

全局优化的神经网络方法

赵华敏, 陈开周

(西安电子科技大学 理学院, 西安 710071)

摘要: 提出了一种解全局优化问题的神经网络模型, 并分析了该模型的收敛性与可行性. 然后, 给出了一个算法, 严格地证明了该算法对优化问题的任意给定的初始点, 都能收敛到它的一个全局极小点. 最后的仿真结果表明, 该算法是有效的.

关键词: 全局优化; 神经网络; 收敛性; 下降点

中图分类号: O221.2

文献标识码: A

Neural network for global optimization

ZHAO Hua-min, CHEN Kai-zhou

(School of Science, Xidian University, Xi'an 710071, China)

Abstract: A neural network model for solving global optimization problems is proposed. The convergence and feasibility of the model are analyzed. Then an algorithm is provided. It is strictly proved that for an arbitrarily given initial point of an optimization problem, the algorithm converges to a global minimizer of the problem. Finally, simulation results are presented to illustrate the effectiveness of the algorithm.

Key words: global optimization; neural network; convergence; descent point

1 引言(Introduction)

在许多学科与领域中, 人们经常遇到大量的全局优化问题, 并需要对它们进行实时求解. 这样, 许多传统的优化方法便不再能够满足人们的要求. 由于人工神经网络具有许多优点, 如大规模并行处理、分布式存储以及网络的计算时间复杂度几乎为零等^[1], 因此, 人工神经网络已成为实时求解优化问题的一条有效的途径^[2~4]. 但是, 由于目前大多数优化神经网络都是基于梯度下降法, 因而, 对于非凸的优化问题, 往往容易陷入局部极小点. 于是, 对全局优化的神经网络方法的研究就显得非常重要.

本文在文献[5]的基础上, 提出了一种解全局优化问题的神经网络方法, 该方法包含两个阶段: 梯度下降阶段和求下降点阶段. 它从优化问题的任意给定的初始点出发, 先在第一阶段中, 进行局部极小化, 得到一个稳定点, 如果该点不是全局极小点, 那么, 利用第二阶段, 求出一个下降点, 并把它作为下次梯度下降阶段的初始点. 交替使用这两个阶段, 最终可得到全局极小点. 该方法不仅简单, 而且能够很快地求出优化问题的全局极小点. 另外, 该方法还可

用于最小二乘等许多问题的求解之中.

2 神经网络模型(Neural network model)

考虑如下的全局优化问题

$$\min f(x), x \in \Omega. \quad (1)$$

其中 $x = (x_1, x_2, \dots, x_n)^T$, $\Omega = \{x \in \mathbb{R}^n \mid a \leq x \leq b, a, b \in \mathbb{R}^n\}$, $f: \Omega \rightarrow \mathbb{R}$ 为二阶连续可微函数. 设 $f(x)$ 在 Ω 上仅有有限个稳定点, 从而只要选择适当的 Ω , 就可以使这些点在 Ω 的内部.

下面, 我们将给出解全局优化问题(1)的神经网络模型并分析其收敛性与可行性.

2.1 梯度下降阶段(Gradient descent phase)

在梯度下降阶段中, 我们构造如下的神经网络模型

$$\begin{cases} dx/dt = -W \nabla f(x), \\ x(0) = \bar{x}, \bar{x} \in \Omega. \end{cases} \quad (2)$$

其中 $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)^T$ 为该阶段的初始点, $W = (w_{ij})_{n \times n}$ 为对称正定矩阵(通常, 取 W 为对角元素均为正数的对角矩阵). 为了便于讨论, 我们令 $x^0 = \bar{x}$.

令 $g(x) = \nabla f(x)$, 且

$$g(x) = (g_1(x), g_2(x), \dots, g_n(x))^T,$$

我们可作出神经网络(2)的框图^[6](如图1所示),其输入为 W, x , 输出为 x .

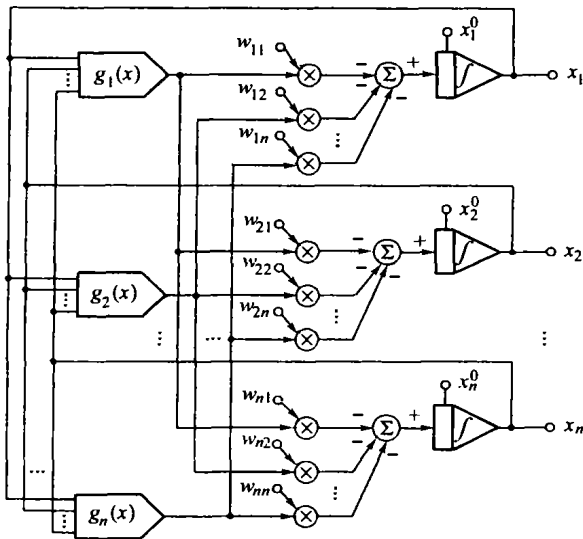


图1 神经网络(2)的框图

Fig. 1 Block diagram of the neural network in (2)

由前面的假设知,初值问题(2)的解存在且唯一.

为了分析神经网络(2)的收敛性,我们定义能量函数 $E(x) = f(x) - f(x_{gm})$, 其中 x_{gm} 为问题(1)的一个全局极小点. 并且, 对 $\forall x^0 \in \Omega$, 记能量函数 $E(x)$ 的水平集为 $L(x^0) = \{x \in \Omega \mid E(x) \leq E(x^0)\}$, 显然, 水平集 $L(x^0)$ 为有界集.

定理1 当 $t \rightarrow +\infty$ 时, 神经网络(2)的轨线 $x = x(t, x^0)$ 趋于网络(2)在 Ω 中的一个平衡点, 即

$$\lim_{t \rightarrow +\infty} x(t, x^0) = x^*, \quad x^0 \in \Omega, \quad (3)$$

其中 $x^* \in \Omega$, 且满足 $\nabla f(x^*) = 0$.

证 1) 任给初始点 $x^0 \in \Omega$, 网络(2)存在唯一轨线 $x = x(t, x^0)$, 沿此轨线有

$$\frac{dE}{dt} = \nabla E(x)^T \frac{dx}{dt} = -\nabla f(x)^T W \nabla f(x) \leq 0,$$

因此, 沿轨线 $x = x(t, x^0)$, $E[x(t, x^0)]$ 是单调不增的. 从而, 对正半轨线 $\gamma^+(x^0) = \{x(t, x^0) \mid t \geq 0\}$, 有 $\gamma^+(x^0) \subseteq L(x^0)$, 且因 $L(x^0)$ 有界, 所以 $\gamma^+(x^0)$ 是有界的.

2) 因 $\gamma^+(x^0)$ 是有界点集, 取严格单调递增数列 $\{\bar{t}_n\}$:

$$0 \leq \bar{t}_1 < \bar{t}_2 < \cdots < \bar{t}_n < \bar{t}_{n+1} < \cdots, \\ \bar{t}_n \rightarrow +\infty,$$

则 $\{x(\bar{t}_n, x^0)\}$ 为有界点集, 因而存在极限点 x^* , 即存在严格单调子数列 $\{t_n\} \subseteq \{\bar{t}_n\}$, $t_n \geq 0$, $t_n \rightarrow +\infty$, 使

$$\lim_{n \rightarrow +\infty} x(t_n, x^0) = x^* \in L(x^0),$$

即 x^* 为正半轨线 $\gamma^+(x^0)$ 的 ω 极限点.

3) 由1)的证明知, $E(x)$ 是开集 $\text{int}\Omega$ (Ω 的内部)上的 Lyapunov 函数, 且

$$\dot{E}[x(t)] = 0 \Leftrightarrow \nabla f(x) = 0.$$

若令 S 为神经网络(2)的平衡点集中的最大不变集, 那么, 由 LaSalle 不变原理知, 当 $t \rightarrow +\infty$ 时, 有 $x(t, x^0) \rightarrow S$, 故

$$\lim_{t \rightarrow +\infty} x(t, x^0) = \lim_{n \rightarrow +\infty} x(t_n, x^0) = x^*,$$

$$\nabla f(x^*) = 0.$$

因此, 任给初始点 x^0 , 网络(2)的轨线 $x = x(t, x^0)$ 趋于它的一个平衡点 x^* .

定理1中的公式(3)表明, 从 Ω 中的初始点 x^0 出发, 通过对神经网络(2)进行求解, 可以得到 $f(x)$ 在 Ω 内的一个稳定点 x^* . 如果 x^* 也是全局极小点, 那么, 我们就得到了问题(1)的一个最优解; 否则, 我们就需要求出它的一个下降点.

2.2 求下降点阶段(Finding descent point phase)

在该阶段中, 我们令

$$h(x) = f(x) - f(x^*) + \epsilon, \quad x \in \Omega. \quad (4)$$

其中 x^* 为由上一阶段所求出的 $f(x)$ 在 Ω 上的一个稳定点, ϵ 是一个很小的正数. 并构造方程

$$h(x) = 0, \quad x \in \Omega. \quad (5)$$

类似于文献[5], 我们可得出如下定理:

定理2 x^* 为 $f(x)$ 的全局极小点的充要条件为对任意给定的 $\epsilon > 0$, 方程(5)无解.

证 必要性. 若 x^* 为 $f(x)$ 的全局极小点, 则对 $\forall x \in \Omega$, 有 $f(x) \geq f(x^*)$. 于是, 对任意给定的 $\epsilon > 0$, 有 $f(x) - f(x^*) + \epsilon > 0$, 故方程(5)无解.

充分性. 若对任意给定的 $\epsilon > 0$, 方程(5)无解, 则 x^* 必为 $f(x)$ 在 Ω 上的一个局部极小点, 从而, 对 $\forall x \in \Omega$, 有 $f(x) - f(x^*) + \epsilon > 0$. 再由 ϵ 的任意性知, 在 Ω 上有 $f(x) - f(x^*) \geq 0$, 即 x^* 为 $f(x)$ 的全局极小点.

定理3 若 x^* 不是 $f(x)$ 的全局极小点, 则必存在 $\epsilon > 0$, 使得方程(5)有解. 若令方程(5)的解为 $\bar{x}$, 则 $f(\bar{x}) < f(x^*)$, 即 $\bar{x}$ 为 $f(x)$ 在 Ω 上的一个下降点.

证 由定理2及式(4), (5)显然可得.

为了对方程(5)进行求解, 我们可建立如下的神经网络模型

$$\begin{cases} dx/dt = -D \nabla h(x) h(x), \\ x(0) = x^* + r. \end{cases} \quad (6)$$

其中 r 为 n 维非零向量, 其分量均为 $[-1, 1]$ 上的随机数, $D = (d_{ij})_{n \times n}$ 为对称正定矩阵. 在网络(6)中, 我们令 $x^0 = x^* + r$.

显然, 在上一节的图 1 中, 若令 $g(x) = \nabla h(x)h(x)$, $W = D$, 则图 1 即为神经网络(6)的框图.

同样, 初值问题(6)的解存在且唯一.

定理 4 当 $t \rightarrow +\infty$ 时, 神经网络(6)的轨线 $x = x(t, x^0)$ 趋于网络(6)的一个平衡点, 即

$$\lim_{t \rightarrow +\infty} x(t, x^0) = \bar{x}, x^0 \in \Omega.$$

其中 $\bar{x} \in \Omega$, 且满足 $h(\bar{x}) = 0$ 或者 $\nabla h(\bar{x}) = 0$.

证 只要对神经网络(6)定义能量函数 $E(x) = 1/2 h^2(x)$, 并仿照定理 1 的证明, 即可得到该定理的结论.

定理 4 表明, 任给初始点 $x^0 \in \Omega$, 网络(6)收敛于它的一个平衡点 $\bar{x}$. 若 $h(\bar{x}) = 0$, 则由定理 3 知, $f(\bar{x}) < f(x^*)$, 从而, 在这一阶段中, 实现了函数值的严格下降, 故 $\bar{x}$ 可作为下一次梯度下降阶段的初始点.

否则, 若对任意的初始点 $x^0 \in \Omega$, 由网络(6)所求得的平衡点总不满足方程(5), 则方程(5)无解, 从而, 由定理 2 知, 当前的稳定点 x^* 即为 $f(x)$ 在 Ω 上的全局极小点.

3 算法及其收敛性 (Algorithm and its convergence)

根据上节中的理论分析, 我们可以给出解全局优化问题(1)的神经网络算法.

第一步 初始化: 令 $t = 0, k = 0$, 任选初始点 $\bar{x} = x(0) \in \Omega$, 取 $\Delta t > 0, \epsilon > 0$, 自然数 K , 矩阵 $W, D > 0$, 精度 $\epsilon_1, \epsilon_2, \epsilon_3 > 0$, 并令 $x = \bar{x}, u = \nabla f(x)$;

第二步 对网络(2)进行状态更新: $x(t + \Delta t) = x - \Delta t W u$;

第三步 计算 $u = \nabla f(x(t + \Delta t))$;

第四步 若 $\|u\|_2 < \epsilon_1$ (此处及以下 $\|\cdot\|_2$ 均指 Euclid 范数), 则令 $x^* = x(t + \Delta t), t = t + \Delta t$, 转第五步; 否则, 令 $x = x(t + \Delta t), t = t + \Delta t$, 转第二步;

第五步 产生 n 维随机向量 r , 令 $x = x^* + r, u = \nabla h(x), v = h(x), k = k + 1$;

第六步 对网络(6)进行状态更新: $x(t + \Delta t) = x - \Delta t D u v$;

第七步 计算 $u = \nabla h(x(t + \Delta t)), v = h(x(t + \Delta t))$;

第八步 若 $\|v\| < \epsilon_2$, 则令 $x = x(t + \Delta t), k = 0, t = t + \Delta t$, 转第二步; 否则, 转第九步;

第九步 若 $\|u\|_2 < \epsilon_2$, 则转第十步; 否则, 令 $x = x(t + \Delta t), t = t + \Delta t$, 转第六步;

第十步 若 $k \geq K$, 则停, 令优化问题(1)的全局极小点 $x_{gm} \approx x^*$, 全局极小值 $f_{gm} \approx f(x^*)$; 否则, 令 $t = t + \Delta t$, 转第五步.

类似于文献[5], 对于该算法, 有如下的收敛性定理:

定理 5 上述算法经过有限次迭代能够求得优化问题(1)的一个全局极小点.

证 由定理 1 知, 从 Ω 中的初始点 $\bar{x}$ 出发, 在梯度下降阶段中, 通过对神经网络(2)进行求解, 可以得到优化问题(1)的一个稳定点 x_1^* . 若 x_1^* 也是全局极小点, 则由定理 2 知, 方程(5)无解, 此时算法终止, 且求得了问题(1)的一个全局极小点.

否则, 若 x_1^* 不是问题(1)的全局极小点, 则算法转入求下降点阶段, 由定理 3 知, 从初始点 $x_1^* + r$ 出发, 解方程 $f(x) - f(x_1^*) + \epsilon = 0$, 可得一个解 $\bar{x}$, 且有 $f(\bar{x}) < f(x_1^*)$. 再以 $\bar{x}$ 为初始点转入第一阶段, 进行局部极小化, 不妨设求得的稳定点为 x_2^* , 则有 $f(x_2^*) \leq f(\bar{x})$, 从而有 $f(x_2^*) < f(x_1^*)$. 又由于 $f(x)$ 在 Ω 上仅有有限个稳定点, 故如此反复迭代, 算法必在有限次迭代后求得问题(1)的一个全局极小点.

4 仿真结果 (Simulation results)

为了验证文中算法的有效性, 我们对许多典型的全局优化试验函数进行了计算机仿真, 结果表明该算法不仅可行, 而且简单、有效. 下面, 仅给出两个试验函数^[7]及其相应的仿真结果.

在算法中, 取 $\Delta t = 0.02, \epsilon = 0.01$, 自然数 $K = 10$, 矩阵 W, D 为单位矩阵, 精度 $\epsilon_1 = 1.0 \times 10^{-4}, \epsilon_2 = 1.0 \times 10^{-5}, \epsilon_3 = 1.0 \times 10^{-4}$.

例 1 Six-hump camelback 函数

$$f(x_1, x_2) =$$

$$(4 - 2.1x_1^2 + \frac{1}{3}x_1^4)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2,$$

$$-3 \leq x_1 \leq 3, -2 \leq x_2 \leq 2.$$

它有 6 个局部极小点, 其中的两个点 $(-0.08983, 0.7126)^T$ 和 $(0.08983, -0.7126)^T$ 为全局极小点, 全局极小值为 -1.03163 .

当取初始点为 $(1.6, 0.6)^T$ 时, 仿真结果为 $(-0.08983, 0.71266)^T$; 当取初始点为 $(1.6, -0.6)^T$

时,仿真结果为 $(0.08985, -0.71266)^T$. 图 2 显示了该问题的计算机仿真轨线.

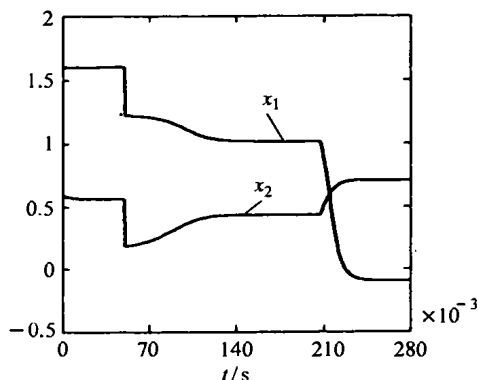
例 2 二维 Shubert 函数

$$f(x_1, x_2) = \left\{ \sum_{i=1}^5 i \cos[(i+1)x_1 + i] \right\} \cdot \left\{ \sum_{i=1}^5 i \cos[(i+1)x_2 + i] \right\},$$

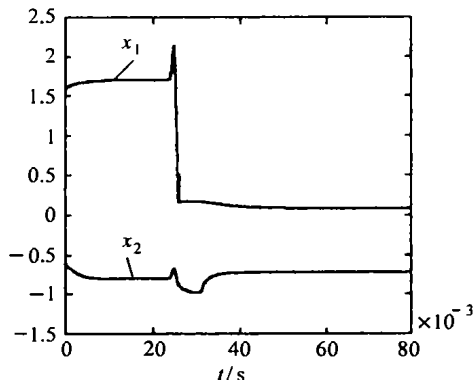
$$-10 \leq x_j \leq 10, j = 1, 2.$$

它有 760 个局部极小点,其中有 18 个点为全局极小点,全局极小值为 -186.73091 .

当取初始点为 $(7, 7)^T$ 时,仿真结果为 $(5.48287,$



(a) 初始点 $(1.6, 0.6)^T$



(b) 初始点 $(1.6, -0.6)^T$

图 2 例 1 的计算机仿真轨线

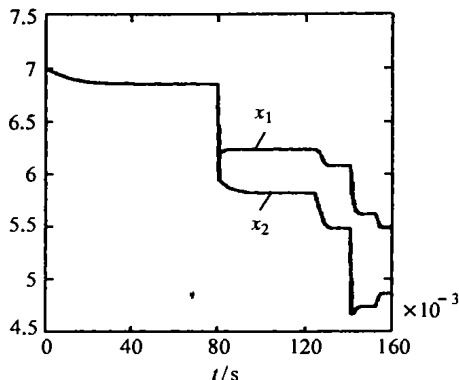
Fig. 2 Computer simulated trajectories for Example 1

5 结论(Conclusion)

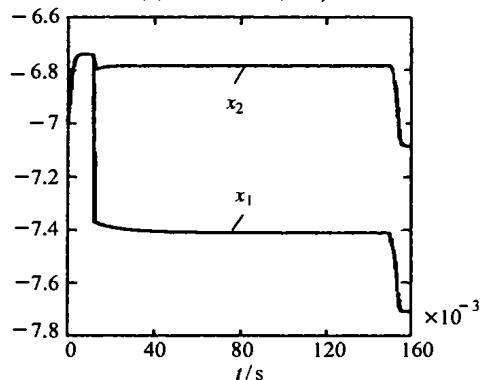
神经网络的大规模并行处理能力使得其应用已越来越广泛,近年来,有许多学者提出了求解各种优化问题的神经网络模型.但由于大多数网络都是以梯度下降法为基础的,因而,对于非凸的优化问题,往往只能得到问题的局部极小点.本文提出了一种解全局优化问题的神经网络方法,该方法不仅克服了梯度下降法的上述缺点,而且,与文献[5]中的方法相比,该方法不需要在 Ω 中布局射线并沿射线方向求方程(5)的根,它通过对网络(6)进行求解,就能

$4.85806)^T$;当取初始点为 $(-7, -7)^T$ 时,仿真结果为 $(-7.70831, -7.08351)^T$. 在仿真中,我们令 $F(x_1, x_2) = 0.01f(x_1, x_2)$,并对函数 $F(x_1, x_2)$ 进行求解.显然,若求出了函数 $F(x_1, x_2)$ 的全局极小点 x_{gm} ,则 x_{gm} 也一定为函数 $f(x_1, x_2)$ 的全局极小点.在图 3 中,我们画出了该问题的计算机仿真轨线.

从图 2 和图 3 中,我们可以看出,梯度下降法(即第一阶段)往往只能求得问题的局部极小点,而本文中的算法对其加以改进,即加上了求下降点阶段,从而,总能求得问题的全局极小点.



(a) 初始点 $(7, 7)^T$



(b) 初始点 $(-7, -7)^T$

图 3 例 2 的计算机仿真轨线

Fig. 3 Computer simulated trajectories for Example 2

直接求得方程(5)的一个实根,从而使得算法的迭代过程更加简单,也减少了计算量.理论分析和仿真结果均表明,该方法能够有效地求出优化问题的一个全局极小点,该方法在最小二乘等问题的求解中的应用还有待于进一步研究.

参考文献(References)

- [1] Jiao Licheng. Systematic Theory of Neural Networks [M]. Xi'an: Xidian University Press, 1996 (in Chinese)
- [2] Tao Qing, Fang Tingjian. The neural network model for solving min-max problems with constraints [J]. Control Theory and Applica-

- tions, 2000, 17(1): 82-84 (in Chinese)
- [3] Jiang Danchi, Wang Jun. A recurrent neural network for real-time semidefinite programming [J]. IEEE Transactions on Neural Networks, 1999, 10(1): 81-93
- [4] Hu Tiesong, Guo Yuanyu. Neural network for multiobjective dynamic programming [J]. Acta Electronica Sinica, 1999, 27(10): 70-73 (in Chinese)
- [5] Chen Kaizhou, Wu Kejian, Xiao Gaoxi. A new algorithm for global minimization[A]. Yue Minyi. Optimization Theory and Its Applications [C]. Xi'an: Xidian University Press, 1994, 384-391 (in Chinese)
- [6] Cichocki A, Unbehauen R. Neural Networks for Optimization and

Signal Processing[M]. New York: John Wiley & Sons. Inc., 1993

- [7] Levy A V and Montalvo A. The tunneling algorithm for the global minimization of functions[J]. SIAM Journal on Scientific and Statistical Computing, 1985, 6(1): 15-29

本文作者简介

赵华敏 1976年生.1997年毕业于西安电子科技大学应用数学系.现为该系博士生.研究方向为神经网络,最优化理论与算法.

Email: huamin_zhao@yahoo.com.cn

陈开周 1930年生.现为西安电子科技大学应用数学系教授,博士生导师.研究领域为神经网络,最优化理论、算法与应用等.

《控制理论与应用》征稿简则

1.《控制理论与应用》(英文名:Control Theory and Applications)是经国家教育委员会批准,由华南理工大学和中国科学院系统科学研究所联合主办的全国性一级学术刊物,为中、英文混合版,双月刊,国内外发行.

2.本刊主要报道系统控制科学中具有新观念、新思想的理论研究成果及其在各个领域中,特别是高科技领域中的应用研究成果和在国民经济有关领域技术开发、技术改造中的应用成果.内容包括:①集中参数控制系统;②分布参数控制系统;③随机控制系统;④离散事件系统;⑤自适应控制;⑥鲁棒控制;⑦智能控制;⑧系统控制科学中的其它重要问题;⑨信息不精确、不确定和部分真实情况下控制理论与应用的研究.读者对象是从事控制理论与应用研究的科技人员、高校师生及其它有关人员.

3.本刊设置的栏目主要有:综述与评论,论文与报告,应用研究,短文,研究简报,问题讨论,书刊评介,读者来信,国内外学术活动信息等.

4.本刊特别欢迎有关系统控制设计和应用方面的投稿.

5.本刊特别欢迎概括性强、有分析、有见解的综述与评论文章.

6.本刊原则上只发表原始性稿件,但不排除刊登已在国内外学术会议上发表或准备发表的文章(对于此种情况,作者必须如实说明)的可能性.但请勿将稿件同时投寄两个公开发行的刊物.

7.来稿格式及要求:

7.1 来稿要论点明确,论证充分,语句通顺,文字简练,书写端正,图表清晰,数据可靠(必要时注意保密),一式三份,并用电脑打字.

7.2 文稿的正文前面均要有不超过200字的中文和相应的英文内容摘要及3~8个关键词.摘要一般应包括研究的目的、方法、结果和结论.文末附作者简介.

7.3 一般定稿时综述不超过8000字,论文不超过6000字,短文不超过3000字;其它栏目的文章视具体内容 by 编辑部决定.

7.4 稿中非标准缩写词(中文或英文)须在首次出现时定义清楚.公式、图、表均须分别用阿拉伯数字全文统一编号.表或图(图一般不超过5幅)应放在文中相应的地方.中文稿的图、表,需同时附上中、英文图题、表题.

7.5 公式中的外文字母必须分清大小写,正斜体,上下标;容易混淆的外文字母及符号,要特别注意用铅笔标明.需要排黑体或花体的字母,请在其下划“~”号.

7.6 计量单位一律用国际单位,即SI单位制.名词术语必须规范化、标准化,前后一致.外国人名、地名、书刊名称已通用者外一律用原文.

7.7 参考文献按文中出现的先后次序排列,文献如为期刊时,按编号,作者(姓在前如Wiener L N, Kalman R E, Wang H.).文章题目,期刊名(外文可根据国际惯例使用缩写词),年份,卷号(期号)、页码顺序编排.文献如图书时,则按编号,作者(姓在前).书名.版次(初版不写).出版地点:出版者,年份,页码顺序排列.文中未引用的文献不得列入参考文献栏目.

7.8 来稿若不符合上述要求,编辑部退还作者重新整理.

8.已被本刊接受发表的稿件,按审查意见和“作者稿件加工须知”修改后一式两份寄编辑部.

9.编辑部有权对来稿进行文字删改,或退作者修改.来稿一经发表,按篇酌致稿酬.不拟刊登的稿件,一般情况半年左右退还作者.退修或退稿一般不寄回原稿,请作者自留底稿.可以电子投稿.

10. 来稿请寄:广州市五山华南理工大学《控制理论与应用》编辑部. 邮政编码:510640 Email: aukzllly@scut.edu.cn