Глобальная устойчивость нейронных сетей с недифференцируемыми активационными функциями

Кисиев Т.А., СПбГУ, Санкт-Петербург st085727@student.spbu.ru,
Мокаев Т.Н., СПбГУ, Санкт-Петербург t.mokaev@spbu.ru

Аннотация

В сообщении представлены результаты, полученные в рамках исследования устойчивости нейронных сетей с негладкими активациями. В процессе исследований были продолжены результаты авторов М. Форти и П. Нистри [9]. В частности, доказаны теоремы, гарантирующие устойчивость нейронной сети Хопфилда для более широкого класса функций активации нейронов.

Введение

Однослойные рекуррентные нейронные сети Хопфилда [5] часто используются для решения оптимизационных задач гладкого линейного и нелинейного программирования, когда целевая функция и ограничения являются непрерывно дифференцируемыми функциями.

Основой для формирования архитектуры сети является метод штрафов, включающий в себя градиентную систему энергетической функции, которая, в свою очередь, складывается из целевой функции и функций из ограничений конкретной оптимизационной задачи. Обусловлено это тем, что по энергетической функции можно задать активационные функции, связи между нейронами и определить входные данные для сети. Вычисление решения заключается в предоставлении неких начальных состояний (напряжений) для нейронов, далее нейронная сеть сойдется к устойчивому состоянию равновесия, которое соответствует минимуму энергетической функции, который, в свою очередь, соответствует минимуму функции затрат. Логично, что для таких систем важно наличие единственного глобально асимптотически устойчивого состояния равновесия.

Необходимым условием для сходимости метода штрафов за конечное время является использование недифференцируемых функций в ограничениях [2]. Эти функции отражены активационными функциями нейронов сети, поэтому возникает вопрос устойчивости сетей с негладкими активациями. В работе М. Форти и П. Нистри [9, см. теоремы 1 и 3] доказано наличие единственного глобально асимптотически устойчивого состояния равновесия у

сети Хопфилда с разрывной, ограниченной и монотонной функцией активации. Можно обобщить эти результаты для неограниченных и немонотонных разрывных функций, тем самым расширяя класс оптимизационных задач, которые можно эффективно и точно решать с помощью таких сетей.

Биологическая и математическая модели сети

Чтобы лучше понять концепции, использованные при создании и описании модели нейронной сети Хопфилда, кратко расскажем о её биологическом прародителе – человеческом мозге.

Как известно из биологии, мозг человека состоит примерно из 10^{12} нейронов. Они суть вычислительные единицы мозга. Между ними есть связующие отростки – аксоны (рис. 1). Через них нейроны передают друг другу нервные импульсы, которые принимаются дендритами. Сигналы передаются за счет разности зарядов на внешней и внутренней сторонах мембраны аксона, которая, в свою очередь определяется разностью в количестве ионов натрия и калия. Следовательно, активация нейрона ассоциируется с распространением импульса электрического напряжения вдоль цилиндрической мембраны аксона. Когда суммарный ток от других нейронов превышает определенный порог, нейрон генерирует собственный импульс, после которого на время становится невосприимчивым к внешним воздействиям (состояние рефрактерности).

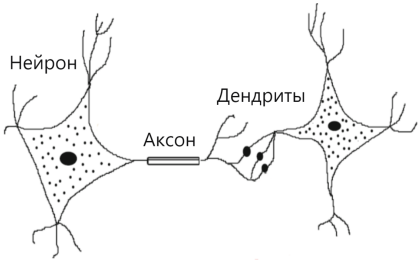

Рис. 1: Биологические нейроны и связи между ними.

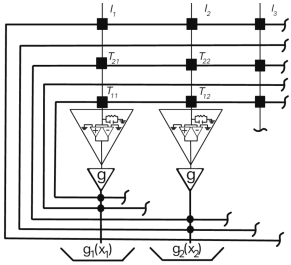

Рис. 2: Электронная цепь, моделирующая нейронную сеть.

Так как биологические нейроны передают друг другу электрические заряды, то аналоговая нейронная сеть, моделирующая абстрактную, может представлять собой электронную цепь Кирхгофа (рис. 2). Такой подход был предложен в 1952г. в работе Ходкина и Хаксли [6].

Цепь состоит из искусственных нейронов, которые, по сути, являются электрическими усилителями. Напряжения на них изменяются с течением

времени, отражая изменения состояний реальных нейронов. Связи между нейронами осуществляются включением в цепь резисторов проводимостью T_{ij} и сопротивлением $R_{ij} = \frac{1}{|T_{ij}|}$. Функция активации $g_i(x_i(t))$ характеризует выходное напряжение на i-ом нейроне относительно поступившего в него напряжения $x_i(t)$. Пауза в восприимчивости нейрона осуществляется за счет конденсатора емкости C_j. Внешние сигналы I_j зависят от постановки конкретной задачи, для решения которой адаптируется сеть.

В системе явно задается динамика, поэтому математически её можно описать как динамическую систему. Для задания векторного поля динамической системы нужны дифференциальные уравнения, вывести которые можно из физических законов. По первому закону Кирхгофа суммарный ток, поступающий в нейрон, равен суммарному току, исходящему из него:

$$\frac{x_j(t)}{R_j} + C_j \frac{\mathbf{d}x_j(t)}{\mathbf{d}t} = \sum_{i=1}^{N} T_{ij} g_i(x_i(t)) + I_j. \qquad (*)$$

Каждое уравнение типа (*) моделирует изменение с течением времени состояния нейрона в сети. Объединив эти уравнения в систему, получим систему дифференциальных уравнений (1), моделирующую нейронную сеть:

$$\dot{x}(t) = f(x(t)) = Bx(t) + Tg(x(t)) + I, \qquad (1)$$

где матрица $B = \mathrm{diag}(-b_1, ..., -b_n) \in M_n(\mathbb{R})$, $b_i > 0$, – включает в себя емкости и сопротивления, тем самым моделирует рефрактерное состояние нейрона, T – матрица связей нейронов.

Определение решений системы

Функция активации в правой части системы (1) в данной теории является разрывной функцией с конечным числом разрывов перового рода, причем для односторонних предельных значений функции в точках разрыва верно соотношение $g_i(\rho_k^+) > g_i(\rho_k^-)$. В дальнейшем такие функции будем считать принадлежащими классу $\mathcal{G}'$. Важно подметить, что в отличие от работы [9], тут от функции не требуется монотонность и ограниченность.

Для определения решений системы с разрывной правой частью избрана теория А. Филиппова [4].

Смысл определения *решений по Филиппову* в том, что касательный вектор к траектории решения, где он существует, должен лежать в замыкании выпуклой оболочки[0] предельных значений векторного поля $f(x(t))$ в сколь

угодно малых окрестностях траектории. Таким образом, система (1) запишется как система дифференциальных включений:

$$\dot{x}(t) \in \varphi\big(f(x(t))\big) = \overline{\mathrm{conv}}\{\lim f(x_i(t)) \mid x_i(t) \to x(t),\ x_i(t) \notin N_f\} = \quad (2)$$
$$= Bx(t) + T\ \overline{\mathrm{conv}}[g(x(t))] + I,$$

Важно, что мы исключаем множества меры нуль N_f, где функция имеет разрыв. Это позволяет определить траектории в точках, в которых само векторное поле, вообще говоря, не определено. Используя определение решений по Филлипову, можно также определить *состояние равновесия системы* (2) как определенное на полуинтервале $[0, +\infty)$ стационарное решение дифференциального включения:

$$0 \in \varphi(x^*) = Bx^* + T\ \overline{\mathrm{conv}}\big(g(x^*)\big) + I.$$

Инструменты для анализа устойчивости

Для анализа глобальной асимптотической устойчивости состояния равновесия системы (2) было решено использовать обобщенный подход Ляпунова [8], который включает в себя исследование знакоопределенности производной в силу системы специально подобранной функции Ляпунова типа Лурье-Постникова, которая также является энергетической функцией нейронной сети Хопфилда [5]. С помощью *обобщенного градиента по Кларку* [3] и *правила цепочки для регулярных функций* [10] можно модифицировать *принцип инвариантности ЛаСалля* [7, Следствие 4.2, стр. 129] и применить его для дифференцируемой почти всюду функции Ляпунова.

Устойчивость состояния равновесия нейронной сети

Условия C1.-C3., гарантируют для нейронной сети (1) существование решения (теорема 1.), а также существование (теорема 2.) единственного глобально асимптотически устойчивого состояния равновесия (теорема 3.).

[9] $\overline{\mathrm{conv}}(g(x))$ – замыкание выпуклой оболочки предельных значений в точке.

Вывод этих условий является одним из основных результатов аналитической работы, представленных в данном сообщении.

C1. Найдутся постоянные $a, b \geqslant 0$, такие, что: $\left\|\overline{\mathrm{conv}}[g(x)]\right\| \leqslant a\|x\| + b$.

C2. Найдется постоянная C_i, такая, что для $\forall \eta_{1,2i} \in \left\|\overline{\mathrm{conv}}[g_i(x_{1,2})]\right\|$:

$$\eta_{1i} - \eta_{2i} \leqslant C_i(x_1 - x_2), \quad \forall x_1, x_2 \in \mathbb{R}, \forall i \in 1:n.$$

C3. Найдется $\alpha = \mathrm{diag}\{\alpha_1, ..., \alpha_n\}$, $\alpha_i > 0$ такая, что матрица $\alpha T + T^{\mathrm{T}}\alpha$ отрицательно определена, и:

$$4C_i\alpha_i b_M^2 \| - B^{-1}T\|_2^2 < -\lambda_M b_m, \quad \forall i \in 1:n,$$

где $\lambda_M = \rho(\alpha T + T^{\mathrm{T}}\alpha)$, $b_m = \min\limits_{1 \leqslant i \leqslant n} b_i$, $b_M = \max\limits_{1 \leqslant i \leqslant n} b_i$.

Перед тем как исследовать глобальную асимптотическую устойчивость, нужно убедиться в том, что решения системы (1) существуют на промежутке $[0, +\infty)$.

Теорема 1. (о существовании и продолжении решений)

Пусть активационная функция нейронной сети принадлежит расширенному классу $g(x) \in \mathcal{G}'$, и выполнено условие C1., тогда интервал существования каждого решения системы (1) с начальными данными $x(0) = x_0$ равен $[0, +\infty)$.

Замечание. Теорема была доказана с помощью неравенства Гронуолла и теорем о существовании и продолжении решений дифференциальных включений А.Ф. Филиппова [4, теоремы 1 и 2, стр. 77-78].

Теорема 2. (о существовании состояния равновесия сети)

Пусть $g \in \mathcal{G}'$ и выполнены условия C1.-C3., тогда у нейронной сети (1) существует состояние равновесия.

Замечание. Доказательство данной теоремы было проведено путем применения результатов теоремы Лере-Шаудера о неподвижной точке [1, стр.90 т.3.2.8]. В статье М. Форти и П. Нистри [9] схожий результат доказывался с помощью теоремы Какутани о неподвижной точке [1, стр. 87, т. 3.2.3]. Однако использование этой теоремы более не даст желаемого результата, так как для её применения была необходима ограниченность функции активации. Поэтому было получено новое доказательство.

---

[1] $\|A\|_2 = \sqrt{\rho(A^{\mathrm{T}}A)}$, где ρ – спектральный радиус.

Теорема 3. (о глобальной асимптотической устойчивости единственного состояния равновесия сети)

Пусть функция активации $g \in \mathcal{G}'$ и выполнены условия С1.-С3., тогда для любого входного $I \in \mathbb{R}^n$ у сети (1) существует единственное глобально асимптотически устойчивое состояние равновесия.

Замечание. Теорема доказана с помощью принципа инвариантности Ла-Салля, модифицированного для регулярных функций Ляпунова [10]. Для определения производной в силу системы функции Ляпунова использовалось понятие обобщенного градиента по Кларку и теорема о правиле цепочки для регулярных функций.

Пример

Подкрепим аналитические результаты теоремы 3. примером. Рассмотрим систему дифференциальных уравнений:

$$\begin{cases} \frac{\mathbf{d}x_1}{\mathbf{d}t} = -x_1 - \frac{1}{4}f_1(x_1) - \frac{1}{8}f_2(x_2), \\ \frac{\mathbf{d}x_2}{\mathbf{d}t} = -x_2 + \frac{1}{8}f_1(x_1) - \frac{1}{4}f_2(x_2). \end{cases} \tag{1'}$$

Где в качестве функции активации возьмем:

$$f_i(x_i) = \tanh(x_i) - \operatorname{sign}(x_i)x_i + \operatorname{sign}(x_i), \quad i \in 1:2.$$

Она имеет разрыв первого рода и не является ограниченной и монотонной (рис. 3). Условия теоремы 3. выполнены, значит, в данном случае начало координат $(0, 0)$ есть глобально асимптотически устойчивое состояние равновесия, что видно по фазовому портрету (рис. 4).

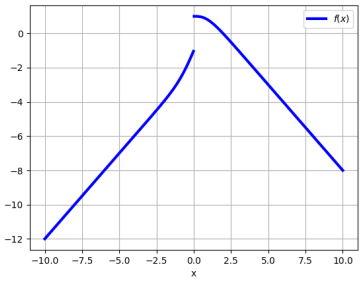

Рис. 3: График функции активации $f(x)$ для сети (1), пример

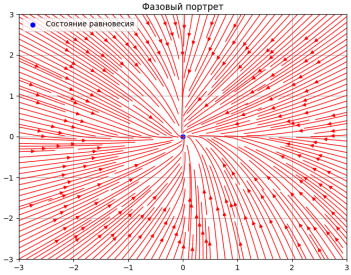

Рис. 4: Фазовый портрет системы ДУ из примера 1.

Заключение.

В процессе работы над материалами сообщения был избран подход для определения решений системы дифференциальных уравнений с разрывной правой частью. Также подобраны теоремы и инструменты для анализа устойчивости негладких систем, на базе которых были выведены и доказаны теоремы, являющиеся обобщением результатов из [9, теоремы 1 и 3] на случай, когда функция активации не является монотонной и ограниченной.

Список литературы

[1] J. P. Aubin. Set-valued analysis. 1990.

[2] D. P. Bertsekas. Necessary and sufficient conditions for a penalty method to be exact. *Mathematical Programming*, 9:87–99, 1975.

[3] F. H. Clarke. Optimization and nonsmooth analysis. 1983.

[4] A. F. Filippov. Differential equations with discontinuous righthand sides. In *Mathematics and Its Applications*, 1988.

[5] J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 81 10:3088–92, 1984.

[6] A. Hodgkin, A. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Bulletin of Mathematical Biology*, 52:25–71, 1952.

[7] H. K. Khalil. Nonlinear systems third edition. 1992.

[8] A. M. Lyapunov. The general problem of the stability of motion. 1892.

[9] M. Forti, P. Nistri. Global convergence of neural networks with discontinuous neuron activations. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 50(11):1421–1435, 2003.

[10] D. W. Shevitz, B. Paden. Lyapunov stability theory of nonsmooth systems. *Proceedings of 32nd IEEE Conference on Decision and Control*, pages 416–421 vol.1, 1993.