



基于Nyström柯西核共轭梯度算法的混沌时间序列预测

齐乐天 王世元 沈明琳 黄刚毅

Prediction of chaotic time series based on Nyström Cauchy kernel conjugate gradient algorithm

Qi Le-Tian Wang Shi-Yuan Shen Ming-Lin Huang Gang-Yi

引用信息 Citation: *Acta Physica Sinica*, 71, 108401 (2022) DOI: 10.7498/aps.71.20212274

在线阅读 View online: <https://doi.org/10.7498/aps.71.20212274>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

基于分数阶最大相关熵算法的混沌时间序列预测

Prediction of chaotic time series based on the fractional-order maximum correntropy criterion algorithm

物理学报. 2018, 67(1): 018401 <https://doi.org/10.7498/aps.67.20171803>

基于大脑情感学习模型和自适应遗传算法的混沌时间序列预测

Chaotic time series prediction based on brain emotional learning model and self-adaptive genetic algorithm

物理学报. 2018, 67(8): 080502 <https://doi.org/10.7498/aps.67.20172104>

非高斯冲激干扰下基于Softplus函数的核自适应滤波算法

Kernel adaptive filtering algorithm based on Softplus function under non-Gaussian impulse interference

物理学报. 2021, 70(2): 028401 <https://doi.org/10.7498/aps.70.20200954>

基于P范数的核最小对数绝对差自适应滤波算法

Kernel least logarithm absolute difference algorithm based on P-norm

物理学报. 2022, 71(4): 048401 <https://doi.org/10.7498/aps.71.20211124>

基于鲁棒极端学习机的混沌时间序列建模预测

Chaotic time series prediction based on robust extreme learning machine

物理学报. 2018, 67(3): 030501 <https://doi.org/10.7498/aps.67.20171887>

一种变尺度S型核分式低次幂自适应滤波算法

A variable-scale S-type kernel fractional low-power adaptive filtering algorithm

物理学报. 2021, 70(15): 158401 <https://doi.org/10.7498/aps.70.20210075>

基于 Nyström 柯西核共轭梯度算法的混沌时间序列预测*

齐乐天¹⁾²⁾ 王世元^{1)2)†} 沈明琳¹⁾²⁾ 黄刚毅¹⁾²⁾

1) (西南大学电子信息工程学院, 重庆 400715)

2) (非线性电路与智能信息处理重庆市重点实验室, 重庆 400715)

(2021 年 12 月 8 日收到; 2022 年 1 月 27 日收到修改稿)

混沌时间序列能够较好反映真实环境的非线性和非平稳性特性, 然而具有二阶统计特性的核自适应滤波器 (kernel adaptive filter, KAF) 在处理含噪声和异常值的混沌时间序列时, 其预测性能显著下降. 为提高核自适应滤波器的鲁棒性, 本文提出了一种用于测量非线性相似度的柯西核损失 (Cauchy kernel loss, CKL), 并采用半平方 (half-quadratic, HQ) 方法保证了 CKL 的全局凸性. 为改善随机梯度下降法收敛速度较慢且容易陷入局部最优的不足, 采用共轭梯度 (conjugate gradient, CG) 方法优化 CKL. 进一步, 为解决核矩阵网络增长的问题, 采取 Nyström 稀疏策略近似核矩阵, 并利用概率密度秩量化 (probability density rank-based quantization, PRQ) 提高逼近精度. 基于此, 本文提出了一种新的基于 Nyström 和 PRQ 的柯西核共轭梯度 (Nyström Cauchy kernel conjugate gradient with PRQ, NCKCG-PRQ) 算法有效实现了混沌时间序列的预测. 基于合成和真实两类混沌时间序列验证了所提 NCKCG-PRQ 算法在稳态性能、鲁棒性和计算存储复杂度上的优势.

关键词: 混沌时间序列预测, 核自适应滤波, 柯西核函数, Nyström 方法**PACS:** 84.30.Vn, 05.45.Tp, 05.10.Ln**DOI:** 10.7498/aps.71.20212274

1 引言

混沌现象^[1]作为一种类似随机的无规律行为, 是由不附加随机因素的确定性非线性动力学系统产生. 由于现实环境常具有非线性和非平稳性, 故混沌系统在很大程度上能够重构现实世界的非线性系统. 然而, 混沌系统对初始值的敏感特性使得其长期行为不可精准预测; 而混沌吸引子的确定性动力学机制使得混沌时间序列具有较为准确的短期预测性^[2]. 因此, 混沌时间序列的有效预测是重构混沌系统的基础, 具有重要的实际意义和挑战性; 精准的预测方法能够被广泛应用于保密通信、情感

学习、电力系统短期负荷预测、天气预测^[3-4]等领域. 然而, 在实际应用中, 混沌信号常受到不同程度的噪声干扰, 因此如何最大限度地各类噪声环境下实现混沌序列的精准预测, 已经成为混沌信号处理的研究热点. 混沌时间序列的预测方法可分为全局预测法^[5]、局部预测法^[6]和自适应预测法^[7]. 其中, 全局预测法采用全部样本数据拟合非线性函数, 但当样本数据量大时运算速度较慢^[8-9]; 局部预测法因仅利用部分数据用于预测, 运算速度较快但预测精度较低^[10-11]. 作为一种全局预测法以及局部预测法的扩展, 自适应预测法将自适应滤波算法融合于前两类方法中^[12-13]. 相比于经典的局部预测法和全局预测法, 自适应预测法因其参数配

* 国家自然科学基金 (批准号: 62071391)、重庆市自然科学基金面上项目 (批准号: cstc2020jcyj-msxmX0234) 和中央高校基本科研业务费 (批准号: 2020jd001) 资助的课题.

† 通信作者. E-mail: wsy@swu.edu.cn

置简单、混沌运动轨迹的自适应跟踪特性、良好的去噪特性、以及在小数据量前提下仍具备较高的预测精度等诸多优点,已在现实环境中得到广泛应用^[14]. 因能够根据混沌序列的变化自动调整模型参数的特性,自适应预测法为混沌时间序列预测提供了一种新的思路,但因采用的自适应滤波算法收敛速度不够快限制了其在实际应用的实时性.

考虑到现实环境中的数据常具有非线性特征,使得基于线性结构的自适应滤波系统在处理非线性问题时存在局限性. 为此,核方法(kernel method)将输入信号非线性映射到高维再生核希尔伯特空间中,然后在特征空间中执行线性滤波,发展出了核自适应滤波(kernel adaptive filter, KAF),如基于均方误差(mean square error, MSE)和最小二乘(least squares, LS)准则的核最小均方算法^[15]和核递归最小二乘算法^[16]. 在自适应滤波算法中,代价函数,权重更新方式和稀疏策略决定了混沌时间序列预测性能.

作为一种只包含二阶统计特性的误差准则,基于MSE和LS准则的自适应滤波算法在现实场景中的非高斯环境,如重尾非高斯噪声以及非高斯脉冲噪声下性能较差,甚至严重退化. 因此,找到一个合适的误差准则来保证算法在非高斯噪声情况下的鲁棒性就显得十分必要. 为此,针对非高斯噪声的鲁棒相似测度应运而生,如高阶误差准则,低阶误差准则,对数误差准则以及相关熵误差准则. 相关熵(correntropic loss, C-Loss)^[17]是一种从信息论学习(information theoretic learning, ITL)^[18]角度出发具有鲁棒性和非线性双重特点的局部相似性度量,已被成功应用于鲁棒KAF^[19]中. 作为对数误差准则的柯西损失(Cauchy loss, CL),因其独特的分布特性以及对抗异常值的鲁棒性,在鲁棒KAF中也有广泛应用^[20]. 受启发于函数映射以及对数误差准则的鲁棒性,将CL和核方法结合生成了柯西核损失(Cauchy kernel loss, CKL)^[21]. CKL因具有低计算复杂度和良好的抗脉冲噪声性能,已在KAF得到了广泛的应用.

优化方法是KAF算法中权向量更新的关键. 从本质上讲,优化方法的选取决定了算法的计算复杂程度和收敛速度. 随机梯度下降(stochastic gradient descent, SGD)^[22]法因其简单,在KAF得到广泛应用. 然而,基于SGD的滤波算法通常容易陷入局部最优,导致性能下降. 作为SGD方

法的延拓,牛顿法和共轭梯度法(conjugate gradient, CG)^[23]能更好地求解误差函数. 但是,在牛顿法中需要保证海森矩阵(Hessian matrix)的正定性并计算海森矩阵的逆,这在一定程度上增加了计算复杂度,限制了其实际应用^[24]. CG法能够平衡收敛速度和计算复杂度. 与SGD法相比,CG法不仅提高了算法收敛速度,并且在计算复杂度较低的情况下取得了与递归算法相当的性能^[23]. 但是,相比于牛顿法,CG法因不需要计算海森矩阵的逆,所以能够获得更加稳定的解.

通常,KAF采用高斯核实现在线学习,即根据每一个新输入的数据对系统参数进行学习. 然而,大规模的数据导致网络尺寸线性增长,增加了计算量和空间存储,这对KAF的在线实际应用提出了巨大挑战. 抑制KAF网络增长的方法主要包括量化^[25]和稀疏化策略. 但是,量化和稀疏化策略均不能预先设定网络结构大小,用以满足特定计算和存储要求. Nyström方法采用近似法重构矩阵^[26],将高维度问题转化为低维度空间,通过固定维网络结构来逼近KAF中的核矩阵,并在逼近精度和计算复杂度两方面实现了平衡. 然而,Nyström方法中常采取的抽样策略是随机抽样^[27]和 k -均值抽样^[28]. 由于两种抽样方法具有对初始样本选择的敏感性,所以其逼近精度通常无法达到预期效果. 基于此,提出了基于概率密度秩的量化(probability density rank-based quantization, PRQ)^[29]采样方法进一步提高近似精度.

受启发于半平方(half-quadratic, HQ)^[24]方法转化CKL后代价函数的全局凸性和鲁棒性,本文根据混沌时间序列的短期可预测性,在CG优化方法上引入了基于PRQ采样的Nyström方法,提出了一种新的基于PRQ采样的Nyström柯西核共轭梯度(Cauchy kernel conjugate gradient, CKCG)算法(NCKCG-PRQ). 所提的NCKCG-PRQ算法在预测麦克格拉斯(Mackey-Glass, MG)混沌时间序列和实际电路采集的蔡氏混沌时间序列的两类实例中,展现了在非高斯噪声环境下预测性能的强鲁棒性、低稳态误差及低计算复杂度的特点.

2 背景介绍

本节首先介绍柯西核代价函数,接着采用HQ方法将柯西核函数转化为具有二次型形式的全局

凸函数, 最后采用 Nyström 法对矩阵进行有效重构.

2.1 柯西核代价函数

给定两个随机变量 $X \in \mathbf{R}^n$ 和 $Y \in \mathbf{R}^n$, 柯西核函数^[30] 被定义为如下形式:

$$L(X, Y) = E \left[\ln \left(1 + \eta \exp \left(\frac{\|X - Y\|^2}{-2\delta^2} \right) \right) \right] \\ = \int \ln \left(1 + \eta \exp \left(\frac{\|X - Y\|^2}{-2\delta^2} \right) \right) dF_{XY}(x, y), \quad (1)$$

式中 $E[\cdot]$ 表示数学期望, η 是一个大于 0 的常数, δ 是柯西核函数带宽. $F_{XY}(x, y)$ 是变量 X 和 Y 的联合分布函数, 通常情况下由于其未知, 导致很难计算 $L(X, Y)$. 事实上, 基于有限数量样本 $\{x_k, y_k\}_{k=1}^N$, (1) 式可近似为

$$\hat{L}(X, Y) = \frac{1}{N} \sum_{i=1}^N \ln \left[1 + \eta \exp \left(\frac{(x_i - y_i)^2}{-2\delta^2} \right) \right], \quad (2)$$

定义误差 $\mathbf{e} = (x_1 - y_1, x_2 - y_2, \dots, x_N - y_N) = (e_1, e_2, \dots, e_N)$. 为了简化 (2) 式, 令 $s_i = e_i^2 / (2\delta^2)$, (2) 式中关于误差的海森矩阵为

$$\mathbf{H}_{\hat{L}(X, Y)}(\mathbf{e}) = \left[\frac{\partial^2 \hat{L}(X, Y)}{\partial e_i \partial e_j} \right] = \text{diag}[\xi_1, \xi_2, \dots, \xi_N], \quad (3)$$

其中, ξ_i 的表达式如下:

$$\xi_i = \frac{-\exp(-s_i)}{N\delta^2[\eta^{-1} + \exp(-s_i)]^2} [\exp(-s_i) + \eta^{-1}(1 - 2s_i)], \quad i = 1, 2, \dots, N, \quad (4)$$

根据 (3) 和 (4) 式可以计算出海森矩阵正定的条件为

$$\min \{|e_i|\}_{i=1}^N > \delta \quad \text{且} \\ \max \left\{ \sqrt{\frac{\delta^2}{e_i^2 - \delta^2} \exp(-s_i)} \right\}_{|e_i| > \delta, i=1}^N \leq \sqrt{\eta^{-1}}, \quad (5)$$

显然, 柯西核函数不是一个全局凸函数.

2.2 半平方法转化柯西核函数

由于 (3) 式的海森矩阵只有在满足一定条件时才是正定的, 因此不能始终保证其全局凸性, 这就限制了柯西核函数在解决凸优化问题中的应用. 实际上, 柯西核函数可以看作一个由具有对数函数形式的柯西函数和指数函数形式的相关熵函数映

射而来的复合函数, 借助两次半平方优化法使柯西核函数转化为全局凸函数. 因转化后的数学表达式为二次型形式, 所以能够有效解决凸优化和共轭梯度优化的问题.

首先, 考虑柯西核函数中的柯西函数和相关熵函数. 不失一般性, 令 $\tilde{g}(x) = \exp(-x)$ 和 $g(x) = -\ln(1+x)$, 其中, $x \in \mathbf{R}$. 根据共轭方程理论^[30], 当 $v < 0$ 时, 可得 $\tilde{g}(x)$ 的共轭方程为 $\tilde{f}(v) = -v \ln(-v) + v$ 和 $g(x)$ 的共轭方程为 $f(v) = -1 - v - \ln(-v)$, 这两个共轭方程均存在各自的上界. 由于共轭方程的凸性和可微性, 可得 $\exp(-s_i) = \sup_{v < 0} \{vs_i - f(v)\}$, 且当 $v = -\exp(-s_i)$ 时达到上确界. 利用熵半二次优化法^[24] 可得:

$$\max_{v < 0} \sum_{i=1}^N (vs_i - \tilde{f}(v)) \\ = \min \sum_{i=1}^N -vs_i = \min \sum_{i=1}^N s_i \exp(-s_i), \quad (6)$$

因此, (2) 式可等价于

$$\hat{L}(X, Y) = \frac{1}{N} \sum_{i=1}^N \ln[1 + \eta s_i \exp(-s_i)]. \quad (7)$$

然而, (7) 式仍然不是二次型形式. 因此借助半平方优化法再次转化使得柯西核函数为二次型形式. 根据 $f(v)$ 的凸性, 可得到如下命题.

命题 1 当 $v < 0$ 时, (2) 式中的对数函数是凸函数 $f(v) = -1 - v - \ln(-v)$ 的共轭函数, 可得 $-\ln(1 + \eta s_i \exp(-s_i)) = \sup_{v < 0} \{v\eta s_i \exp(-s_i) - f(v)\}$, 当 $v = -(1 + \eta s_i \exp(-s_i))^{-1}$ 时达到上确界.

证明: 根据共轭函数定义^[31], $f(v)$ 的共轭函数由 $f^*(u)$ 给出, 即 $f^*(u) = \sup_{v < 0} \{vu - f(v)\} = \sup_{v < 0} \{vu + 1 + v + \ln(-v)\}$. 令 $g(v) = vu + 1 + v + \ln(-v)$, 当 $v = -(1+u)^{-1}$ 时, 有 $f^*(u) = \sup_{v < 0} \{vu + 1 + v + \ln(-v)\} = -\ln(1+u)$. 紧接着, 用 $\eta s_i \exp(-s_i)$ 替代 u , 同时将 $u = \eta s_i \exp(-s_i)$ 代入 $f^*(u)$, 可得:

$$f^*(\eta s_i \exp(-s_i)) = \sup_{v < 0} \{v\eta s_i \exp(-s_i) + 1 + v + \ln(-v)\} \\ = -\ln[1 + \eta s_i \exp(-s_i)], \quad (8)$$

当 $v_i = -(1 + \eta s_i \exp(-s_i))^{-1}$ 时达到上确界. 此命题证毕.

此时, (2) 式中求解柯西核函数的最优解问题

等价于求解如下问题:

$$\max_{v_i < 0} \sum_{i=1}^N [v_i \eta s_i \exp(-s_i) - f(v_i)], \quad (9)$$

其中, $v_i \eta s_i \exp(-s_i) - f(v_i)$ 是关于 v_i 独立函数. (9) 式的求解可以通过交替优化方法实现. 借助 (8) 式可以解得 v_i 的解析解是 $v_i = -[1 + \eta s_i \exp(-s_i)]^{-1}$, 其中 v_i 是一个负数. 给定 v_i , (9) 式的优化问题等价于如下形式:

$$\min \sum_{i=1}^N \left[-v_i \frac{\eta \exp(-s_i) e_i^2}{2\delta^2} \right], \quad (10)$$

其中, $v_i = -[1 + \eta s_i \exp(-s_i)]^{-1}$. 代入 v_i 并将 (10) 式写成共轭梯度优化方法所需的二次型形式, 即:

$$\min \sum_{i=1}^N \left\{ \frac{\exp(-s_i)}{2\delta^2 [\eta^{-1} + s_i \exp(-s_i)]} \right\} e_i^2, \quad (11)$$

不难看出, (11) 式是一个全局凸函数且其中包含了共轭梯度优化方法所需要的二次型形式. (11) 式中加权最小二乘问题的海森矩阵如下:

$$\mathbf{H}(\mathbf{e}) = \text{diag} \left[-v_1 \frac{\eta s_1}{2\delta^2}, \dots, -v_N \frac{\eta s_N}{2\delta^2} \right]. \quad (12)$$

由于 v_i 小于 0 且 ηs_i 大于 0, 所以海森矩阵是正定的, 保证了全局凸性. (11) 式的解可通过半二次方法求解加权最小二乘问题得到. 由此可知, 半平方法对柯西核函数的转化打破了海森矩阵中参数 δ 和 η 关于误差 e 的限制条件, 进一步促进了柯西核函数在凸优化问题中的应用.

2.3 Nyström 方法

将原始输入 $\mathbf{u}(i)$ 映射到再生核希尔伯特空间中, 可得输入 $\varphi(\mathbf{u}(i))$ 的数据矩阵 $\Phi \in \mathbf{R}^{D \times n}$ 为

$$\Phi = [\varphi(\mathbf{u}(1)), \varphi(\mathbf{u}(2)), \dots, \varphi(\mathbf{u}(n))]. \quad (13)$$

核矩阵在数据矩阵 Φ 的基础上可以表示为 Gram 矩阵, 即 $\mathbf{G} = \Phi^T \Phi$ 且 $\mathbf{G} \in \mathbf{R}^{n \times n}$. 显而易见, 随着维数 n 的增大, 矩阵 \mathbf{G} 的空间复杂度和计算复杂度也随之增高, 运算成本显著增加. 因此, 有必要采用稀疏的 Nyström 方法逼近庞大的数据矩阵. 通过在 n 维数据集里选取 m 维 ($m < n$) 数据 $\{\mathbf{u}(g(i))\}_{i=1}^m$ 当作其子集, 得到矩阵 $\hat{\mathbf{K}} = [\kappa(\mathbf{u}(g(i)), \mathbf{u}(g(j)))]_{m \times m}$ 和矩阵 $\mathbf{K}_c = [\kappa(\mathbf{u}, \mathbf{u}(g(j)))]_{n \times m}$, 其中 $g(i)$ 是矩阵 \mathbf{K}_c 的索引. 则 \mathbf{G} 矩阵可近似表示为

$$\mathbf{G} \approx \mathbf{K}_c \hat{\mathbf{K}}^\dagger \mathbf{K}_c^T, \quad (14)$$

其中, $\hat{\mathbf{K}}^\dagger$ 是 $\hat{\mathbf{K}}$ 的伪逆, $\hat{\mathbf{K}}\mathbf{W} = \mathbf{W}\Lambda$, 且 $\Lambda = \text{diag}[\lambda_1^{(m)}, \lambda_2^{(m)}, \dots, \lambda_m^{(m)}]$ 指的是由 m 个降序排列的特征值组成的对角矩阵, $\mathbf{W} = [\mathbf{w}_1^{(m)}, \mathbf{w}_2^{(m)}, \dots, \mathbf{w}_m^{(m)}]$ 是对应的特征向量构造的列正交矩阵. 将伪逆 $\hat{\mathbf{K}}^\dagger$ 代入可得重构后的低秩矩阵:

$$\mathbf{G} \approx \mathbf{K}_c \mathbf{W} \Lambda^{-1} \mathbf{W}^T \mathbf{K}_c^T, \quad (15)$$

有关核函数 $\kappa(\mathbf{u}(i), \mathbf{u}(j))$ 的近似表达可写为如下形式:

$$\begin{aligned} \kappa(\mathbf{u}(i), \mathbf{u}(j)) &\approx (\mathbf{K}_c(i) \mathbf{W} \Lambda^{-1/2}) (\mathbf{K}_c(j) \mathbf{W} \Lambda^{-1/2})^T \\ &= \mathbf{z}^T(\mathbf{u}(i)) \mathbf{z}(\mathbf{u}(j)), \end{aligned} \quad (16)$$

其中, $\mathbf{K}_c(i) = [\kappa(\mathbf{u}(i), \mathbf{u}(g(1))), \dots, \kappa(\mathbf{u}(i), \mathbf{u}(g(m)))]$, $\mathbf{z}(\cdot) \in \mathbf{R}^m$ 的表达式如下:

$$\mathbf{z}(\cdot) = \Lambda^{-1/2} \mathbf{W}^T [\kappa(\cdot, \mathbf{u}(g(1))), \dots, \kappa(\cdot, \mathbf{u}(g(m)))]^T. \quad (17)$$

因此, 采用 Nyström 方法生成固定维数为 m 的特征空间, 可有效降低计算和空间复杂度.

3 基于 PRQ 采样的 Nyström 柯西核共轭梯度

3.1 在线 NCKCG 算法

在自适应滤波预测的在线学习中, 根据特征空间中变换后的输入 $\mathbf{z}(\mathbf{u}_k)$, 将非线性系统表示为

$$d_k = (\Omega^z)^T \mathbf{z}(\mathbf{u}_k) + q_k, \quad (18)$$

其中, $\Omega^z \in \mathbf{R}^m$ 是特征空间中的最优权向量, $\mathbf{z}(\mathbf{u}_k)$ 是在离散时间 k 时刻的输入, q_k 是系统噪声, $(\cdot)^z$ 表示在特征空间 \mathbf{z} 中的变量. 为了表达的简洁性, 定义 $\mathbf{z}_k = \mathbf{z}(\mathbf{u}_k)$. (11) 式中的最小值问题可采用权向量 Ω^z 表示为

$$\min_{\Omega^z} \frac{-v_k \eta \exp(-s_k)}{2\delta^2} E \left| \left(d_k - (\Omega^z)^T \mathbf{z}(\mathbf{u}_k) \right) \right|^2, \quad (19)$$

其等价于

$$\min_{\Omega^z} \frac{1}{2} (\Omega^z)^T \mathbf{R}^z \Omega^z - (\mathbf{c}^z)^T \Omega^z, \quad (20)$$

式中, \mathbf{R}^z 是 \mathbf{z}_k 的相关矩阵, \mathbf{c}^z 是 \mathbf{z}_k 的相关向量. 在线算法中, (20) 式中的 \mathbf{R}^z 和 \mathbf{c}^z 可采用指数衰减或滑动窗法以迭代形式估计:

$$\mathbf{R}_{k+1}^z = \lambda \mathbf{R}_k^z + \zeta_{k+1} \mathbf{z}_{k+1} \mathbf{z}_{k+1}^T, \quad (21)$$

$$\mathbf{c}_{k+1}^z = \lambda \mathbf{c}_k^z + \zeta_{k+1} d_{k+1} \mathbf{z}_{k+1}, \quad (22)$$

其中, $\lambda \in (0, 1)$ 是遗忘因子, $\zeta_k = \exp(-s_k) / \{\delta^2 [\eta^{-1} + s_k \exp(-s_k)]\}$.

则残差向量的递归形式为:

$$\begin{aligned} \mathbf{r}_{k+1}^z &= \mathbf{c}_{k+1}^z - \mathbf{R}_{k+1}^z \boldsymbol{\Omega}_{k+1}^z \\ &= \lambda \zeta_k^z + \zeta_{k+1} d_{k+1} \mathbf{z}_{k+1} \\ &\quad - (\lambda \mathbf{R}_k^z + \zeta_{k+1} \mathbf{z}_{k+1} \mathbf{z}_{k+1}^T) (\boldsymbol{\Omega}_{k+1}^z + \alpha_k \mathbf{p}_k^z) \\ &= \lambda \mathbf{r}_k^z - \alpha_k \mathbf{R}_{k+1}^z \mathbf{p}_k^z + \zeta_{k+1} \mathbf{z}_{k+1} \mathbf{e}_{k+1}. \end{aligned} \quad (23)$$

共轭梯度法的关键是选择共轭方向向量, 即连续梯度的共轭形式^[32]. 通常, 初始方向向量 \mathbf{p}_0^z 是负梯度方向. 则权重和方向向量的更新形式为

$$\boldsymbol{\Omega}_{k+1}^z = \boldsymbol{\Omega}_k^z + \alpha_k \mathbf{p}_k^z, \quad (24)$$

$$\mathbf{p}_{k+1}^z = \mathbf{r}_{k+1}^z + \beta_k \mathbf{r}_k^z, \quad (25)$$

其中, $\mathbf{r}_k^z = -\mathbf{g}_k^z$ 是残差向量, α_k 和 β_k 是步长更新因子, 其迭代方程为

$$\alpha_k = \frac{(\mathbf{p}_k^z)^T \mathbf{r}_k^z}{(\mathbf{p}_k^z)^T \mathbf{R}_{k+1}^z \mathbf{p}_k^z}, \quad (26)$$

$$\beta_k = \frac{(\mathbf{r}_{k+1}^z)^T (\mathbf{r}_{k+1}^z - \mathbf{r}_k^z)}{(\mathbf{r}_k^z)^T \mathbf{r}_k^z}. \quad (27)$$

最后, 结合 (20)–(27) 式, 得到 NCKCG 算法.

3.2 概率密度秩的量化采样

在 Nyström 稀疏策略^[29]中, 采样方法的选取对矩阵近似起着至关重要的作用, 将直接影响低秩矩阵的近似精度. 直接选取任意 m 列训练数据重构的低维矩阵可能在很大程度上不能反映数据集的整体特征, 在数据存在有大异常值情况下, 随机采样方法并不能获得预期的近似精度. 此外, k -均值采样法采用预先设定的阈值距离来获取迭代所对

应的样本点, 容易选取少数无效的边界值或离群值, 降低了近似精度. 为了保证近似精度, 采用 PRQ 的自适应采样法, 选择信息量更大的列, 使得所选数据包含更多的有效信息. PRQ 采样通过将原始数据替换为对应聚类中的地标点, 有效避免了随机采样和 k -均值采样均存在的初始点选取问题.

为进一步提高近似精度, 将 PRQ 采样方法应用于 NCKCG, 得到基于 PRQ 采样的 NCKCG-PRQ 算法. 给定 N 个的样本 $\mathbf{U} = \{\mathbf{u}(i)\}_{i=1}^N$, 采用帕尔森窗口方法^[29]估计 \mathbf{u} 点的概率密度:

$$p(\mathbf{u}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{(2\pi\epsilon^2)^{D/2}} \exp\left[-\|\mathbf{u} - \mathbf{u}(i)\|^2 / (2\epsilon^2)\right], \quad (28)$$

其中, D 代表数据维度, ϵ 是控制窗口大小的平滑因子.

根据西尔弗曼规则^[33], 合适的 ϵ 值可以通过下式选出:

$$\epsilon = \epsilon_u \{4/[N(2D+1)]\}^{1/(D+4)}, \quad (29)$$

其中, $\epsilon_u^2 = \frac{1}{D} \sum_{i=1}^N U_{ii}$ 且 U_{ii} 为输入样本协方差矩阵的对角元素. 结合 (28) 和 (29) 式, 可得估计的数据概率密度 $\{p_1, p_2, \dots, p_N\}$. 将原始数据序列 $\mathbf{U} = \{\mathbf{u}(i)\}_{i=1}^N$ 对应的概率密度大小按照降序排列, 重新排列后为 $\hat{\mathbf{U}} = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \dots, \hat{\mathbf{u}}_N\}$. 为了满足算法的稀疏要求, 将 $\hat{\mathbf{U}}$ 平均分为 m 个簇, 选取每个聚类中的第一个样本 (即概率密度最高的样本) 重构相应的地标点 $\{\hat{\mathbf{u}}(i)\}_{i=1}^m$.

最后, 用 $\{\hat{\mathbf{u}}(i)\}_{i=1}^m$ 在 (17) 式中替换 $\mathbf{u}(g(i))$, $i = 1, 2, \dots, m$. 表 1 总结了所提的 NCKCG-PRQ 算法.

表 1 NCKCG-PRQ 算法
Table 1. NCKCG-PRQ algorithm.

输入: 输入输出对 $\{\mathbf{u}_k, d_k\}$, $k = 1, 2, \dots$
初始化参数: $\hat{\mathbf{u}}(i)$ 为 PRQ 采样后样本点; $\boldsymbol{\Lambda}$ 和 \mathbf{W} 分别为有关 $\hat{\mathbf{U}}$ 核矩阵的特征值降序排列的对角矩阵和对应特征向量构成列正交矩阵; $\mathbf{K}_c(1) = [\kappa(\mathbf{u}(1), \hat{\mathbf{u}}(1)), \dots, \kappa(\mathbf{u}(1), \hat{\mathbf{u}}(m))]$ 为初始核向量; 映射输入 $\mathbf{z}_1 = \boldsymbol{\Lambda}^{-1/2} \mathbf{W}^T \mathbf{K}_c(1)^T$; 权重 $\boldsymbol{\Omega}_1^z = \mathbf{0}$; 加权函数 $\zeta_1 = \exp(-s_1) / \{\delta^2 [\eta^{-1} + s_1 \exp(-s_1)]\}$; 期望 $d_1 = (\boldsymbol{\Omega}_1^z)^T \mathbf{z}_1$; 相关矩阵 $\mathbf{R}_1^z = \zeta_1 \mathbf{z}_1 \mathbf{z}_1^T$; 互相关向量 $\mathbf{c}_1^z = \zeta_1 d_1 \mathbf{z}_1$; 冗余向量 $\mathbf{r}_1^z = \mathbf{c}_1^z - \mathbf{R}_1^z \boldsymbol{\Omega}_1^z$; 方向向量 $\mathbf{p}_1^z = \mathbf{r}_1^z$, 遗忘因子 $\lambda = 0.999$
循环 ($k = 2, 3, \dots$):
1. 输入核向量 $\mathbf{K}_c(i) = [\kappa(\mathbf{u}(i), \hat{\mathbf{u}}(1)), \dots, \kappa(\mathbf{u}(i), \hat{\mathbf{u}}(m))]$;
2. 映射输入 $\mathbf{z}(\cdot) = \boldsymbol{\Lambda}^{-1/2} \mathbf{W}^T [\kappa(\cdot, \hat{\mathbf{u}}(1)), \dots, \kappa(\cdot, \hat{\mathbf{u}}(m))]^T$;
3. 误差更新 $\mathbf{e}_{k+1} = d_{k+1} - (\boldsymbol{\Omega}_k^z)^T \mathbf{z}_{k+1}$, 加权函数 $\zeta_k = \exp(-s_k) / \{\delta^2 [\eta^{-1} + s_k \exp(-s_k)]\}$
4. 自相关矩阵更新 $\mathbf{R}_{k+1}^z = \lambda \mathbf{R}_k^z + \zeta_{k+1} \mathbf{z}_{k+1} \mathbf{z}_{k+1}^T$, 计算步长 $\alpha_k = \frac{(\mathbf{p}_k^z)^T \mathbf{r}_k^z}{(\mathbf{p}_k^z)^T \mathbf{R}_{k+1}^z \mathbf{p}_k^z}$;
5. 权重更新 $\boldsymbol{\Omega}_{k+1}^z = \boldsymbol{\Omega}_k^z + \alpha_k \mathbf{p}_k^z$, 残差向量更新 $\mathbf{r}_{k+1}^z = \lambda \mathbf{r}_k^z - \alpha_k \mathbf{R}_{k+1}^z \mathbf{p}_k^z + \zeta_{k+1} \mathbf{z}_{k+1} \mathbf{e}_{k+1}$;
6. 计算步长 $\beta_k = \frac{(\mathbf{r}_{k+1}^z)^T (\mathbf{r}_{k+1}^z - \mathbf{r}_k^z)}{(\mathbf{r}_k^z)^T \mathbf{r}_k^z}$, 共轭方向更新 $\mathbf{p}_{k+1}^z = \mathbf{r}_{k+1}^z + \beta_k \mathbf{r}_k^z$
循环终止

4 仿真结果及分析

为验证本文所提 NCKCG-PRQ 算法在混沌时间序列预测的有效性, 选择两类混沌时间序列, 即数值仿真的 MG 混沌时间序列^[34] 和由真实蔡氏电路产生的混沌时间序列^[31].

4.1 仿真实验模型

4.1.1 混沌时间序列模型

MG 混沌时间序列因其具有周期性和混沌动力学特性, 被广泛地应用于非线性系统建模. 因此, 考虑使用这个例子来验证本文所提出的 NCKCG-PRQ 算法在非高斯噪声环境下的优越性. MG 混沌时间序列由下式产生:

$$\frac{ds(t)}{dt} = -0.1s(t) + \frac{0.2s(t-\tau)}{1+s(t-\tau)^{10}}, \quad (30)$$

式中, 时滞参数 τ 影响了系统的混沌特性. τ 大于 17 时, 系统呈现混沌, 且其值越大, 混沌程度越高, 本文 τ 设置为 30. 在采样周期为 6 s 时, 对时间序列进行离散化. 为实现混沌时间序列的预测, 选取之前的 7 个样本点 $[s_{t-1}, s_{t-2}, \dots, s_{t-7}]$ 预测当前值 s_t .

4.1.2 蔡氏混沌时间序列模型

蔡氏电路可以产生一个现实存在的混沌时间序列. 为获取真实世界的混沌时间序列, 建立了如图 1 所示的电路系统. 电源模块将 220 V 电压转换为 12 V 电压用于蔡氏电路. 系统搭建中主要采用 LM358 运放, 100 nF 和 10 nF 各一个以及功率直插电感 20 mH, 详细的电路原理图如图 2 所示.

选取图 2 中电容 C_2 之间的电压值作为蔡氏混沌时间序列, 即图 1 中示波器的绿色波形所示. 在此次仿真实验中, 预测 C_2 之间的电压值并将其标准化为 0—0.1 用以保证实验预测精度. 之前最近的 5 个电压值被用作输入来预测当前的电压值.

4.2 噪声环境

为了验证算法在非高斯环境下的鲁棒性, 采用如下的混合高斯噪声模型, 即:

$$q(i) = b(i)q_b(i) + (1 - b(i))q_a(i), \quad (31)$$

噪声模型中, $q_a(i)$ 是零均值且方差为 $\sigma_q^2 = 0.0016$ 的高斯噪声; $q_b(i)$ 是用于建模脉冲噪声的 α -stable

分布^[35], 其参数为 $V_{\alpha\text{-stable}}(0.8, 0, 0.1, 0)$; $b(i)$ 是伯努利二项分布发生的概率, $\Pr\{b(i) = 1\} = c$ 且 $\Pr\{b(i) = 0\} = 1 - c$, 仿真中选取 $c = 0.1$. 在仿真实验中一个包含 2000 个样本点的数据集添加非高斯噪声后用作训练, 另外干净的 200 个样本点用作测试, 并另选 200 个样本点用于 Nyström 近似. 为了比较预测精度, 测试采用如下所示的 MSE(单位为 dB):

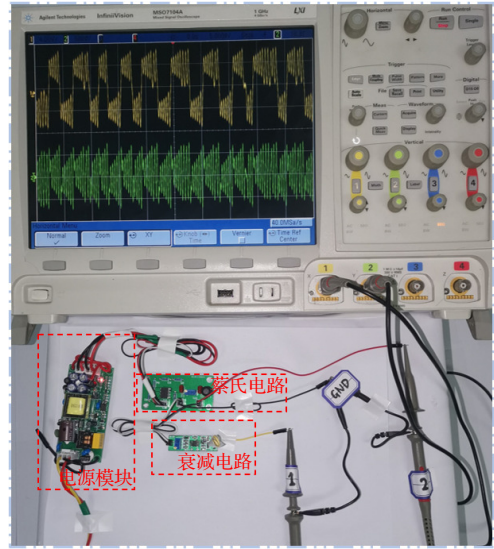


图 1 示波器示意图

Fig. 1. Schematic diagram of oscilloscope.

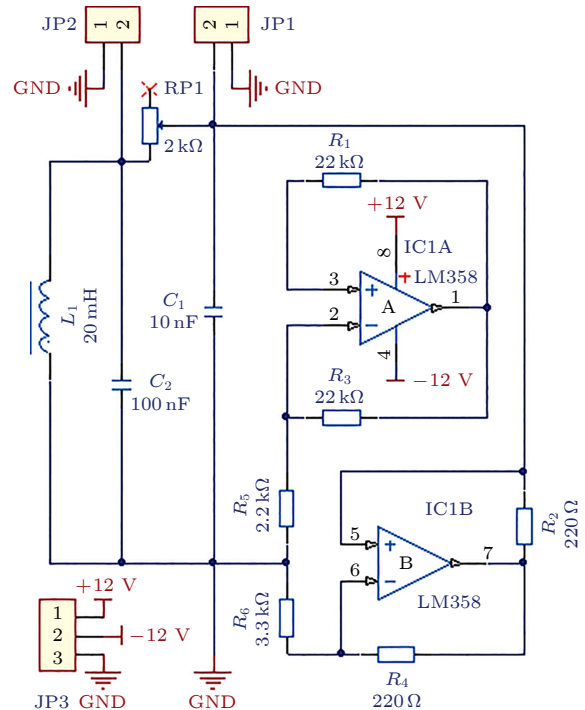


图 2 蔡氏电路原理图

Fig. 2. Schematic diagram of the Chua's circuit.

$$\text{MSE} = 10 \lg \left(\sum_{i=1}^N \frac{1}{N} [d(i) - \hat{d}(i)]^2 \right), \quad (32)$$

其中, $\hat{d}(i)$ 是 $d(i)$ 的估计值, 测试数据长度 $N = 200$. 为有效降低单次测试的不确定, 最终的测试 MSE 由 50 次独立的蒙特卡洛仿真在所有 200 个测试数据上获得的 MSE 的平均值构成. 为体现算法的稳态性能, 稳态 MSE 只取测试 MSE 最后 100 个数据的平均值. 所有进行比较算法中的高斯核带宽 h 设为 1. 为了体现算法比较的公平性, 在不同的仿真模型中, 参与比较算法的参数均相同.

4.3 仿真结果及分析

图 3 所示为不同采样点个数对 NCKCG-PRQ 算法的稳态 MSE 值和平均运算时间的影响. 从图 3(a) 可以看出: 1) 平均运算时间随 m 的增大而延长; 2) NCKCG-PRQ 算法滤波精度可以通过增大 m 的方法在一定程度上提高, 当 m 为 80 时, 其滤波精度趋于稳定; 3) m 维数越大, 滤波精度越高, 但计算时间越长. 当 m 为 60 时, 运算时间约为 1.5 s; 而当 m 为 80 时, 运算时间约为 2.5 s. 当 m 从 60 增至 80 时, 滤波精度提高所带来的优势不及运算时间增加带来的计算负担. 因此, 综合滤波精度和计算时间两方面, 在 MG 混沌时间序列仿真中将采样点设置为 60. 为公平起见, 所有参与比较的稀疏算法的采样点个数均设置为 60. 从图 3(b) 可以看出, 随着 m 增大, 滤波精度先增大而后减小, 但运算时间是一直延长的. 在 $m = 13$ 时, 滤波精度达到最高. 因此, 同时考虑滤波精度和运算时间两方面, 在蔡氏混沌时间序列仿真中最优的采样点个数为 13. 故参与比较的稀疏算法采样点数目也同样设置为 13.

比较算法包含 CKCG^[30] 算法, 随机傅里叶特征柯西共轭梯度 (random Fourier features Cauchy conjugate gradient, RFFCCG)^[36] 算法, Nyström 核递归广义最大相关熵 (Nyström kernel recursive generalized maximum correntropy with PRQ sampling, NKRGMCP-PRQ)^[29] 算法, Nyström 核共轭梯度 (Nyström kernel conjugate gradient based on k-means sampling, NKCG-KM)^[28] 算法. 其中, CKCG 算法是采用具有鲁棒特性的柯西核函数作为代价函数, 通过 CG 有效优化该代价函数, 但随着输入数据的不断加入, 核矩阵规模不断增

大, 计算存储负担加重. CKCG 算法因其鲁棒性强、滤波精度高和运算时间长的特点, 是未经过稀疏处理的典型代表. RFFCCG 算法是采用鲁棒柯西函数作为代价函数, 通过 CG 方法优化该代价函数, 利用随机傅里叶特征和均匀随机数来近似核函数, 因此具备低计算复杂度且与传统核算法相当的性能, 是固定维度稀疏算法的代表. NKRGMCP-PRQ 算法采用信息论学习 (information theoretic learning, ITL) 准则中广泛采用的广义最大相关熵准则作为代价函数, 通过经典的递归更新方式、基于 Nyström 低秩近似的核矩阵以及 PRQ 采样方法进一步提高了滤波精度. NKCG-KM 因采用传统二阶误差准则作为代价函数在非高斯环境下不具备鲁棒性. 因此, 为了全面体现所提 NCKCG-PRQ 算法的优越性, 本文是从代价函数的鲁棒性、优化方法、稀疏策略以及采样方法选择比较算法. 为了体现比较的公平性, 对所有参与比较的算法参数进行设置, 使其都达到理想的滤波精度.

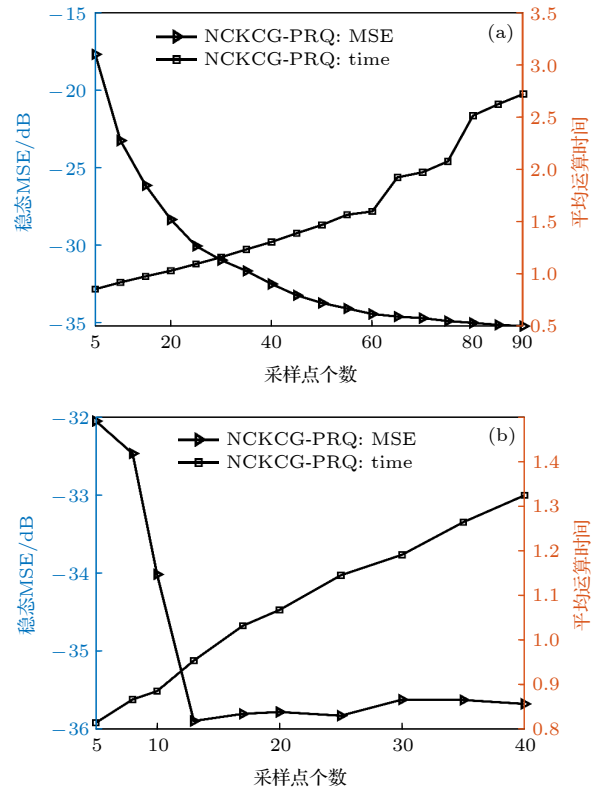


图 3 不同采样点个数 m 对 NCKCG-PRQ 算法的稳态 MSE 值和平均运算时间的影响 (a) MG 混沌时间序列; (b) 蔡氏混沌时间序列

Fig. 3. Influence of different number of sampling points on steady-state MSE value and average operation time of NCKCG-PRQ algorithm: (a) MG chaotic time series; (b) chaotic time series based on Chua's circuit.

算法的学习曲线和具体的参数选择如图 4 所示. 其中, $\sigma = 0.5$ 是 CKCG 中柯西核代价函数的核带宽, $\lambda = 0.25$ 是柯西核函数参数; $\alpha = 2$ 和 $\sigma = 0.6$ 是 NKRGM-PRQ 参数; $\gamma = 0.3$ 为 RFFCCG 中柯西代价函数参数; $\lambda = 0.999$ 是 NKCG-KM 算法遗忘因子; $\delta = 0.5$ 和 $\eta = 0.25$ 是 NCKCG-PRQ 算法柯西核代价函数参数.

各个算法的平均字典个数大小 (即采样点数目), 平均计算时间和稳态 MSE 的详细仿真结果如表 2 和表 3 所示. 从图 4 可以看出, 由于脉冲噪声的存在, NKCG-KM 算法的性能出现发散, 由此可知基于二阶统计测度的算法在脉冲噪声下不具备鲁棒性. 利用柯西损失来惩罚噪声项, 使基于柯西核代价函数的 NCKCG-PRQ 算法对非高斯噪声, 特别是重尾噪声有很好的抑制效果. 从表 2 和表 3

可知, 与稀疏算法 NKRGM-PRQ 和 RFFCCG 相比, 从稳态 MSE 角度比较, 所提的 NCKCG-PRQ 算法具有最好的滤波精度; 从运算时间角度比较, 所提的 NCKCG-PRQ 算法具有最短的运算时间. NCKCG-PRQ 算法在低计算复杂度下最大程度逼近了高计算复杂度 CKCG 算法的滤波精度. 图 5 是显示了测试集中验证所提算法的预测结果, 其中, 纵坐标已作归一化处理. 从图 5 可知, 在合成数据的 MG 混沌时间序列中所提算法在该仿真实验中能够有效地实现混沌序列的预测; 在真实采集的蔡氏电路混沌时间序列中, 所提算法基本实现了真实值和预测的输出值的拟合. 由此可知, 无论是在仿真的 MG 混沌时间序列还是真实蔡氏电路产生的混沌时间序列, 所提的 NCKCG-PRQ 算法在鲁棒性, 稳态性能和计算复杂度上均具有较大的优势.

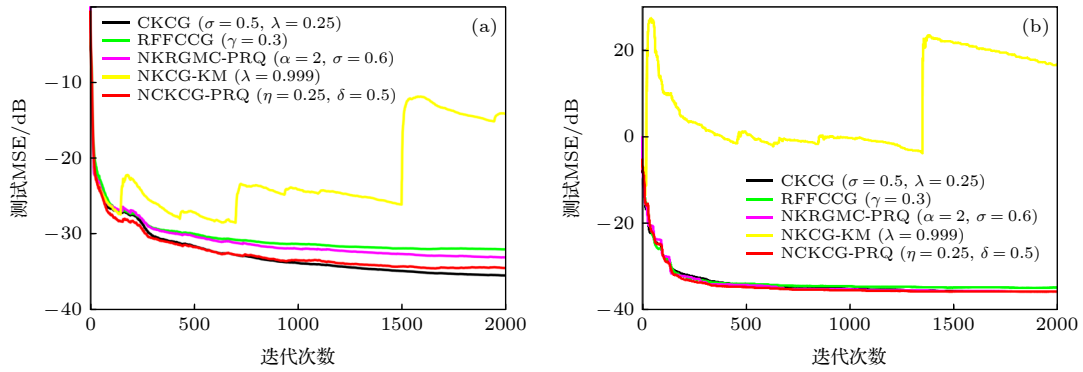


图 4 在脉冲噪声环境下不同算法的测试 MSE 学习曲线 (a) MG 时间序列; (b) 蔡氏混沌时间序列

Fig. 4. Testing MSE learning curves of different algorithms in impulsive noise environment: (a) MG chaotic time series; (b) chaotic time series based on Chua's circuit.

表 2 不同算法在 MG 混沌时间序列中的仿真结果

Table 2. Simulation results of different algorithms in MG chaotic time series.

实验模型	算法	字典数目	运算时间/s	稳态MSE/dB
MG混沌时间序列	CKCG	2000	41.486	-35.443
	RFFCCG	60	2.095	-32.006
	NKRGM-PRQ	60	2.383	-33.068
	NKCG-KM	60	3.745	N/A
	NCKCG-PRQ	60	1.584	-34.446

表 3 不同算法在蔡氏电路混沌时间序列中的仿真结果

Table 3. Simulation results of different algorithms in chaotic time series based on Chua's circuit.

实验模型	算法	字典数目	运算时间/s	稳态MSE/dB
蔡氏混沌时间序列	CKCG	2000	42.618	-35.840
	RFFCCG	13	1.128	-34.926
	NKRGM-PRQ	13	0.988	-35.819
	NKCG-KM	13	1.146	N/A
	NCKCG-PRQ	13	0.957	-35.865

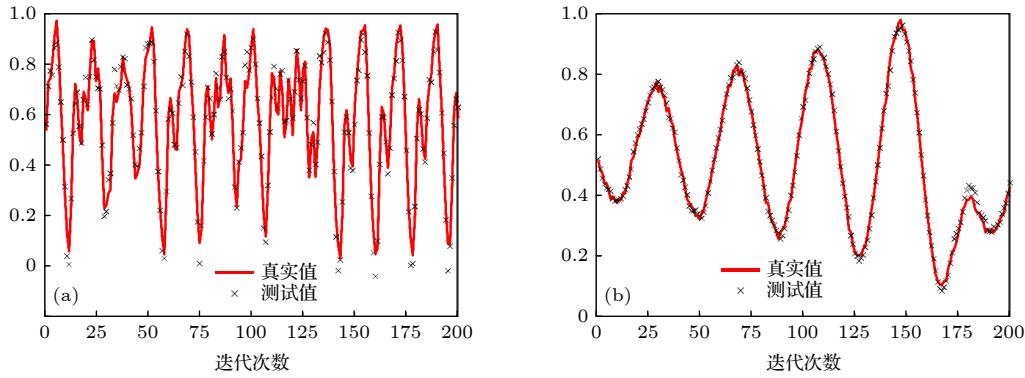


图5 NCKCG-PRQ算法对测试数据的最终预测结果 (a) MG混沌时间序列; (b) 蔡氏混沌时间序列

Fig. 5. Final predicted results of NCKCG-PRQ algorithm for the test sets: (a) MG chaotic time series; (b) chaotic time series based on Chua's circuit.

5 结论

本文将基于PRQ采样的Nyström方法作为稀疏策略应用于核共轭梯度优化法,有效地抑制了核共轭梯度法的核矩阵网络规模增长.进一步,采用半平方法转化柯西核函数生成全局凸的鲁棒函数作为核共轭梯度法的代价函数,提出了一种新的具有鲁棒性的基于PRQ采样的NCKCG-PRQ算法.NCKCG-PRQ在存储需求较低的情况下,与没有采取稀疏策略的CKCG算法拥有近乎相当的滤波精度.在数值仿真的MG混沌时间序列和由蔡氏电路真实产生的混沌时间序列的预测中,结果表明在存在脉冲噪声的情况下,NCKCG-PRQ与其他固定维度的自适应滤波器相比,具有较低计算复杂度和较高的预测精度,有效证实了所提NCKCG-PRQ算法在混沌时间序列预测中的性能优势.

参考文献

- [1] Lin Y, Liu W B, Shen Q 2018 *Acta Phys. Sin.* **67** 230502 (in Chinese) [林毅, 刘文波, 沈骞 2018 物理学报 **67** 230502]
- [2] Wang M J, Wu Z T, Feng J C 2015 *Acta Phys. Sin.* **64** 040503 (in Chinese) [王梦蛟, 吴中堂, 冯久超 2015 物理学报 **64** 040503]
- [3] Tang Z J, Ren F, Peng T, Wang W B 2014 *Acta Phys. Sin.* **63** 050505 (in Chinese) [唐舟进, 任峰, 彭涛, 王文博 2014 物理学报 **63** 050505]
- [4] Mei Y, Tan G Z, Liu Z T, Wu H 2018 *Acta Phys. Sin.* **67** 080502 (in Chinese) [梅英, 谭冠政, 刘振焘, 武鹤 2018 物理学报 **67** 080502]
- [5] Wang X Y, Han M, Wang Y N 2013 *Acta Phys. Sin.* **62** 050504 (in Chinese) [王新迎, 韩敏, 王亚楠 2013 物理学报 **62** 050504]
- [6] Wang S Y, Shi C F, Qian G B, Wang W L 2018 *Acta Phys. Sin.* **67** 018401 (in Chinese) [王世元, 史春芬, 钱国兵, 王万里 2018 物理学报 **67** 018401]
- [7] Peng L B, Li X F, Bi D J, Xie Y L 2018 *Signal Process. Lett.* **25** 1335
- [8] Zhao Y P, Zhang L Y, Li D C, Wang L F, Jiang H Z 2013 *Acta Phys. Sin.* **62** 120511 (in Chinese) [赵永平, 张丽艳, 李德才, 王立峰, 蒋洪章 2013 物理学报 **62** 120511]
- [9] Zhang J S, Dang J L, Li H C 2007 *Acta Phys. Sin.* **56** 67 (in Chinese) [张家树, 党建亮, 李恒超 2007 物理学报 **56** 67]
- [10] Zhang H B, Sun X D, He Y L 2014 *Acta Phys. Sin.* **63** 040505 (in Chinese) [张洪宾, 孙小端, 贺玉龙 2014 物理学报 **63** 040505]
- [11] Tang Z J, Peng T, Wang W B 2014 *Acta Phys. Sin.* **63** 130504 (in Chinese) [唐舟进, 彭涛, 王文博 2014 物理学报 **63** 130504]
- [12] Huo Y L, Wang D F, Long X Q, Lian P J, Qi Y F 2021 *Acta Phys. Sin.* **70** 158401 (in Chinese) [火元莲, 王丹凤, 龙小强, 连培君, 齐永锋 2021 物理学报 **70** 158401]
- [13] Huo Y L, Wang D F, Long X Q, Lian P J, Qi Y F 2021 *Acta Phys. Sin.* **70** 028401 (in Chinese) [火元莲, 王丹凤, 龙小强, 连培君, 齐永锋 2021 物理学报 **70** 028401]
- [14] Wu Z, Shi J, Xie Z, Ma W 2015 *Signal Process.* **117** 11
- [15] Liu W F, Pokharel P P, Príncipe J C 2008 *IEEE Trans. Signal Process.* **56** 543
- [16] Engel Y, Mammor S, Meir R 2004 *IEEE Trans. Signal Process.* **52** 2275
- [17] Chen B D, Príncipe J C 2012 *Signal Process. Lett.* **19** 491
- [18] Li C G, Shen P C, Liu Y, Zhang Z Y 2013 *IEEE Trans. Signal Process.* **61** 4011
- [19] Chen B D, Xing L, Zhao H Q, Zheng N N, Príncipe J C 2016 *IEEE Trans. Signal Process.* **64** 3376
- [20] Li X L, Lu Q M, Dong Y S, Tao D C 2019 *Trans. Neural Netw. Learn. Syst.* **30** 2067
- [21] Shi W, Xiong K, Wang S Y 2019 *IEEE Access.* **7** 120548
- [22] Lei Y W, Hu T, Li G Y, Tang K 2020 *Trans. Neural Netw. Learn. Syst.* **31** 4394
- [23] Zhang M, Wang X J, Chen X M, Zhang A X 2018 *IEEE Trans. Signal Process.* **66** 4377
- [24] Xiong K, Herbert H C, Wang S Y 2021 *IEEE Trans. Cybern.* **51** 5497
- [25] Chen B D, Zhao S L, Zhu P P, Príncipe J C 2013 *Trans. Neural Netw. Learn. Syst.* **24** 1484
- [26] Zhang T, Wang S Y, Huang X W, Jia L 2020 *Signal Process. Lett.* **27** 361
- [27] Zhang T, He F L, Zheng Z, Wang S Y 2020 *IEEE Trans.*

- Circuits Syst. Express Briefs* **67** 2772
- [28] He F L, Xiong K, Wang S Y 2020 *IEEE Access.* **8** 18716
- [29] Zhang T, Wang S Y 2020 *Signal Process. Lett.* **27** 1535
- [30] Qi L T, Shen M L, Wang D L, Wang S Y 2021 *Signal Process. Lett.* **28** 1011
- [31] Zhang H N, Yang B, Wang L, Wang S Y 2021 *IEEE Trans. Signal Process.* **69** 1859
- [32] Xiong K, Wang S Y 2019 *Signal Process. Lett.* **26** 740
- [33] Qin Z D, Chen B D, GU Y T, Zheng N N, Principe J C 2020 *IEEE Trans. Neural Netw. Learn. Syst.* **31** 3100
- [34] Zheng Y F, Wang S Y, Feng J C, Tse C K 2016 *Digit. Signal Process.* **48** 130
- [35] Weng B W, Barner K E 2005 *IEEE Trans. Signal Process.* **53** 2588
- [36] Huang X W, Wang S Y, Xiong K 2019 *Symmetry* **11** 1323

Prediction of chaotic time series based on Nyström Cauchy kernel conjugate gradient algorithm*

Qi Le-Tian¹⁾²⁾ Wang Shi-Yuan^{1)2)†} Shen Ming-Lin¹⁾²⁾
Huang Gang-Yi¹⁾²⁾

1) (College of Electronic Information Engineering, Southwest University, Chongqing 400715, China)

2) (Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing, Chongqing 400715, China)

(Received 8 December 2021; revised manuscript received 27 January 2022)

Abstract

Chaotic time series can well reflect the nonlinearity and non-stationarity of real environment changes. The traditional kernel adaptive filter (KAF) with second-order statistical characteristics suffers performance degeneration dramatically for predicting chaotic time series containing noises and outliers. In order to improve the robustness of adaptive filters in the presence of impulsive noise, a nonlinear similarity measure named Cauchy kernel loss (CKL) is proposed, and the global convexity of CKL is guaranteed by the half-quadratic (HQ) method. To improve the convergence rate of stochastic gradient descent and avoid a local optimum simultaneously, the conjugate gradient (CG) method is used to optimize CKL. Furthermore, to address the issue of kernel matrix network growth, the Nyström sparse strategy is adopted to approximate the kernel matrix and then the probability density rank-based quantization (PRQ) is used to improve the approximation accuracy. To this end, a novel Nyström Cauchy kernel conjugate gradient with PRQ (NCKCG-PRQ) algorithm is proposed for the prediction of chaotic time series in this paper. Simulations on prediction of synthetic and real-world chaotic time series validate the advantages of the proposed algorithm in terms of filtering accuracy, robustness, and computational storage complexity.

Keywords: prediction of chaotic time series, kernel adaptive filtering, Cauchy kernel loss, Nyström method

PACS: 84.30.Vn, 05.45.Tp, 05.10.Ln

DOI: 10.7498/aps.71.20212274

* Project supported by the National Nature Science Foundation of China (Grant No. 62071391), Natural Science Foundation of Chongqing (Grant No. cstc2020jcyj-msxmX0234), and Fundamental Research Funds for the Central Universities (Grant No. 2020jd001).

† Corresponding author. E-mail: wsy@swu.edu.cn