

基于罚函数内点法的泄露积分型回声状态网的参数优化

伦淑娴¹ 胡海峰²

摘要 为了提升泄露积分型回声状态网 (Leaky integrator echo state network, Leaky-ESN) 的性能, 提出利用罚函数内点法优化 Leaky-ESN 的全局参数, 如泄漏率、内部连接权矩阵谱半径、输入比例因子等, 这克服了通过反复试验法选取参数值而降低了 Leaky-ESN 模型的优越性和性能. Leaky-ESN 的全局参数必须保障回声状态网满足回声状态特性, 因此它们之间存在不等式约束条件. 有学者提出利用随机梯度下降法来优化内部连接权矩阵谱半径、输入比例因子、泄露率三个全局参数, 一定程度上提高了 Leaky-ESN 的逼近精度. 然而, 随机梯度下降法是解决无约束优化问题的基本算法, 在利用随机梯度下降法优化参数时, 没有考虑参数必须满足回声特性的约束条件 (不等式约束条件), 致使得到的参数值不是最优解. 由于罚函数内点法可以求解具有不等式约束的最优化问题, 应用范围广, 收敛速度较快, 具有很强的全局寻优能力. 因此, 本文提出利用罚函数内点法优化 Leaky-ESN 的全局参数, 并以时间序列预测为例, 检验优化后的 Leaky-ESN 的预测性能, 仿真结果表明了本文提出方法的有效性.

关键词 回声状态网, 时间序列预测, 有约束优化, 罚函数内点法, 牛顿法

引用格式 伦淑娴, 胡海峰. 基于罚函数内点法的泄露积分型回声状态网的参数优化. 自动化学报, 2017, 43(7): 1160–1168

DOI 10.16383/j.aas.2017.c160541

Parameter Optimization of Leaky Integrator Echo State Network with Internal-point Penalty Function Method

LUN Shu-Xian¹ HU Hai-Feng²

Abstract To improve leaky integrator echo state network (Leaky-ESN) performance, internal-point penalty function (IPF) method is used to optimize the global parameters of Leaky-ESN, such as leakage rate, spectral radius of internal connection weight matrix, scaling of input, etc., which overcomes loss of superiority and performance of Leaky-ESN because of using trial and error method to select parameter values. The global parameters of Leaky-ESN have to guarantee the echo state network to meet the echo state property, thus inequality constraints exist between them. Some researchers put forward the method using the stochastic gradient descent (GD) to optimize leakage rate, spectral radius of internal connection weight matrix, and scaling of input, which can improve the approximation precision of the Leaky-ESN to some certain extent. However, the stochastic gradient descent method is a basic algorithm to solve unconstrained optimization problems. Without considering parameters which need satisfy the constraint conditions of the echo state property (inequality constraints) during using stochastic gradient descent method, the parameter value is not the optimal solution. Internal-point penalty function method can solve the optimized problem with inequality constraints, a wide scope of application, fast convergence speed, strong ability of global optimization. Therefore, in this paper, internal-point penalty function method is used to optimize the global parameters of Leaky-ESN, and time series prediction is selected as an example to examine the performance of the optimized Leaky-ESN. Simulation results show the effectiveness of the proposed approach.

Key words Echo state network (ESN), time series prediction, constrained optimization, internal-point penalty function (IPF) method, Newton method

Citation Lun Shu-Xian, Hu Hai-Feng. Parameter optimization of leaky integrator echo state network with internal-point penalty function method. *Acta Automatica Sinica*, 2017, 43(7): 1160–1168

收稿日期 2016-07-22 录用日期 2016-10-09

Manuscript received July 22, 2016; accepted October 9, 2016
国家自然科学基金 (61573072, 21506014), 辽宁省自然科学基金 (2014020143), 2011 年辽宁省第一批科学技术计划项目 (2011402001), 辽宁省教育厅科学技术研究项目 (L2015008) 资助

Supported by National Nature Science Foundation of China (61573072, 21506014), Nature Science Foundation of Liaoning Province (2014020143), the First Batch of Science and Technology Projects of Liaoning Province (2011402001), Science and Technology Research Projects of the Department of Education of Liaoning Province (L2015008)

本文责任编辑 魏庆来

回声状态网 (Echo state network, ESN) 是由 Jaeger 教授于 2001 年提出的一种新型递归神经网络模型^[1–2], 不仅能够解决递归神经网络存在的问题, 而且利用其预测著名的 Mackey-Glass 混沌时

Recommended by Associate Editor WEI Qing-Lai

1. 渤海大学新能源学院 锦州 121013 2. 渤海大学工学院 锦州 121013

1. College of New Energy, Bohai University, Jinzhou 121013
2. College of Engineering, Bohai University, Jinzhou 121013

间序列, 与以往报道的预测结果相比, 预测精度提高 2400 倍, 成为 ESN 研究的标志性成果^[3]. 回声状态网在系统辨识^[4]、动态模式识别^[5]、滤波或者控制^[6-8]、预测^[9-10]等领域获得成功应用. 与传统的递归神经网络相比, ESN 具有如下显著特点: 1) ESN 由随机稀疏连接的神经元组成储备池作为隐层. 储备池的生成过程独立于回声状态网络的训练过程, 保证了 ESN 在训练过程中的稳定性. 2) ESN 采用线性回归或最小二乘法训练储备池至输出层的权值, 使网络的训练过程得以简化, 保证了权值的全局最优性. 然而, 除了输出权值外, ESN 的全局参数还包括储备池处理单元个数、内部连接矩阵谱半径、输入比例因子、输出反馈比例因子、泄露率等. 这些全局参数之间存在复杂的非线性关系, 通常在训练之前根据经验选取, 因此使得 ESN 的逼近能力不能达到最优. 文献 [11] 提出了一种 ESN 的改进模型, 称为泄露积分型回声状态网 (Leaky integrator echo state network, Leaky-ESN). Leaky-ESN 利用随机梯度下降法来优化内部连接矩阵谱半径、输入比例因子、泄露率三个参数, 一定程度上提高了 Leaky-ESN 的预测精度. 众所周知, 回声状态网需要满足回声特性. 然而, 随机梯度下降法是解决无约束优化问题的基本算法, 在利用随机梯度下降法优化参数时, 没有考虑参数必须满足回声特性的约束条件 (不等式约束条件). 因此, 从某种意义上说, 得到的参数值不是最优解. 罚函数内点法可以求解具有不等式约束的最优化问题, 应用范围广, 收敛速度较快, 具有很强的全局寻优能力. 因此, 本文以 Leaky-ESN 为例, 提出利用罚函数内点法优化回声状态网的全局参数, 进一步提升回声状态网的预测性能. 利用罚函数内点法实质上是将不等式约束条件下的优化问题转换为无约束的优化问题, 进而通过牛顿法来求解无约束优化问题, 获得回声状态网的全局最优参数.

1 泄露积分型回声状态网

Leaky-ESN 网络是 ESN 网络的一种改进模型, 能够学习慢变动态系统等. Leaky-ESN 与 ESN 的拓扑结构相同, 其储备池是由泄露积分型神经元组成, 如图 1 所示^[11]. 在图 1 中, 输入层有 K 个输入节点; 储备池由 N 个内部节点以及稀疏的节点连接权值构成; 输出层有 L 个输出节点. 图中实线表示了网络的必要连接, 而虚线表示了不同情况下可能存在的连接. 假定网络的输入 $u(n) = [u_1(n), u_2(n), \dots, u_K(n)]^T$, 储备池的神经元单元状态为 $x(n) = [x_1(n), x_2(n), \dots, x_N(n)]^T$, 网络的输出 $y(n) = [y_1(n), y_2(n), \dots, y_L(n)]^T$. 那么, Leaky-

ESN 网络的状态更新方程和输出方程为

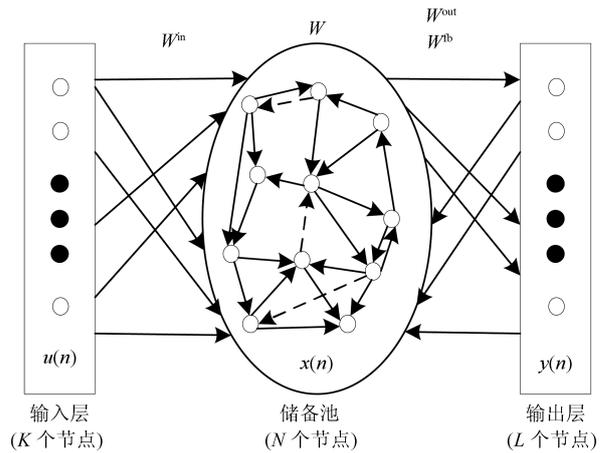


图 1 回声状态网络的拓扑结构

Fig. 1 Structure of echo state network

$$x(n+1) = (1-a)x(n) + f(s^{\text{in}}W^{\text{in}}u(n+1) + (\rho W)x(n) + W^{\text{fb}}y(n)) \quad (1)$$

$$y(n) = g(W^{\text{out}}[x(n); u(n)]) \quad (2)$$

其中, n 是离散时刻, a 是泄露率, $a \in [0, 1]$; ρ 是储备池连接权值 W 的谱半径; s^{in} 是归一化之后输入单元的比例因子, 称为输入比例因子; f 是储备池单元的激励函数 (通常是 \tanh 函数或者是 S 型函数); g 表示输出激励函数 (典型的是单位函数和 \tanh 函数), 根据问题的不同 g 可以取线性函数或者 S 型函数; $[\cdot; \cdot]$ 表示二个向量级联; W^{in} 、 W 、 W^{out} 和 W^{fb} 分别表示输入、储备池、输出、输出反馈的连接权矩阵. 在整个学习和训练过程中, W 、 W^{in} 和 W^{fb} 在随机生成之后就固定不变, W^{out} 是通过网络训练计算得到. Leaky-ESN 模型的训练过程就是计算输出权值 W^{out} 的过程, 就是在均方根误差的意义上的最小化. 通常采用 Tikhonov 正则化 (岭回归) 的方法获得 W^{out} ^[12], 即:

$$W^{\text{out}} = YX^T (XX^T + \theta I)^{-1} \quad (3)$$

式中, θ 是一正则化系数. 假设 $u_{\text{teach}}(n)$ 是训练样本的输入信号, $y_{\text{teach}}(n)$ 是训练样本的期望输出信号. 把所有的 $u_{\text{teach}}(n)$ 和 $x(n)$ 储存起来并用矩阵 $X \in \mathbf{R}^{(K+N) \times T}$ 的行来表示, 对应的目标输出值 $y(n)$ 表示成 $Y \in \mathbf{R}^{L \times T}$ 的行向量. 为了评价 Leaky-ESN 模型训练效果, 通常采用标准均方根误

差 ($NRMSE$), 其表达式为:

$$NRMSE(y, y_{teach}) = \sqrt{\frac{\langle \|y(n) - y_{teach}(n)\|^2 \rangle}{\langle \|y_{teach}(n) - \langle y_{teach}(n) \rangle\|^2 \rangle}} \quad (4)$$

式中, $\|\cdot\|$ 表示的是欧氏距离; $\langle \cdot \rangle$ 表示的是均值函数.

2 罚函数内点法在 Leaky-ESN 全局参数优化中的应用

2.1 全局参数优化问题的数学模型

在 Leaky-ESN 的训练过程中, 本文除了优化 W^{out} , 还优化泄漏率 a , 储备池连接权矩阵的谱半径 ρ , 输入比例因子 s^{in} . 因此, 需要建立这些全局参数优化问题的数学模型. 由于这些参数必须保障 Leaky-ESN 具有回声状态特性, 它们之间存在不等式约束. 因此, 具有不等式约束的优化问题表示为:

$$\begin{aligned} \min \quad & \bar{f}(\zeta) \\ \text{s.t.} \quad & \bar{g}_i(\zeta) \geq 0, \quad i = 1, 2, 3, \dots, m \\ & \zeta \in \bar{X} \end{aligned} \quad (5)$$

式中 \bar{f}, \bar{g}_i 是定义在 \mathbf{R} 上的函数, $\bar{f}(\zeta)$ 是目标函数, $\bar{g}_i(\zeta) \geq 0$ 是不等式约束条件; $\zeta = (\zeta_1, \zeta_2, \dots, \zeta_n)$ 是一个 n 维向量; 集合约束 \bar{X} 是 \mathbf{R} 上的一个子集. 其可行域记为

$$S = \{\zeta | \bar{g}_i(\zeta) \geq 0, i = 1, 2, 3, \dots, m, \zeta \in \bar{X}\}$$

罚函数法就是利用目标函数和约束函数构造具有惩罚性质的函数 $P(\zeta) = \bar{P}(\bar{f}(\zeta), \bar{g}_i(\zeta))$, 使原约束优化问题转化为求 $P(\zeta)$ 最优解的无约束优化问题. 罚函数内点法, 又称内罚函数法, 它是一类保持严格可行性的方法, 在求解无约束优化问题时, 严格要求迭代点在可行域的内部移动. 当迭代点由可行域的内部接近可行域的边界时, 将有无穷大的障碍, 迫使迭代点返回可行域的内部^[12].

2.2 目标函数和约束条件的建立

Leaky-ESN 全局参数优化的目的是希望网络输出值与期望值之间尽可能的接近. 因此, 在本文中建立了如下的目标函数:

$$L(n) = \frac{1}{2} \|y_{teach}(n) - y(n)\|^2 \quad (6)$$

只有在满足回声状态特性的情况下, 回声状态网才是有效的, 能够具有很好的性能, 比如对时间序列的预测性能. 回声状态特性就是随着迭代次数的增加, 回声状态网受到初始条件的影响以一定的速率越来越

越小直至消失, 对于输入驱动来说, 就是不受之前输入和内部状态的影响, 网络达到渐近稳定状态. 递归神经网络的大多数应用是纯输入驱动情况 (例如动态模式识别、时间序列预测、滤波或控制). 对于这些应用, 回声状态网模型通常忽略输出反馈, 即 $W^{fb} = 0$.

下面给出文献 [11] 的回声状态特性充分条件 (定理 1) 和必要条件 (定理 2).

定理 1^[11]. 对于 Leaky-ESN 模型 (1) 和 (2), 如果满足如下条件:

- 1) f 是 \tanh 函数;
- 2) g 是有界函数 (比如 \tanh 函数) 或没有输出反馈 ($W^{fb} = 0$);
- 3) $|1 - (a - \sigma(W))| < 1$, $\sigma(\cdot)$ 表示矩阵的最大奇异值;

那么 Leaky-ESN 模型对所有输入都满足回声状态特性.

定理 2^[11]. 对于 Leaky-ESN 模型 (1) 和 (2), 如果满足如下条件:

- 1) f 是 \tanh 函数;
- 2) 矩阵的谱半径 $|\lambda|_{\max}(\bar{W}) > 1$, 其中 $\bar{W} = (\rho W) + (1 - a)I$, (I 是单位矩阵);

那么 Leaky-ESN 模型不具有回声状态特性.

由文献 [11] 的定理 1 和定理 2 可知, 在 Leaky-ESN 不考虑输出反馈情况下 (即 $W^{fb} = 0$), 如果矩阵 $\bar{W} = (\rho W) + (1 - a)I$ 的谱半径 $|\lambda|_{\max}(\bar{W})$ 大于 1, 那么 Leaky-ESN 模型就不存在回声状态特性, 换句话说就是必须小于等于 1 才行. $|\lambda|_{\max}(\bar{W}) \leq 1$ 被称为有效谱半径, 能够保证对所有的任务都满足回声状态特性. 因此根据 $|\lambda|_{\max}(\bar{W}) \leq 1$ 可得约束条件 $1 - a + \rho \leq 1$. 因此, Leaky-ESN 全局参数优化问题的约束条件为

$$a - \rho \geq 0 \quad (7)$$

$$1 - a > 0 \quad (8)$$

$$\rho > 0 \quad (9)$$

2.3 罚函数内点法的障碍函数建立

罚函数内点法需要通过构建障碍函数来解决约束优化问题. 所谓障碍函数, 就是在可行域边界设置一道屏障, 以保证算法在进行优化时始终在可行域内. 罚函数内点法的障碍函数形式如下:

$$\varphi(\zeta, M^{(k)}) = \bar{f}(\zeta) + M^{(k)} \sum_{i=1}^m G[\bar{g}_i(\zeta)] \quad (10)$$

式中, $k = 1, 2, \dots, l$ 是内点法惩罚因子的递归次数; ζ 是集合约束; m 是集合约束中不等式条件的个数; $M^{(k)}$ 是第 k 步惩罚因子 ($M^{(k)} > 0$). 惩罚函数

$G[\bar{g}_i(\zeta)]$ 通常具有如下两种形式:

$$G[\bar{g}_i(\zeta)] = \frac{1}{\bar{g}_i(\zeta)} \quad (11)$$

$$G[\bar{g}_i(\zeta)] = -\ln(\bar{g}_i(\zeta)) \quad (12)$$

本文采用倒数形式的惩罚项, 即式 (11). 因此, 构造的内点法障碍函数形式如下:

$$\varphi(\zeta, M^{(k)}) = \bar{f}(\zeta) + M^{(k)} \sum_{i=1}^m \frac{1}{\bar{g}_i(\zeta)} \quad (13)$$

式中 $\zeta = [a; \rho]$, $m = 3$. 在优化过程中 $M^{(k)}$ 是一个严格递减序列, $M^{(0)} > M^{(1)} > \dots > M^{(k)}$. 因此在实际操作过程中令

$$M^{(k+1)} = cM^{(k)} \quad (14)$$

式中 c 为缩减系数 ($0 < c < 1$). 初始的惩罚因子 $M^{(0)}$ 对于整个障碍函数的性能起着很大的影响作用. 如果太小, 那么它起的作用就会很小; 如果太大, 那么需要花费大量的时间来得到最优值, 因此采用如下的形式来计算:

$$M^{(0)} = \left| \frac{\bar{f}(\zeta(0))}{\sum_{i=1}^m \frac{1}{\bar{g}_i(\zeta(0))}} \right| \quad (15)$$

式中, $\bar{f}(\zeta(0))$ 是目标函数初始值; $\zeta(0)$ 是参数的初始优化值.

通过式 (13) 可知, 随着迭代次数 k 的增加, 在障碍函数中, 惩罚因子越来越小, 因此障碍项也越来越小, 换句话说就是目标函数越来越接近最优值, 那么此时得到的解越来越接近最优解.

Leaky-ESN 模型的惩罚项可以表示成:

$$G[g_i(a, \rho)] = \{a, \rho | a - \rho \geq 0, 1 - a > 0, \rho > 0\} \quad (16)$$

因此, Leaky-ESN 模型的障碍函数为:

$$\varphi(X, M^{(k)}) = \sum_{n=1}^T L(n) + M^{(k)} \left(\frac{1}{a - \rho} + \frac{1}{1 - a} + \frac{1}{\rho} \right) \quad (17)$$

式中, T 是批量计算中一个样本点所含的离散时刻数.

通过式 (17), 我们把非线性不等式约束优化问题转换成了无约束优化问题.

2.4 罚函数内点法的优化过程及步骤

本文采用牛顿法来求解无约束优化问题的最优参数值. 牛顿法的基本思想是利用二次近似多项式

的极值点求法给出原函数的极值点的求法. 其迭代公式如下^[13]:

$$\zeta(n+1) = \zeta(n) - [\nabla^2 \varphi(\zeta(n))]^{-1} \nabla \varphi(\zeta(n)) \quad (18)$$

根据式 (18) 可知, 需要计算泄露率 a 和谱半径 ρ 关于障碍函数的一阶偏导数和二阶偏导数. 为了简化符号的表达, 我们用 $\varphi_1, \varphi_2, \varphi_{11}, \varphi_{12}, \varphi_{21}$ 和 φ_{22} 来代替 $\frac{\partial \varphi}{\partial a}, \frac{\partial \varphi}{\partial \rho}, \frac{\partial^2 \varphi}{\partial a^2}, \frac{\partial^2 \varphi}{\partial a \partial \rho}, \frac{\partial^2 \varphi}{\partial \rho \partial a}$ 和 $\frac{\partial^2 \varphi}{\partial \rho^2}$ ($\partial \varphi = \partial \varphi(\zeta, M^{(k)})$), 用 $0_u = (0, \dots, 0)^T$ 来表示 $\frac{\partial u(n)}{\partial q}$ ($q \in \{a, \rho\}$). 令 $\ell(n) = y_{\text{teach}}(n) - y(n)$, 所有偏导数的推导如下所示:

$$\varphi_1(n) = -\ell^T(n) W^{\text{out}} \left[\frac{\partial x(n)}{\partial a}; 0_u \right] - M^{(k)} \left(\frac{1}{(\rho - a)^2} - \frac{1}{(a - 1)^2} \right) \quad (19)$$

$$\varphi_2(n) = -\ell^T(n) W^{\text{out}} \left[\frac{\partial x(n)}{\partial \rho}; 0_u \right] + M^{(k)} \left(\frac{1}{(\rho - a)^2} - \frac{1}{\rho^2} \right) \quad (20)$$

$$\varphi_{11}(n) = \left(W^{\text{out}} \left[\frac{\partial x(n)}{\partial a}; 0_u \right] \right)^2 - \ell^T(n) W^{\text{out}} \left[\frac{\partial^2 x(n)}{\partial a^2}; 0_u \right] - M^{(k)} \left(\frac{2}{(\rho - a)^3} + \frac{1}{(a - 1)^3} \right) \quad (21)$$

$$\varphi_{12}(n) = \left(W^{\text{out}} \left[\frac{\partial x(n)}{\partial \rho}; 0_u \right] \cdot W^{\text{out}} \left[\frac{\partial x(n)}{\partial a}; 0_u \right] \right) - \ell^T(n) W^{\text{out}} \left[\frac{\partial^2 x(n)}{\partial a \partial \rho}; 0_u \right] + M^{(k)} \frac{3}{(\rho - a)^3} \quad (22)$$

$$\varphi_{21}(n) = \left(W^{\text{out}} \left[\frac{\partial x(n)}{\partial a}; 0_u \right] \cdot W^{\text{out}} \left[\frac{\partial x(n)}{\partial \rho}; 0_u \right] \right) - \ell^T(n) W^{\text{out}} \left[\frac{\partial^2 x(n)}{\partial \rho \partial a}; 0_u \right] + M^{(k)} \frac{3}{(\rho - a)^3} \quad (23)$$

$$\varphi_{22}(n) = \left(W^{\text{out}} \left[\frac{\partial x(n)}{\partial \rho}; 0_u \right] \right)^2 - \ell^T(n) W^{\text{out}} \left[\frac{\partial^2 x(n)}{\partial \rho^2}; 0_u \right] - M^{(k)} \left(\frac{1}{(\rho - a)^3} - \frac{2}{\rho^3} \right) \quad (24)$$

通过式 (19)~(24) 可知, 我们还必须求得偏导数 $\frac{\partial x(n)}{\partial a}$, $\frac{\partial x(n)}{\partial \rho}$, $\frac{\partial^2 x(n)}{\partial a^2}$, $\frac{\partial^2 x(n)}{\partial a \partial \rho}$, $\frac{\partial^2 x(n)}{\partial \rho \partial a}$ 和 $\frac{\partial^2 x(n)}{\partial \rho^2}$. 其中, 为了表示方便, 令 $\bar{\zeta}(n) = s^{\text{in}} W^{\text{in}} u(n) + (\rho W)x(n)$, 可以得到如下偏导数:

$$\frac{\partial x(n)}{\partial a} = (1-a) \frac{\partial x(n-1)}{\partial a} - x(n-1) + f'(\bar{\zeta}(n)) \cdot \left((\rho W) \frac{\partial x(n-1)}{\partial a} \right) \quad (25)$$

$$\frac{\partial x(n)}{\partial \rho} = (1-a) \frac{\partial x(n-1)}{\partial \rho} + f'(\bar{\zeta}(n)) \cdot \left((\rho W) \frac{\partial x(n-1)}{\partial \rho} + Wx(n-1) \right) \quad (26)$$

$$\begin{aligned} \frac{\partial^2 x(n)}{\partial a^2} &= (1-a) \frac{\partial^2 x(n-1)}{\partial a^2} - 2 \cdot \frac{\partial x(n-1)}{\partial a} + \\ & f''(\bar{\zeta}(n)) \cdot \left((\rho W) \frac{\partial x(n-1)}{\partial a} \right)^2 + \\ & f'(\bar{\zeta}(n)) \cdot \left((\rho W) \frac{\partial^2 x(n-1)}{\partial a^2} \right) \quad (27) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 x(n)}{\partial a \partial \rho} &= (1-a) \frac{\partial^2 x(n-1)}{\partial a \partial \rho} - \\ & \frac{\partial x(n-1)}{\partial \rho} + f''(\bar{\zeta}(n)) \cdot \\ & \left((\rho W) \frac{\partial x(n-1)}{\partial \rho} + Wx(n-1) \right) \cdot \\ & \left((\rho W) \frac{\partial x(n-1)}{\partial a} \right) + f'(\bar{\zeta}(n)) \cdot \\ & \left(W \frac{\partial x(n-1)}{\partial a} + (\rho W) \frac{\partial^2 x(n-1)}{\partial a \partial \rho} \right) \quad (28) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 x(n)}{\partial \rho \partial a} &= (1-a) \frac{\partial^2 x(n-1)}{\partial \rho \partial a} - \\ & \frac{\partial x(n-1)}{\partial \rho} + f''(\bar{\zeta}(n)) \cdot \\ & \left((\rho W) \frac{\partial x(n-1)}{\partial \rho} + Wx(n-1) \right) \cdot \\ & \left((\rho W) \frac{\partial x(n-1)}{\partial a} \right) + f'(\bar{\zeta}(n)) \cdot \\ & \left(W \frac{\partial x(n-1)}{\partial a} + (\rho W) \frac{\partial^2 x(n-1)}{\partial \rho \partial a} \right) \quad (29) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 x(n)}{\partial \rho^2} &= (1-a) \frac{\partial^2 x(n-1)}{\partial \rho^2} + f''(\bar{\zeta}(n)) \cdot \\ & \left((\rho W) \frac{\partial x(n-1)}{\partial \rho} + Wx(n-1) \right)^2 + \\ & f'(\bar{\zeta}(n)) \cdot \left(W \frac{\partial x(n-1)}{\partial \rho} + \right. \\ & \left. (\rho W) \frac{\partial^2 x(n-1)}{\partial \rho^2} + W \frac{\partial x(n-1)}{\partial \rho} \right) \quad (30) \end{aligned}$$

式 (25)~(30) 提供了一个由 $\frac{\partial x(n-1)}{\partial q}$, $\frac{\partial^2 x(n-1)}{\partial q^2}$ 到 $\frac{\partial x(n)}{\partial q}$, $\frac{\partial^2 x(n)}{\partial q^2}$ 的递归计算方法. 这些递归方程的初始值可以是任意值 (通常我们都设为零矩阵, 即 $\frac{\partial x(n-1)}{\partial q} = 0$ 和 $\frac{\partial^2 x(n-1)}{\partial q^2} = 0$). 在递归方程中, f 函数都是双曲正切函数 (\tanh 函数), 因此我们有

$$f'(\bar{\zeta}) = \frac{4}{2 + e^{2\bar{\zeta}} + e^{-2\bar{\zeta}}} \quad (31)$$

$$f''(\bar{\zeta}) = \frac{8(e^{-2\bar{\zeta}} - e^{2\bar{\zeta}})}{(2 + e^{2\bar{\zeta}} + e^{-2\bar{\zeta}})^2} \quad (32)$$

下面给出利用罚函数内点法寻找 Leaky-ESN 全局参数 (主要是泄漏率 a 和谱半径 ρ) 的最优值的具体实现步骤.

步骤 1. 对罚函数内点法初始化. 初始参数 $\zeta(0)$ 的值需要选取在可行域内; 根据式 (12) 可以计算得到 M^0 ; 缩减因子 c 的选取; 收敛精度; 迭代步数 $k = 0$.

步骤 2. 利用牛顿法求解无约束优化问题. 首先要判断 $\|\varphi_1(n) - \varphi_2(n)\| \leq \varepsilon$ 是否成立, 如果满足收敛条件, 那么此时经过优化得到的参数值就是牛顿优化算法下的最优值, 若不然, 我们就利用牛顿法来继续优化参数, 即利用式 (32) 优化参数 ζ 直到满足收敛条件.

$$\begin{aligned} \zeta(n) &= \zeta(n-1) - \begin{bmatrix} \varphi_{11}(n-1) & \varphi_{12}(n-1) \\ \varphi_{21}(n-1) & \varphi_{22}(n-1) \end{bmatrix}^{-1} \times \\ & \begin{bmatrix} \varphi_1(n-1) \\ \varphi_2(n-1) \end{bmatrix}^T \quad (33) \end{aligned}$$

步骤 3. 罚函数内点法寻找全局最优值. 当步骤 2 收敛条件满足时, 此时的最优值 $\zeta(n)$ 可能只是相对的最优值, 并不一定是最终的全局最优值. 因此, 我们还需要判断 $\|\zeta(n) - \zeta(n-1)\| \leq \varepsilon$ 是否成立, 如果收敛条件成立, 那么此时的参数值就是要寻找的全局最优值, 否则就是相对最优值. 因此, 我们将此时的相对最优值作为下次迭代的初始值, 令

$M^{(k+1)} = cM^{(k)}$, $k = k + 1$, 然后重复上述两步, 直到满足所有的收敛条件为止.

步骤 4. 现在通过前面 3 个步骤就可以得到全局最优参数了, 同时根据式 (3) 和式 (4) 得到输出权值 W^{out} 和训练误差.

3 仿真与分析

为了评价罚函数内点法的优化效果, 本文以时间序列预测为例, 与利用文献 [10] 中的随机梯度下降法 (Gradient descent, GD) 优化 Leaky-ESN 模型参数进行比较和分析. 程序运行总的步长是 20000 步, 500 次为一个样本点 ($T = 500$), 总共 40 个样本点. 当 Leaky-ESN 模型学习完成之后, 利用 50 个样本数据来评估优化后的模型的预测精度. 为了保证数据的科学性与合理性, 训练误差 NRMSE 和测试 NRMSE, 预测误差等都是经过 50 次随机测试之后所得出的平均值.

3.1 第一个时间序列

时间序列模型为 $u(n) = \sin(n) + \sin(0.51n) + \sin(0.1002n) + \sin(0.054n)$, 其期望输出 $d(n) = u(n - 5)$. Leaky-ESN 模型的初始值设置为 $[a(0); \rho(0)] = \zeta(0) = [0.43; 0.2]$, $s^{in} = 0.3$, $W^{fb} = 0$. W^{in} 是由具有均匀伪随机分布的 $[-1, 1]$ 元素构成的矩阵. W 是由具有均匀伪随机分布的非零元素产生具有单位谱半径的随机稀疏矩阵, 其稀疏连接度为 0.5. 连接输入权值 W^{in} 和内部连接权值 W 在整个训练过程中不变, 储备池元素个数 $N = 20$. 罚函数内点法的惩罚因子 $M^{(0)} = 6$, 缩减因子 $c = 0.7$; 收敛精度 $\varepsilon = 10^{-6}$. 梯度下降法的初始值选取相同.

图 1 是采用随机梯度下降法和罚函数内点法分别优化 Leaky-ESN 时的训练误差变化曲线. 由图 1 可以看出, 罚函数内点法和随机梯度下降法均能收敛, 利用内点法优化 Leaky-ESN 比随机梯度下降法具有更小的 NRMSE. 另外, 图 1 表明, 在总迭代步数中, 运用罚函数内点法优化后的模型收敛的有效迭代步数 (优化过程中趋于收敛真正需要的迭代步数) 更少, 说明罚函数内点法在优化时具有更快的收敛速度. 图 2 和图 3 是随机测试中某次时间序列预测的结果. 从图 2 和图 3 中可以看出, 罚函数内点法优化的 Leaky-ESN 模型预测结果精度更高, 因此与它们相对应的误差与零值线 (虚线) 的距离更为接近. 表 1 是采用随机梯度下降法和罚函数内点法分别优化 Leaky-ESN 时的测试 NRMSE (50 次测试的统计平均值). 由表 1 可以看出, 利用罚函数内点法和随机梯度下降法优化后的 Leaky-ESN 模型均具有良好的预测性能, 其中罚函数内点法的测试误

差明显要比随机梯度下降法小. 图 4 是第 5000 步数据的 NRMSE 的分布区间. 其中, 柱形图纵坐标

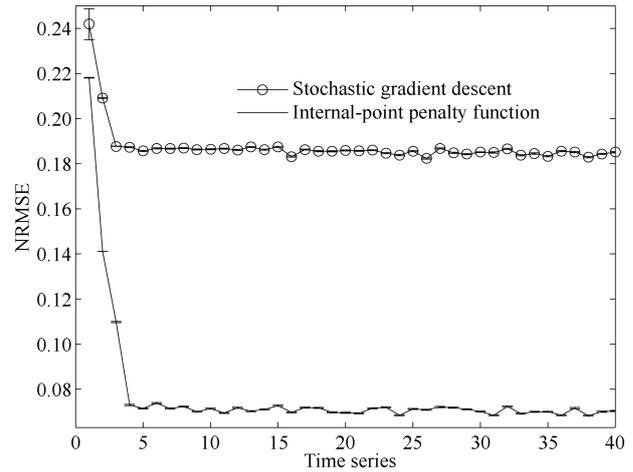


图 2 第一时间序列的训练误差

Fig. 2 Training errors for the first time series

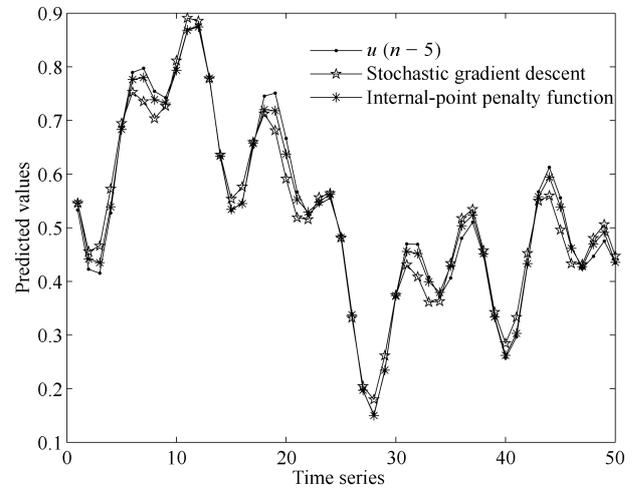


图 3 第一时间序列的预测

Fig. 3 Predicted values for the first time series

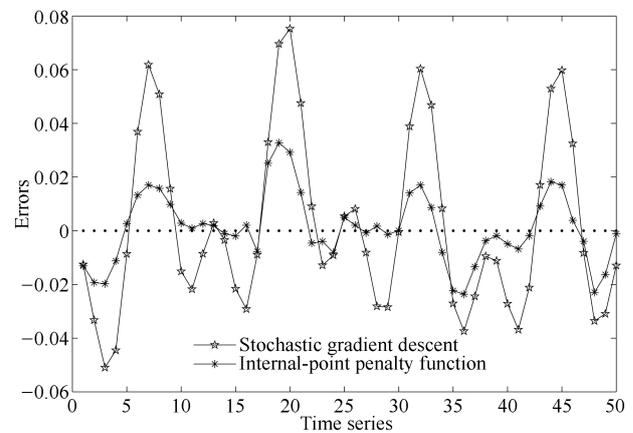


图 4 第一时间序列的预测误差

Fig. 4 Predicted errors for the first time series

表示出现在每个区间总共的次数, 横坐标上数字 1 代表的误差出现的区间是 [0, 0.05] (因为在此统计区间误差出现的次数为零, 所以人为地将第一个统计区间扩大, 避免在图中空置), 数字 2 代表的区间是 [0.051, 0.071], 数字 2 和数字 3 的区间间隔为 0.02, 以此类推. 从图中可知, 罚函数内点法优化模型误差在区间 [0.072, 0.092] 出现次数最多达到 23 次, 随机梯度下降法在区间 [0.198, 0.218] 出现最多为 25 次, 因此罚函数内点法优化模型的标准均方根误差出现在更小区间的次数更多, 进一步说明了优化后的模型具有更高的预测精度和更好的预测稳定性.

3.2 第二个时间序列

Mackey-Glass 混沌时间序列模型:

$$x(n+1) = x(n) + \Delta T \left(\frac{\alpha x \left(t - \frac{\tau}{\Delta T} \right)}{1 + x \left(t - \frac{\tau}{\Delta T} \right)^{10}} + \gamma x(n) \right) \quad (34)$$

式中, $\alpha = 0.2$, $\tau = 17$, $\gamma = -0.1$, $\Delta T = 1/10$. τ 表示延迟因子, 当 $\tau > 16.8$ 时序列 (34) 具有混沌特性. Leaky-ESN 模型, 罚函数内点法初始值设置与第一个时间序列一样. 通过仿真可知, 与随机梯度下降法相比, 用罚函数内点法优化的 Leaky-ESN 模型仍具有良好的逼近性能. 图 5 描述了随机梯度下降法和罚函数内点法的训练误差 NRMSE. 从图 5 中可知, 罚函数内点法优化的模型具有更小的 NRMSE. 图 6 和图 7 是随机测试中某次时间序列预测的结果. 从图 6 和图 7 中可以看出罚函数内点

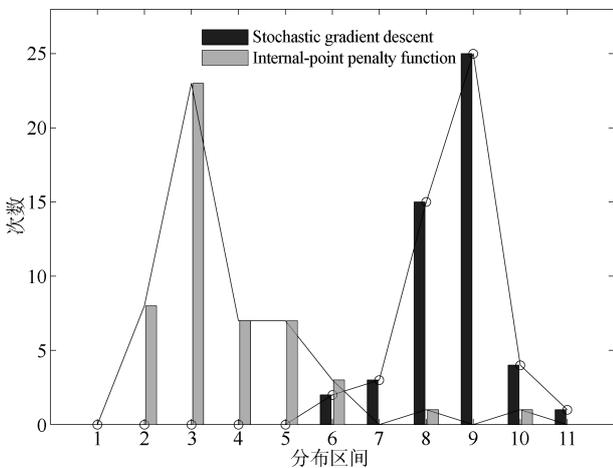


图 5 第一时间序列预测的 NRMSE 分布图
Fig. 5 NRMSE of predicted values for the first time series

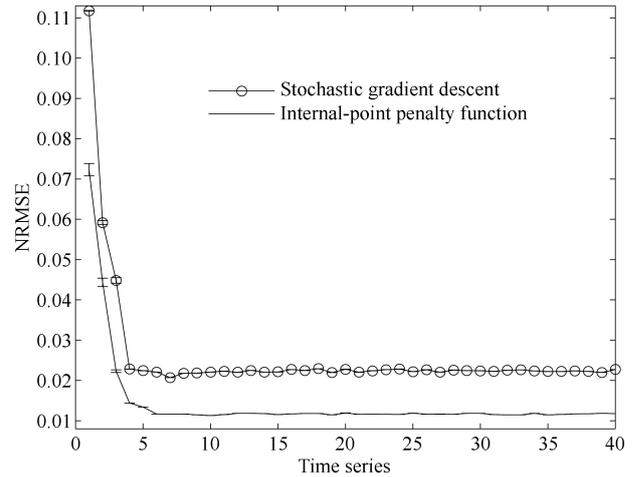


图 6 第二时间序列的训练误差

Fig. 6 Training errors for the second time series

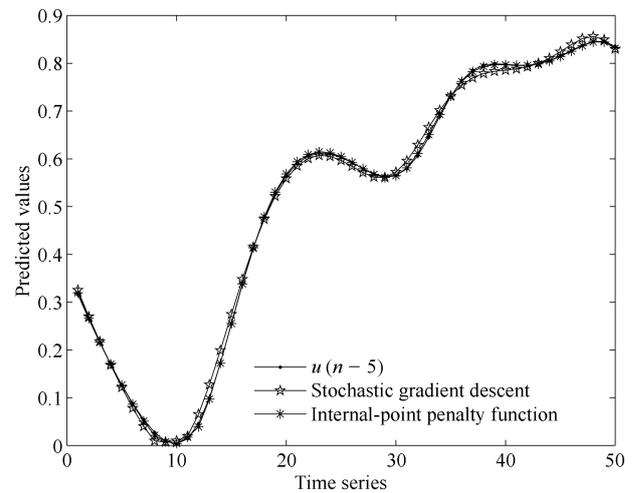


图 7 第二时间序列的预测

Fig. 7 Predicted values for the second time series

表 1 Leaky-ESN 的测试 NRMSE

Table 1 Testing NRMSE of Leaky-ESN

方法	第一时间序列	第二时间序列
随机梯度下降法	0.21240	0.17710
罚函数内点法	0.08698	0.01397

法优化的模型预测结果精度更高. 由表 1 也可以看出, 与随机梯度下降法相比, 利用罚函数内点法优化后的 Leaky-ESN 模型均具有良好的预测性能. 图 8 是第 5000 步数据的 NRMSE 的分布区间. 其中, 柱形图纵坐标表示 NRMSE 出现在每个区间总共的次数, 数字 1 表示误差的统计区间 [0, 0.01] (同样地, 为了避免过多的在图中空置, 因此我们就将第一个统计区间人为地放大一点), 数字 2 和数字 3 的区间间隔为 0.004, 以此类推. 罚函数内点法优化的模型误差主

要集中在区间 $[0.011, 0.015]$, 总共出现 22 次, 并且分布比较紧凑, 说明稳定性相对比较好; 而随机梯度下降法优化的模型主要集中在后面几个区间, 而且分布相对比较均匀, 说明模型训练时会有一定的随机性和不稳定性。

由以上两个仿真例子可以说明, 与随机梯度下降法优化的 Leaky-ESN 相比, 罚函数内点法优化的 Leaky-ESN 具有更好的逼近精度. 一旦参数的初始值满足可行域要求, 罚函数内点法将始终使参数工作在可行域, 因此不用像随机梯度下降法那样去实时判断参数是否满足回声特性条件, 避免算法被强行中断重新开始, 这样不仅节约了整个优化过程的时间, 而且对可行域中的任意一点初值都具有很好的收敛性和稳定性。

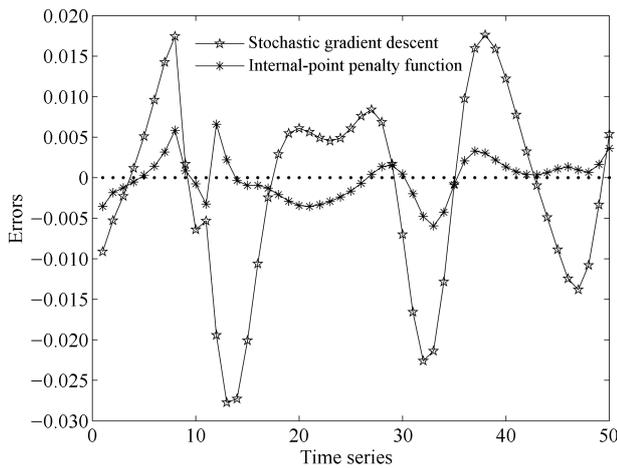


图 8 第二时间序列的预测误差

Fig. 8 Predicted errors for the second time series

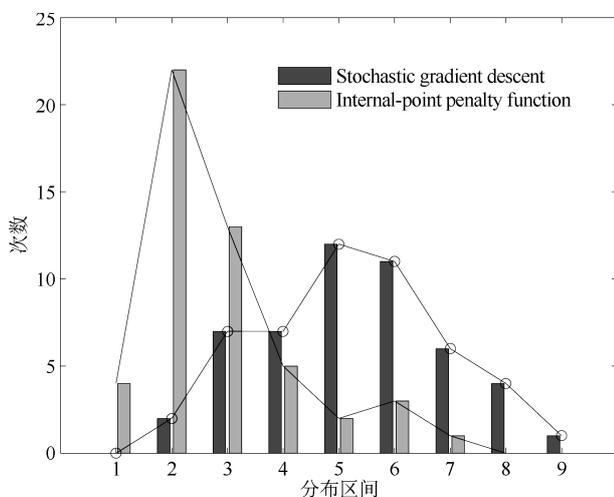


图 9 第二时间序列预测的 NRMSE 分布图

Fig. 9 NRMSE of predicted values for the second time series

4 结论

为了保障 Leaky-ESN 模型的回声特性, Leaky-ESN 的泄漏率和储备池连接权谱半径等全局参数之间满足不等式约束条件. 通常需要合理地选择 Leaky-ESN 的全局参数值, 有效地提升 Leaky-ESN 模型的性能. 因此, 本文利用罚函数内点法在参数的可行域内进行搜索, 充分利用了目标函数的二阶导数信息, 能够获取 Leaky-ESN 模型的最优全局参数值, 提高了 Leaky-ESN 模型的预测精度. 另外, 在参数修正过程中增加了目标函数的二阶导数信息, 使得计算量增大, 但与随机梯度下降法相比, 罚函数内点法将始终使参数工作在可行域内, 不用实时判断参数是否满足回声特性条件, 避免算法被强行中断重新开始, 实际上节约了整个优化过程的时间, 而且对可行域中的任意一点初值都具有很好的收敛性和稳定性。

References

- 1 Jaeger H. Tutorial on Training Recurrent Neural Networks, Covering BPTT, RTRL, EKF, and the “Echo State Network” Approach. Technical Report GMD Report 159, German National Research Center for Information Technology, German, 2002.
- 2 Jaeger H. The “Echo State” Approach to Analysing and Training Recurrent Neural Networks. Technical Report GMD report 148, German National Research Center for Information Technology, German, 2001.
- 3 Jaeger H, Haas H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science*, 2004, **304**(5667): 78–80
- 4 Lun S X, Wang S, Guo T T, Du C J. An I-V model based on time warp invariant echo state network for photovoltaic array with shaded solar cells. *Solar Energy*, 2014, **105**: 529–541
- 5 Skowronski M D, Harris J G. Noise-robust automatic speech recognition using a predictive echo state network. *IEEE Transactions on Audio, Speech, and Language Processing*, 2007, **15**(5): 1724–1730
- 6 Han S I, Lee J M. Precise positioning of nonsmooth dynamic systems using fuzzy wavelet echo state networks and dynamic surface sliding mode control. *IEEE Transactions on Industrial Electronics*, 2013, **60**(11): 5124–5136
- 7 Li G Q, Niu P F, Zhang W P, Zhang Y. Control of discrete chaotic systems based on echo state network modeling with an adaptive noise canceler. *Knowledge-Based Systems*, 2012, **35**(15): 35–40
- 8 Song R Z, Xiao W D, Sun C Y. A new self-learning optimal control laws for a class of discrete-time nonlinear systems

based on ESN architecture. *Science China Information Sciences*, 2014, **57**(6): Article No. 068202

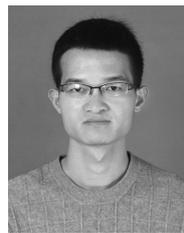
- 9 Lun S X, Yao X S, Qi H Y, Hu H F. A novel model of leaky integrator echo state network for time-series prediction. *Neurocomputing*, 2015, **159**(1): 58–66
- 10 Bianchi F M, Scardapane S, Uncini A, Rizzi A, Sadeghian A. Prediction of telephone calls load using echo state network with exogenous variables. *Neural Networks*, 2015, **71**(C): 204–213
- 11 Jaeger H, Lukoševičius M, Popovici D, Siewert U. Optimization and applications of echo state networks with leaky-integrator neurons. *Neural Networks*, 2007, **20**(3): 335–352
- 12 Lukoševičius M. A practical guide to applying echo state networks. *Neural Networks: Tricks of the Trade* (Second Edition). Berlin Heidelberg: Springer-Verlag, 2012. 659–686
- 13 Nocedal J, Wright S J. *Numerical Optimization* (Second Edition). New York: Springer, 2006. 30–31



伦淑娴 渤海大学教授。2005 年获得东北大学控制理论与控制工程专业博士学位。2011 年在中国科学院自动化研究所从事博士后研究。主要研究方向为神经网络, 能源互联网, 光伏发电系统建模与能量管理。本文通信作者。

E-mail: jzlunzi@163.com

(**LUN Shu-Xian** Professor at Bohai University. She received her Ph. D. degree from Northeastern University in 2005. She did her postdoctor research at the Institute of Automation, Chinese Academy of Sciences in 2011. Her research interest covers neural network, energy internet, modeling and energy management for photovoltaic power generation systems. Corresponding author of this paper.)



胡海峰 渤海大学硕士研究生。主要研究方向为神经网络。

E-mail: 15504991421@163.com

(**HU Hai-Feng** Master student at Bohai University. His main research interest is neural network.)