorenz time series prediction[C]. *The 12th International Bhurban Conference on Applied Sciences and Technology*, Islamabad, Pakistan, 2015: 218–221. doi: [10.1109/IBCAST.2015.7058507](https://doi.org/10.1109/IBCAST.2015.7058507).
- [13] 胡站伟, 焦立国, 徐胜金, 等. 基于多尺度重采样思想的类指数核函数构造[J]. *电子与信息学报*, 2016, 38(7): 1689–1695. doi: [10.11999/JEIT150157](https://doi.org/10.11999/JEIT150157).

- 10.11999/JEIT151101.
- HU Zhanwei, JIAO Liguang, XU Shengjin, *et al.* Design of an exponential-like kernel function based on multi-scale resampling[J]. *Journal of Electronics & Information Technology*, 2016, 38(7): 1689–1695. doi: 10.11999/JEIT151101.
- [12] PUAL T K and OGUNFUNMI T. A kernel adaptive algorithm for quaternion-valued inputs[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 26(10): 2422–2439. doi: 10.1109/TNNLS.2014.2383912.
- [13] NISHIKAWA K and NAKAZATO H. Mixture structure of kernel adaptive filters for improving the convergence characteristics[C]. 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference, Hollywood, USA, 2012: 1–6.
- [14] POKHAREL R, SETH S, and PRINCIPE J C. Mixture kernel least mean square[C]. 2013 International Joint Conference on Neural Networks, Dallas, USA, 2013: 1–7. doi: 10.1109/IJCNN.2013.6706867.
- [15] VAN VAERENBERGH S, SCARDAPANE S, and SANTAMARIA I. Recursive multikernel filters exploiting nonlinear temporal structure[C]. The 25th European Signal Processing Conference, Kos, Greece, 2017: 2674–2678. doi: 10.23919/EUSIPCO.2017.8081696.
- [16] SILVA M T M, CANDIDO R, ARENAS-GARCIA J, *et al.* Improving multikernel adaptive filtering with selective bias[C]. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing, Calgary, Canada, 2018: 4529–4533. doi: 10.1109/ICASSP.2018.8461290.
- [17] ISHIDA T and TANAKA T. Multikernel adaptive filters with multiple dictionaries and regularization[C]. 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Kaohsiung, China, 2013: 1–6. doi: 10.1109/APSIPA.2013.6694279.
- [18] TODA O and YUKAWA M. On kernel design for online model selection by Gaussian multikernel adaptive filtering[C]. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Siem Reap, Cambodia, 2014: 1–5. doi: 10.1109/APSIPA.2014.7041802.
- [19] LIU Weifeng and PRÍNCIPE J. Kernel affine projection algorithms[J]. *EURASIP Journal on Advances in Signal Processing*, 2008, 2008: 784292. doi: 10.1155/2008/784292.
- [20] RICHARD C, BERMUDEZ J C M, and HONEINE P. Online prediction of time series data with kernels[J]. *IEEE Transactions on Signal Processing*, 2009, 57(3): 1058–1066. doi: 10.1109/TSP.2008.2009895.
- [21] GAO Wei, CHEN Jie, RICHARD C, *et al.* Kernel LMS algorithm with forward-backward splitting for dictionary learning[C]. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, Canada, 2013: 5735–5739. doi: 10.1109/ICASSP.2013.6638763.
- [22] TAKIZAWA M A and YUKAWA M. An efficient sparse kernel adaptive filtering algorithm based on isomorphism between functional subspace and Euclidean space[C]. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing, Florence, Italy, 2014: 4508–4512. doi: 10.1109/ICASSP.2014.6854455.
- 李群生：男，1977年生，博士，研究方向为滤波信号处理，组合导航技术。
- 赵 刻：男，1956年生，教授，研究方向为惯性技术，信号处理技术。
- 寇 磊：女，1971年生，高级工程师，研究方向为惯性技术。
- 王进达：男，1989年生，博士，研究方向为滤波信号处理，组合导航技术。