

连续时间 Hopfield 网络模型数值实现分析*

叶世伟^{1,2+}, 郑宏伟³, 王文杰¹, 马琳¹, 史忠植²

¹(中国科学院 研究生院 信息科学与工程学院,北京 100039)

²(中国科学院 计算技术研究所 智能信息处理重点实验室,北京 100080)

³(四川师范大学 数学系,四川 成都 610066)

Analysis for Numerical Implementation of Continuous Time Hopfield Network Model

YE Shi-Wei^{1,2+}, ZHENG Hong-Wei³, WANG Wen-Jie¹, MA Lin¹, SHI Zhong-Zhi²

¹(School of Information Science and Engineering, Graduate School, The Chinese Academy of Sciences, Beijing 100039, China)

²(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100080, China)

³(Department of Mathematics, Sichuan Normal University, Chengdu 610066, China)

+ Corresponding author: Phn: +86-10-88256016, Fax: +86-10-88256056, E-mail: shwye@gscas.ac.cn, <http://www.gscas.ac.cn>

Received 2003-09-02; Accepted 2003-10-15

Ye SW, Zheng HW, Wang WJ, Ma L, Shi ZZ. Analysis for numerical implementation of continuous time Hopfield network model. *Journal of Software*, 2004,15(6):881~890.

<http://www.jos.org.cn/1000-9825/15/881.htm>

Abstract: The choices of discrete time step for Euler method and trapezoidal method and terminating condition of iteration in trapezoidal method are discussed in this paper for numerical implementation of continuous time Hopfield network. The decreasing conditions of an energy function are investigated by the use of convex function. By utilization of the primal convex function, the conditions are analyzed under which its conjugate function minus a quadratic function is also convex. Based on the analysis of the proof for convergence of the continuous time Hopfield network model, a generalized model is proposed. For the common Euler and trapezoidal methods, the choice of their discrete time step is discussed for numerical implementation of the continuous time Hopfield network. As the trapezoidal method is an implicit scheme, its realization needs an iterated procedure. The conditions to terminate the iterated procedure are analyzed. According to the special forms of the continuous time Hopfield network model, an improved iterated algorithm for trapezoidal method is proposed and analyzed. The numerical

* Supported by the Presidential Foundation of Graduate School of the Chinese Academy of Sciences under Grant No.YZJJ200206 (中国科学院研究生院院长基金); the Opening Foundation of Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, the Chinese Academy of Sciences under Grant No.IIP 2003-4(中国科学院计算技术研究所智能信息处理重点实验室开放基金)

作者简介: 叶世伟(1968—),男,四川仁寿人,博士,副教授,主要研究领域为神经网络理论,机器学习;郑宏伟(1966—),男,讲师,主要研究领域为泛函分析,人工智能;王文杰(1964—),男,博士,副教授,主要研究领域为人工智能,多 Agent 系统;马琳(1978—),男,硕士生,主要研究领域为神经网络理论,机器学习;史忠植(1941—),研究员,博士生导师,主要研究领域为人工智能,多 Agent 系统,神经计算,机器学习。

results show that choosing a suitably large discrete time step will be helpful not only to accelerate the numerical implementation but also to improve the optimization performance.

Key words: convex function; Hopfield network; numerical implementation; Euler method; trapezoidal method

摘要: 讨论使用 Euler 方法和梯形方法在数值求解连续时间的 Hopfield 网络模型时,离散时间步长的选择和迭代停止条件问题.利用凸函数的定义研究了能量函数下降的条件,根据凸函数的性质分析它的共轭函数减去二次函数之差仍为凸函数的条件.分析连续时间 Hopfield 网络模型的收敛性证明,提出了一个广义的连续时间 Hopfield 网络模型.对于常用的 Euler 方法和梯形方法数值求数值实现连续时间 Hopfield 网络,讨论了离散时间步长的选择.由于梯形方法为隐式方法,分析了它的迭代求算法的停止条件.根据连续时间 Hopfield 网络的特点,提出改进的迭代算法,并对其进行了分析.数值实验的结果表明,较大的离散时间步长不仅加速了数值实现,而且有利于提高优化性能.

关键词: 凸函数;Hopfield 网络;数值实现;Euler 方法;梯形方法

中图法分类号: TP18

文献标识码: A

神经网络系统研究的重要意义已为许多科学家所承认,将它看成智能信息处理发展的一个主流方向.20 世纪 80 年代中期以来,神经网络重新引起了许多科技工作者的兴趣,形成近代非线性科学和计算智能研究的主要内容之一.特别是神经网络经历了近 20 年的迅速发展,它的独特知识表示结构和信息处理的原则,使其在许多应用领域中取得了显著效果^[1],能够为解决一些传统计算机极难求解的问题提供满意解或者为寻求满意解提供全新的思路.

连续时间的 Hopfield 网络(continuous time Hopfield network,简称 CTHN)一般用于求解优化问题,包括非线性规划和离散的组合优化问题.原则上,CTHN 都应该使用硬件实现,以充分体现神经网络运行的并行性和计算的分布式处理特点.但是,从已有的研究和实现结果来看,大多采用软件数值模拟^[2]实现.这主要有两个方面的原因,首先,硬件实现受目前现有条件的限制;其次,利用数值模拟实现可以得到新的求解问题的算法.因此,研究 CTHN 的离散化显得特别有意义.在讨论 CTHN 的离散化时,我们必须注意到这样一个事实,当 CTHN 模型用于优化计算时,我们仅需要网络演化的最终状态,而不关心详细的状态演化轨迹.这一点非常重要,它使得我们可以取得较大的时间步长,而不担心数值模拟的中间结果是否接近一条状态演化轨迹.这是 CTHN 离散化研究和一般的常微分方程组数值求解的最根本的区别,这也从一个侧面说明,已有的常微分方程组数值求解理论不适合 CTHN 离散化研究的原因,同时也是本文的研究有意义的一个现实背景.在 CTHN 的离散化实现中,常用的是 Euler 方法^[3]和梯形^[2]方法.在这两种方法中,关于时间步长的选取目前尚没有理论指导和分析,在具体实现中总是取不同的离散时间步长进行实验.时间步长的选择成为一个公开的理论问题.本文利用凸优化理论讨论和分析了 CTHN 的离散化方法的时间步长的选择.所得结果可以用于指导选择 CTHN 的离散化方法中时间步长的选择.对 CTHN 进行离散化后,我们就得到了离散时间的 Hopfield 网络(discrete time Hopfield network,简称 DTHN)模型^[3],当离散时间步长取为 1 时,这时我们得到标准的 DTHN 模型,因此本文的结论可用于标准的 DTHN 收敛性分析.

本文是这样安排的.考虑到优化理论是 CTHN 离散化分析的基础,所以我们首先简单介绍了凸函数的定义和性质^[4],其中着重介绍了次梯度和广义梯度以及凸函数的共轭函数的定义和性质.其次,我们简单介绍了 CTHN 的定义和在此基础上的广义的 CTHN 模型.设计了 CTHN 的两种离散化方法:Euler 方法和梯形方法.当然还有其他离散化方法,但是由文献[2]中的数值实验结果我们知道,高阶方法并不适合 CTHN 的求解,因此,在本文中我们仅仅局限于讨论这两种 CTHN 的离散化方法.利用凸优化理论讨论和分析了 CTHN 的离散化方法的时间步长的选择,用数值实验研究了具体的实现过程,最后给出了结论.

1 凸优化理论

为了后面的讨论方便,首先引入凸函数的定义和性质,它们都可以在文献[4]中找到,所以我们省略所有的证

明.对于 $f(\cdot)$ 为 $R^n \rightarrow (-\infty, \infty]$ 的凸函数,记 $P(f) = \{(x, \alpha) : x \in R^n, \alpha \in R, f(x) \leq \alpha\}$ 和 $ED(f) = \{x : x \in R^n, f(x) < +\infty\}$.

我们说, $f(\cdot)$ 为正常凸函数,如果它无处取 $-\infty$ 且不恒等于 $+\infty$.

定义凸函数 $f(\cdot)$ 支撑集如下:

$$L(f) = \{(a, b) : a \in R^n, b \in R, \forall x \in R^n \text{ 有 } a^T x - b \leq f(x)\}.$$

同时定义它的闭为

$$\text{cl } f(x) = \sup_{(a,b) \in L(f)} \{a^T x - b\}.$$

$f(\cdot)$ 为闭凸函数的条件为

$$\text{cl } f(x) = f(x).$$

本文讨论的凸函数都是闭凸函数,这里不再一一注明.

定义 1.1. 若凸函数 $f(\cdot)$ 在 x 处 $f(x) < +\infty$, 定义 $f(\cdot)$ 在 x 处的次梯度:

$$\xi \in \partial f(x) \Leftrightarrow \text{对任给 } y \text{ 有 } f(y) \geq f(x) + \xi^T (y - x) \quad (1)$$

特别地,若 $f(x)$ 在点 x 可微,则次梯度 $\partial f(x)$ 仅包含梯度 $\nabla f(x)$, 这时我们又将它看为一个向量.

性质 1.1. $\partial f(x)$ 为一个闭的凸集.

定义 1.2. 设 $f(x)$ 为 $R^n \rightarrow (-\infty, +\infty]$ 的凸函数,定义 $f(x)$ 的共轭函数如下:

$$f^*(\xi) = \max_{x \in R^n} \{x^T \xi - f(x)\} \quad (2)$$

性质 1.2. 设 $f(x)$ 为 $R^n \rightarrow (-\infty, +\infty]$ 的凸函数,则共轭函数 f^* 也是凸函数,且 $P(f^*) = L(f)$, 此外

$$f^{**} = \text{cl}(f).$$

性质 1.3. 设 $f(x)$ 为 R^n 上的凸函数,则 $\xi \in \partial f(x)$ 的充要条件为 $f^*(\xi) = \xi^T x - f(x)$.

性质 1.4. 设 $f(x)$ 为 R^n 上的凸函数,若 $\xi \in \partial f(x)$, 则 $x \in \partial f^*(\xi)$. 若 $x \in \partial f^*(\xi)$ 且 $f(x)$ 在 x 点是闭的, 则 $\xi \in \partial f(x)$.

定义 1.3. 设 $\|\cdot\|$ 是定义在 R^n 上的范数. 若 $f(x)$ 在 x 附近是 Lipschitz 的, 即 $|f(x) - f(y)| \leq K \|\cdot\| \|x - y\|$, 则对任何 $d \in R^n$, 定义 $f(x)$ 在 x 处沿方向 d 的广义方向导数为

$$f^\circ(x, d) = \limsup_{y \rightarrow x, t \downarrow 0} \frac{f(y + td) - f(y)}{t}.$$

定义 1.4. 若 $f(x)$ 在 x 附近是 Lipschitz 的, 则我们称集合 $\{\xi \in X^* : f^\circ(x, d) \geq \xi^T d, \forall d \in R^n\}$ 是函数 $f(x)$ 在点 x 的广义梯度, 记为 $\partial f(x)$.

函数 $f(x)$ 的广义梯度和次梯度在 $f(x)$ 为凸函数是吻合的, 这也是我们用同一记号的理由.

为了简便起见, 在 $\partial f(x)$ 以后的使用中, 有时我们用它表示次梯度中的任意一个向量.

引理 1.1 (广义梯度或次梯度的基本性质).

- 1) 若 $f(x)$ 是凸函数, $g(x)$ 在 x 附近可微, 则有 $\partial f(x) + \nabla g(x) \subset \partial(f+g)(x)$;
- 2) 若 $f(x), g(x)$ 是凸集 Ω 上的凸函数且存在点 $x_0 \in \Omega$ 使得它们中的一个在 x_0 处连续, 则

$$\partial f(x) + \partial g(x) \subset \partial(f+g)(x).$$

证明: 参见文献[4].

这些关于凸函数的基本定义和性质, 也可参见文献[4].

定理 1.1. 设 $h(x) = x^T A x$, $E(x) = f(x) - g(x)$, 其中, A 为对称矩阵, $f(x), g(x) + h(x)$ 和 $f(x) - h(x)$ 都为凸集 Ω 上的凸函数. 对任给向量 $x \in \Omega$ 取 $\xi \in \partial g(x)$, 若 $y \in \Omega$ 使得 $\xi \in \partial f(y)$, 则 $E(y) \leq E(x)$. 特别地, 若 $g(x) + h(x)$ 或 $f(x) - h(x)$ 有一个为严格凸函数, 且 $x \neq y$, 则 $E(y) < E(x)$.

证明: 由于 $g(x) + h(x)$ 为凸函数, 由次梯度的定义有

$$g(y) + h(y) \geq g(x) + h(x) + (y-x)^T \partial(g+h)(x).$$

因为 $h(x)$ 为可微, 同时考虑到引理 1.1 的 1), 即 $\partial g(x) + \nabla h(x) \subset \partial(g+h)(x)$, 从而有

$$g(y) - g(x) \geq h(x) - h(y) + (y-x)^T \nabla h(x) + (y-x)^T \xi \quad (3)$$

同样, 由于 $f(x) - h(x)$ 为凸函数, 由次梯度的定义有

$$f(x) - h(x) \geq f(y) - h(y) + (x-y)^T \partial(f-h)(y).$$

因为 $h(x)$ 为可微,同时考虑到引理 1.1 的 2), $\partial f(y) - \nabla h(y) \subset \partial(f-h)(y)$, 从而有

$$f(x) - f(y) \geq h(x) - h(y) - (x-y)^T \nabla h(y) + (x-y)^T \xi \quad (4)$$

由式(3)和式(4), 有 $E(x) - E(y) \geq 2[h(x) - h(y) - (x-y)^T (\nabla h(x) + \nabla h(y))/2] = 0$.

由此可证明 $E(y) \leq E(x)$. 若 $g(x) + h(x)$ 或 $f(x) - h(x)$ 有一个为严格凸函数, 并且 $x \neq y$, 则式(3)或式(4)成为严格不等式, 从而 $E(y) < E(x)$ 成立. \square

若次梯度仅是近似, 则上述定理并不能直接应用. 但我们有如下的定理:

定理 1.2. 设 $E(x) = f(x) - g(x)$, 其中 $f(x)$ 为凸集 Ω 上的凸函数, $g(x)$ 为凸集 Ω 上局部 Lipschitz 连续函数, 存在 $h_i(x) = x^T A_i x, i=1, 2$, 使得 $f(x) - h_1(x)$ 和 $g(x) + h_2(x)$ 都为凸函数且其中至少有一个为严格凸函数, $h(x) = h_1(x) - h_2(x)$. 对任给向量 $x \in \Omega$ 取 $\xi \in \partial g(x)$, 若 $y \in \Omega$ 使得 $\eta \in \partial f(y)$, 满足 $h(y-x) + (\eta - \xi)^T (y-x) \geq 0$, 则 $E(y) < E(x)$.

证明: 由题设, 类似定理 1.1 的证明过程可以得到 $E(x) - E(y) > h(y-x) + (\eta - \xi)^T (y-x) \geq 0$. \square

在神经元模型中常用的输出方程为 $y_i = \varphi(u_i)$, 其中 u_i 为神经元的净输入, 而 y_i 为输出, 激活函数 $\varphi(\cdot)$ 为单调不减的连续函数. 这时定义 $\Phi(x) = \int_0^x \varphi(s) ds$, 因为 $\varphi(\cdot)$ 为连续函数, 所以 $\Phi(x)$ 为闭凸函数. 在本文中, 我们讨论的 $\Phi(x)$ 和 $\Phi^*(\xi)$ 都为正常闭凸函数.

引理 1.2. 设 $\varphi(x)$ 为连续单调非减的一元函数, $k = \sup_{x \neq y} \frac{\varphi(y) - \varphi(x)}{y-x} < +\infty$, 令 $\Phi(x) = \int_0^x \varphi(s) ds$, 则 $\frac{k}{2} x^2 - \Phi(x)$ 为凸函数.

证明: 由凸函数的定义直接验证即可. \square

下面研究共轭函数的性质, 这个关于凸共轭函数的性质是很有用的, 它是我们分析 CTHN 的离散化的基础.

定理 1.3. 设 $f(x)$ 为 $R^n \rightarrow (-\infty, +\infty)$ 的闭凸函数, $f^*(\xi)$ 为 $R^n \rightarrow (-\infty, +\infty]$ 的闭凸函数, 设 $ED(f^*) = \{\xi: f^*(\xi) < +\infty, \xi \in R^n\} = \Omega$ 为 R^n 的凸集, 矩阵 A 为正定对称矩阵. 那么, $f(x) - \frac{1}{2} x^T A x$ 为凸函数的充要条件是 $\frac{1}{2} \xi^T A^{-1} \xi - f^*(\xi)$ 为在 Ω 内的凸函数.

证明: 参见文献[6]中推论 1. \square

定理 1.1~定理 1.3 和引理 1.2 是我们第 2 节需要应用的理论基础. 这些结论在已有的有关凸函数的文献中未见报道, 所以我们给出了它们的详细说明.

2 连续时间 Hopfield 模型的离散化分析

Hopfield 网络(模型)包含一组神经元和一组相应的单位延迟, 构成了一个多回路反馈系统. 在标准的 CTHN 中, 神经元之间的连接权值矩阵 W 是对称的, 它的运行方程如下:

$$\frac{d}{dt} v_j(t) = -v_j(t) + \sum_{i=1}^N w_{ji} \varphi_i(v_i(t)), \quad j=1, 2, \dots, N \quad (5)$$

我们假定: 每个神经元有它自己的非线性单调不减的连续激活函数, 因此在式(5)中使用 $\varphi_i(\cdot)$. 基于此, 我们将每个神经元的不同偏置包含在激活函数之中, 从而在表达式(5)中没有明显地表示它们. 我们定义 n 元函数为

$\Phi(x) = \sum_{i=1}^n \int_0^{x_i} \varphi_i(s) ds$, 从而定义标准 CTHN 的能量(Lyapunov)函数定义为

$$E(x) = \Phi^*(x) - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N w_{ji} x_i x_j \quad (6)$$

可以证明结论: 能量函数 E 是连续 Hopfield 模型的 Lyapunov 函数. 根据 Lyapunov 稳定性理论, CTHN 模型是稳定的. 同时, Hopfield 网络的能量函数 E 是时间的单调减函数. 因此, Hopfield 网络是全局渐近稳定的; 吸引子固定点是能量函数的局部最小值.

2.1 广义的连续 Hopfield 网络模型

根据连续 Hopfield 网络模型的收敛性分析, 可以重新改写连续 Hopfield 网络模型描述方程(5)如下:

$$\frac{dv}{dt} = -\nabla_x E(x) \quad (7)$$

$$\mathbf{x} = \nabla_v \Phi(\mathbf{v}) \quad (8)$$

其中 $\Phi(\mathbf{v})$ 为连续可微凸函数, $E(\mathbf{x})$ 为连续可微的能量函数.

下面,我们分别讨论利用 Euler 方法和梯形方法对标准 CTHN 和广义 CTHN 进行离散化时离散时间步长的选择.

2.2 Euler方法对连续Hopfield网络模型的离散化分析

利用定理 1.1,我们来分析在连续 Hopfield 网络模型的离散化过程中,时间步长 h 的选择问题.通过 Euler 方法对连续时间 Hopfield 网络模型进行离散化,可以得到离散时间的 Hopfield 模型如下:

$$\mathbf{v}(t+1) = \mathbf{v}(t) - h \nabla_x E(\mathbf{x}(t)) \quad (9)$$

$$\mathbf{x}(t+1) = \nabla_v \Phi(\mathbf{v}(t+1)) \quad (10)$$

在 CTHN 网络的离散化过程中,需要讨论的就是离散时间步长 h 的选择问题.

定理 2.1. 对于由式 (7) 和式 (8) 描述的广义 CTHN 模型, 设存在二次函数 $H(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$, 使得 $\Phi^*(\mathbf{x}) - H(\mathbf{x}), 2 \cdot H(\mathbf{x}) - hE(\mathbf{x})$ 都为凸函数且其中一个为严格凸函数, 其中 \mathbf{A} 为对称矩阵. 进一步地, 我们假定 $\mathbf{x}(t)$ 为有界的, 则由式 (9) 和式 (10) 定义的 DTHN 收敛.

证明: 在定理 1.1 中, 我们取 $f(\mathbf{x}) = \Phi^*(\mathbf{x}), g(\mathbf{x}) = \Phi^*(\mathbf{x}) - hE(\mathbf{x})$, 由题设存在二次函数 $H(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$, 使得 $f(\mathbf{x}) - H(\mathbf{x}), g(\mathbf{x}) + H(\mathbf{x})$ 都为凸函数且其中一个为严格凸函数, 其中 \mathbf{A} 为对称矩阵, 根据定理 1.1 的结论, 则由式 (9) 和式 (10) 定义的 DTHN 模型的状态对于能量函数 $hE(\mathbf{x})$ 单调下降. 进一步地, 我们假定 $\mathbf{x}(t)$ 为有界的, 则由式 (9) 和式 (10) 定义的 DTHN 收敛. \square

定理 2.2. 对于由式 (5) 描述的标准 CTHN, 设神经元 i 的激活函数 $\varphi_i(\cdot)$ 为连续单调非减的一元函数, $k_i = \sup_{x \neq y} \frac{\varphi_i(y) - \varphi_i(x)}{y - x}$, 记 $\mathbf{K} = \text{diag}[k_1, k_2, \dots, k_n]$, 若 $(2-h) \cdot \mathbf{K}^{-1} + h \cdot \mathbf{W}$ 为正定矩阵, 进一步地, 我们假定 $\mathbf{x}(t)$ 为有界的, 则标准 CTHN 模型的 Euler 方法离散化模型收敛.

证明: 我们定义 n 元函数为 $\Phi(\mathbf{x}) = \sum_{i=1}^n \int_0^{x_i} \varphi_i(s) ds$, 根据矩阵 \mathbf{K} 的定义和引理 1.2, $\frac{1}{2} \mathbf{x}^T \mathbf{K} \mathbf{x} - \Phi(\mathbf{x})$ 为凸函数, 从而由定理 1.3 的结论可得 $\Phi^*(\mathbf{x}) - \frac{1}{2} \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}$ 为凸函数. 此时, $E(\mathbf{x}) = \Phi^*(\mathbf{x}) - \frac{1}{2} \mathbf{x}^T \mathbf{W} \mathbf{x}$. 所以, 在定理 1.1 中取 $f(\mathbf{x}) = \Phi^*(\mathbf{x})$, $H(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}, g(\mathbf{x}) = (1-h) \Phi^*(\mathbf{x}) + \frac{h}{2} \mathbf{x}^T \mathbf{W} \mathbf{x}$. 由题设 $(2-h) \cdot \mathbf{K}^{-1} + h \cdot \mathbf{W}$ 为正定矩阵, 此时 $g(\mathbf{x}) + H(\mathbf{x})$ 为凸函数. 由定理 2.1 的结果, 由式 (6) 描述的能量函数 $E(\mathbf{x})$ 单调下降. 由于假定 $\mathbf{x}(t)$ 为有界的, 我们可得定理结论. \square

在定理 2.1 和定理 2.2 中, 令 $h=1$, 则它们的条件就是文献 [5] 中离散时间连续状态 Hopfield 网络并行收敛的条件.

2.3 梯形方法对连续Hopfield网络模型的离散化分析

对由式 (7) 和式 (8) 描述的 CTHN 模型, 如果我们利用梯形方法进行离散化, 可以得到如下的离散时间 Hopfield 模型:

$$\mathbf{v}(t+1) = \mathbf{v}(t) - h(\nabla_x E(\mathbf{x}(t)) + \nabla_x E(\mathbf{x}(t+1)))/2 \quad (11)$$

$$\mathbf{x}(t+1) = \nabla_v \Phi(\mathbf{v}(t+1)) \quad (12)$$

在对于梯形方法进行离散化过程的研究中, 我们首先讨论离散时间步长 h 的选择问题.

定理 2.3. 如果存在二次函数 $H_i(\mathbf{x}) = \mathbf{x}^T \mathbf{A}_i \mathbf{x}, i=1, 2, 3$, 使得 $\Phi^*(\mathbf{x}) - H_1(\mathbf{x}), H_2(\mathbf{x}) - E(\mathbf{x}), E(\mathbf{x}) + H_3(\mathbf{x})$ 都为凸函数且其中一个为严格凸函数, 其中 \mathbf{A}_i 为对称矩阵. 选择离散化时间步长 h 使得函数 $4H_1(\mathbf{x}) - h \cdot (H_2(\mathbf{x}) + H_3(\mathbf{x}))$ 为凸函数, 进一步地, 我们假定 $\mathbf{x}(t)$ 为有界的, 则由式 (11) 和式 (12) 定义的梯形方法收敛.

证明: 在定理 1.1 中, 我们取 $f(\mathbf{x}) = \Phi^*(\mathbf{x}) + h \cdot E(\mathbf{x})/2, g(\mathbf{x}) = \Phi^*(\mathbf{x}) - h \cdot E(\mathbf{x})/2$, 由题设存在二次函数 $H(\mathbf{x}) = \mathbf{x}^T \mathbf{A}_1 \mathbf{x} - h \cdot H_2(\mathbf{x})/2$, 使得 $f(\mathbf{x}) - H(\mathbf{x}), g(\mathbf{x}) + H(\mathbf{x})$ 都为凸函数且其中一个为严格凸函数, 其中 \mathbf{A} 为对称矩阵. 根据定理 1.1 的结论, 则由式 (11) 和式 (12) 定义的 DTHN 模型的状态对于能量函数 $E(\mathbf{x})$ 单调下降. 同时, 我们假定 $\mathbf{x}(t)$ 为有界的, 则由式 (11) 和式 (12) 定义的 DTHN 收敛. \square

定理 2.4. 对于由式(5)描述的标准的 CTHN 模型,设神经元 i 的激活函数 $\varphi_i(\cdot)$ 为连续单调非减的一元函数, $k_i = \sup_{x \neq y} \frac{\varphi_i(y) - \varphi_i(x)}{y - x}$, 记 $\mathbf{K} = \text{diag}[k_1, k_2, \dots, k_n]$, 则对任意步长 h , 由式(11)和式(12)描述的 CTHN 模型的梯形方法离散化模型收敛.

证明:我们定义 n 元函数为 $\Phi(\mathbf{x}) = \sum_{i=1}^n \int_0^{x_i} \varphi_i(s) ds$, 根据矩阵 \mathbf{K} 的定义和引理 1.2, $\frac{1}{2} \mathbf{x}^T \mathbf{K} \mathbf{x} - \Phi(\mathbf{x})$ 为凸函数, 从而由定理 1.3 的结论可得 $\Phi^*(\mathbf{x}) - \frac{1}{2} \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}$ 为凸函数. 此时, $E(\mathbf{x}) = \Phi^*(\mathbf{x}) - \frac{1}{2} \mathbf{x}^T \mathbf{W} \mathbf{x}$. 在定理 1.1 中取 $f(\mathbf{x}) = (1 + \frac{h}{2}) \cdot \Phi^*(\mathbf{x}) - \frac{h}{4} \mathbf{x}^T \mathbf{W} \mathbf{x}$, $H(\mathbf{x}) = (\frac{1}{2} + \frac{h}{4}) \cdot \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x} - \frac{h}{4} \mathbf{x}^T \mathbf{W} \mathbf{x}$, $g(\mathbf{x}) = (1 - \frac{h}{2}) \cdot \Phi^*(\mathbf{x}) + \frac{h}{4} \mathbf{x}^T \mathbf{W} \mathbf{x}$. 显然, 对于任意的时间步长 h , $f(\mathbf{x}) - H(\mathbf{x})$ 为凸函数, $g(\mathbf{x}) + H(\mathbf{x})$ 为严格凸函数. 由定理 1.1 的结果, 我们可得结论. \square

2.4 梯形方法的迭代公式收敛条件

由式(11)和式(12)描述的梯形方法为隐式的, 可用迭代法求解. 梯形方法所用的迭代公式为

$$\mathbf{v}^0(t+1) = \mathbf{v}(t) - h \nabla_{\mathbf{x}} E(\mathbf{x}(t)) \quad (13)$$

$$\mathbf{x}^0(t+1) = \nabla_{\mathbf{v}} \Phi(\mathbf{v}^0(t+1)) \quad (14)$$

$$\mathbf{v}^{k+1}(t+1) = \mathbf{v}(t) - h (\nabla_{\mathbf{x}} E(\mathbf{x}(t)) + \nabla_{\mathbf{x}} E(\mathbf{x}^k(t+1))) / 2 \quad (15)$$

$$\mathbf{x}^{k+1}(t+1) = \nabla_{\mathbf{v}} \Phi(\mathbf{v}^{k+1}(t+1)), \quad (16)$$

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

这个迭代过程我们从理论上需要一直进行到收敛为止. 实际上, 由于我们仅需得到 $\mathbf{x}(t+1)$ 的一个估计, 从而式(13)所示的迭代过程只需迭代几次, 使得能量函数 $E(\mathbf{x})$ 在 $\mathbf{x}(t+1)$ 估计值 $\mathbf{x}^k(t+1)$ 处的取值 $E(\mathbf{x}^k(t+1))$ 小于 $E(\mathbf{x}(t))$ 即可. 这时我们需要讨论如何控制迭代次数的问题. 下面我们将利用定理 1.2 的结果进行深入分析, 得到初步的结果.

定理 2.5. 如果存在二次函数 $H_i(\mathbf{x}) = \mathbf{x}^T \mathbf{A}_i \mathbf{x} / 2, i=1, 2, 3$, 使得 $\Phi^*(\mathbf{x}) - H_1(\mathbf{x}), H_2(\mathbf{x}) - E(\mathbf{x}), E(\mathbf{x}) + H_3(\mathbf{x})$ 都为凸函数且其中一个为严格凸函数, 其中 \mathbf{A}_i 为对称矩阵.

① 对给定离散化时间步长 h , 令 $H(\mathbf{x}) = 2 \cdot H_1(\mathbf{x}) - h \cdot (H_2(\mathbf{x}) + H_3(\mathbf{x}))$, 在时刻 t , 迭代公式(13)~(16)的停止条件为

$$h \cdot (\mathbf{v}^{k+1}(t+1) - \mathbf{v}^k(t+1))^T (\mathbf{x}^k(t+1) - \mathbf{x}(t)) + H(\mathbf{x}^k(t+1) - \mathbf{x}(t)) \geq 0 \quad (17)$$

此时, 置 $\mathbf{x}(t+1) = \mathbf{x}^k(t+1)$ 我们有 $E(\mathbf{x}(t+1)) < E(\mathbf{x}(t))$.

② 若 $H_1(\mathbf{x}) - h \cdot H_2(\mathbf{x}) / 2$ 为正定二次函数, 迭代公式 (13)~(16)收敛. 此时当 $H(\mathbf{x})$ 为正定二次函数, 式(17)可在有限步内达到.

证明: ① 在定理 1.2 中, 我们取 $f(\mathbf{x}) = \Phi^*(\mathbf{x}) + h \cdot E(\mathbf{x}) / 2, g(\mathbf{x}) = \Phi^*(\mathbf{x}) - h \cdot E(\mathbf{x}) / 2$, 由题设存在二次函数 $G_1(\mathbf{x}) = H_1(\mathbf{x}) - h \cdot H_3(\mathbf{x}) / 2, G_2(\mathbf{x}) = h \cdot H_2(\mathbf{x}) / 2 - H_1(\mathbf{x})$ 使得 $f(\mathbf{x}) - G_1(\mathbf{x}), g(\mathbf{x}) + G_2(\mathbf{x})$ 都为凸函数且其中一个为严格凸函数. 这时 $H(\mathbf{x}) = G_1(\mathbf{x}) - G_2(\mathbf{x})$. 根据式(16)我们可得

$$\begin{aligned} \nabla_{\mathbf{x}} f(\mathbf{x}^k(t+1)) &= \nabla_{\mathbf{x}} \Phi^*(\mathbf{x}^k(t+1)) + h \cdot \nabla_{\mathbf{x}} E(\mathbf{x}^k(t+1)) / 2 = \mathbf{v}^k(t+1) + h \cdot \nabla_{\mathbf{x}} E(\mathbf{x}^k(t+1)) / 2 \\ &= \mathbf{v}(t) - h \cdot (\nabla_{\mathbf{x}} E(\mathbf{x}(t)) + \nabla_{\mathbf{x}} E(\mathbf{x}^{k-1}(t+1))) / 2 + h \cdot \nabla_{\mathbf{x}} E(\mathbf{x}^k(t+1)) / 2 \end{aligned} \quad (18)$$

$$\nabla_{\mathbf{x}} g(\mathbf{x}(t)) = \nabla_{\mathbf{x}} \Phi^*(\mathbf{x}(t)) - h \cdot \nabla_{\mathbf{x}} E(\mathbf{x}(t)) / 2 = \mathbf{v}(t) - h \cdot \nabla_{\mathbf{x}} E(\mathbf{x}(t)) / 2 \quad (19)$$

根据定理 1.2, 这时 $E(\mathbf{x}^k(t+1)) < E(\mathbf{x}(t))$ 成立的条件为

$$(\nabla_{\mathbf{x}} f(\mathbf{x}^k(t+1)) - \nabla_{\mathbf{x}} g(\mathbf{x}(t)))^T (\mathbf{x}^k(t+1) - \mathbf{x}(t)) + H(\mathbf{x}^k(t+1) - \mathbf{x}(t)) \geq 0 \quad (20)$$

将式(18)和式(19)代入式(20)可得

$$h \cdot (\nabla_{\mathbf{x}} E(\mathbf{x}^k(t+1)) - \nabla_{\mathbf{x}} E(\mathbf{x}^{k-1}(t+1)))^T (\mathbf{x}^k(t+1) - \mathbf{x}(t)) / 2 + H(\mathbf{x}^k(t+1) - \mathbf{x}(t)) \geq 0.$$

根据式(16)可得

$$\nabla_{\mathbf{x}} E(\mathbf{x}^k(t+1)) - \nabla_{\mathbf{x}} E(\mathbf{x}^{k-1}(t+1)) = \mathbf{v}^{k+1}(t+1) - \mathbf{v}^k(t+1),$$

由此可得式(17).

② 在定理 1.1 中, 我们取 $f(\mathbf{x}) = \Phi^*(\mathbf{x}), g(\mathbf{x}) = \mathbf{x}^T [\mathbf{v}(t) - h \nabla_{\mathbf{x}} E(\mathbf{x}(t))] / 2 - h \cdot E(\mathbf{x}) / 2$. 此时 $f(\mathbf{x}) - H_1(\mathbf{x})$ 为凸函数. 由假设,

当 $H_1(\mathbf{x})-h \cdot H_2(\mathbf{x})/2$ 为正定二次函数时, $g(\mathbf{x})+H_1(\mathbf{x})$ 也为凸函数, 注意在 $g(\mathbf{x})$ 的表达式中, $\mathbf{v}(t)-h \nabla_{\mathbf{x}} E(\mathbf{x}(t))/2$ 为常数向量. 由定理 1.1 的结论, 我们知道, 迭代公式(15)收敛. 设 $\mathbf{v}^k(t+1) \rightarrow \mathbf{v}(t+1)$, 则有

$$\mathbf{v}^{k+1}(t+1)-\mathbf{v}^k(t+1) \rightarrow 0, H(\mathbf{x}^k(t+1)-\mathbf{x}(t)) \rightarrow H(\mathbf{x}(t+1)-\mathbf{x}(t)),$$

由题设 $H(\mathbf{x}(t+1)-\mathbf{x}(t)) > 0$, 从而停止条件(17)在有限步内一定能达到. \square

定理 2.6. 对于由式(5)描述的标准的 CTHN 模型, 设神经元 i 的激活函数 $\varphi_i(\cdot)$ 为连续单调非减的一元函数,

$$k_i = \sup_{x \neq y} \frac{\varphi_i(y) - \varphi_i(x)}{y - x}, \text{ 记对角矩阵 } \mathbf{K} = \text{diag}[k_1, k_2, \dots, k_n].$$

① 对给定离散化时间步长 h , 令 $H(\mathbf{x}) = \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}$, 在时刻 t , 迭代公式(13)~(16)的停止条件为

$$h \cdot (\mathbf{v}^{k+1}(t+1) - \mathbf{v}^k(t+1))^T (\mathbf{x}^k(t+1) - \mathbf{x}(t)) + H(\mathbf{x}^k(t+1) - \mathbf{x}(t)) > 0.$$

此时, 置 $\mathbf{x}(t+1) = \mathbf{x}^k(t+1)$, 我们有 $E(\mathbf{x}(t+1)) < E(\mathbf{x}(t))$.

② 当 $(2-h) \cdot \mathbf{K}^{-1} + h \cdot \mathbf{W}$ 为正定矩阵时, 迭代公式(13)~(16)收敛. 此时, 式(17)可在有限步内达到.

证明: ① 我们定义 n 元函数为 $\Phi(\mathbf{x}) = \sum_{i=1}^n \int_0^{\varphi_i} \varphi_i(s) ds$, 根据矩阵 \mathbf{K} 的定义和引理 1.2, $\frac{1}{2} \mathbf{x}^T \mathbf{K} \mathbf{x} - \Phi(\mathbf{x})$ 为凸函数,

从而由定理 1.3 的结论可得 $\Phi^*(\mathbf{x}) - \frac{1}{2} \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}$ 为凸函数. 在定理 1.2 中取

$$\begin{aligned} f(\mathbf{x}) &= (1 + \frac{h}{2}) \cdot \Phi^*(\mathbf{x}) - \frac{h}{4} \mathbf{x}^T \mathbf{W} \mathbf{x}, H_1(\mathbf{x}) = (\frac{1}{2} + \frac{h}{4}) \cdot \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x} - \frac{h}{4} \mathbf{x}^T \mathbf{W} \mathbf{x}, \\ g(\mathbf{x}) &= (1 - \frac{h}{2}) \cdot \Phi^*(\mathbf{x}) + \frac{h}{4} \mathbf{x}^T \mathbf{W} \mathbf{x}, H_2(\mathbf{x}) = (\frac{1}{2} - \frac{h}{4}) \cdot \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x} - \frac{h}{4} \mathbf{x}^T \mathbf{W} \mathbf{x}. \end{aligned}$$

显然, $f(\mathbf{x})-H_1(\mathbf{x}), g(\mathbf{x})-H_2(\mathbf{x})$ 为凸函数. 此时 $H(\mathbf{x}) = \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x} = H_1(\mathbf{x}) - H_2(\mathbf{x})$. 根据定理 2.5 的推导, 我们有 $\nabla_{\mathbf{x}} f(\mathbf{x}^k(t+1)) - \nabla_{\mathbf{x}} g(\mathbf{x}(t)) = h(\nabla_{\mathbf{x}} E(\mathbf{x}^k(t+1)) - \nabla_{\mathbf{x}} E(\mathbf{x}^{k-1}(t+1)))/2 = h(\mathbf{v}^{k+1}(t+1) - \mathbf{v}^k(t+1))/2$, 此处 $E(\mathbf{x})$ 由式(6)定义. 由定理 1.2 的结果, 可得迭代公式(13)~(16)在时刻 t 的停止条件为

$$h \cdot (\mathbf{v}^{k+1}(t+1) - \mathbf{v}^k(t+1))^T (\mathbf{x}^k(t+1) - \mathbf{x}(t)) + H(\mathbf{x}^k(t+1) - \mathbf{x}(t)) > 0.$$

此时, 置 $\mathbf{x}(t+1) = \mathbf{x}^k(t+1)$, 我们有 $E(\mathbf{x}(t+1)) < E(\mathbf{x}(t))$.

② 在定理 1.1 中, 我们取 $f(\mathbf{x}) = \Phi^*(\mathbf{x}), g(\mathbf{x}) = \mathbf{x}^T [\mathbf{v}(t) - h \nabla_{\mathbf{x}} E(\mathbf{x}(t))/2] - h \cdot E(\mathbf{x})/2$, 其中 $E(\mathbf{x})$ 由式(2.2)定义. 此时 $f(\mathbf{x}) - H(\mathbf{x})/2$ 为凸函数. 由假设当 $(2-h) \cdot \mathbf{K}^{-1} + h \cdot \mathbf{W}$ 为正定矩阵时, $g(\mathbf{x}) + H(\mathbf{x})/2$ 也为凸函数, 注意, 在 $g(\mathbf{x})$ 的表达式中, $\mathbf{v}(t) - h \nabla_{\mathbf{x}} E(\mathbf{x}(t))/2$ 为常数向量. 由定理 1.1 的结论, 我们知道, 迭代公式(15)收敛. 设 $\mathbf{v}^k(t+1) \rightarrow \mathbf{v}(t+1)$, 则有 $\mathbf{v}^{k+1}(t+1) - \mathbf{v}^k(t+1) \rightarrow 0, H(\mathbf{x}^k(t+1) - \mathbf{x}(t)) \rightarrow H(\mathbf{x}(t+1) - \mathbf{x}(t)) > 0$, 从而停止条件(17)在有限步内能达到. \square

由定理 2.6 的结论可知, 如果每个神经元的激活函数除去一个平移外都相同, 这正是在利用 CTHN 进行组合优化问题时常见的情形, 这时定理 2.6 中 $H(\mathbf{x}) = K \cdot \|\mathbf{x}\|^2$, 其中 K 为神经激活函数导数的最大值. 迭代的停止条件(17)变为

$$K/h + (\mathbf{v}^{k+1}(t+1) - \mathbf{v}^k(t+1))^T (\mathbf{x}^k(t+1) - \mathbf{x}(t)) / \|\mathbf{x}^k(t+1) - \mathbf{x}(t)\|^2 > 0,$$

这实际上变得非常易于检验. 注意, 对于某些步长值 h 能够找到满足停止条件的 $\mathbf{x}^k(t+1)$ 的值, 但有可能迭代过程(13)~(16)并不收敛. 所以在定理 2.5 和定理 2.6 中 ① 给出的条件更易满足, 而 ② 给出的迭代收敛条件更强.

2.5 标准 CTHN 模型的梯形方法迭代公式的改进

由式(5)描述的标准 CTHN 模型对式(11)和式(12)描述的梯形方法, 得到公式如下:

$$\begin{aligned} \mathbf{v}(t+1) &= \mathbf{v}(t) - h[\mathbf{v}(t) - \mathbf{W} \mathbf{x}(t) + \mathbf{v}(t+1) - \mathbf{W} \mathbf{x}(t+1)]/2, \\ \mathbf{x}(t+1) &= \nabla_{\mathbf{v}} \mathcal{Q}(\mathbf{v}(t+1)). \end{aligned}$$

整理后可得

$$\begin{aligned} \mathbf{v}(t+1) &= (1-2p)\mathbf{v}(t) + p \cdot \mathbf{W}(\mathbf{x}(t) + \mathbf{x}(t+1)), \\ \mathbf{x}(t+1) &= \nabla_{\mathbf{v}} \mathcal{Q}(\mathbf{v}(t+1)), \end{aligned}$$

其中 $p = h/(2+h)$. 由于它为隐式公式, 所以我们一个很自然的想法就是使用如下的迭代公式以替代由式(13)~(16)给出的迭代过程

$$\mathbf{v}^0(t+1)=(1-2p)\mathbf{v}(t)+2p\cdot\mathbf{W}\mathbf{x}(t) \quad (21)$$

$$\mathbf{x}^0(t+1)=\nabla_{\mathbf{v}}\Phi(\mathbf{v}^0(t+1)) \quad (22)$$

$$\mathbf{v}^{k+1}(t+1)=(1-2p)\mathbf{v}(t)+p\cdot\mathbf{W}(\mathbf{x}(t)+\mathbf{x}^k(t+1)) \quad (23)$$

$$\mathbf{x}^{k+1}(t+1)=\nabla_{\mathbf{v}}\Phi(\mathbf{v}^{k+1}(t+1)) \quad (24)$$

其中 $p=h/(2+h)$, 显然 $0 < p < 1$. 这个迭代过程理论上需要一直进行到收敛为止. 类似于定理 2.6, 由式(21)~(24)描述的迭代过程如何控制迭代次数的问题, 我们有如下的结论.

定理 2.7. 对于由式(5)描述的标准的 CTHN 模型, 设神经元 i 的激活函数 $\varphi_i(\cdot)$ 为连续单调非减的一元函数,

$$k_i = \sup_{x \neq y} \frac{\varphi_i(y) - \varphi_i(x)}{y - x}, \text{ 记对角矩阵 } \mathbf{K} = \text{diag}[k_1, k_2, \dots, k_n].$$

① 对给定离散化时间步长 h , 令 $H(\mathbf{x}) = \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}$, 在时刻 t , 迭代公式(21)~(24)的停止条件为

$$h \cdot (\mathbf{x}^k(t+1) - \mathbf{x}(t))^T \mathbf{W} (\mathbf{x}^{k-1}(t+1) - \mathbf{x}^k(t+1)) / 2 + H(\mathbf{x}^k(t+1) - \mathbf{x}(t)) > 0 \quad (25)$$

此时, 置 $\mathbf{x}(t+1) = \mathbf{x}^k(t+1)$, 我们有 $E(\mathbf{x}(t+1)) < E(\mathbf{x}(t))$.

② 当 $(2+h) \cdot \mathbf{K}^{-1} + h \cdot \mathbf{W}$ 为正定矩阵时, 迭代公式(21)~(24)收敛. 此时, 式(25)可在有限步内达到.

证明: ① 我们定义 n 元函数为 $\Phi(\mathbf{x}) = \sum_{i=1}^n \int_0^{x_i} \varphi_i(s) ds$, 根据矩阵 \mathbf{K} 的定义和引理 1.2, $\frac{1}{2} \mathbf{x}^T \mathbf{K} \mathbf{x} - \Phi(\mathbf{x})$ 为凸函数,

从而由定理 1.3 的结论可得 $\Phi^*(\mathbf{x}) - \frac{1}{2} \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}$ 为凸函数. 此时, $E(\mathbf{x}) = \Phi^*(\mathbf{x}) - \frac{1}{2} \mathbf{x}^T \mathbf{W} \mathbf{x}$. 在定理 1.2 中取

$$\begin{aligned} f(\mathbf{x}) &= (1 + \frac{h}{2}) \cdot \Phi^*(\mathbf{x}) - \frac{h}{4} \mathbf{x}^T \mathbf{W} \mathbf{x}, H_1(\mathbf{x}) = (\frac{1}{2} + \frac{h}{4}) \cdot \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x} - \frac{h}{4} \mathbf{x}^T \mathbf{W} \mathbf{x}, \\ g(\mathbf{x}) &= (1 - \frac{h}{2}) \cdot \Phi^*(\mathbf{x}) + \frac{h}{4} \mathbf{x}^T \mathbf{W} \mathbf{x}, H_2(\mathbf{x}) = (\frac{1}{2} - \frac{h}{4}) \cdot \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x} - \frac{h}{4} \mathbf{x}^T \mathbf{W} \mathbf{x}. \end{aligned}$$

显然 $f(\mathbf{x}) - H_1(\mathbf{x}), g(\mathbf{x}) - H_2(\mathbf{x})$ 为凸函数. 此时 $H(\mathbf{x}) = \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x} = H_1(\mathbf{x}) - H_2(\mathbf{x})$. 根据式(23)并且注意到 $p = h/(2+h)$, 当 $k > 0$ 时我们有

$$\begin{aligned} \nabla_{\mathbf{x}} f(\mathbf{x}^k(t+1)) &= (1 + \frac{h}{2}) \cdot \mathbf{v}^k(t+1) - \frac{h}{2} \cdot \mathbf{W} \mathbf{x}^k(t+1) \\ &= (1 + \frac{h}{2}) [(1-2p)\mathbf{v}(t) + p\mathbf{W}(\mathbf{x}(t) + \mathbf{x}^{k-1}(t+1))] - \frac{h}{2} \cdot \mathbf{W} \mathbf{x}^k(t+1) \\ &= (1 - \frac{h}{2}) \mathbf{v}(t) + \frac{h}{2} \cdot \mathbf{W} \mathbf{x}(t) + \frac{h}{2} \cdot \mathbf{W} (\mathbf{x}^{k-1}(t+1) - \mathbf{x}^k(t+1)). \end{aligned}$$

显然, $\nabla_{\mathbf{x}} g(\mathbf{x}(t)) = (1-h/2)\mathbf{v}(t) + (h/2) \cdot \mathbf{W} \mathbf{x}(t)$, 从而 $\nabla_{\mathbf{x}} f(\mathbf{x}^k(t+1)) - \nabla_{\mathbf{x}} g(\mathbf{x}(t)) = \frac{h}{2} \cdot \mathbf{W} (\mathbf{x}^{k-1}(t+1) - \mathbf{x}^k(t+1))$.

由定理 1.2 的结果可得, 迭代公式(21)~(24)在时刻 t 的停止条件为

$$h \cdot (\mathbf{x}^k(t+1) - \mathbf{x}(t))^T \mathbf{W} (\mathbf{x}^{k-1}(t+1) - \mathbf{x}^k(t+1)) / 2 + H(\mathbf{x}^k(t+1) - \mathbf{x}(t)) > 0.$$

此时, 置 $\mathbf{x}(t+1) = \mathbf{x}^k(t+1)$ 我们有 $E(\mathbf{x}(t+1)) < E(\mathbf{x}(t))$.

② 在定理 1.1 中, 我们取 $f(\mathbf{x}) = \Phi^*(\mathbf{x}), g(\mathbf{x}) = \mathbf{x}^T [(1-2p)\mathbf{v}(t) + p\mathbf{W}\mathbf{x}(t)] + p \cdot \mathbf{x}^T \mathbf{W} \mathbf{x} / 2$. 此时 $f(\mathbf{x}) - H(\mathbf{x}) / 2$ 为凸函数. 由假设, 当 $(2+h) \cdot \mathbf{K}^{-1} + h \cdot \mathbf{W}$ 为正定矩阵时, $g(\mathbf{x}) + H(\mathbf{x}) / 2$ 也为凸函数. 注意, 在 $g(\mathbf{x})$ 的表达式中, $[(1-2p)\mathbf{v}(t) + p\mathbf{W}\mathbf{x}(t)]$ 为常数向量. 由定理 1.1 的结论, 迭代公式(23)收敛. 设 $\mathbf{v}^k(t+1) \rightarrow \mathbf{v}(t+1)$, 则有 $\mathbf{v}^{k+1}(t+1) - \mathbf{v}^k(t+1) \rightarrow 0$. 由于 $H(\mathbf{x})$ 为正定二次函数, $H(\mathbf{x}^k(t+1) - \mathbf{x}(t)) \rightarrow H(\mathbf{x}(t+1) - \mathbf{x}(t)) > 0$, 从而停止条件(25)在有限步内一定能达到. \square

比较定理 2.6 的结果②和定理 2.7 的结果②, 可见由式(23)描述的迭代法收敛比由式(15)描述的迭代法收敛的条件弱. 这样从一个侧面说明求解梯形方法的迭代公式(23)确实比式(15)有所改进, 从而更易满足.

3 数值实验

为了简单说明理论分析的结果, 我们对由式(5)描述的标准 CTHN 模型进行了数值实现. 神经元数目为 50 个, 每个神经元 i 的激活函数 $\varphi_i(\cdot)$ 都相同, 它们都为简单的 Sigmoid 函数, 即 $\varphi_i(s) = 1/(1+e^{-s})$. 神经元之间的连接权值和神经元的自反馈连接权值以及偏置都为 $[-5, +5]$ 之间均匀分布的随机数. 我们对不同的网络权值和偏置进

行了几次实验,数值实现的结果基本都相似.下面我们给出由其中一组连接权值和偏置给出的数值实现.由于梯形方法和 Euler 方法的数值实验结果基本相似,限于篇幅,我们仅仅给出利用 Euler 方法的实验结果.

对于给定的这组连接权值矩阵,我们首先计算出它的最大特征值 λ_{\max} 和最小特征值 λ_{\min} .由于神经元激活函数的选择,这时 $\mathbf{K}=0.25 \cdot \mathbf{I}$.当利用 Euler 方法实现时,根据定理 2.2 的结果,取 $h=8/(4-\lambda_{\min})$.在本实验中,由于 $\lambda_{\min}=-29.8160$,从而得到 $h=0.2366$.CTHN 状态的初值 $x(0)=(0.5,0.5,\dots,0.5)^T$.当 $h=0.2366$,图 1 和图 2 对应于 Euler 的神经元状态和由式(6)描述的网络能量函数二者随时间变化的图形.图 1 所示 $h=0.2366$ 时,Euler 方法得到的 CTHN 状态的神经元输出随时间变化的图形.其中的每一条线表示一个神经元的输出演化轨迹.从图中可以发现,大约 40 步已经收敛.图 2 是当 $h=0.2366$ 时,Euler 方法得到的能量演化图形.从图 2 中可以看出,能量在 15 步时变化已经明显缓慢,这是从图 1 可见大部分输出已经收敛.为了便于比较,在图 3 和图 4 中我们给出了当 $h=0.0473$ 利用 Euler 方法的数值实现.图 3 是当 $h=0.0473$ 时,Euler 方法得到的 CTHN 状态的神经元输出随时间变化的图形,其中每一条线表示一个神经元的输出演化轨迹.从图中可见,在 100 步时尚未收敛.图 4 是当 $h=0.0473$ 时,Euler 方法得到的能量演化图形,从图 4 中可以看出,能量在 40 步时变化已经明显缓慢,这时从图 3 可见大部分输出已经收敛.比较图 1 和图 3,我们发现较大的时间步有利于加速网络状态的收敛,由图 2 和图 4 给出的网络能量函数的演化图形也说明了这一点.图 5 给出了在 10 个不同状态初值条件下利用不同离散时间步长得到的最终网络能量函数图,由图可见,对不同初值,较大步长对应的最终能量较小.从图中还可以看到,对于所有的初值,较大的离散时间步长都不同程度地提高了优化性能.这里需要特别说明的是,这些结论对梯形方法也是成立的.

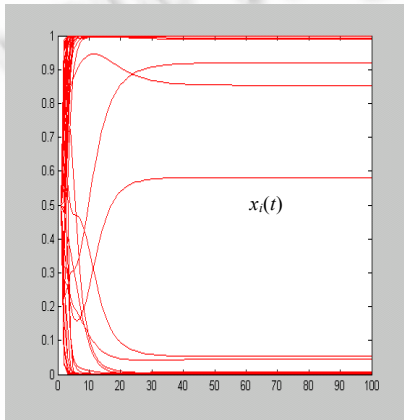


Fig.1
图 1

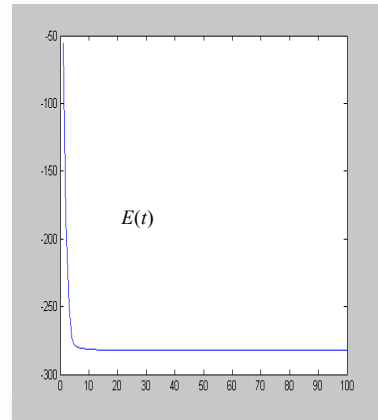


Fig.2
图 2

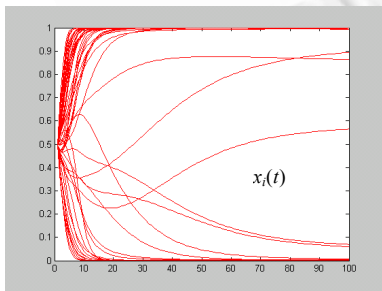


Fig.3
图 3

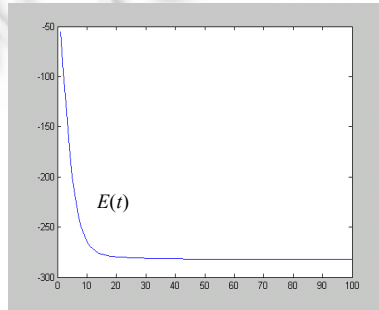


Fig.4
图 4

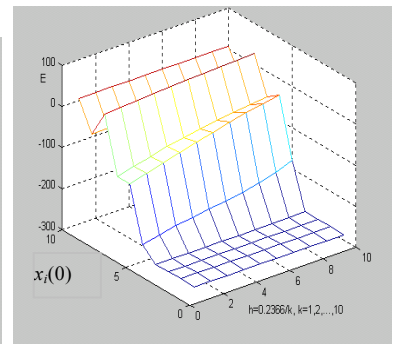


Fig.5
图 5

4 结 语

在本文中,我们利用凸函数次梯度性质,讨论 CTHN 进行最优化问题求解中利用数值模拟时离散化的时间步长的选择.推广连续 Hopfield 网络模型,引入了一个广义 CTHN 模型.对于由式(7)和式(8)描述的广义的 CTHN 和由式(5)描述的标准 CTHN,分别详细讨论了 Euler 方法和梯形方法离散化的时间步长选择条件.对于标准的 CTHN,由定理 2.4 得出了精确实现的梯形方法对于任意时间步长都是收敛的,这个结论出人意外.由于梯形方法为隐式方法,每一时间步我们需要进行迭代求解,这时我们研究了利用迭代求解时停止的条件.从我们的分析可以看出,标准 CTHN 的离散化方法的时间步长的选取一方面取决于 Hopfield 网络连接矩阵的最小特征值,同时还取决于网络每个神经元的激活函数的导数的大小,与神经元的偏置无关.这些结论,一方面有助于在进行 CTHN 离散化时时间步长的选择,另一方面也有助于我们选择神经元的激活函数.数值实验表明,较大的步长对应的最终能量较小.因此,我们寻找较大步长应该还是值得的.

References:

- [1] Haykin S. Neural Networks: A Comprehensive Foundation. Prentice Hall, 1999.
- [2] Atencia M. Numerical implementation of continuous Hopfield networks for optimization. In: Proc. of the ESANN 2001. Bruges, 2001. 359~364. <http://www.dice.ucl.ac.be/Proceedings/esann/esannpdf/es2001-40.pdf>
- [3] Wang LP. Discrete-Time convergence theory and updating rules for neural networks with energy functions. IEEE Trans. on NN, 1997,8:455~447.
- [4] Boyd S, Vandenberghe L. Convex Optimization. Cambridge University Press, 2004. <http://www.ee.ucla.edu/~vandenbe/cvxbook.html>
- [5] Ye SW, Zheng H, Wang WJ, Ma L, Shi ZZ. Stability analysis for discrete time Hopfield network. Journal of Software, 2003,14(5):930~935 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/930.htm>
- [6] Ye SW, Zheng HW, Wang WJ, Ma L, Shi ZZ. Analysis for the dynamics of discrete time Hopfield network. Journal of Computer Research and Development, 2003,40(10):1414~1418 (in Chinese with English abstract).

附中文参考文献:

- [5] 叶世伟,郑宏伟,王文杰,马琳,史忠植.离散时间 Hopfield 网络稳定性分析.软件学报,2003,14(1):930~935.<http://www.jos.org.cn/1000-9825/14/930.htm>
- [6] 叶世伟,郑宏伟,王文杰,马琳,史忠植.离散时间 Hopfield 网络动力系统分析.计算机研究与发展,2003,40(10):1414~1418.