非线性连续神经网络全局指数稳定的定量特征 [1]

谈 正 王利生

(西安交通大学电信学院信息工程研究所 西安 710049)

摘 要 引入一个特征函数用以定量刻画最优化计算神经网络的全局指数稳定性。利用该特征函数得到了一族范数下神经网络全局指数稳定的判定条件及下三角神经网络全局指数稳定的充分必要条件。所得结果推广了 Bouzerdoum， Killy， Sugawara 等人的结论。

关键词 神经网络，全局指数稳定性，最优化计算

中图号 TN-052

1 引 言

非线性连续神经网络的一个重要应用是最优化计算。与联想记忆神经网络有多个平衡点不同，对于最优化计算神经网络，理想情形是有且只有一个全局渐进的平衡点，而且应用中网络常设计为按指数收敛 [1,2] 。因此最优化计算神经网络稳定性分析的目的是判断在何种条件下，网络具有全局指数稳定性并分析网络收敛速度。网络收敛速度由网络收敛指数决定。这方面的一些研究结果见文献 [1−4] 。考虑最常见的一类网络模型:

$$\mathrm{d}x_i/\mathrm{d}t = -x_i + \sum_j w_{ij} \cdot g_j(x_j) + s_i \tag{1}$$

其中 $0 < \sup g_j'(z) \le 1$ 。 (1) 式的矩阵表示为 $\mathrm{d}x/\mathrm{d}t = -x + WG(x) + S$ 。文献 [3,4] 证明当 $\|W\|_p < 1$ 时， (1) 式全局指数稳定， $p \in \{1, 2, \cdots, \infty\}$ 。很自然存在如下问题：在任一矩阵范数 $\|\ \|$ 下， $\|W\| < 1$ 可否推出 (1) 式全局指数稳定?

(1) 式中 $g_i(x_i)$ 连续可微的条件在实际应用中不一定满足，如文献 [1, 2] 构造的最优化计算神经网络中， $g_i(x_i)$ 为单调非减的线性函数 $\begin{cases} a, & x_i > a \\ x_i, & a \ge x_i \ge b \\ b, & x_i < b \end{cases}$ 。因此有必要研究 $g_i(x_i)$ 条件减弱情形下 (1) 式的全局指数稳定性。比较合理且具有一般性的条件可以设为 $g_i(x_i)$ 单调非减且满足 Lipschitz 条件

$$|g_i(z_1) - g_i(z_2)| \le r_i \cdot |z_1 - z_2|, \quad \forall z_1, z_2 \in R \tag{2}$$

不失一般性，可设 $r_i = 1$ 。当 $g_i(x_i)$ 连续可微且 $g_i'(x_i) < \infty$ 时， $g_i(x_i)$ 满足 (2) 式；当 $g_i(x_i)$ 为单调非减的分段线性函数时， $g_i(x_i)$ 满足条件 (2) 式。本文将讨论网络模型 (1) 式满足条件 (2) 式时的全局稳定性。

构造 Lyapunov 函数是许多文献证明网络稳定性的常用方法，不同的范数需构造不同的 Lyapunov 函数，比较复杂。与此不同，本文通过计算 lub-Dahlquist 数 [5] 分析非线性连续神经网络的稳定性。 lub-Dahlquist 数小于零，则网络全局指数稳定。该方法可以给出一族矩阵范数下神经网络全局指数稳定的判定条件。对于 (1) 式，本文证明：若矩阵范数 $\|\ \|$ 对于任一对角阵 $D = \mathrm{diag}(d_1, d_2, \cdots, d_n)$ 有 $\|D\| = \max_i\{|d_i|\}$ ，则当 $\|W\| < 1$ 时， (1) 式全局指数稳定，收敛指数为 $\|W\| - 1$ 。如果 $g_i(x_i)$ 的条件减弱为 (2) 式且 $\|\ \|$ 为 R^n 上的单调向量

[1] 1998-11-25 收到， 1999-12-09 定稿

范数 [6]，则当从属矩阵范数 $\|W\| = \sup_{\|x\|=1}\|Wx\| < 1$ 时，(1) 式全局指数稳定，收敛指数为 $\|W\| - 1$。本文结论推广了文献 [2-4] 中结果，而且证明非常简单。

下三角神经网络是另一类重要的神经网络 [7]，本文给出了这类神经网络全局指数稳定的充要条件。

2 主要结果及讨论

令 $L(R^n)$ 表示 R^n 中 Lipschitz 连续算子全体，即 $\forall T \in L(R^n), T: R^n \to R^n$ 满足条件:

$$\|Tx - Ty\| \le K\|x - y\|, \qquad \forall x, y \in R^n$$

这里，K 为大于零的常数。易知，满足条件 (2) 式的任意非线性算子属于 $L(R^n)$。记 $L(T) = \sup\limits_{\substack{x \ne y \\ x,y\in R^n}} \frac{\|Tx-Ty\|}{\|x-y\|}$，$M(T) = \lim_{t\to\infty} \frac{L(I+tT)-1}{t}$，$I$ 为恒等算子，则称 $M[T]$ 为 T 的 lub-Dahlquist 数 [5]。$L(T)$ 和 $M[T]$ 与 R^n 上范数选取有关。对任意 T，$G \in L(R^n)$，$M[T+G] \le M[T] + M[G]$，而且 $L(T) \ge M[T]$。关于 $M[T]$ 有如下定理:

定理 1　设 $T \in L(R^n)$。若存在等价范数使得 $M[T] \le \alpha < 0$，则初值问题

$$x(t) = T(x(t)), \quad x(0) \in R^n$$

有唯一临界点 x^*，而且存在常数 $L > 0$，使得 $\|x(t) - x^*\|_2 \le L\|x(0) - x^*\|_2 e^{t\alpha}$。显然，由初值问题决定的 $x(t)$ 全局指数稳定且按指数 α 收敛于 x^*。

证明　设在等价范数 $\|\ \|^*$ 下 $M[T] \le \alpha < 0$，记 $l(T) = \inf_{x\ne y}\|Tx - Ty\|^*/\|x - y\|^*$，则 $l(T) \ge -M[T] > 0$，因而 T 满同胚，存在唯一 x^* 使得 $T(x^*) = 0$[5]。令 $x(t)$，$y(t)$ 为初值问题的两个不同解，则 $v(t) = x(t) - y(t)$ 满足 $\mathrm{d}\|v\|^*/\mathrm{d}t \le M[T]\|v\|^*$，$\|v\|^* \le \|v(0)\|^* e^{tM[T]} \le \|v(0)\|^* e^{t\alpha}$[5]。取 $y(t) = x^*$，则 $\|x(t) - x^*\|^* \le \|x(0) - x^*\|^* e^{t\alpha}$。据范数的等价性定义知，存在 $C_2 \ge C_1 > 0$，$\forall u \in R^n$，$C_1\|u\|_2 \le \|u\|^* \le C_2\|u\|_2$。因此 $\|x(t) - x^*\|_2 \le \frac{C_2}{C_1}\|x(0) - x^*\|_2 e^{t\alpha}$，$\lim_{t\to\infty}\|x(t) - x^*\| = 0$。　　证毕

注 1　求出某一范数下的收敛指数后，当选取不同等价范数时，只改变 L 不改变收敛指数。

注 2　初值问题的全局指数稳定性分析转化为寻找等价范数使得 $M[T] \le \alpha < 0$。收敛速度可通过计算全体等价范数下最小的 $M[T]$ 进行估计。以 $\alpha(T)$ 表示全体等价范数下最小的 $M[T]$，$\alpha(T)$ 可以刻画收敛速度，$\alpha(T)$ 越小，收敛越快。

算子发生扰动时是否改变算子的稳定性是人们非常关心的问题，下面定理表明算子发生小扰动，不改变稳定性。

定理 2　若 $T, G \in L(R^n), M[T] < 0$，则当 $L(T - G) < -M[T]$ 时，$M(T) < 0$。

证明　$M[G] = M[G - T + T] \le M[T] + M(T - G) \le M[T] + L(T - G) < 0$。　　证毕

网络模型 (1) 式可看作初值问题的一个特例。令 $T = -I + WG$，则 $T \in L(R^n)$。显然，只要证明 $M(T) < 0$ 则 (1) 式全局指数稳定。

定理 3　若矩阵范数 $\|\ \|$ 对于任一对角阵 $D = \mathrm{diag}(d_1, d_2, \cdots, d_n)$ 有 $\|D\| = \max_i\{|d_i|\}$，则当 $\|W\| < 1$ 时，非线性连续神经网络模型 (1) 式全局指数稳定且收敛指数为 $\|W\| - 1$。

证明　$\forall x, G(x) = \mathrm{diag}(g_1'(x_1), g_2'(x_2), \cdots, g_n'(x_n))$ 为对角阵。因为 $\sup_z g_i'(z) \le 1$，所以 $\|G'(x)\| \le 1$, $L(G) \le \sup_{x\in R^n}\{\|G'(x)\|\} \le 1$[5]。$M[T] = M[-I + WG] \le \mu(-I) + M[WG] \le -1 + L(WG) \le -1 + \|W\|L(G) \le -1 + \|W\|$。据定理 1，当 $\|W\| < 1$ 时，$M[T] \le \|W\| - 1 < 0$，(1) 式全局指数稳定，且收敛指数为 $\|w\| - 1$。　　证毕

据定理 2，当 W 发生小扰动 ΔW 时，只要 $\|\Delta W\| < 1 - \|W\|$，则小扰动不影响 (1) 式的全局指数稳定性。

对于任意 $x = (x_1, x_2, \cdots, x_n) \in R^n$，记 $|x| = (|x_1|, |x_2|, \cdots, |x_n|)$, R^n 上向量范数 $\|\ \|$ 称为是单调范数，若从 $|x| \le |y|$ 可得 $\|x\| \le \|y\|$。$\|\ \|$ 为单调范数当且仅当 $\|x\| = \||x|\|$[6]。当 R^n 上向量范数为单调范数时，其从属矩阵范数均满足 $\|D\| = \max_i\{|d_i|\}$。注意到 R^n 上向量范数 $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$，$p \in \{1, 2, 3, \cdots, \infty\}$ 均为单调范数，因此定理 3 推广了文献 [3,4] 中结果。

定理 4 在 (1) 式中，设 $g_i(x_i)$ 单调非减且满足 Lipschitz 条件：$|g_i(z_1) - g_2(z_2)| \le |z_1 - z_2|$, $\forall z_1, z_2 \in R$。若 $\|\ \|$ 为 R^n 上单调向量范数，而且矩阵范数 $\|W\| = \sup_{\|x\|=1}\{\|Wx\|\} < 1$，则 (1) 式全局指数稳定且收敛指数为 $\|W\| - 1$。

证明 $\forall x, y, G(x) - G(y) = (g_1(x_1) - g_1(y_1), g_2(x_2) - g_2(y_2), \cdots, g_n(x_n) - g_n(y_n))^T$，因为 $|g_i(x_i) - g_i(y_i)| \le |x_i - y_i|$，所以 $\|Gx - Gy\| \le \|x - y\|$，这推出 $L(G) \le 1$。令 $T = -I + WG$，$M[T] \le -1 + \|W\|$。据定理 1，当 $\|W\| < 1$ 时，$M[T] \le \|W\| - 1 < 0$，(1) 式全局指数稳定且收敛指数为 $\|W\| - 1$。 证毕

文献 [2] 在 $g_i(x_i)$ 为单调非减分段线性函数且 $\|W\|_2 < 1$ 的条件下证明了上述结论。

文献 [7] 讨论了下三角神经网络模型 $x = Ax + G(x)$ 的稳定性。其中，A 为下三角矩阵，$G(x) = \mathrm{diag}(g_1(x), g_2(x), \cdots, g_n(x))$，$g_i(x)$ 为连续的多元函数，$g_1(x) = C$，$g_i(x) = g_i(x_1, x_2, \cdots, x_{i-1})$，$C$ 为常数。文献 [7] 证明若 A 为稳定矩阵，则网络有唯一平衡点，而且全局渐近稳定。令 $T = A + G$，下面证明，当 $G(x)$ 为 Frechet 可导 (即，每一 $g_i(x)$ 可微) 的 Lipschitz 连续算子时，网络全局指数稳定的充要条件是 A 为稳定矩阵 (即，$\max_i\{a_{ii}\} < 0$)。

定理 5 如果 G：$R^n \to R^n$ 为 Frechet 可导的 Lipschitz 连续算子，则下三角神经网络模型中等价范数下最小的 lub-Dahlquist 数 $\alpha(T)$ 小于等于 $\max_i\{a_{ii}\}$。特别，下三角神经网络全局指数稳定的充要条件是下三角阵 A 为稳定矩阵。这时，收敛指数为 $\max_i\{a_{ii}\} + \varepsilon$, ε 可以为任意小整数。

证明 记 $\Lambda = \mathrm{diag}(a_{11}, \cdots, a_{nn})$，则 $B = A - \Lambda$ 为严格下三角矩阵，而且 $B + G \in L(R^n)$。由中值定理，$\forall x, y, \|(B+G)x - (B+G)y\| \le \sup_{z \in R^n}\|B + G'(z)\| \cdot \|x - y\|$，因此，在任一等价范数下 $L(B+G) \le \sup_{z \in R^n}\{\|B + G'(z)\|\}$。因为 $\{B + G'(z) : z \in R^n\}$ 为有界的严格下三角矩阵集合，所以等价范数下可同时逼近它们的谱半径 (谱半径均为 0)，即 $\forall \varepsilon > 0$, $\exists \delta > 0$, $P = \mathrm{diag}(1, \delta, \cdots, \delta^{n-1})$，使得 $\sup_{z \in R^n}\{\|P(B + G'(x))P^{-1}\|_1\} \le \varepsilon$。定义等价范数 $\|u\|_\varepsilon = \|Pu\|_1, \forall u \in R^n$。新范数 $\|\ \|_\varepsilon$ 下 $L((B+G)) \le \sup_{z \in R^n}\{\|P(B+G'(x))P^{-1}\|_1\} \le \varepsilon$[5]，而且 $M[\Lambda] = \max_i\{a_{ii}\}$。因而在 $\|\ \|_\varepsilon$ 下，$M[T] \le M[\Lambda] + M[G] \le \max_i\{a_{ii}\} + L((B+G)) \le \max_i\{a_{ii}\} + \varepsilon$。这推出 $\alpha(T) \le \max_i\{a_{ii}\} + \varepsilon$。让 $\varepsilon \to 0$，则有 $\alpha(T) \le \max_i\{a_{ii}\}$。

如果 A 为下三角稳定矩阵，则 A 的对角元 $a_{ii} < 0, i = 1, \cdots, n$。因而 $\forall \varepsilon > 0, \exists$ 等价范数 $\|\ \|_\varepsilon$，使得 $M[T] \le \max_i\{a_{ii}\} + \varepsilon < 0$。据定理 1，下三角神经网络全局指数稳定，而且收敛指数为 $\max_i\{a_{ii}\} + \varepsilon$。

反之，设下三角神经网络全局指数稳定，x^* 为稳定点。记 $y(t) = x(t) - x^*$，$F(y(t)) = T(y(t) + x^*)$，则下三角神经网络在 x^* 的指数稳定特性与初值问题：

$$y(t) = F(y(t)), \quad y(0) \in R^n$$

在 $y = 0$ 处的稳定性等价。因为 T 为 Frechet 可导算子，所以 F 也 Frechet 可导。注意到 $F(0) = 0$，由可导的定义知，$\lim_{y \to 0}\|Fy - F'(0)y\|/\|y\| = 0$。写 $F = F'(0) + (F - F'(0))$，

由文献 [8]， $F'(0)$ 的特征值实部全部小于零。因为 $F'(0)=T'(x^*)$, $T'(x^*)=A+G'(x^*)$，而且 $G'(x^*)$ 为严格下三角矩阵，所以 $F'(0)$ 的特征值为 A 的对角元 $\{a_{ii}: i=1,2,\cdots,n\}$。显然 $a_{ii}<0, i=1,2,\cdots,n$，即，A 为下三角稳定矩阵。　　证毕

参 考 文 献

[1] S. I. Sublrsanan, M. K. Sundareshan, Exponential stabiliy and a systematic synthesis of a neural network for quadratic minimization, Neural Network, 1991, 6(4), 599–613.

[2] A. Bouzerdoum, T. R. Pattison, Neural network for quadratic optimization with bound constraints, IEEE Trans. on Neural Networks, 1993, 4(2), 293–303.

[3] D. G. Killy, Stability in contractive nonlinear neural networks, IEEE Trans. on Biomedical Engineering, 1990, 37(4), 231–242.

[4] K. Sugawara, M. Harao, On the stability of equilibrium states of analogue neural networks, Transactions of the Institute of Electronics and Communication Engineers of Japan , 1983, J-66-A, 258–265.

[5] G. Soderlind, On nonlinear difference and differential equations, BIT, 1984, 24(4), 667–680.

[6] F. Bauer, J. Stoer, C. Witzgall, Absolute and monotonic norms, Numer. Math, 1961, 3, 257–264.

[7] G. Avitabile, M. Forti, *et al.*, On a class of nonsymmetrical neural networks with application to ADC, IEEE Trans. on CAS, 1991, 38(2), 202–208.

[8] G. Seifert, On a converse result for perron's theorem for asymptotic stability for nonlinear differentiable equation, Proc. of Amer. math. Soc., 1987, 99(4), 733–736.

THE QUANTITATIVE PROPERTIES OF EXPONENTIAL STABILITY OF NONLINEAR CONTINUOUS NEURAL NETWORK

Tan Zheng　　Wang Lisheng

(*School of Electronics and Information, Xi'an Jiaotong University, Xi'an* 710049 China)

Abstract　In the paper, a characteristic function is introduced and used to characterize quantitatively global exponential stability of nonlinear continuous neural network . Some sufficient conditions for the network to be exponentially stable under all monotone norms are presented, and the sufficient and necessary condition for lower-triangle neural network to be globally exponentially stable is obtained also.

Key words　Neural network, Exponential stability, Optimization computation

谈　正：　男，1939 年生，教授，博士生导师，主要研究方向为图像、图形处理，多媒体技术，虚拟现实等.

王利生：　男，1969 年生，工程师，博士生，主要研究方向为神经网络、数据可视化及图像图形处理.