

DOI: 10.13718/j.cnki.xdzk.2022.04.023

基于稀疏系统辨识的 广义递归核风险敏感算法

王代丽, 王世元, 张涛, 齐乐天

西南大学 电子信息工程学院/非线性电路与智能信息处理重庆市重点实验室, 重庆 400715

摘要: 为了降低非高斯噪声对系统性能的影响, 核风险敏感损失函数(Kernel Risk-Sensitive Loss, KRSL)因其较高的凸性而被广泛应用为自适应滤波器的代价函数。基于此, 为了提高非高斯情况下系统的滤波精度, 本文采用广义高斯密度(Generalized Gaussian Density, GGD)函数作为KRSL的核函数, 进而提出了一种广义核风险敏感损失函数(Generalized Kernel Risk-Sensitive Loss, GKRLS), 并给出了GKRLS的重要性质。为了进一步识别稀疏系统, 结合GKRLS的优点, 采用递归更新方式提出了一种基于稀疏惩罚约束的广义递归核风险敏感损失(Generalized Recursive Kernel Risk-Sensitive Loss with Sparse Penalty Constraint, GRKRLS-SPC)算法。仿真结果表明, GRKRLS-SPC算法能够显著提高非高斯噪声下系统的滤波精度和鲁棒性。

关键词: 广义相关熵; 核风险敏感损失函数; 稀疏系统; 辨识;
自适应滤波

中图分类号: TN911.7

文献标志码: A

开放科学(资源服务)标识码(OSID):



文章编号: 1673-9868(2022)04-0196-10

Generalized Recursive Kernel Risk-Sensitive Loss Algorithm Based on Sparse System Identification

WANG Daili, WANG Shiyuan, ZHANG Tao, QI Letian

College of Electronic Information Engineering, Southwest University/Chongqing Key Laboratory of
Nonlinear Circuits and Intelligent Information Processing, Chongqing 400715, China

Abstract: The kernel risk-sensitive loss (KRSL) is widely used as the cost function of adaptive filters to reduce the influence of non-Gaussian noises on system performance owing to its high convexity. In this paper, a generalized kernel risk-sensitive loss (GKRLS) is proposed by using a generalized Gaussian density (GGD) function as the kernel of KRSL to improve the filtering accuracy of system in non-Gaussian noises.

收稿日期: 2021-06-22

基金项目: 国家自然科学基金项目(62071391); 重庆市自然科学基金项目(cstc2020jcyj-msxmX0234); 中央高校基本科研业务费项目(2020jd001)。

作者简介: 王代丽, 硕士研究生, 主要从事信号处理。

通信作者: 王世元, 教授, 博士研究生导师。

The important properties of GKRS� are presented for optimization. Furthermore, combined with the advantages of the GKRS� under the sparse penalty constrain, the recursive updating method is used to generate a novel generalized recursive kernel risk-sensitive loss with the sparse penalty constrain (GRKRS�-SPC) algorithm for identification of sparse systems. The superiorities of GRKRS�-SPC from the aspects of accuracy and robustness are verified by Monte Carlo simulations.

Key words: generalized correntropy; kernel risk-sensitive loss; sparse system; identification; adaptive filtering

自适应滤波是统计信号处理的重要组成部分,而稀疏自适应滤波是自适应滤波领域中不可或缺的部分,其显著特征是脉冲响应的大部分分量是零或者接近于零.在实际场景中,存在大量的稀疏系统,例如数字电视传输通道^[1]、回波路径^[2]、信道估计^[3]等.由于稀疏系统通常是不确定的,因此需采用基于稀疏系统的自适应滤波算法对其进行辨识^[4-5].

应用在稀疏系统中的自适应滤波算法通常采用与稀疏性相关的范数作为稀疏惩罚约束项(Sparse Penalty Constraint, SPC)^[6-7],如 l_1 范数, l_p 范数和 l_0 范数,其中基于 l_1 范数的最小均方自适应滤波算法包括零吸引最小均方(Zero-attracting Least Mean Square, ZA-LMS)算法^[8]和加权零吸引最小均方(Reweighted Zero-attracting Least Mean Square, RZA-LMS)算法^[9]等,但是 ZA-LMS 和 RZA-LMS 的收敛速率较慢,因此,为了提高收敛速率提出了基于零吸引的递归最小二乘(Zero-attracting Recursive Least Squares, ZA-RLS)算法^[10].通常由于 l_1 范数存在零点处的非光滑性的缺点,会导致算法性能降低,因此,引入 l_p 范数作为 SPC 提高稀疏系统的滤波精度,进而提出了基于平方根变步长 l_p 范数的 LMS 算法^[11],以变步长的形式分析了稀疏系统中稳态均方误差和收敛速率之间的关系.通常最小任何 l_p 范数($0 < p < 1$)可等效为最小 l_0 范数,但这是一个非凸优化问题,因此为了解决 l_0 范数中的(Non-deterministic Polynomial, NP)难问题提出了相关熵诱导度量(Correntropy Induced Metric, CIM)用来近似 l_0 范数^[12],其典型应用是具有 CIM 的最大相关熵(Maximum Correntropy Criterion with CIM, CIMMCC)算法^[13],CIMMCC 算法能够提高非高斯环境下稀疏自适应滤波算法的鲁棒性.

因为非高斯噪声在自然界中是普遍存在的,所以高斯噪声环境下的稀疏算法在非高斯噪声环境下会产生性能不稳定或退化等问题.从信息理论学习(Information Theoretic Learning, ITL)^[14]的观点出发,为解决非高斯噪声对算法性能的影响,提出了广义相关熵(Generalized Correntropic, GC)准则.GC 准则本质上是定义在特征空间中相似性度量的一种方法,利用数据的高阶统计特性消除非高斯噪声,其在自适应滤波中最经典的应用是广义相关熵损失(Generalized Correntropic Loss, GC-Loss)算法^[15].而在稀疏系统中,利用广义最大相关熵准则(Generalized Maximum Correntropy Criterion, GMCC),同时采用 CIM 作为稀疏惩罚约束项进而提出了具有稀疏惩罚约束的递归广义最大相关熵(Recursive Generalized Maximum Correntropy Criterion with Sparse Penalty Constraint, RGMCC-SPC)算法^[16],该算法在 CIMMCC 算法基础上采用广义递归的更新方式提高了收敛速率和滤波精度.为了进一步解决 RGMCC-SPC 算法中非零均值误差在零处误差识别精度较差的问题,提出了可变中心的 RGMCC-SPC(Variable Center RGMCC-SPC, RGMCCVC-SPC)算法^[17].然而由于 GC-Loss 的性能表面具有高度非凸的特性,导致算法的收敛性能较差.为了解决这个问题,引入定义在特征空间的核风险敏感损失(Kernel Risk-Sensitive Loss, KRS�)函数^[18],使其在非高斯噪声中的性能优于 GC-Loss.

启发于 KRS� 和 GMCC,本文提出了一种新的应用在稀疏系统下的广义自适应滤波算法,该算法以广义高斯密度(Generalized Gaussian Density, GGD)^[19]函数作为 KRS� 函数中的核函数,结合稀疏惩罚约束项,以递归的方式进行更新,进而为稀疏系统辨识设计出具有稀疏惩罚约束项的广义递归核风险敏感损失(Generalized Recursive Kernel Risk-Sensitive Loss with Sparse Penalty Constraint, GRKRS�-SPC)算法.

所提出的 GRKRSL-SPC 算法利用了特征空间中映射数据非二阶统计量的特征,以指数形式强调较大误差的相似性,使得算法在非高斯噪声环境下,能够同时具有强鲁棒性和高滤波精度的特性.

1 背景介绍

本节在核风险敏感损失函数和广义核风险敏感损失函数的基础上,提出了最小化广义核风险敏感损失函数准则.

1.1 核风险敏感损失函数

核风险敏感损失函数是指定义在核空间中两个随机变量 X 和 Y 的相似度测量,其表达式如下所示:

$$J_{\lambda}(X, Y) = \frac{1}{\lambda} \mathbb{E}[\exp(\lambda(1 - k_{\sigma}(X - Y)))] = \frac{1}{\lambda} \int \exp(\lambda(1 - k_{\sigma}(x - y))) dF_{XY}(x, y) \quad (1)$$

其中 $\lambda > 0$ 是风险敏感参数, \mathbb{E} 表示期望符号, $F_{XY}(x, y)$ 表示关于随机变量 X 和 Y 的联合分布函数, $k_{\sigma}(\cdot)$ 是指带宽为 σ 的 Mercer 核. 其表达式为

$$k_{\sigma}(x - y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - y)^2}{2\sigma^2}\right) \quad (2)$$

本质上,公式(1)表示核风险敏感损失函数的指数形式只包含二阶统计量,实际上,不同阶的统计量在实际应用中更为普遍.因此,可以利用广义高斯密度函数作为核风险敏感损失函数的核函数并以此设计广义核风险敏感损失函数.

1.2 广义核风险敏感损失函数

给定如下具有零均值的广义高斯密度函数^[19]:

$$G_{a,b}(x) = \frac{\alpha}{2b\Gamma(1/\alpha)} \exp\left(-\left|\frac{x}{b}\right|^{\alpha}\right) = \gamma_{a,b} \exp(-\beta|x|^{\alpha}) \quad (3)$$

其中, $\Gamma(\cdot)$ 表示伽马函数, $b > 0$ 是带宽, $\alpha > 0$ 是形状参数, $\beta = 1/b^{\alpha}$ 表示核参数. 从公式(3)可以看出广义高斯密度具有普适性. 当 $\alpha = 1$ 时, 广义高斯密度为拉普拉斯分布; 当 $\alpha = 2$ 时, 则为高斯分布; 而当 $\alpha \rightarrow \infty$ 时, 则为均匀分布. 根据文献[15], 定义广义相关熵损失函数的表达式为

$$J_{GC}(X, Y) = \frac{1}{\lambda} \mathbb{E}[\lambda(\gamma_{a,b} - G_{a,b}(X - Y))] = \gamma_{a,b} - \mathbb{E}[G_{a,b}(X - Y)] \quad (4)$$

结合公式(1)和公式(4), 可得广义核风险敏感损失函数的表达式为

$$J_G(X, Y) = \frac{1}{\lambda} \mathbb{E}[\exp(\lambda(\gamma_{a,b} - G_{a,b}(X - Y)))] = \frac{1}{\lambda} \int \exp(\lambda(\gamma_{a,b} - G_{a,b}(x - y))) dF_{XY}(x, y) \quad (5)$$

其中, 当 $0 < \alpha \leq 2$ 时, $G_{a,b}(\cdot)$ 表示为 Mercer 核, 所以在此范围内, 广义核风险敏感损失函数可用传统风险敏感损失函数的类似形式重新表示为

$$J_G(x, y) = \frac{1}{\lambda} \mathbb{E} \left[\exp \left(\lambda \left(\frac{1}{\alpha} \|\Phi(x) - \Phi(y)\|_{\mathbf{H}}^{\alpha} \right) \right) \right] \quad (6)$$

其中, $\Phi(\cdot)$ 表示的是将数据从原始空间转换到核空间 \mathbf{H} 的非线性映射算子, $\|\cdot\|_{\mathbf{H}}$ 是核空间中的范数. 通过比较式(4)和式(5)可知, 式(5)可以看作是广义相关熵损失函数的指数形式. 当 $\alpha = 2$ 时, 核风险敏感损失函数是广义核风险损失函数的一个特例, 所以, 广义核风险损失函数具有普适性. 然而, 因为 $F_{XY}(x, y)$ 通常未知, 所以式(5)和式(6)难以计算. 当有 L 个样本 $\{x(i), y(i)\}_{i=1}^L$ 可用时, 通过计算该 L 个样本

的平均值来获得广义核风险损失函数的近似值,即

$$\hat{J}_G(X, Y) = \frac{1}{L\lambda} \sum_{i=1}^L \exp(\lambda(\gamma_{a,b} - G_{a,b}(x(i) - y(i)))) \quad (7)$$

1.3 广义核风险敏感损失函数性质

根据式(5)和式(7),关于广义核风险敏感损失函数的重要性质如下:

1) $J_G(X, Y)$ 是对称、正定、有界的.

2) $J_G(X, Y)$ 是具有广义特性的损失函数,在特定情况下可以转换为核风险敏感损失函数、广义熵损失函数和均值 P 幂误差^[20].

证 当 $\alpha = 2$ 时,广义核风险敏感损失函数退化为非广义的形式,风险敏感参数由 λ 转变为 $\lambda/\gamma_{a,b}$. 因为存在极限 $\lim_{x \rightarrow 0^+} \exp(x) \rightarrow 1 + x$, 当 $\lambda \rightarrow 0^+$ 和 $\beta \rightarrow 0^+$ 时, $J_G(X, Y)$ 可分别近似于 $1/\lambda + J_{G,c}(X, Y)$ 和 $1/\lambda + \beta E[(X - Y)^\alpha]$, 即意味着广义核风险敏感损失函数可近似于广义相关熵损失函数和均值 P 幂误差函数.

3) 令 $\mathbf{e} = X - Y = [e(1), e(2), \dots, e(L)]^T$, 其中 $e(i) = x(i) - y(i)$, $i = 1, 2, \dots, L$. 当 $\alpha > 1$ 且 $\beta > 0$ 时, 在任意的 $|e(i)| \leq [(\alpha - 1)/(\alpha\beta)]^{1/\alpha}$ 处 $\hat{J}_G(X, Y)$ 均具有凸性. 同样地, 如果 λ 的取值超过某一值时, $\hat{J}_G(X, Y)$ 在任意的非零 \mathbf{e} 处也具有凸性. 然而, 当 $\beta \rightarrow 0^+$ 时, 在 $\alpha \in (0, 1]$ 情况下, $\hat{J}_G(X, Y)$ 对任意 \mathbf{e} 都是凹面, 而当 $\alpha > 1$, $\hat{J}_G(X, Y)$ 对任意 \mathbf{e} 都是凸面.

证 黑塞矩阵如下:

$$\mathbf{H}_{\hat{J}}(\mathbf{e}) = \text{diag}[h_1, h_2, \dots, h_L] \quad (8)$$

其中, $h_i = \xi_i((\alpha - 1) - \alpha\beta |e(i)|^\alpha) + \alpha\beta\lambda \exp(-\beta |e(i)|^\alpha) |e(i)|^\alpha$, $\xi_i = \frac{\alpha\beta\gamma_{a\beta}}{L} |e(i)|^{\alpha-2} \times \exp(-\beta |e(i)|^\alpha) \times \exp(\gamma_{a\beta}(1 - \exp(-\beta |e(i)|^\alpha)))$. 当 $\alpha > 1$ 且 $\beta > 0$, 有 $h_i > 0$; 继而对任意 \mathbf{e} , 当 $|e(i)| \leq [(\alpha - 1)/(\alpha\beta)]^{1/\alpha}$, 有 $\mathbf{H}_{\hat{J}}(\mathbf{e}) \geq 0$.

性质(3)阐述了 $\hat{J}_G(X, Y)$ 的凸性是由参数 α, β, λ 和 \mathbf{e} 决定的. 风险敏感参数 λ 控制核风险敏感损失函数的凸范围, λ 取值越大, 凸范围就越大, 通常选择合适 β 可保证核风险敏感损失函数的凸性.

因此根据 $\hat{J}_G(X, Y)$ 的凸性, 核风险敏感损失函数更加适合设计自适应滤波器.

1.4 最小化广义核风险敏感损失函数准则

给定一个数学模型, 如下所示:

$$d(i) = \mathbf{\Omega}_o^T \mathbf{X}(i) + \nu(i) \quad (9)$$

其中, $d(i)$ 表示第 i 时刻的期望输出, $\mathbf{\Omega}_o = [\Omega_o^0, \Omega_o^1, \dots, \Omega_o^{L-1}]^T \in R^{L \times 1}$ 表示未知系统的最优权重, $\mathbf{X}(i) = [x(i), x(i-1), \dots, x(i-L+1)]^T \in R^{L \times 1}$ 是输入信号, $\nu(i)$ 则代表噪声. 采用 $\mathbf{\Omega}$ 表示权重的估计值, 基于公式(7), 可得如下广义核风险敏感损失函数准则^[21]:

$$J_{GKRS L}(\mathbf{\Omega}) = \frac{1}{L\lambda} \sum_{i=1}^L \exp(\lambda(\gamma_{a,b} - G_{a,b}(e(i)))) = \frac{1}{L\lambda} \sum_{i=1}^L \exp(\lambda(1 - \exp(-\beta |d(i) - \mathbf{\Omega}^T \mathbf{X}(i)|^\alpha))) \quad (10)$$

通过最小化公式(10)可以获得广义核风险敏感损失函数的最优解, 称之为最小化广义核风险敏感损失函数准则, 其表达式如下:

$$\mathbf{\Omega}_{\text{opt}} = [\mathbf{R}_{\text{xx}}^g]^{-1} \mathbf{R}_{\text{dx}}^g \quad (11)$$

其中, $\mathbf{R}_{\text{xx}}^g = \sum_{i=1}^L g(e(i)) \mathbf{X}(i) \mathbf{X}(i)^T$ 是输入信号的加权自相关矩阵; $\mathbf{R}_{\text{dx}}^g = \sum_{i=1}^L g(e(i)) d(i) \mathbf{X}(i)$ 是期望输

出与输入信号之间的加权互相关向量, 这里 $g(e(i)) = \exp(\lambda(1 - \exp(-\beta |e(i)|^\alpha)))\exp(-\beta |e(i)|^\alpha) |e(i)|^{\alpha-2}$.

最后, 图 1 显示了不同 α 下对广义核风险敏感损失函数曲线平滑度的影响. 从图中可以看出: 当误差较小且 α 值越大时表面越光滑, 表明广义核风险敏感损失函数的精度高于非广义形式.

2 算法

2.1 稀疏惩罚约束

这一小节主要介绍近似 l_0 范数的稀疏惩罚约束项. 实际上, 在寻找最优稀疏项时需要最小化 l_0 范数, 而这是一个 NP 难问题. 通常采用

近似 l_0 范数的方法来解决, 一种是将其转化为无约束的 l_1 范数正则化问题, 可获得近似 l_0 范数的解, 但代价是增加采样过程中的测量次数^[22]; 另一种则是采用 CIM 来近似 l_0 范数, 减少了测量中的计算消耗^[12], 其表达式如下:

$$C_{\text{CIM}}^2(\mathbf{\Omega}(i), 0) = h(i) = \frac{1}{L\sqrt{2\pi}\sigma} \sum_{k=1}^L \left(1 - \exp\left(-\frac{\Omega_k^2}{2\sigma^2}\right)\right) \quad (12)$$

其中, σ 是核宽. 此外, 存在其他近似 l_0 范数的稀疏惩罚约束项为 $\|\mathbf{\Omega}\|_0 \approx \sum_{i=0}^{L-1} (1 - e^{-\beta|\Omega_i|})$ ^[23-24], 该方法与 CIM 之间最显著的区别是指数部分是否为二阶统计量. CIM 是具有二阶统计特性, 其指数部分权重向量能够保证整个函数具有凸性. 基于公平性原则, 本文选择了与比较算法相同稀疏惩罚约束项的 CIM. 另外, $h(i)$ 梯度的向量形式如下:

$$\mathbf{h}'(i) = \frac{1}{L\sqrt{2\pi}\sigma^3} \mathbf{\Omega}_i \cdot \exp\left(-\frac{\mathbf{\Omega}_i \cdot \mathbf{\Omega}_i}{2\sigma^2}\right) \quad (13)$$

2.2 GRKRSLS-SPC 算法

定义如下带有稀疏惩罚约束项的广义核风险敏感损失函数为成本函数:

$$\hat{J}(\mathbf{\Omega}) = \frac{1}{\lambda} \sum_{i=1}^L \mu^{L-i} \exp(\lambda(1 - \exp(-\beta |e(i)|^\alpha))) + \rho h(i) \quad (14)$$

其中, μ 表示遗忘因子, $\rho h(i)$ 代表稀疏惩罚约束项, $\rho > 0$ 是控制权重向量的稀疏惩罚约束程度的正则化参数. 采用梯度下降法最小化该成本函数, 可得:

$$\begin{aligned} \frac{\partial \hat{J}(\mathbf{\Omega})}{\partial \mathbf{\Omega}(i)} &= \sum_{i=1}^L \mu^{L-i} \exp(\lambda(1 - \exp(-\beta |e(i)|^\alpha))) \exp(-\beta |e(i)|^\alpha) |e(i)|^{\alpha-2} e(i) \mathbf{X}(i) - \rho \mathbf{h}'(i) = \\ & \sum_{i=1}^L \mu^{L-i} M(i) (d(i) - \mathbf{\Omega}^T(i) \mathbf{X}(i)) \mathbf{X}(i) - \rho \mathbf{h}'(i) \end{aligned} \quad (15)$$

其中,

$$M(i) = \exp(\lambda(1 - \exp(-\beta |e(i)|^\alpha))) \exp(-\beta |e(i)|^\alpha) |e(i)|^{\alpha-2}$$

令公式(15)的梯度等于 0 可得权重 $\mathbf{\Omega}$ 的解为

$$\sum_{i=1}^L \mu^{L-i} M(i) d(i) \mathbf{X}(i) - \rho \mathbf{h}'(i) = \sum_{i=1}^L \mu^{L-i} M(i) \mathbf{X}(i) \mathbf{X}^T(i) \mathbf{\Omega}(i) \quad (16)$$

根据公式(16), 定义 $\mathbf{Y}(i)$ 和 $\mathbf{\Theta}(i)$ 为

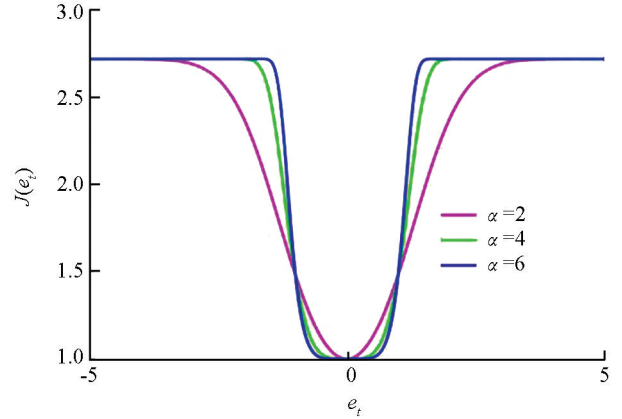


图 1 不同 α 对 GRKRSLS 误差的影响

$$\begin{cases} \Upsilon(i) = \sum_{i=1}^L \mu^{L-i} M(i) d(i) \mathbf{X}(i) \\ \Theta(i) = \sum_{i=1}^L \mu^{L-i} M(i) \mathbf{X}(i) \mathbf{X}^T(i) \end{cases} \quad (17)$$

根据公式(17)将公式(16)改写为矩阵形式,其权向量可以表示为

$$\boldsymbol{\Omega}(i) = \Theta^{-1}(i) (\Upsilon(i) - \rho \mathbf{h}'(i)) \quad (18)$$

$\Theta(i)$ 通过递归形式进行更新可得:

$$\begin{aligned} \Theta(i) &= \mu \sum_{i=1}^{L-1} \mu^{L-1-i} M(i) \mathbf{X}(i) \mathbf{X}^T(i) + M(i) \mathbf{X}(i) \mathbf{X}^T(i) = \\ &= \mu \Theta(i-1) + M(i) \mathbf{X}(i) \mathbf{X}^T(i) \end{aligned} \quad (19)$$

为了避免计算矩阵的逆运算,根据矩阵求逆引理^[25]: $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{C}^{-1} + \mathbf{D} \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{D} \mathbf{A}^{-1}$.

在公式(19)中令 $\mathbf{A} = \mu \Theta(i-1)$, $\mathbf{B} = \mathbf{X}(i)$, $\mathbf{C} = M(i)$, $\mathbf{D} = \mathbf{X}^T(i)$, 可得:

$$\begin{aligned} \Theta^{-1}(i) &= (\mu \Theta(i-1) + M(i) \mathbf{X}(i) \mathbf{X}^T(i))^{-1} = \\ &= \mu^{-1} (\Theta^{-1}(i-1) - M(i) \mathbf{G}(i) \mathbf{X}^T(i) \Theta^{-1}(i-1)) \end{aligned} \quad (20)$$

其中, $\mathbf{G}(i) = \frac{\Theta^{-1}(i-1) \mathbf{X}(i)}{\mu + M(i) \mathbf{X}^T(i) \Theta^{-1}(i-1) \mathbf{X}(i)}$ 为卡尔曼增益. 令 $\mathbf{P}(i) = \Theta^{-1}(i)$, 公式(20)则可重新表示为

$$\mathbf{P}(i) = \mu^{-1} (\mathbf{P}(i-1) - M(i) \mathbf{G}(i) \mathbf{X}^T(i) \mathbf{P}(i-1)) \quad (21)$$

所以

$$\mathbf{G}(i) = \frac{\mathbf{P}(i-1) \mathbf{X}(i)}{\mu + M(i) \mathbf{X}^T(i) \mathbf{P}(i-1) \mathbf{X}(i)}$$

采用同样的方法可以得到下列表达式:

$$\Upsilon(i) = \mu \Upsilon(i-1) + M(i) \mathbf{X}(i) d(i) \quad (22)$$

将公式(22)两端同时减去 $\rho \mathbf{h}'(i)$, 可得:

$$\begin{aligned} \Upsilon(i) - \rho \mathbf{h}'(i) &= \mu \Upsilon(i-1) + M(i) \mathbf{X}(i) d(i) - \rho \mathbf{h}'(i) = \\ &= \mu \Upsilon(i-1) - \mu \rho \mathbf{h}'(i-1) + M(i) \mathbf{X}(i) d(i) - \rho \mathbf{h}'(i) + \mu \rho \mathbf{h}'(i-1) = \\ &= \mu (\Upsilon(i-1) - \rho \mathbf{h}'(i-1)) + M(i) \mathbf{X}(i) d(i) - \rho (1 - \mu) \mathbf{h}'(i-1) \end{aligned} \quad (23)$$

当迭代至算法性能稳定时, 权向量几乎无变化. 即当 $i \rightarrow \infty$ 时, 有 $\mathbf{h}'(i-1) \approx \mathbf{h}'(i)$. 所以公式(23)成立.

将公式(21)和(23)代入公式(18)可得权重向量更新式为

$$\boldsymbol{\Omega}(i) = \boldsymbol{\Omega}(i-1) + \mathbf{G}(i) M(i) e(i) - \rho \mu^{-1} (1 - \mu) [I - M(i) \mathbf{G}(i) \mathbf{X}^T(i)] \mathbf{P}(i-1) \mathbf{h}'(i) \quad (24)$$

最后, 根据上述的推导过程, 总结 GRKRSL-SPC 算法如表1所示.

表1 GRKRSL-SPC 算法更新过程

初始化参数: $\boldsymbol{\Omega}(0) = 0$, $\mathbf{P}(0) = 0$, μ , α , λ , β , σ , ρ

当 $\{\mathbf{X}(i), d(i)\}$, $i = 1, 2, \dots, L$ 可用时, 按下列步骤计算:

步骤1: $e(i) = d(i) - \boldsymbol{\Omega}^T(i-1) \mathbf{X}(i)$

步骤2: $M(i) = \exp(\lambda(1 - \exp(-\beta |e(i)|^\alpha))) \exp(-\beta |e(i)|^\alpha) |e(i)|^{\alpha-2}$

步骤3: $\mathbf{G}(i) = \frac{\mathbf{P}(i-1) \mathbf{X}(i)}{\mu + M(i) \mathbf{X}^T(i) \mathbf{P}(i-1) \mathbf{X}(i)}$

步骤4: $\boldsymbol{\Omega}(i) = \boldsymbol{\Omega}(i-1) + \mathbf{G}(i) M(i) e(i) - \rho \mu^{-1} (1 - \mu) [I - M(i) \mathbf{G}(i) \mathbf{X}^T(i)] \mathbf{P}(i-1) \mathbf{h}'(i)$

步骤5: $\mathbf{P}(i) = \mu^{-1} (\mathbf{P}(i-1) - M(i) \mathbf{G}(i) \mathbf{X}^T(i) \mathbf{P}(i-1))$

2.3 计算复杂度分析

本节分析 GRKRSL-SPC 算法的计算复杂度, 这里考虑每次迭代过程中的加法、除法以及乘法次数. 以 $\alpha=4$ 为例, 各种算法的计算复杂度比较如表 2 所示, 其中, D 表示输入数据的长度, 比较算法为基于稀疏惩罚约束的递归广义最大相关熵 (Recursive Generalized Maximum Correntropy Criterion with SPC, RGMCC-SPC) 算法^[16] 和基于稀疏惩罚约束的递归广义最大相关熵变中心 (Recursive Generalized Maximum Correntropy Criterion with Variable Center under Sparsity Constrained, RGMCCVC-SPC) 算法^[17]. 从表 2 中可知, 3 种算法具有相同的除法次数, 而在乘法和加法运算上, GRKRSL-SPC 算法的计算量小于 RGMCCVC-SPC 算法, 但高于 RGMCC-SPC 算法.

表 2 RGMCC-SPC 和 RGMCCVC-SPC 及 GRKRSL-SPC 算法每次迭代的计算复杂度比较

| 算法 | 加法次数 | 除法次数 | 乘法次数 |
|-------------|-----------------|------|------------------|
| RGMCC-SPC | $5D^2 - 2D + 4$ | 5 | $9D^2 + 5D + 9$ |
| RGMCCVC-SPC | $6D^2 - 2D + 7$ | 5 | $9D^2 + 7D + 8$ |
| GRKRSL-SPC | $5D^2 - 2D + 6$ | 5 | $9D^2 + 5D + 13$ |

3 仿真结果和分析

本节将采用蒙特卡洛仿真验证提出的 GRKRSL-SPC 算法在稀疏系统辨识下的有效性. 未知的稀疏系统中的结构如图 2 所示, 其中最优权重表示为 $\Omega_o = [\Omega_o^0, \Omega_o^1, \dots, \Omega_o^{L-1}]^T \in R^{L \times 1}$ 的列向量, 输入信号为 $\mathbf{X}(i) = [x(i), x(i-1), \dots, x(i-L+1)]^T \in R^{L \times 1}$ 的列向量, $x(i)$ 表示为第 i 时刻输入的的第一个值, $x(i-L+1)$ 则为第 i 时刻的第 L 个值, Z^{-1} 表示系统的单位延迟. 首先假定自适应滤波器的权重向量是由 16 个抽头随机产生, 将第 5 个值设置为 1, 其他值均设置为 0, 可得稀疏度为 1/16, 表述为 $\Omega_o = [0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]^T$ ^[10]. 其次, 设置滤波器的长度为 30, 系统稀疏度为 5/30^[16]. 假定自适应滤波器的初始权重向量是零向量, 输入信号为单位方差的零均值高斯序列. 噪声模型为: $v(i) = (1-b(i))v_1(i) + b(i)v_2(i)$, 其中 $v_1(i)$ 是常见的噪声, $v_2(i)$ 为较大方差的脉冲噪声, 表示大的离群值; $b(i)$ 是由伯努利随机过程引起的发生概率, 其中概率分布为 $Pr\{b(i)=1\} = c$, $Pr\{b(i)=0\} = 1-c$, $0 \leq c \leq 1$, 仿真中选择 $c=0.01$. $v_2(i)$ 是方差为 9 均值为 0 的高斯序列. $v_1(i)$ 考虑两种噪声分布: ① 在 $\{1, -1\}$ 范围内的二进制分布, 其概率分布为 $Pr\{x=1\} = Pr\{x=-1\} = 0.5$. ② $2\sin(\omega)$ 的正弦波, ω 满足 $[0, 2\pi]$ 的均匀分布. 所有仿真结果是在上述混合噪声环境下执行 200 次蒙特卡洛获得的. 计算机软硬件配置分别为 Windows10 和 Intel(R) Core(TM) i7-8700 CPU 3.20GHz, RAM 8.0G.

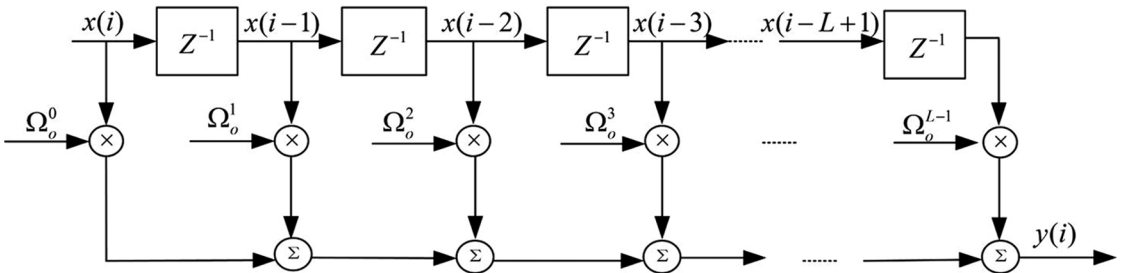


图 2 稀疏系统的结构图

在进行实验仿真过程中, 为了确保算法比较的公平性, 选择在非高斯噪声环境下具有鲁棒性且采用递归更新方式的 RGMCC-SPC^[16] 和 RGMCCVC-SPC 算法^[17] 作为比较算法, 当 RGMCCVC-SPC 算法的变中心变为 0 时退化为 RGMCC-SPC 算法, 设置比较算法的参数使得所有算法具有一致的收敛速率. 为了进一步定量评价滤波精度, 定义稳态均方偏差 (Mean Square Deviation, MSD) 如下, 用 M_{MSD} 表示:

$$M_{\text{MSD}} = E[\|\tilde{\Omega}(i)\|^2] \quad (25)$$

其中, $\tilde{\Omega}(i) = \Omega_o - \Omega(i)$, Ω_o 为稀疏系统辨识中的最优权重, $\Omega(i)$ 则是其第 i 时刻的权重估计值.

图 3 显示了 GRKRSL-SPC 算法在二进制噪声下的学习曲线和所有算法的参数设置, 其中, 图 3(a) 和图 3(b) 的稀疏度分别为 1/16, 5/30. 从图 3 中可知 3 个算法在保持几乎一致的收敛速率下, GRKRSL-SPC 算法的滤波精度显著高于 RGMCC-SPC 和 RGMCCVC-SPC 算法, 尤其当 $\alpha=4$ 和 $\alpha=6$ 时, 表现更明显. 同样地, 在稀疏度为 1/16 和 5/30 的正弦噪声环境下, GRKRSL-SPC 算法提高了稀疏系统辨识中的精度, 并且 $\alpha=4$ 和 $\alpha=6$ 的滤波性能优于 $\alpha=2$ 的性能, 仿真结果如图 4 所示.

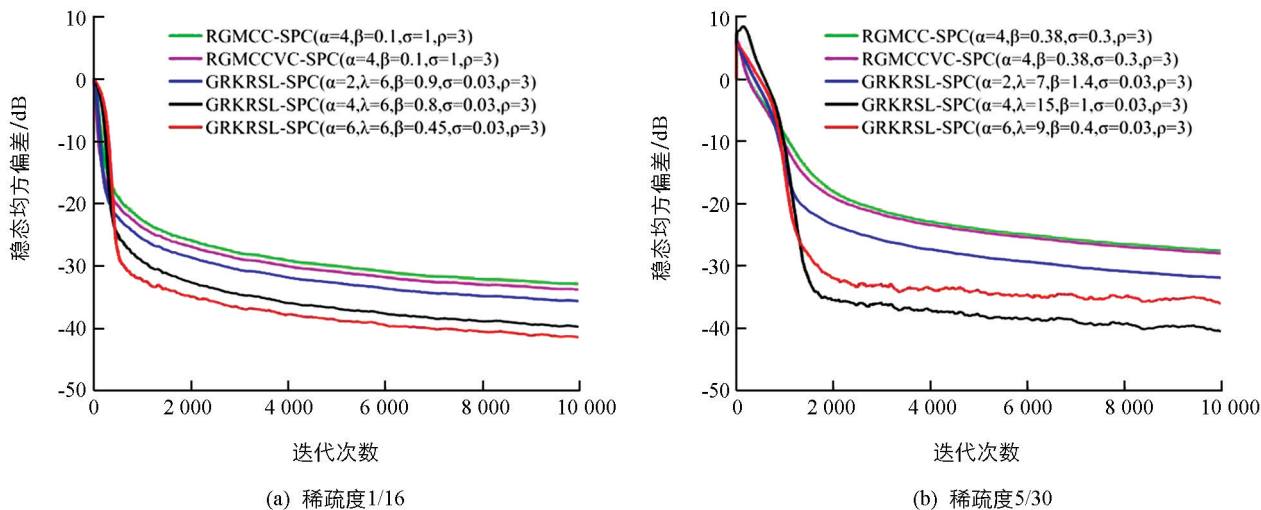


图 3 在二进制噪声下不同算法的稳态均方偏差学习曲线

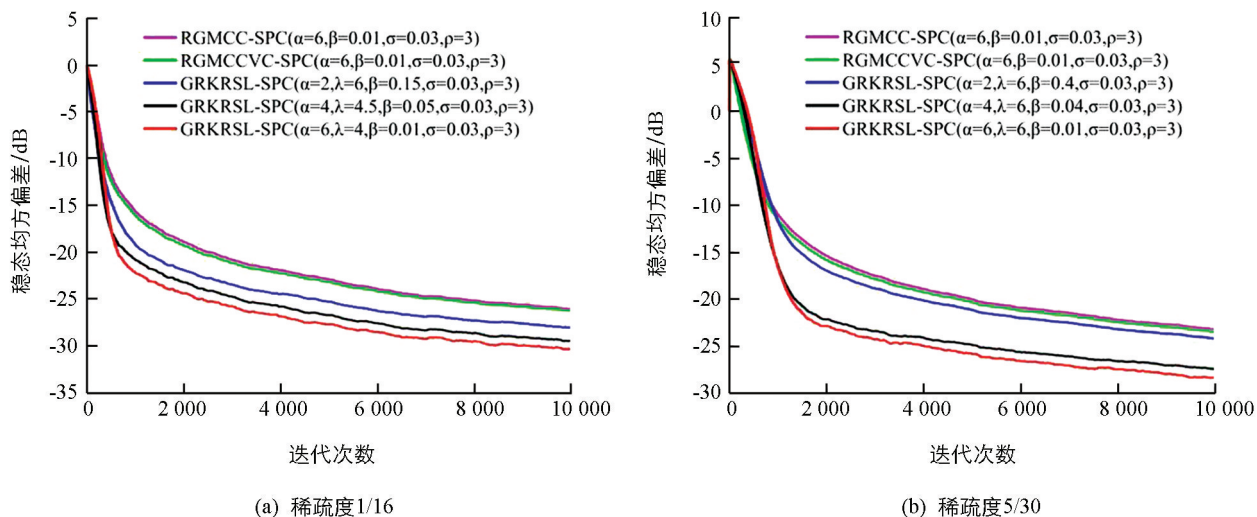


图 4 在正弦噪声下不同算法的稳态均方偏差学习曲线

表 3 显示了在两种稀疏度以及两种噪声环境下算法的消耗时间.

表 3 3 种算法在稀疏度为 1/16 和 5/30 和两种噪声下稀疏系统识别的仿真结果

| 噪声 | 算法 | 时间/ms | |
|-----|---------------------------|----------|----------|
| | | 稀疏度 1/16 | 稀疏度 5/30 |
| 二进制 | RGMCC-SPC($\alpha=4$) | 0.889 | 1.200 |
| | RGMCCVC-SPC($\alpha=4$) | 1.200 | 1.600 |
| | GRKRSL-SPC($\alpha=4$) | 0.917 | 1.500 |
| 正弦 | RGMCC-SPC($\alpha=6$) | 0.896 | 1.400 |
| | RGMCCVC-SPC($\alpha=6$) | 1.200 | 1.800 |
| | GRKRSL-SPC($\alpha=6$) | 0.910 | 1.400 |

从表 3 中可以得出:不论是在二进制噪声还是正弦噪声环境中,GRKRSL-SPC 算法比 RGMCCVC-SPC 算法消耗更少的时间,比 RGMCC-SPC 算法消耗更多的时间,此结论与表 2 中计算复杂度结果相一致.总而言之,从稳态均方偏差和计算复杂度两方面而言,GRKRSL-SPC 算法性能优于 RGMCCVC-SPC 算法和 RGMCC-SPC 算法.

4 结论

本文利用广义高斯密度(GGD)函数作为核函数,提出了一种定义在核空间的非线性相似度量方法,即广义核风险敏感损失函数(GRKRSL).进一步结合递归更新方式提出了应用在稀疏系统模型中的基于稀疏惩罚约束的广义核递归风险敏感(GRKRSL-SPC)算法.从计算复杂度和滤波精度两个方面去验证了 GRKRSL-SPC 算法在非高斯噪声环境中的有效性和滤波精度. GRKRSL-SPC 算法在保持与 RGMCC-SPC 和 RGMCCVC-SPC 算法相同计算复杂度的前提下,提高了稀疏系统的滤波性能,尤其是当 $\alpha=4$ 和 $\alpha=6$ 时滤波精度明显提高.蒙特卡洛仿真结果验证了 GRKRSL-SPC 算法对稀疏系统识别精度优于其他的鲁棒稀疏自适应滤波算法.

参考文献:

- [1] SCHREIBER W F. Advanced Television Systems for Terrestrial Broadcasting: Some Problems and Some Proposed Solutions [J]. Proceedings of the IEEE, 1995, 83(6): 958-981.
- [2] DENG H Y, DOROSLOVACKI M. Proportionate Adaptive Algorithms for Network Echo Cancellation [J]. IEEE Transactions on Signal Processing, 2006, 54(5): 1794-1803.
- [3] GUI G, PENG W, ADACHI F. Improved Adaptive Sparse Channel Estimation Based on the Least Mean Square Algorithm [C]//2013 IEEE Wireless Communications and Networking Conference (WCNC). April 7-10, 2013, Shanghai, China. IEEE, 2013: 3105-3109.
- [4] KALOUPSIDIS N, MILEOUNIS G, BABADI B, et al. Adaptive Algorithms for Sparse System Identification [J]. Signal Processing, 2011, 91(8): 1910-1919.
- [5] 周千, 马文涛, 桂冠. 基于 l_1 范数约束的递归互相关熵的稀疏系统辨识 [J]. 信号处理, 2016, 32(9): 1079-1086.
- [6] 李少东, 杨军, 胡国旗. 一种改进的压缩感知信号重构算法 [J]. 信号处理, 2012, 28(5): 744-749.
- [7] 杨秀杰. 基于深度学习稀疏测量的压缩感知图像重构 [J]. 西南师范大学学报(自然科学版), 2020, 45(1): 42-47.
- [8] 金坚, 谷源涛, 梅顺良. 用于稀疏系统辨识的零吸引最小均方算法 [J]. 清华大学学报(自然科学版), 2010, 50(10): 1656-1659.
- [9] CHEN Y L, GU Y T, HERO A O. Sparse LMS for System Identification [C]//2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). April 19-24, 2009, Taipei, Taiwan, China. IEEE 2009: 3125-3128.
- [10] HONG X, GAO J B, CHEN S. Zero-Attracting Recursive Least Squares Algorithms [J]. IEEE Transactions on Vehicular Technology, 2017, 66(1): 213-221.
- [11] 周其玉, 张爱华, 曹文周, 等. 平方根变步长 l_p 范数 LMS 算法的稀疏系统辨识 [J]. 电讯技术, 2020, 60(2): 137-141.
- [12] SETH S, PRINCIPE J C. Compressed Signal Reconstruction Using the Correntropy Induced Metric [C]//2008 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). March 31-April 4, 2008, Las Vegas, NV, USA. IEEE, 2008: 3845-3848.
- [13] MA W T, QU H, GUI G, et al. Maximum Correntropy Criterion Based Sparse Adaptive Filtering Algorithms for Robust Channel Estimation under Non-Gaussian Environments [J]. Journal of the Franklin Institute, 2015, 352(7): 2708-2727.

- [14] PRINCIPE J C. Information Theoretic Learning: Renyi's Entropy and Kernel Perspective [M]. New York: Springer, 2010: 366-376.
- [15] CHEN B D, XING L, ZHAO H Q, et al. Generalized Correntropy for Robust Adaptive Filtering [J]. IEEE Transactions on Signal Processing, 2016, 64(13): 3376-3387.
- [16] MA W T, DUAN J D, CHEN B D, et al. Recursive Generalized Maximum Correntropy Criterion Algorithm with Sparse Penalty Constraints for System Identification [J]. Asian Journal of Control, 2017, 19(3): 1164-1172.
- [17] SUN Q, ZHANG H, WANG X F, et al. Sparsity Constrained Recursive Generalized Maximum Correntropy Criterion with Variable Center Algorithm [J]. IEEE Transactions on Circuits and System II: Express Briefs, 2020, 67(12): 3517-3521.
- [18] CHEN B D, XING L, XU B, et al. Kernel Risk-Sensitive Loss: Definition, Properties and Application to Robust Adaptive Filtering [J]. IEEE Transactions on Signal Processing, 2017, 65(11): 2888-2901.
- [19] SONG K S. Asymptotic Relative Efficiency and Exact Variance Stabilizing Transformation for the Generalized Gaussian Distribution [J]. IEEE Transactions on Information Theory, 2013, 59(7): 4389-4396.
- [20] PEI S C, TSENG C C. Least Mean P-Power Error Criterion for Adaptive FIR Filter [J]. IEEE Journal on Selected Areas in Communications, 1994, 12(9): 1540-1547.
- [21] LO J T, WANNER T. Existence and Uniqueness of Risk-sensitive Estimates [J], IEEE Transactions on Automatic Control, 2002, 47(11): 1945-1948.
- [22] 王文东, 王尧, 王建军. 一类光滑加权 l_1 算法的收敛性分析与数值仿真实验 [J]. 西南大学学报(自然科学版), 2014, 36(5): 72-77.
- [23] WU F Y, YANG K D, HU Y. Sparse Estimator with l_0 -Norm Constraint Kernel Maximum Correntropy Criterion [J], IEEE Transactions on Circuits and System-II: Express Briefs, 2020, 67(2): 400-404.
- [24] DAS R L. l_0/l_1 Regularized Conjugate Gradient Based Sparse Adaptive Algorithms [C]//2020 International Conference on Signal Processing and Communications (SPCOM). July 19-24, 2020. Bangalore, India. IEEE, 2020: 1-5.
- [25] LIU W F, PRINCIPE J C, HAYKIN S. Kernel Adaptive Filtering [M]. New Jersey: John Wiley & Sons Inc, 2010: 95-96.

责任编辑 汤振金