

一种基于梯度搜索的全局优化新算法

郑立辉 郭亚军 潘德惠

(东北大学工商管理学院·沈阳, 110006)

摘要: 本文以神经网络为背景, 提出了一种以“惯性搜索”为核心的全局优化新算法, 在证明了算法的稳定性、最优性及可行性之后, 给出了算法的具体步骤和电路实现模型. 最后的仿真结果也表明, 本算法能够克服梯度下降法停留在局部极小值的缺点.

关键词: 全局优化; 梯度搜索; 降能; 神经网络

1 引言

最优化技术是自动控制、管理科学等领域的核心技术之一, 始终是专家、学者和工程技术人员们的研究热点. 自从 1985 年 Hopfield 把 Lyapunov 函数引入到神经网络系统, 并成功地解决了旅行商问题^[1]后, 应用神经网络的大规模并行处理特点, 来研究复杂的最优化问题似乎成了克服该领域中困难的“新的突破口”^[2]. 但是, Hopfield 网络以及后来由 Chua 等人提出的非线性规划网络^[3,4]都是以梯度下降法为基础的, 因而都存在着陷入局部极小值的危险. 为了消除这一弊端, 各种随机搜索算法越来越受到人们的重视, 如贝叶斯抽样法, 模拟退火技术等等^[5]. 随机搜索的思想是使搜索方向与目标函数的下降方向按统计规律相联系, 从而存在着跳出局部极小点的可能性.

与随机搜索不同, 本文将提出一种以“惯性搜索”为核心的全局优化方法. 本方法可以看成是具有“惯性”的梯度下降法. 形象地说, 就象沿着斜面滚落的小球, 能够跃过小的低谷而继续向前滚动一样, 本算法使搜索点在沿目标函数的梯度下降的过程中, 获得一定的“动能”, 算法依靠这种动能(或惯性)能够逃出局部极小点. 然后, 通过引入适当的降能策略, 就可以使算法最终到达全局最小值. 因此惯性搜索算法在继承了梯度下降法的优点的同时, 能够搜索到目标函数的全局最优解.

2 惯性搜索算法模型

一个无约束非线性规划问题可以表示为

$$\min_{x \in \mathbb{R}^n} \{g(x)\}. \quad (1)$$

这里只考虑映射 $g: \mathbb{R}^n \rightarrow \mathbb{R}$ 关于各个分量连续可微的情形.

定义 1 如果对任意 $x \in \mathbb{R}^n$, 点 x^* 满足

$$g(x^*) \leq g(x),$$

则称 x^* 为优化问题(1)的全局最优解.

定义 2 设 $\alpha \in \mathbb{R}$, 集合 $H[\alpha] := \{x \in \mathbb{R}^n \mid g(x) < \alpha\}$ 称为目标函数 $g(x)$ 的 α 水平集.

对目标函数作如下规定:

H) 假设对任意 $\alpha \in \mathbb{R}$, 目标函数 $g(x)$ 的 α 水平集 $H[\alpha]$ 为有界集合.

2.1 惯性搜索的基本原理

为了得到非线性规划问题(1)的全局最优解,对梯度下降法作如下改进,并表示成神经网络的形式

$$\begin{cases} C \frac{du_i}{dt} = -\frac{\partial g(x)}{\partial x_i}, & u_i(0) = u_i^0, \\ L \frac{dx_i}{dt} = u_i, & x_i(0) = x_i^0, \end{cases} \quad i = 1, 2, \dots, n. \quad (2a)$$

$$(2b)$$

式中 C 和 L 分别为正的常数, $x := [x_1, x_2, \dots, x_n]^T$, $u := [u_1, u_2, \dots, u_n]^T$, 均为时间 t 的函数, x_i^0 和 u_i^0 分别为算法的初值.

可以看出,本算法与梯度下降法的不同之处是,其搜索方向是通过一个线性系统间接地与目标函数的梯度 $\nabla_x g$ 相联系的(如图1).

若定义系统在 t 时刻的能量为

$$E(t) := \frac{1}{2} LC \sum_{i=1}^n \dot{x}_i^2(t) + g(x). \quad (3)$$

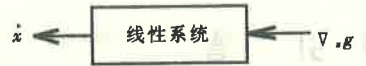


图1 搜索方向与梯度的间接关系

并称其前一项为动能,后一项为势能,那么这种改进后的梯度下降法有如下性质.

定理1 神经网络系统(2)的能量函数 $E(t)$ 在任何时刻均不改变,且等于其初始能量

$$E_0 = \frac{C}{2L} u_0^T u_0 + g(x_0). \quad (4)$$

其中 $x_0 := [x_1^0, x_2^0, \dots, x_n^0]$, $u_0 := [u_1^0, u_2^0, \dots, u_n^0]$ 为算法初值向量.

证 将式(2a)和(2b)的左右两端交叉相乘,并关于 i 求和,

$$C \sum_{i=1}^n \mu_i \frac{du_i}{dt} = -L \sum_{i=1}^n \frac{\partial g(x)}{\partial x_i} \frac{dx_i}{dt}.$$

上式两端从 0 到 t 对时间积分,得

$$\frac{1}{2} \sum_{i=1}^n [u_i^2(t) - u_i^2(0)] = -\frac{L}{C} \{g[x(t)] - g(x_0)\}.$$

将式(2b)代入,整理得

$$\frac{1}{2} LC \sum_{i=1}^n \dot{x}_i^2(t) + g(x) = \frac{C}{2L} u_0^T u_0 + g(x_0).$$

由于系统所能搜索到的点必须满足在该点处动能不小于 0,由定理1知,该点处势能必小于 E_0 ,因此系统的搜索区域等于目标函数 E_0 水平集的闭包的某连通子集(称集合 A 为连通的,若对 A 中任意两点,在 A 中存在连接该两点的连续曲线).

定义3 对于任意给定的初值 x_0 和初始能量 E_0 ,水平集 $H[E_0]$ 中包含 x_0 的最大连通子集的闭包,称为以 x_0 和 E_0 为初值的能达域,记为 $D(x_0, E_0)$.

在假设 H) 下,能达域为有界闭集,这使系统(2)所对应的搜索点在 $D(x_0, E_0)$ 内无能量损耗地振荡.

2.2 惯性搜索的收敛问题

神经网络系统(2)的能量守恒特性使搜索速度 \dot{x} 与目标函数值的大小之间建立了直接联系,并使网络在目标函数的全局最小值处具有最大的动能.这样就有可能通过适当的降能方法使系统收敛到此全局最小值.为此,对式(2a)加以改动,得到如下神经网络系统

$$\begin{cases} C \frac{du_i}{dt} = -\frac{\partial g(x)}{\partial x_i} - ru_i f[g(x_0) - g(x)], & u_i(0) = u_i^0, \\ L \frac{dx_i}{dt} = u_i, & x_i(0) = x_i^0, \end{cases} \quad i = 1, 2, \dots, n. \quad (5a)$$

$$i = 1, 2, \dots, n. \quad (5b)$$

其中 r 为正的常数, 并取

$$f(v) = \begin{cases} v, & v \geq 0. \\ 0, & v < 0. \end{cases} \quad (6)$$

定理 2 对于任意给定的初值 x_0 和 u_0 , 神经网络系统(5)在 Lyapunov 意义下是全局稳定的, 并且在水平集 $H[g(x_0)]$ 内是渐近稳定的.

证 仿定理 1 的证明, 将(5a)和(5b)左右两端交叉相乘, 关于 i 求和, 并从 0 到 t 对时间积分, 整理得

$$E(t) = E_0 - \frac{r}{L} \int_0^t \sum_{i=1}^n u_i^2(\tau) f[g(x_0) - g[x(\tau)]] d\tau.$$

其中 $E(t)$ 和 E_0 仍然由式(3)、(4)定义, 于是有

$$\dot{E}(t) = -\frac{r}{L} \sum_{i=1}^n u_i^2(t) f[g(x_0) - g(x)] \leq 0. \quad (7)$$

取 $E(t)$ 为系统(5)的 Lyapunov 函数, 由式(3)知, 若 $g(x)$ 有全局最小值, 则 $E(t)$ 有下界. 再由(7)式知系统是全局稳定的.

另外, 当系统对应的搜索点在水平集 $H[g(x_0)]$ 内运动时, 有 $g(x) < g(x_0)$, 且至少存在一个 i , 使 $u_i^2(t) > 0$. 由式(7), $\dot{E}(t) < 0$, 因此系统是渐近稳定的.

由定理 2 的证明过程知, 由式(3)定义的能量函数 $E(t)$ 对于系统(5)不再守恒了. 给定初值 x_0 及初始能量 E_0 后, 如果

$$D(x_0, E_0) \cap H[g(x_0)] \neq \emptyset, \quad (8)$$

即系统所对应的搜索点能够进入非空的渐近稳定区域时, 系统的能量将下降, 并逐渐地收敛到某一平衡点, 系统(5)的渐近稳定平衡点有如下性质.

定理 3 设给定初值 (x_0, u_0) 后, 系统(5)存在渐近稳定平衡点 (x_e, u_e) , 则

$$1) \frac{\partial g(x)}{\partial x_i} \Big|_{x=x_e} = 0, \quad i = 1, 2, \dots, n.$$

$$2) g(x_e) < g(x_0).$$

证 由于系统(5)平衡的充要条件是 $\dot{x}_i = 0, \dot{u}_i = 0 (i = 1, 2, \dots, n)$. 由式(5b)知 $u_i = 0$, 将 \dot{u}_i 和 u_i 代入(5a), 知命题 1) 成立.

若 (x_e, u_e) 是系统的渐近稳定平衡点, x_e 一定在系统(5)的渐近稳定域内, 由定理 2 知, $x_e \in H[g(x_0)]$, 因此 $g(x_e) < g(x_0)$, 即命题 2) 成立.

定理 4 对任意初值 x_0 , 存在 E_0 , 使 $H[g(x_0)] \subseteq D(x_0, E_0)$ 成立.

证 由 $g(x)$ 的连续性, 对任意 $x_0 \in \mathbb{R}^n, g(x_0) \in \mathbb{R}$. 根据假设 H), 存在连通的有界闭集 S , 使

$$H[g(x_0)] = \{x \in \mathbb{R}^n \mid g(x) \leq g(x_0)\} \subseteq S. \quad (9)$$

其中 \bar{A} 表示集合 A 的闭包, 显然 $x_0 \in S$. 由 $g(x)$ 在有界区域 S 内的有界性, 设 $d = \sup_{x \in S} g(x)$. 取 $E_0 > d$, 则有 $S \subseteq H[E_0]$, 即 S 是 $H[E_0]$ 中包括 x_0 的一个连通子集. 由能达域的定义知 $S \subseteq D(x_0, E_0)$, 结合式(9)知定理成立.

推论 对任意 x_0 , 如果存在满足条件 $g(x) < g(x_0)$ 的 x , 那么, 一定存在 E_0 使式(8)成立.

证 对任意 x_0 , 如果存在 x , 满足 $g(x) < g(x_0)$, 则 $H[g(x_0)] \neq \emptyset$. 由定理4, 一定存在 E_0 , 使 $D(x_0, E_0) \cap H[g(x_0)] = H[g(x_0)] \neq \emptyset$.

综上所述, 对于任意给定的初值 x_0 , 若 x_0 不是优化问题(1)的最优解, 即 $H[g(x_0)] \neq \emptyset$, 则一定存在 E_0 , 使 $D(x_0, E_0) \cap H[g(x_0)] \neq \emptyset$. 此时, 系统(5)在给定初值后能量下降, 并收敛于某一渐近稳定平衡点 x_e , 满足 $g(x_e) < g(x_0)$. 若 $H[g(x_0)] = \emptyset$, 则以此 x_e 为初值继续搜索, 直至所得 x_e^* 满足 $H[g(x_e^*)] = \emptyset$. 此 x_e^* 即为优化问题(1)的全局最优解.

3 算法的具体步骤和电路实现模型

根据上节中的理论分析, 我们可以给出算法的具体步骤:

- ① 任选初值 x_0 , 并选取 E_0 , 使式(8)成立.
- ② 按下式计算初值 u_0 ,

$$u_i^0 = \sqrt{\frac{2L}{nC} [E_0 - g(x_0)]}, \quad i = 1, 2, \dots, n.$$

把 x_0 和 u_0 代入式(5a)和(5b)中, 迭代求解.

- ③ 若由式(3)定义的能量函数 $E(t)$ 下降, 系统收敛于平衡点 x_e , 则以此点为初值 x_0 , 转①.

- ④ 若网络能量 $E(t)$ 始终不变, 则有以下两种情况:

a) 初始能量 E_0 不够大, 使式(8)中的条件没有满足, 此时应该重新选取 E_0 使式(8)中条件满足, 再转②.

b) 水平集 $H[g(x_0)]$ 为空集. 此无论怎么增大 E_0 , 网络能量 $E(t)$ 仍然不降, 此 x_0 即为优化问题(1)的全局最优解.

为了用神经网络实现惯性搜索, 给出系统(5)的电路模型如图2, 它由受控电流源、受控电压源、电容和电感组成. 其中受控电流源的电流为

$$I_{ai} = \frac{\partial g(x)}{\partial x_i}, \quad I_{bi} = ru_i f[g(x_0) - g(x)], \quad i = 1, 2, \dots, n.$$

受控电压源的电压 $u_{si}(t)$ 等于电容两端的电压 $u_i(t)$.

4 仿真实验

最后, 用著名的 Six-hump-camel-back 函数进行计算机仿真, 该函数的表达式为

$$g(x, y) = (4 - 2.1x^2 + \frac{x^4}{3})x^2 + xy + (-4 + 4y^2)y^2.$$

它有6个局部极小点, 其中的两个 $(-0.0898, 0.7126)$ 和 $(0.0898, -0.7126)$ 是全局极小点, 最小值为 -1.0316 .

在仿真中, 将(5a)和(5b)分别离散化为

$$\begin{cases} u_i(k+1) = u_i(k) - \frac{1}{C} \left\{ \frac{\partial g[x(k)]}{\partial x_i} + ru_i(k) f[g(x_0) - g(x(k))] \right\}, \\ x_i(k+1) = x_i(k) + \frac{1}{L} u_i(k), \end{cases}$$

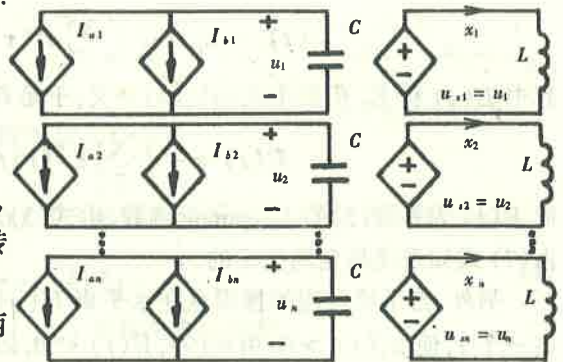


图2 惯性搜索算法的电路模型

$$k = 0, 1, 2, \dots,$$

并取 $C = 500, L = 500, r = 0.005$. 为了检验算法的效果, 我们把它与梯度下降法进行了对比, 结果表明, 梯度下降法在多数情况下都陷于局部极小点, 而惯性搜索算法在多数情况下都收敛于全局最小点, 即使停在局部极小点, 经过恢复能量再次搜索, 也能很快地收敛到全局最优解. 用 386 运行本算法的 C 语言程序, 在两分钟内便可搜索到最优解, 具体结果见附表(其中“*”为全局最优解).

附表 计算机仿真结果

初值	梯度下降法		惯性搜索法		
	平衡点	极小值	第一次的平衡点	极小值	第二次的平衡点
(6, 6)	(1.6071, 0.5687)	2.1043	(-0.0898, 0.7127)	-1.0316*	
(6, -6)	(1.7036, -0.7961)	-0.2155	(-0.0899, 0.7127)	-1.0316*	
(-6, 6)	(-1.7036, 0.7961)	-0.2155	(1.7036, -0.7961)	-0.2155	(0.0898, -0.7126)*
(-6, -6)	(-1.6071, -0.5687)	2.1043	(-0.0898, 0.7127)	-1.0316*	
(8, 2)	(1.6070, 0.5687)	2.1043	(0.0898, -0.7127)	-1.0316*	
(-1, 10)	(-0.0898, 0.7127)	-1.0316*	(-0.0898, 0.7127)	-1.0316*	
(-10, 2)	(-1.7036, 0.7961)	-0.2155	(1.7036, -0.7961)	-0.2155	(0.0898, -0.7126)*
(-1, -8)	(0.0898, -0.7126)	-1.0316*	(0.0898, -0.7127)	-1.0316*	

5 结束语

近年来, 神经网络的并行处理能力, 使它在人工智能、自动控制及管理科学等领域显示了巨大的应用前景. 但由于大多数神经网络都是以梯度下降法为基础的, 因而只能达到局部最优. 本文对梯度下降法加以改进, 提出了一种惯性搜索策略, 克服了梯度下降法的这一弱点, 从而有可能大大地提高神经网络的使用效果, 惯性搜索算法在其它以梯度下降法为基础的神经网络(如 BP 网)中的应用效果还有待于进一步研究.

参 考 文 献

- 1 Hopfield, J.J. and Tank, D.W.. "Neural" computation of decisions in optimization problems. *Biol. Cybern.*, 1985, 52: 141 - 152
- 2 吴沧浦. 系统优化方法中的并行处理与神经网络技术. *控制理论与应用*, 1992, 9(3): 309 - 312
- 3 Kennedy, M.P. and Chua, L.O.. Unifying the tank and hopfield linear programming circuit of Chua and Lin. *IEEE Trans. Circuit & Systems*, 1987, 34(2): 210 - 214
- 4 Kennedy, M.P. and Chua, L.O.. Networks for nonlinear programming. *IEEE Trans. Circuit & Syst.*, 1988, 35(5): 554 - 562
- 5 Bruce, E.S. and Eric, E.E.. A Comparison of bayesian/sampling global optimization techniques. *IEEE Trans. Syst. Man and Cybern.*, 1992, 22(5): 1024 - 1032

A New Global Optimization Algorithm Based upon Gradient Search

ZHENG Lihui, GUO Yajun and PAN Dehui

(Faculty of Business Administration, Northeastern University, Shenyang, 110006, PRC)

Abstract: In this paper a new global optimization algorithm centered around inertial search is put forward in the

background of neural network. First the stability, optimality and feasibility of the algorithm are proved. Then concrete steps as well as a circuit model for the algorithm are provided. In the end, an example is given to show the effectiveness of the algorithm in overcoming the drawback of being trapped by local minimum which often occurs in gradient search.

Key words: global optimization; gradient search; energy reduction; neural network

本文作者简介

郑立辉 1968年生. 1992年在沈阳大学获经济学学士学位. 同年考入东北大学工商管理学院攻读硕士学位, 1993年提前转为该校自动控制系博士研究生. 现在华中理工大学系统工程研究所做博士后研究. 主要研究兴趣是线性系统的降价逼近, 鲁棒控制与辨识, 优化算法, 以及这些理论、方法在经济分析中的应用.

郭亚军 1952年生. 1976年毕业于东北大学数学系, 1984年和1990年分别获工学硕士和博士学位. 现任东北大学工商管理学院教授, 经济学系主任. 研究方向为分布参数系统理论及其在管理科学、数量经济学中的应用.

潘德惠 1928年生. 1949年毕业于东北大学理学院. 现为东北大学工商管理学院教授, 自动控制系统控制理论与应用专业博士生导师. 中国数学会理事, 曾多年从事应用数学的研究, 现在的研究领域是分布参数系统的模型辨识与最优控制.

学术会议消息

1 全国(第三届)自动化应用技术交流会将于今年8月6日~12日在四川省阿坝藏族、羌族自治州的九寨沟召开。

2 中国(第五届)交流电机调速传动学术会议将于今年11月9日~13日在海南省三亚市召开。

3 全国(第三届)工业控制系统及应用学术会议将于今年11月17日~22日在云南省西双版纳召开。

欢迎对上述会议感兴趣的同志光临会议, 并请与100071北京919信箱学会办公室联系, 并说明拟参加哪次会议. 以便在会前(约40天)把会议通知寄给您。

中国自动化学会应用专业委员会
中国金属学会冶金自动化专业委员会