Восстановление выпуклых функций класса CPL нейронными сетями над ReLU-базисами

В. Г. Шишляков[1]

В работе рассматривается вопрос о классах функций, получаемых при использовании нейронных сетей над базисами с нелинейностями типа max. Сначала в работе рассматриваются некоторые свойства непрерывных кусочно-линейных функций и порождающих их классов эквивалентности. Затем, на базе этих свойств доказывается теорема о том, что нейронные сети, построенные над базисом, состоящим из всех линейных функций и максимумов от любого числа аргументов в качестве нелинейностей, могут в точности восстанавливать любую выпуклую непрерывную кусочно-линейную функцию.

Затем в работе рассматривается переход к RELU-базису, который является частным случаем базисов с нелинейностями типа max и доказывается теорема, аналогичная теореме, упомянутой выше. Также в работе обсуждается вопрос об оценке количества нейронов и слоев в полученных архитектурах.

Доказательство всех упомянутых теорем конструктивно, то есть в них явно строятся архитектуры нейронных сетей, удовлетворяющих вышеописанным свойствам.

Ключевые слова: Нейронные сети, архитектура, восстановление функций, выразимость функций, выпуклые функции, кусочно-линейные функции, функция ReLU, функция максимума.

1. Введение

Вопросы выразимости функций нейронными сетями начали изучаться в работе [1]. В указанном исследовании были впервые сформулированы и использованы некоторые определения, связывающие понятия нейронных сетей с классическими понятиями дискретной математики, такие, как схемы функциональных элементов и автоматные функции [2].

Базовой идеей в [1], которая используется и в данном исследовании, являлось рассмотрение нейронной сети с точки зрения схем функциональных элементов над определенным заранее фиксированным на-

[1] *Шишляков Владимир Геннадьевич* — аспирант каф. математической теории интеллектуальных систем мех.-мат. ф-та МГУ, e-mail: bolotmaks@yandex.ru.

Shishlyakov Vladimir Gennad'evich — graduate student, Lomonosov Moscow State University, Faculty of Mechanics and Mathematics, Chair of Mathematical Theory of Intellectual Systems.

бором функциональных элементов (базисом) и сравнение (с теоретико-множественной точки зрения) множества функций, соответствующих всем таким схемам, с классом функций, обладающих определенными свойствами.

Упомянутые выше схемы называются нейронными схемами и, по сути, являются синонимами термина нейронная сеть (нейронную сеть можно получить из нейронной схемы, группируя определенные комбинации функциональных элементов в нейроны, что подробно описано в работах [3] и [4]).

Также в [1] были впервые введены две числовые характеристики схем, именуемые нелинейной сложностью и глубиной, которые являются примерными оценками числа нейронов и числа слоев, соответственно, в нейронной сети, к которой можно преобразовать рассматриваемую нейронную схему.

В работе [1] было показано, что класс функций, реализуемых классическими нейронными сетями Мак-Каллока и Питтса (с функциями активации Хевисайда) [5] в точности совпадает с классом кусочно-параллельных функций. Однако, данный класс функций является достаточно узким классом, при помощи которого невозможно полноценно решать, например, задачи регрессии.

Есть и другая классическая функция активации, которая часто используется в современных архитектурах – это функция $RELU$. Возникает теоретический интерес – выяснить, какой класс функций выражается при помощи нейронных сетей с указанной функцией активации.

Учитывая, что базис для построения нейронных сетей с $RELU$ функциями активации содержит лишь кусочно-линейные непрерывные функции, а классы кусочно-линейных и непрерывных функций замкнуты по операции суперпозиции [1],[6], то возникает гипотеза, что, скорее всего, класс функций, представимых нейронными сетями над базисами с функциями активации $RELU$, совпадает с классом кусочно-линейных непрерывных функций.

Данная гипотеза частично была исследована в работе [6]. В указанной работе было доказано, что нейронные сети с $RELU$ функциями активации могут восстанавливать произвольную кусочно-линейную непрерывную функцию от одной и двух переменных. Но вопрос исследования функций с бóльшим числом переменных в работе [6] затронут не был.

Итоговый результат мною уже доказан, но будет публиковаться частями из-за большого объема материала. Данная статья открывает серию таких публикаций. В рассматриваемой работе приводится промежуточный, но очень важный результат, в котором показывается, что любая выпуклая (и, соответственно, вогнутая) кусочно-линейная непрерывная функция может быть представлена в виде нейронной сети с функциями активации типа $RELU$.

При доказательстве данного результата был исследован более общий базис, состоящий из линейных функций и единственного класса нелинейностей – $\max(x_1, ..., x_n)$, а основной результат был доказан уже, как следствие результата, полученного над общим базисом.

Также в работе оценивается нелинейная сложность и нелинейная глубина полученных нейронных схем.

2. Основные понятия

Для начала определим основные понятия, которые используются в данной статье.

Нейронные сети можно рассматривать с двух точек зрения. С одной стороны, на них можно смотреть, как на функции с большим количеством подбираемых параметров (весов), а с другой стороны – как на схемы, реализующие эти функции. Поэтому при рассмотрении искусственных нейронных сетей со схематической точки зрения естественным образом возникает понятие базиса нейронной сети. Введем понятие базиса и связанные с ним понятия, следуя работам [1], [2] и [3].

Определение 1. *Базисом будем называть некоторый набор функциональных элементов, где каждый функциональный элемент представляет из себя пару $(S, f(x_1, ..., x_n))$, в которой $f(x_1, ..., x_n) : \mathbb{R}^n \to \mathbb{R}$, а S – сопоставленный ей графический объект с n входными стрелками и одной выходной (кратко – входы и выход объекта S). Входам объекта S приписаны слева направо переменные $x_1, ..., x_n$, являющиеся аргументами функции f, выходу приписан выход функции f.*

В данной работе будем рассматривать два базиса (1) и (2), приведенных ниже.

$$B_1 = \{c, \gamma \cdot x, \sum_n (x_1, ..., x_n), \max(x_1, ..., x_n)\} \tag{1}$$

$$B_2 = \{c, \gamma \cdot x, \sum_n (x_1, ..., x_n), RELU(x)\} \tag{2}$$

В базисах (1) и (2) используются следующие классы функций:

- Сумматор — каждая функция данного класса суммирует определенное количество входных аргументов и обозначается $\sum_n (x_1, ..., x_n)$.

- Максимум — каждая функция данного класса выбирает максимум из n аргументов $x_1, ..., x_n$.
- Константа — каждая функция данного класса выдает константу (у данной функции нет входных аргументов, каждый раз, когда на схему приходят входные данные, константа выдает одинаковое, заранее определенное, значение). Предполагается, что для каждого действительного числа c есть свой элемент-константа c из данного класса.
- Усилитель (умножение на константу) — в данном классе каждая функция умножает пришедший на вход аргумент x на фиксированную константу γ. Предполагается, что для каждого действительного числа γ есть свой элемент-усилитель $\gamma \cdot x$ из данного класса.
- RELU — определяется следующим образом: $RELU(x) = \max(x, 0)$.

Графические изображения данных элементов приведены на рис. 1 (а, б, в, г, д).

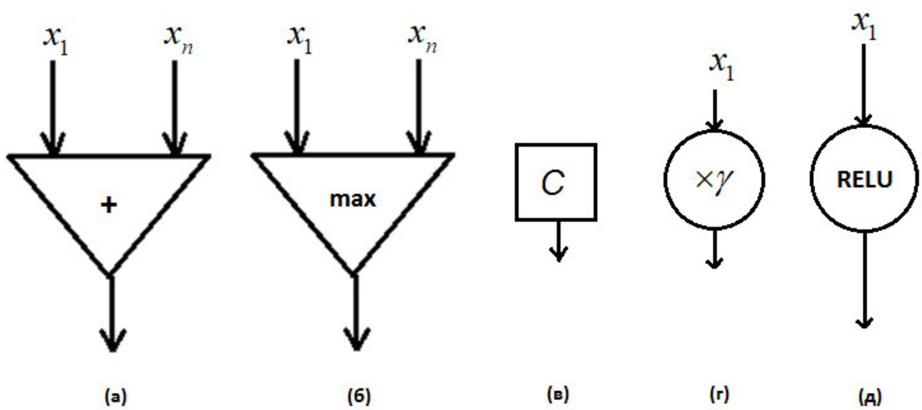

Рис. 1. Функциональные элементы рассматриваемых базисов

Функциональные элементы max и $RELU$ из базисов (1) и (2) будем называть нелинейными. Все остальные элементы данных базисов будем называть линейными.

Подробнее о том, как строятся схемы функциональных элементов из элементов любого из рассматриваемых в данной работе базисов, описано в [1]. Получаемые схемы называются нейронными схемами. Множество нейронных схем, получаемых из элементов некоторого базиса B, будем обозначать $[B]$.

Будем говорить, что функция $f : \mathbb{R}^n \to \mathbb{R}$ представима (выражается, восстанавливается) нейронной схемой над некоторым базисом B, если $\exists g \in [B] : \forall \overline{x} \in \mathbb{R}^n \; f(\overline{x}) = g(\overline{x})$.

Определение 2. *Путем в нейронной схеме назовем последовательность функциональных элементов* $(S_1, f_1), ..., (S_k, f_k)$ *схемы, таких, что:*

- *Один из входов элемента* (S_1, f_1) *является входом схемы*
- *Выход элемента* (S_k, f_k) *является выходом схемы*
- *Один из входов элемента* (S_i, f_i) *является выходом элемента* (S_{i-1}, f_{i-1}) *при* $i \in \{2, ..., k\}$

Определение 3. *Длиной пути называется число нелинейных элементов, содержащихся в нем.*

Определение 4. *Нелинейной глубиной схемы называется длина самого длинного пути в данной схеме. Нелинейной сложностью схемы называется число всех нелинейных элементов в схеме.*

Далее введем важнейшие понятия для данной работы — это понятия гиперплоскости, класса эквивалентности, кусочно-линейной функции, функции класса CPL, а также связанные с ними понятия.

Определение 5. *Пусть дано пространство* $\mathbb{R}^n$. *Гиперплоскостью (плоскостью размерности* $n-1$*) будем называть геометрическое место точек* $l = \{(x_1, ..., x_n) \in \mathbb{R}^n | a_1x_1 + ... + a_nx_n + c = 0\}$ *(где* $a_1, ..., a_n, c$ *– константы, причем* $a_1, ..., a_n$ *не равны нулю одновременно).*

Пусть $\overline{x} = (x_1, ..., x_n) \in \mathbb{R}^n$. *Тогда под значением* $l(\overline{x})$ *будем понимать значение* $a_1x_1 + ... + a_nx_n + c$. *Будем говорить, что точка* $\overline{x}$ *лежит на гиперплоскости* l, *если* $l(\overline{x}) = 0$. *Таким образом, гиперплоскость – это множество решений уравнения* $l(\overline{x}) = 0$.

Так как по уравнению вида $a_1x_1 + ... + a_nx_n = 0$ *можно в точности восстановить гиперплоскость, то в дальнейшем будем иногда определять гиперплоскость через ее уравнение, а само уравнение будем называть уравнением, определяющим гиперплоскость.*

Вместе с гиперплоскостью $l = \{(x_1, ..., x_n) \in \mathbb{R}^n | a_1x_1 + ... + a_nx_n + c = 0\}$ *естественным образом определяются два полупространства* $l^+ = \{(x_1, ..., x_n) \in \mathbb{R}^n | a_1x_1 + ... + a_nx_n + c > 0\}$ *и* $l^- = \{(x_1, ..., x_n) \in \mathbb{R}^n | a_1x_1 + ... + a_nx_n + c < 0\}$. *Очевидно, что* $l^+ \cup l^- \cup l = \mathbb{R}^n$ *для любой гиперплоскости* l.

Определение 6. *Гиперплоскостью размерности $n-k$ будем называть геометрическое место точек, полученное, как пересечение k гиперплоскостей размерности $n-1$.*

Определение 7. *Пусть $\overline{x} = (x_1,...,x_n) \in \mathbb{R}^n$, а $l_1,...,l_k$ – некоторые гиперплоскости, определяемые выражениями $l_i = \{(x_1,...,x_n) \in \mathbb{R}^n | a_{i1}x_1 + ... + a_{in}x_n + c_i = 0\}$, $i = 1,...,k$.*

Тогда все пространство $\mathbb{R}^n$ разбивается гиперплоскостями $l_1,...,l_k$ на классы эквивалентности. Рассмотрим функцию $\sigma(\overline{x}) = (sgn(a_{11}x_1 + ... + a_{1n}x_n + c_1),...,sgn(a_{k1}x_1 + ... + a_{kn}x_n + c_k))$, определяющую – в каком куске пространства лежит точка $\overline{x} \in \mathbb{R}^n$.

Вектор-функция $\sigma(\overline{x})$ называется сигнатурой вектора $\overline{x}$ [1]. Каждая ее компонента показывает расположение точки $\overline{x}$ относительно каждой из гиперплоскостей $l_1,...,,l_k$.

Определение 8. *Будем считать точки $\overline{x}, \overline{y} \in \mathbb{R}^n$ эквивалентными относительно гиперплоскостей $l_1,...,l_k$ и обозначать это $\overline{x} \sim \overline{y}$, если $\sigma(\overline{x}) = \sigma(\overline{y})$.*

Очевидно, что отношение $\overline{x} \sim \overline{y}$ является отношением эквивалентности [1]. Следовательно, гиперплоскости разбивают $\mathbb{R}^n$ на множество непересекающихся классов $R^1,...,R^s$, дающих в объединении все множество $\mathbb{R}^n$.

Определение 9. *Множество классов эквивалентности $\{R^1,...,R^s\}$, полученное разбиением пространства $\mathbb{R}^n$ некоторым набором гиперплоскостей $\{l_1,...,l_k\}$ при помощи введенного выше отношения эквивалентности $\overline{x} \sim \overline{y}$, будем называть разбиением пространства $\mathbb{R}^n$ или просто разбиением.*

Определение 10. *Сигнатурой класса $R^i \in \{R^1,...,R^s\}$ назовем значение $\sigma(R^i) = \sigma(\overline{x})$, где $\overline{x}$ – произвольный элемент класса R^i. Отметим, что определение корректно, так как по построению $\forall \overline{x}, \overline{y} \in R^i$ $\sigma(\overline{x}) = \sigma(\overline{y})$ [1].*

Определение 11. *Класс эквивалентности R^i назовем плоским, если $\exists l_j \in \{l_1,...,l_k\} : R^i \subset l_j$.*

Определение 12. *Класс эквивалентности R^i назовем объемным, если он не является плоским.*

Очевидно, что если R^i – непустой и объемный, то он имеет сигнатуру $\sigma = (\sigma_1,...,\sigma_k) : \sigma_j \neq 0$ $\forall j \in \{1,...,k\}$. Если же класс R^i – непустой и плоский, то он имеет сигнатуру $\sigma = (\sigma_1,...,\sigma_k)$, в которой $\exists j \in \{1,...,k\} : \sigma_j = 0$.

Определение 13. *Пусть дан объемный класс эквивалентности $R^i \subset \mathbb{R}^n$, имеющий сигнатуру $\sigma = (\sigma_1, ..., \sigma_k)$. Тогда контуром размерности $n - l, l \in \{1, ..., n\}$ данного класса назовем множество всех классов эквивалентности, имеющих сигнатуры σ', содержащиеся во множестве векторов, полученных из сигнатуры $\sigma = (\sigma_1, ..., \sigma_k)$ занулением всех возможных комбинаций из l компонент вектора σ.*

Контуром будем называть объединение контуров размерностей $n - 1, n - 2, ..., 0$. Контур класса R^i обозначим $\partial_{lin}\ R^i$. Контур класса R^i размерности k будем обозначать $\partial_{lin}^k\ R^i, k \in \{0, ..., n-1\}$.

Определение 14. *Пусть пространство $\mathbb{R}^n$ разбивается на классы $R^1, ..., R^s$ гиперплоскостями $l_1, ..., l_k$. Будем говорить, что $f(\overline{x})$ является кусочно-линейной [1] и обозначать это $f(\overline{x}) \in PL$, если $\forall R^i \in \{R^1, ..., R^s\}$ верно, что $\forall \overline{x} \in R^i\ f(\overline{x}) = \overline{b_i} \cdot \overline{x} + d_i$, где $\overline{b_i}, d_i$ – константы.*

Всякую функцию $f_{R^i}(\overline{x}) = \overline{b_i} \cdot \overline{x} + d_i$ будем называть линейной частью функции $f(\overline{x})$ на классе R^i.

Иногда, чтобы сразу ввести кусочно-линейную функцию и разбиение пространства $\mathbb{R}^n$, над которым она задана, будем говорить, что функция $f(\overline{x})$ - кусочно-линейная функция, заданная над (разбиением, классами) $\{R^1, ..., R^s\}$.

Определение 15. *CPL-функцией назовем всякую кусочно-линейную непрерывную функцию. Множество всех непрерывных кусочно-линейных функций обозначим CPL.*

Наконец, напомним несколько определений и свойств из топологии и выпуклого анализа, которые будут активно использоваться в дальнейшем.

Определение 16. *Отрезком, соединяющим две произвольные точки $\overline{x}, \overline{y} \in \mathbb{R}^n$ назовем множество $[\overline{x}; \overline{y}] = \{\overline{\xi} = (1 - \alpha) \cdot \overline{x} + \alpha \cdot \overline{y} \mid \alpha \in [0; 1]\}$. Аналогично вводится понятие интервала: $(\overline{x}; \overline{y}) = \{\overline{\xi} = (1 - \alpha) \cdot \overline{x} + \alpha \cdot \overline{y} \mid \alpha \in (0; 1)\}$.*

Определение 17. *Множество $M \subset \mathbb{R}^n$ называется выпуклым, если выполняется, что $\forall \overline{x}, \overline{y} \in M\ [\overline{x}; \overline{y}] \subset M$.*

Определение 18. *Функция $f : U \to \mathbb{R}$ называется выпуклой, если $U \subset \mathbb{R}^n$ – выпуклое множество и $\forall \overline{x_1}, \overline{x_2} \in U\ \forall \alpha \in (0; 1)$ выполняется, что $f((1 - \alpha) \cdot \overline{x_1} + \alpha \cdot \overline{x_2}) \leqslant (1 - \alpha) \cdot f(\overline{x_1}) + \alpha \cdot f(\overline{x_2})$.*

Определение 19. *ε-окрестностью точки $\overline{x} \in \mathbb{R}^n$ ($\varepsilon > 0$) будем называть шар $O_\varepsilon(\overline{x}) = \{\overline{x'} \in \mathbb{R}^n : ||\overline{x} - \overline{x'}|| < \varepsilon\}$. Иногда удобно рассуждать в терминах окрестностей вида $O(\overline{x})$. Под такой окрестностью понимается окрестность какого-то радиуса $\varepsilon > 0$, без уточнения конкретного* *значения ε. Обычно такая запись употребляется вместе с кванторами всеобщности или существования, когда конкретное значение ε становится избыточным.*

Точка $\overline{x} \in M \subset \mathbb{R}^n$ называется внутренней точкой множества M, если $\exists O(\overline{x}) \subset M$. Множество всех внутренних точек множества M называется внутренностью множества M и обозначается $\operatorname{int} M$.

Определение 20. *Множество $M \subset \mathbb{R}^n$ называется открытым, если $\forall \overline{x} \in M\ \exists O(\overline{x}) : O(\overline{x}) \subset M$. Или, другими словами, если любая точка данного множества – внутренняя.*

Определение 21. *Точкой прикосновения множества $M \subset \mathbb{R}^n$ называется всякая точка $\overline{x'} \in \mathbb{R}^n$ для которой выполняется, что $\exists \{\overline{x_n}\}_{n\in\mathbb{N}} \subset M : \lim\limits_{n\to+\infty} \overline{x_n} = \overline{x'}$.*

Замыканием множества M назовем множество $\overline{M}$ всех точек прикосновения множества M.

Определение 22. *Граничной точкой множества $M \subset \mathbb{R}^n$ называется всякая точка $\overline{x} \in \mathbb{R}^n$, для которой выполняется, что $\forall O(\overline{x})\ \exists \overline{x'}, \overline{x''} \in O(\overline{x}) : \overline{x'} \notin M \ \&\ \overline{x''} \in M$. Множество всех граничных точек множества M называется границей множества M и обозначается ∂M.*

Из курса общей топологии [7], [8] известно, что если множество M – открыто, то это эквивалентно тому, что $M = \operatorname{int} M$, а также, что $\overline{M} = \operatorname{int} M \cup \partial M$ и для любых множеств $A_1, ..., A_n$ верно, что $\overline{A_1 \cap ... \cap A_n} \subset \overline{A_1} \cap ... \cap \overline{A_n}$.

3. Основные результаты

Далее представлены основные результаты данной работы. Сначала представлены несколько вспомогательных утверждений (технических лемм), в которых изучаются различные свойства гиперплоскостей, классов эквивалентности, линейных и кусочно-линейных функций, необходимых для доказательства основного результата. Затем доказывается основное утверждение о том, что любую выпуклую CPL-функцию можно выразить нейронной схемой над базисом B_1. Наконец, доказывается следствие, в котором основное утверждение переносится на базис B_2.

Лемма 1. *Пусть $f(\overline{x}) = f(x_1, ..., x_n) = a_1 \cdot x_1 + ... + a_n \cdot x_n + d$ – линейная функция и имеются две точки $\overline{x}, \overline{y} \in \mathbb{R}^n$. Тогда выполняется $\forall \alpha \in [0; 1]\ f((1-\alpha) \cdot \overline{x} + \alpha \cdot \overline{y}) = (1-\alpha) \cdot f(\overline{x}) + \alpha \cdot f(\overline{y})$.*

Доказательство. Пусть $\overline{x} = (x_1, ..., x_n)$ и $\overline{y} = (y_1, ..., y_n)$. Рассмотрим выражение $(1-\alpha) \cdot f(\overline{x}) + \alpha \cdot f(\overline{y})$ и распишем f по определению. Получим

$(1-\alpha)\cdot(a_1\cdot x_1+...+a_n\cdot x_n+d)+\alpha\cdot(a_1\cdot y_1+...+a_n\cdot y_n+d)$. Но тогда, раскрывая в последнем выражении скобки и приводя подобные, получаем выражение (3).

$$((1-\alpha)\cdot a_1\cdot x_1+\alpha\cdot a_1\cdot y_1)+...\\ +((1-\alpha)\cdot a_n\cdot x_n+\alpha\cdot a_n\cdot y_n)+((1-\alpha)\cdot d+\alpha\cdot d) \quad (3)$$

Наконец, учитывая, что $(1-\alpha)\cdot d+\alpha\cdot d=d$, выражение (3) можно записать в виде $a_1\cdot((1-\alpha)\cdot x_1+\alpha\cdot y_1)+...+a_n\cdot((1-\alpha)\cdot x_n+\alpha\cdot y_n)+d$. Последнее выражение по определению равно значению $f((1-\alpha)\cdot\overline{x}+\alpha\cdot\overline{y})$, что и доказывает данное утверждение. □

Лемма 2. *Пусть l – гиперплоскость, определяемая уравнением $a_1\cdot x_1+...+a_n\cdot x_n+d=0$, а также l^+ и l^- – соответствующие полупространства, определяемые данной гиперплоскостью. Тогда верны следующие утверждения:*

1) $l=\partial l^+, l=\partial l^-$

2) l^+, l^- *– открытые множества*

3) $\overline{l^+}=l^+\cup l$

4) $l^+, l^-, \overline{l^+}, \overline{l^-}, l$ *– выпуклые множества*

Доказательство. 1. Покажем, что $l=\partial l^+$.

Докажем включение $l\subset\partial l^+$.

Возьмем $\forall\overline{x}=(x_1,...,x_n)\in l$. Тогда для взятого $\overline{x}$, по определению l, выполняется равенство $a_1\cdot x_1+...+a_n\cdot x_n+d=0$. Затем возьмем $\forall\varepsilon>0$ и рассмотрим $O_\varepsilon(\overline{x})$. Так как, по определению, $l^+=\{(x_1,...,x_n)|a_1\cdot x_1+...+a_n\cdot x_n+d>0\}$, то $\overline{x}\notin l^+$.

Теперь обозначим $\overline{a}=(a_1,...,a_n)\neq\overline{0}$, где $\overline{0}=(0,...,0)\in\mathbb{R}^n$ и определим точку $\overline{x'}=\overline{x}+\frac{\varepsilon}{2\cdot||\overline{a}||}\cdot\overline{a}$, где $||\overline{a}||=\sqrt{a_1^2+...+a_n^2}$.

Отметим, что $\overline{x'}\in l^+$. Действительно, рассмотрим значение $l(\overline{x}+\frac{\varepsilon}{2\cdot||\overline{a}||}\cdot\overline{a})$. По определению функции $l(\overline{x})$, данное значение представимо в следующем виде:

$$a_1\cdot(x_1+\frac{\varepsilon}{2\cdot||\overline{a}||}\cdot a_1)+...+a_n\cdot(x_n+\frac{\varepsilon}{2\cdot||\overline{a}||}\cdot a_n)+d \quad (4)$$

Раскрывая в (4) скобки и группируя слагаемые, получаем выражение (5).

$$a_1\cdot x_1+...+a_n\cdot x_n+d+a_1^2\cdot\frac{\varepsilon}{2\cdot||\overline{a}||}+...+a_n^2\cdot\frac{\varepsilon}{2\cdot||\overline{a}||} \quad (5)$$

Заменяя в (5) значение $a_1 \cdot x_1 + ... + a_n \cdot x_n + d$ на $l(\overline{x})$ и $a_1^2 + ... + a_n^2$ на $||\overline{a}||^2$, получаем выражение (6).

$$l(\overline{x}) + ||\overline{a}||^2 \frac{\varepsilon}{2 \cdot ||\overline{a}||} \tag{6}$$

Причем $l(\overline{x}) = 0$, так как по условию $\overline{x} \in l$. Поэтому полученное выражение (6) можно преобразовать следующим образом:

$$l(\overline{x}) + ||\overline{a}||^2 \frac{\varepsilon}{2 \cdot ||\overline{a}||} = 0 + ||\overline{a}||^2 \frac{\varepsilon}{2 \cdot ||\overline{a}||} = \frac{\varepsilon \cdot ||\overline{a}||}{2} \tag{7}$$

Но $\frac{\varepsilon \cdot ||\overline{a}||}{2} > 0$, откуда, учитывая равенство выражений $l(\overline{x} + \frac{\varepsilon}{2 \cdot ||\overline{a}||} \cdot \overline{a})$ и (4), (5), (6), (7), получаем, что $l(\overline{x} + \frac{\varepsilon}{2 \cdot ||\overline{a}||} \cdot \overline{a}) > 0$. А последнее означает, по определению, что $\overline{x'} \in l^+$.

При этом, из определения $\overline{x'}$ следует, что $||\overline{x'} - \overline{x}|| = \|\overline{x} + \frac{\varepsilon \cdot \overline{a}}{2 \cdot ||\overline{a}||} - \overline{x}\| = \|\frac{\varepsilon \cdot \overline{a}}{2 \cdot ||\overline{a}||}\| = \frac{\varepsilon}{2 \cdot ||\overline{a}||} \cdot \|\overline{a}\| = \frac{\varepsilon}{2} < \varepsilon$, откуда следует, что $\overline{x'} \in O_\varepsilon(\overline{x})$.

Таким образом, имеем, что $\exists \overline{x'} \in O_\varepsilon(\overline{x}) \cap l^+$ и $\exists \overline{x} \in O_\varepsilon(\overline{x}) \cap (\mathbb{R}^n \setminus l^+)$. В силу произвольности выбора $\varepsilon > 0$ получаем, что в любой окрестности точки $\overline{x} \in l$ есть как точки из l^+, так и точки не из l^+. Последнее означает, по определению, что $\overline{x} \in \partial l^+$. Откуда, в силу выбора $\overline{x}$, немедленно вытекает доказываемое включение $l \subset \partial l^+$.

Осталось показать обратное включение $\partial l^+ \subset l$.

Возьмем $\forall \overline{x} \in \partial l^+$. Тогда по определению ∂l^+ и учитывая, что выражения $\overline{x'} \in l^+, \overline{x''} \notin l^+$ эквивалентны выражениям $l(\overline{x'}) > 0, l(\overline{x''}) \leqslant 0$, соответственно, верно следующее:

$$\forall \varepsilon > 0 \; \exists \overline{x'}, \overline{x''} \in O_\varepsilon(\overline{x}) : \; l(\overline{x'}) > 0, l(\overline{x''}) \leqslant 0 \tag{8}$$

Далее рассмотрим последовательность $\varepsilon_n = \frac{1}{n}, n = 1, 2, ...$ Тогда из (8) следуют выражения (9) и (10).

$$\exists \{\overline{x'_n}\}_{n \in \mathbb{N}} : \overline{x'_n} \in O_{\frac{1}{n}}(\overline{x}) \& l(\overline{x'_n}) > 0 \tag{9}$$

$$\exists \{\overline{x''_n}\}_{n \in \mathbb{N}} : \overline{x''_n} \in O_{\frac{1}{n}}(\overline{x}) \& l(\overline{x''_n}) \leqslant 0 \tag{10}$$

Из (9) и (10) с одной стороны следует, что $\{\overline{x'_n}\} \underset{n \to \infty}{\to} \overline{x}$ и $\{\overline{x''_n}\} \underset{n \to \infty}{\to} \overline{x}$, а с другой стороны верно, что $\forall n \in \mathbb{N} \;\; l(\overline{x'_n}) > 0$ и $l(\overline{x''_n}) \leqslant 0$.

Но тогда по теореме о предельном переходе в неравенстве получаем, что $l(\overline{x}) \geqslant 0$ и $l(\overline{x}) \leqslant 0$ одновременно. Следовательно, $l(\overline{x}) = 0$.

Итак, взяв $\forall \overline{x} \in \partial l^+$, получили, что $l(\overline{x}) = 0$, то есть, что $\overline{x} \in l$. Следовательно, $\partial l^+ \subset l$.

Из доказанных включений вытекает, что $l = \partial l^+$. Полностью аналогично доказывается, что $l = \partial l^-$.

2. Покажем, что l^+ открыто. Рассмотрим $\forall \overline{x} \in l^+$. Последнее означает, что $l(\overline{x}) > 0$. Возьмем также $\varepsilon = \frac{l(\overline{x})}{2 \cdot n^2 \cdot ||\overline{a}||} > 0$, где $\overline{a} = (a_1, ..., a_n)$ и $||\overline{a}||$ имеют тот же смысл, что и при доказательстве свойства 1.

Но тогда, воспользовавшись неравенством Минковского, а затем неравенством Гёльдера, получаем, что $\forall \overline{x'} \in O_\varepsilon(\overline{x})$ выражение $|l(\overline{x'}) - l(\overline{x})|$ имеет следующую оценку сверху:

$$\begin{gathered}|l(\overline{x'}) - l(\overline{x})| = |a_1 \cdot (x'_1 - x_1) + ... + a_n \cdot (x'_n - x_n)| \leqslant \\ |a_1| \cdot |x'_1 - x_1| + ... + |a_n| \cdot |x'_n - x_n| \leqslant \\ |a_1| \cdot (|x'_1 - x_1| + ... + |x'_n - x_n|) + ... + |a_n| \cdot (|x'_1 - x_1| + ... + |x'_n - x_n|) = \\ (|a_1| + ... + |a_n|) \cdot (|x'_1 - x_1| + ... + |x'_n - x_n|) \leqslant \\ n \cdot ||\overline{a}|| \cdot n \cdot ||\overline{x'} - \overline{x}|| = n^2 \cdot ||\overline{a}|| \cdot ||\overline{x'} - \overline{x}|| < n^2 \cdot ||\overline{a}|| \cdot \varepsilon = \\ n^2 \cdot ||\overline{a}|| \cdot \frac{l(\overline{x})}{2 \cdot n^2 \cdot ||\overline{a}||} = \frac{l(\overline{x})}{2} \end{gathered} \quad (11)$$

Раскрывая в полученной оценке (11) модуль $|l(\overline{x'}) - l(\overline{x})|$, получаем следующее выражение:

$$-\frac{l(\overline{x})}{2} < l(\overline{x'}) - l(\overline{x}) < \frac{l(\overline{x})}{2} \quad (12)$$

Но тогда, используя нижнюю оценку в (12) и тот факт, что $l(\overline{x}) > 0$, получаем следующее:

$$l(\overline{x'}) = l(\overline{x'}) + l(\overline{x}) - l(\overline{x}) = l(\overline{x}) + (l(\overline{x'}) - l(\overline{x})) > l(\overline{x}) - \frac{l(\overline{x})}{2} = \frac{l(\overline{x})}{2} > 0 \quad (13)$$

Из (13) следует, что $l(\overline{x'}) > 0$, а последнее означает, что $\overline{x'} \in l^+$.

Таким образом, показано, что для любой точки $\overline{x} \in l^+$ найдется некоторая окрестность $O_\varepsilon(\overline{x})$ такая, что $\forall \overline{x'} \in O_\varepsilon(\overline{x})$ верно $\overline{x'} \in l^+$. Из сказанного немедленно вытекает открытость множества l^+.

Аналогично доказывается, что l^- – открыто.

3. Известным фактом [7] является следующее выражение:

$$\overline{l^+} = \text{int } l^+ \cup \partial l^+ \quad (14)$$

Учитывая, что, по пункту 1 данной леммы, $\partial l^+ = l$, а также еще один общеизвестный факт [7], что $\text{int } l^+ = l^+$ в силу открытости множества l^+, из (14) получаем доказываемое.

4. Покажем, что l^+ – выпукло. Рассмотрим $\forall \overline{x} = (x_1, ..., x_n), \overline{y} = (y_1, ..., y_n) \in l^+$. Тогда, по определению, $l(\overline{x}) = a_1 \cdot x_1 + ... + a_n \cdot x_n + d > 0$, $l(\overline{y}) = a_1 \cdot y_1 + ... + a_n \cdot y_n + d > 0$.

Определим для произвольного $\alpha \in [0;1]$ величину $\overline{\xi} = (1-\alpha) \cdot (x_1, ..., x_n) + \alpha \cdot (y_1, ..., y_n) = ((1-\alpha) \cdot x_1 + \alpha \cdot y_1, ..., (1-\alpha) \cdot x_n + \alpha \cdot y_n)$.

Тогда из леммы 1 следует, что $l(\overline{\xi}) = (1-\alpha) \cdot l(\overline{x}) + \alpha \cdot l(\overline{y})$.

Поэтому, если $\alpha \in (0;1)$, то $l(\overline{\xi}) > 0$, в силу того, что $\alpha > 0, (1-\alpha) > 0$, а также $l(\overline{x}) > 0$ и $l(\overline{y}) > 0$.

Если $\alpha = 0$, то получаем, что $1 - \alpha = 1$ и, следовательно, $l(\overline{\xi}) = (1-\alpha) \cdot l(\overline{x}) + \alpha \cdot l(\overline{y}) = 1 \cdot l(\overline{x}) + 0 \cdot l(\overline{y}) = l(\overline{x}) > 0$.

Аналогично, если $\alpha = 1$, то $l(\overline{\xi}) = l(\overline{y}) > 0$.

Таким образом, показано, что вместе с любыми двумя точками $\overline{x}, \overline{y} \in l^+$ произвольная точка $\overline{\xi} \in [\overline{x}; \overline{y}]$ также содержится в l^+. Отсюда и следует выпуклость l^+.

Выпуклость множеств $l^-, \overline{l^+}, \overline{l^-}, l$ доказывается аналогично.

□

Лемма 3. *Для всякого разбиения $\{R^1, ..., R^s\}$ пространства $\mathbb{R}^n$ выполнены следующие свойства:*

1) *Любой непустой объемный класс R^i является выпуклым открытым множеством.*

2) *Для любого непустого объемного класса R^i верно, что $\overline{R^i} = R^i \cup \partial_{lin} R^i$.*

3) *Любой непустой плоский класс содержится в замыкании как минимум двух непустых объемных классов эквивалентности.*

Доказательство. 1. Если класс R^i является объемным, то он имеет сигнатуру $\sigma(R^i) = (\sigma_1, ..., \sigma_k)$ такую, что $\forall j \in \{1, ..., k\}$ $\sigma_j \neq 0$. По определению это означает, что класс R^i является решением системы строгих неравенств $\begin{cases} \sigma_1 \cdot l_1(\overline{x}) > 0 \\ ... \\ \sigma_k \cdot l_k(\overline{x}) > 0 \end{cases}$, где $l_j = \{(x_1, ..., x_n) \in \mathbb{R}^n | a_{j1} \cdot x_1 + ... + a_{jn} \cdot x_n + d_j = 0\}, j = 1, ..., k$. Другими словами, R^i является пересечением всех геометрических мест точек $(\sigma_1 \cdot l_1)^+, ..., (\sigma_k \cdot l_k)^+$, где $(\sigma_j \cdot l_j)^+ = \{(x_1, ..., x_n) \in \mathbb{R}^n | \sigma_j \cdot a_{j1} \cdot x_1 + ... + \sigma_j \cdot a_{jn} \cdot x_n + \sigma_j \cdot d_j > 0\}, j = 1, ..., k$.

Далее, так как, по лемме 2, все $(\sigma_j \cdot l_j)^+, j = 1, ..., k$ являются открытыми и выпуклыми множествами, то R^i является открытым и выпуклым множеством, как пересечение конечного числа открытых и выпуклых множеств.

2. Аналогично пункту 1, класс R^i совпадает с множеством решений системы $\begin{cases} \sigma_1 \cdot l_1(\overline{x}) > 0 \\ ... \\ \sigma_k \cdot l_k(\overline{x}) > 0 \end{cases}$, которое назовем A.

Рассмотрим теперь множество $\overline{A}$, являющееся замыканием множества A, и покажем, что $\overline{A}$ совпадает со множеством всех решений системы $\begin{cases} \sigma_1 \cdot l_1(\overline{x}) \geqslant 0 \\ ... \\ \sigma_k \cdot l_k(\overline{x}) \geqslant 0 \end{cases}$, которое обозначим B. Доказательство будем вести методом двойного включения.

Докажем включение $\overline{A} \subset B$. Отметим сначала, что верно [8] включение (15).

$$\overline{A} = \overline{(\sigma_1 \cdot l_1)^+ \cap ... \cap (\sigma_k \cdot l_k)^+} \subset \overline{(\sigma_1 \cdot l_1)^+} \cap ... \cap \overline{(\sigma_k \cdot l_k)^+} \qquad (15)$$

Но по лемме 2 верно, что $\overline{(\sigma_j \cdot l_j)^+} = (\sigma_j \cdot l_j)^+ \cup (\sigma_j \cdot l_j),\ j = 1, ..., k$, откуда получаем, что $\overline{(\sigma_j \cdot l_j)^+} = \{\overline{x} \in \mathbb{R}^n | \sigma_j \cdot l_j(\overline{x}) \geqslant 0\},\ j \in 1, ..., k$. А отсюда следует, что $\overline{(\sigma_1 \cdot l_1)^+} \cap ... \cap \overline{(\sigma_k \cdot l_k)^+} = \{\overline{x} \in \mathbb{R}^n | \sigma_1 \cdot l_1(\overline{x}) \geqslant 0 \& ... \& \sigma_k \cdot l_k(\overline{x}) \geqslant 0\}$, то есть $\overline{(\sigma_1 \cdot l_1)^+} \cap ... \cap \overline{(\sigma_k \cdot l_k)^+}$ – это множество решений системы $\begin{cases} \sigma_1 \cdot l_1(\overline{x}) \geqslant 0 \\ ... \\ \sigma_k \cdot l_k(\overline{x}) \geqslant 0 \end{cases}$. Откуда и следует доказываемое включение.

Докажем теперь обратное включение $B \subset \overline{A}$. Для начала возьмем $\forall \overline{x} \in B$. Так как $A \neq \varnothing$, то $\exists \overline{y} \in A$. Причем, так как по пункту 1 данной леммы A – открыто, то вместе с точкой $\overline{y}$ во множестве A содержится целая окрестность $O_\delta(\overline{y})$, поэтому можно выбрать $\overline{y} \in A : \overline{y} \neq \overline{x}$ (см. рис. 2).

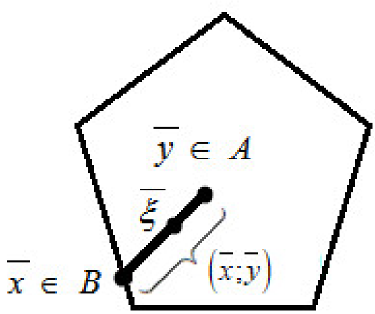

Рис. 2. Построение отрезка, уходящего из $\forall \overline{x} \in B$ внутрь множества A

Рассмотрим множество $N(\overline{x}) = \{j | l_j(\overline{x}) \neq 0\} \subset \{1, ..., k\}$ – множество неактивных ограничений в точке $\overline{x}$. Так как класс R^i – объемный и $\overline{y} \in A = R^i$, то $N(\overline{y}) = \{1, ..., k\}$.

Покажем, что $\forall \overline{\xi} \in (\overline{x}; \overline{y})\ N(\overline{\xi}) = N(\overline{x}) \cup N(\overline{y})$ и $\overline{\xi} \in B$ (см. рис. 2).

Для этого рассмотрим вектор-функцию $r(\overline{x}) = (\sigma_1 \cdot l_1(\overline{x}), ..., \sigma_k \cdot l_k(\overline{x}))$. По построению, для $\forall \overline{x} \in B$ верно, что $r_j(\overline{x}) \geqslant 0, j \in \{1, ..., k\}$. Причем, также для $\forall \overline{x} \in B$, выполняется, что $r_j(\overline{x}) > 0$ тогда и только тогда, когда $l_j(\overline{x}) \neq 0$, то есть $j \in N(\overline{x})$.

Рассмотрим теперь точку $\overline{\xi} \in (\overline{x}; \overline{y})$. По определению это означает, что $\exists t \in (0; 1)\ \overline{\xi} = (1 - t) \cdot \overline{x} + t \cdot \overline{y}$.

Далее, по лемме 1, выполнено равенство (16).

$$r_j(\overline{\xi}) = (1 - t) \cdot r_j(\overline{x}) + t \cdot r_j(\overline{y}), j = 1, ..., k \tag{16}$$

Причем, согласно выбору t, верно, что $t > 0,\ 1 - t > 0$. А так как $\overline{x}, \overline{y} \in B$, то, по определению, $r_j(\overline{x}) \geqslant 0,\ r_j(\overline{y}) \geqslant 0,\ j = 1, ..., k$. Но тогда, в силу (16), выполняется, что $r_j(\overline{\xi}) \geqslant 0, j = 1, ..., k$. Отсюда немедленно следует, что $\overline{\xi} \in B$.

При этом, если хотя бы одна из величин $r_j(\overline{x})$ или $r_j(\overline{y})$ отлична от нуля (учитывая, что данные величины неотрицательны, то сказанное означает, что $r_j(\overline{x}) > 0$ или $r_j(\overline{y}) > 0$), то из (16) следует, что $r_j(\overline{\xi}) > 0$.

Очевидно и обратное, то есть, если $r_j(\overline{\xi}) > 0$, то $r_j(\overline{x}) > 0$ или $r_j(\overline{y}) > 0$.

Из доказанного соотношения $r_j(\overline{\xi}) > 0 \Leftrightarrow r_j(\overline{x}) > 0 \vee r_j(\overline{y}) > 0,\ j = 1, ..., k$ следует, что $\forall \overline{\xi} \in (\overline{x}; \overline{y})\ N(\overline{\xi}) = N(\overline{x}) \cup N(\overline{y})$.

Но тогда, в силу того, что $N(\overline{y}) = \{1, ..., k\}$, получаем, что $N(\overline{\xi}) = \{1, ..., k\}$. Откуда, учитывая дополнительно, что $\overline{\xi} \in B$, получаем, что $\overline{\xi} \in A$.

Теперь рассмотрим $\forall \varepsilon > 0$ и возьмем $t = \frac{\min(\varepsilon, \|\overline{y} - \overline{x}\|)}{2\|\overline{y} - \overline{x}\|} \in (0; 1)$. Ему соответствует некоторая точка $\overline{\xi} = (1 - t) \cdot \overline{x} + t \cdot \overline{y} \in (\overline{x}; \overline{y})$, для которой выполнено, что $\|\overline{\xi} - \overline{x}\| = \|(1 - t) \cdot \overline{x} + t \cdot \overline{y} - \overline{x}\| = \|(1 - t) \cdot \overline{x} + t \cdot \overline{y} - (1 - t) \cdot \overline{x} - t \cdot \overline{x}\| = \|t \cdot (\overline{y} - \overline{x})\| = t \cdot \|\overline{y} - \overline{x}\| \leqslant \frac{\varepsilon}{2\|\overline{y} - \overline{x}\|} \cdot \|\overline{y} - \overline{x}\| = \frac{\varepsilon}{2} < \varepsilon$.

Таким образом, какая бы малая окрестность $O_\varepsilon(\overline{x})$ не была взята, в ней всегда найдется точка $\overline{\xi} \in A$ (см. рис. 3). Отсюда следует, что $\overline{x} \in \overline{A}$.

Итак, показано, что $\overline{A} = B$. Множество B отличается от множества A тем, что в нем добавились классы со всевозможными занулениями сигнатур, то есть классы, обладающие сигнатурами $(\sigma'_1, ..., \sigma'_k) : \sigma'_j \in \{\sigma_j, 0\},\ j = 1, ..., k$. Следовательно, $\overline{A} = A \cup \partial_{lin} A$.

3. Рассмотрим произвольный непустой плоский класс эквивалентности R^i. Без ограничения общности, будем считать, что $\sigma(R^i) = (\sigma_1, ..., \sigma_k)$, где $\sigma_1, ..., \sigma_{k'} \neq 0$, а $\sigma_{k'+1}, ..., \sigma_k = 0$. Причем $k' \in \{1, ..., k-1\}$, так как в противном случае взятый класс является объемным. Тогда,

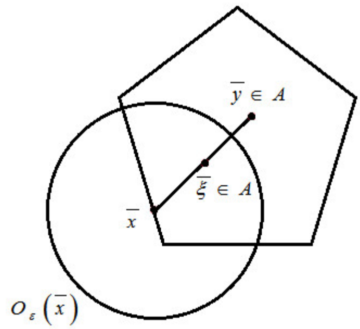

Рис. 3. Иллюстрация того, что $\forall O_\varepsilon(\overline{x}) \exists \overline{\xi} \in O_\varepsilon(\overline{x}) \cap (\overline{x}; \overline{y})$

по определению, R^i является решением системы равенств и неравенств (17), которую можно переписать в виде (18).

$$\begin{cases} \sigma_1 \cdot l_1(\overline{x}) > 0 \\ \ldots \\ \sigma_{k'} \cdot l_{k'}(\overline{x}) > 0 \\ \sigma_{k'+1} \cdot l_{k'+1}(\overline{x}) = 0 \\ \ldots \\ \sigma_k \cdot l_k(\overline{x}) = 0 \end{cases} \tag{17}$$

$$\begin{cases} \sigma_1 \cdot l_1(\overline{x}) > 0 \\ \ldots \\ \sigma_{k'} \cdot l_{k'}(\overline{x}) > 0 \\ \sigma_{k'+1} \cdot l_{k'+1}(\overline{x}) \geqslant 0 \\ \ldots \\ \sigma_k \cdot l_k(\overline{x}) \geqslant 0 \\ \sigma_{k'+1} \cdot l_{k'+1}(\overline{x}) \leqslant 0 \\ \ldots \\ \sigma_k \cdot l_k(\overline{x}) \leqslant 0 \end{cases} \tag{18}$$

Рассмотрим теперь систему (19), полученную из системы (18) заменой всех строгих неравенств на нестрогие.

$$
\begin{cases}
\sigma_1 \cdot l_1(\overline{x}) \geqslant 0 \\
... \\
\sigma_{k'} \cdot l_{k'}(\overline{x}) \geqslant 0 \\
\sigma_{k'+1} \cdot l_{k'+1}(\overline{x}) \geqslant 0 \\
... \\
\sigma_k \cdot l_k(\overline{x}) \geqslant 0 \\
\sigma_{k'+1} \cdot l_{k'+1}(\overline{x}) \leqslant 0 \\
... \\
\sigma_k \cdot l_k(\overline{x}) \leqslant 0
\end{cases} \tag{19}
$$

Очевидно, что множество решений системы (19) содержится в пересечении замыканий объемных классов R^{j_1}, R^{j_2}, задаваемых системами (20) и (21), соответственно.

$$
\begin{cases}
\sigma_1 \cdot l_1(\overline{x}) > 0 \\
... \\
\sigma_{k'} \cdot l_{k'}(\overline{x}) > 0 \\
\sigma_{k'+1} \cdot l_{k'+1}(\overline{x}) > 0 \\
... \\
\sigma_k \cdot l_k(\overline{x}) > 0
\end{cases} \tag{20}
$$

$$
\begin{cases}
\sigma_1 \cdot l_1(\overline{x}) > 0 \\
... \\
\sigma_{k'} \cdot l_{k'}(\overline{x}) > 0 \\
\sigma_{k'+1} \cdot l_{k'+1}(\overline{x}) < 0 \\
... \\
\sigma_k \cdot l_k(\overline{x}) < 0
\end{cases} \tag{21}
$$

Таким образом, получаем, что класс R^i, являющийся множеством решений системы (18), содержится во множестве решений системы (19), которое, как было сказано выше, содержится в пересечении $\overline{R^{j_1}} \cap \overline{R^{j_2}}$. Отсюда следует, что $R^i \subset \overline{R^{j_1}} \cap \overline{R^{j_2}}$. Причем, если хотя бы один из классов R^{j_1} или R^{j_2} оказался пуст, то тогда и $\overline{R^{j_1}} \cap \overline{R^{j_2}}$ было бы пусто, а, следовательно, R^i также был бы пустым, что противоречит условию. $\square$

Лемма 4. *Пусть $f : \mathbb{R}^n \to \mathbb{R}$ – выпуклая непрерывная кусочно-линейная функция, заданная над разбиением $\{R^1, ..., R^s\}$ пространства $\mathbb{R}^n$, а R^i, R^j – два произвольных непустых объемных класса из данного разбиения.*

Тогда верно, что $\forall \overline{x} \in R^j$ $f(\overline{x}) \geqslant f_{R^i}(\overline{x})$ *(см. рис. 4).*

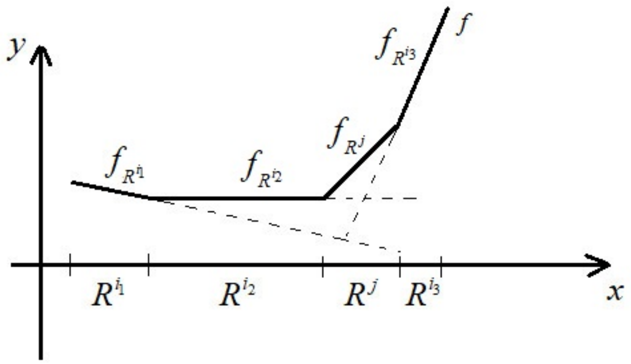

Рис. 4. Пример-иллюстрация утверждения леммы (одномерный случай). В качестве R^i можно взять, например, $R^{i_1}, R^{i_2}, R^{i_3}$

Доказательство. Предположим противное, что нашлись такие непустые объемные классы R^i, R^j, что верно выражение (22).

$$\exists \overline{x} \in R^j \; f(\overline{x}) < f_{R^i}(\overline{x}) \tag{22}$$

Очевидно, что $i \neq j$, так как в противном случае выполняется, что $\forall \overline{x} \in R^j$ $f(\overline{x}) = f_{R^i}(\overline{x})$, что противоречит предположению (22).

Так как по условию $\forall \overline{x} \in R^j$ $f(\overline{x}) = f_{R^j}(\overline{x})$, то неравенство (22) можно записать, в виде (23).

$$\exists \overline{x} \in R^j \; f_{R^j}(\overline{x}) < f_{R^i}(\overline{x}) \tag{23}$$

Зафиксируем $\overline{x} \in R^j$, для которого выполняется предположение (22) и $\forall \overline{y} \in R^i$, после чего соединим их отрезком $[\overline{x}; \overline{y}]$.

Обозначим $f(t) = f((1-t) \cdot \overline{x} + t \cdot \overline{y})$, $f_{R^i}(t) = f_{R^i}((1-t) \cdot \overline{x} + t \cdot \overline{y})$ и $f_{R^j}(t) = f_{R^j}((1-t) \cdot \overline{x} + t \cdot \overline{y})$, где $t \in \mathbb{R}$, сечения функций $f(\overline{x}), f_{R^i}(\overline{x}), f_{R^j}(\overline{x})$ прямыми, проходящими через зафиксированные точки $\overline{x}, \overline{y}$.

Отметим, что $f(t)$ – выпуклая, так как она является сужением выпуклой функции $f(\overline{x})$ на выпуклое множество, а функции $f_{R^i}(t)$ и $f_{R^j}(t)$ являются линейными функциями одного аргумента $t \in \mathbb{R}$.

Рассмотрим функцию $A(t) = f_{R^i}(t) - f_{R^j}(t)$, которая, очевидно, также линейна, причем в силу (23) выполняется (24).

$$A(0) > 0 \tag{24}$$

Далее рассмотрим два случая:

• Пусть $\forall t \in [0;1)\ A(t) \neq 0$. Но тогда, в силу (24) и непрерывности функции $A(t)$ получаем, что $\forall t \in [0;1)\ A(t) > 0$.

По определению, $\forall \overline{x} \in R^j\ f(\overline{x}) = f_{R^j}(\overline{x})$, $\forall \overline{y} \in R^i\ f(\overline{y}) = f_{R^i}(\overline{y})$. При этом R^i, R^j – выпуклые, открытые и $R^j \cap R^i = \varnothing$. Поэтому $\exists t_1, t_2 \in (0;1)$: $t_2 > t_1$ и при этом $\forall t \in [0;t_1]\ f(t) = f_{R^j}(t)$, а также $\forall t \in [t_2;1]\ f(t) = f_{R^i}(t)$ (см. рис. 5).

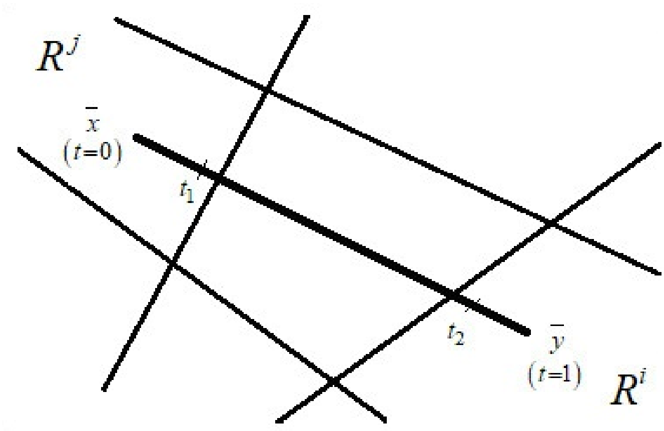

Рис. 5. Иллюстрация к утверждению о том, что $\forall t \in [0;t_1]\ f(t) = f_{R^j}(t)$ и $\forall t \in [t_2;1]\ f(t) = f_{R^i}(t)$

Рассмотрим отрезок $[t_1;1] \subset \mathbb{R}$ и выразим t_2 в виде следующей комбинации границ данного отрезка: $t_2 = (1-\alpha)\cdot t_1 + \alpha \cdot 1$, где, очевидно, $\alpha \in (0;1)$. Отметим, что по лемме 1 верно выражение (25).

$$\forall \alpha \in [0;1]\ (1-\alpha)\cdot f_{R^i}(t_1) + \alpha \cdot f_{R^i}(1) = f_{R^i}((1-\alpha)\cdot t_1 + \alpha \cdot 1) \quad (25)$$

Учитывая (25) и тот факт, что $f_{R^i}(t_1) > f_{R^j}(t_1)$, который следует из предположения $\forall t \in [0;1)\ A(t) > 0$, сделанного в начале доказательства этого случая, получаем, что верна следующая цепочка преобразований:

$$\begin{aligned}(1-\alpha)\cdot f(t_1) + \alpha\cdot f(1) &\stackrel{def}{=} (1-\alpha)\cdot f_{R^j}(t_1) + \alpha\cdot f_{R^i}(1) < \\ &< (1-\alpha)\cdot f_{R^i}(t_1) + \alpha\cdot f_{R^i}(1) = f_{R^i}((1-\alpha)\cdot t_1 + \alpha\cdot 1) = \\ &= f_{R^i}(t_2) = f(t_2) = f((1-\alpha)\cdot t_1 + \alpha\cdot 1) \quad (26)\end{aligned}$$

Полученное неравенство (26) противоречит выпуклости функции $f(t)$.

• Теперь рассмотрим случай, когда $\exists t^* \in [0;1):\ A(t^*) = 0$. Отметим, что $A(0) > 0$. Отсюда и из предыдущего немедленно следует, что данная линейная функция одного аргумента имеет ненулевой угловой коэффициент и является строго монотонной. Поэтому $\forall t : t > t^*$ выполняется, что $A(t) < 0$. В частности, $A(1) < 0$, откуда следует, что $f_{R^i}(1) < f_{R^j}(1)$.

Но тогда для $\forall t \in (0;1)$ выполняется, что $(1-t)\cdot f(0) + t\cdot f(1) = (1-t)\cdot f_{R^j}(0) + t\cdot f_{R^i}(1) < (1-t)\cdot f_{R^j}(0) + t\cdot f_{R^j}(1) = f_{R^j}((1-t)\cdot 0 + t\cdot 1) = f((1-t)\cdot 0 + t\cdot 1)$.

Таким образом, снова имеем противоречие с выпуклостью функции $f(t)$.

Учитывая, что всегда выполняется один из рассмотренных случаев, получаем доказываемое. □

Теорема 1 (о выпуклых CPL-функциях). *Любую выпуклую непрерывную кусочно-линейную функцию $f : \mathbb{R}^n \to \mathbb{R}$, заданную над разбиением $\{R^1, ..., R^s\}$, можно выразить нейронной схемой нелинейной сложности 1 и нелинейной глубины 1 над базисом B_1.*

Доказательство. Отметим сначала, что при задании функции f в действительности используются лишь непустые классы эквивалентности. Это происходит в силу того, что определения вида $\forall \overline{x} \in R^j\ f(\overline{x}) = f_{R^j}(\overline{x})$ попросту не имеют смысла, если класс эквивалентности R^j – пуст.

Таким образом, если показать равенство функции f и некоторой функции g на точках всех непустых классов эквивалентности, то сразу же будет показано равенство данных функций.

Рассмотрим функцию $\max\limits_{j\in\{1,...,s'\}} f_{R^j}(\overline{x})$, где $R^1, ..., R^{s'}$ – все непустые объемные классы эквивалентности разбиения.

Далее возьмем произвольный непустой класс эквивалентности R^j. Возможны два случая:

• R^j – объемный. Тогда, так как функция $f(\overline{x})$ выпукла, то из леммы 4 следует, что $\forall \overline{x} \in R^j\ f(\overline{x}) \geqslant f_{R^{j'}}(\overline{x}), j' \in \{1, ..., s'\}$.

При этом $\exists j' = j : \forall \overline{x} \in R^j\ f(\overline{x}) = f_{R^j}(\overline{x})$, то есть среди объемных классов $\{R^1, ..., R^{s'}\}$ существует класс R^j, на котором достигается равенство функции $f(\overline{x})$ и функции $f_{R^j}(\overline{x})$, соответствующей данному классу. Отсюда следует, что $f(\overline{x}) = \max\limits_{j\in\{1,...,s'\}} f_{R^j}(\overline{x})$.

• R^j – плоский. По лемме 3 данный класс содержится в замыкании некоторого непустого объемного класса эквивалентности R^i. Последнее означает, что $\forall \overline{x} \in R^j\ \exists \{\overline{x_n}\}_{n\in\mathbb{N}} \subset R^i : \overline{x} = \lim\limits_{n\to+\infty} \overline{x_n}$.

По ранее доказанному, на любом непустом объемном классе эквивалентности выполняется, что $f(\overline{x}) = \max\limits_{j \in \{1,\dots,s'\}} f_{R^j}(\overline{x})$, откуда следует выражение (27).

$$\forall n \in \mathbb{N}\ f(\overline{x_n}) = \max_{j \in \{1,\dots,s'\}} f_{R^j}(\overline{x_n}) \tag{27}$$

Переходя в выражении (27) к пределу при $n \to +\infty$, получаем выражение (28).

$$\lim_{n \to +\infty} f(\overline{x_n}) = \lim_{n \to +\infty} \max_{j \in \{1,\dots,s'\}} f_{R^j}(\overline{x_n}) \tag{28}$$

В выражении (28), в силу непрерывности функций $f(\overline{x})$, $\max\limits_{j \in \{1,\dots,s'\}} f_{R^j}(\overline{x_n})$, можно перейти от пределов функций к пределу аргументов данных функций. В результате данного перехода получаем выражение (29).

$$f(\lim_{n \to +\infty} \overline{x_n}) = \max_{j \in \{1,\dots,s'\}} f_{R^j}(\lim_{n \to +\infty} \overline{x_n}) \tag{29}$$

Но $\lim\limits_{n \to +\infty} \overline{x_n} = \overline{x}$, поэтому, переходя к пределу в выражении (29), получаем выражение (30).

$$f(\overline{x}) = \max_{j \in \{1,\dots,s'\}} f_{R^j}(\overline{x}) \tag{30}$$

Таким образом, показано, что для любого непустого класса эквивалентности выполняется выражение (30). Откуда и следует доказываемое утверждение. □

Следствие 1. *Если в условиях теоремы 1 заменить базис B_1 на B_2, то в новом базисе любую выпуклую непрерывную кусочно-линейную функцию $f : \mathbb{R}^n \to \mathbb{R}$, заданную над разбиением пространства $\mathbb{R}^n$, в котором содержится s' непустых объемных классов эквивалентности, можно выразить нейронной схемой нелинейной глубины $\lceil \log_2 s' \rceil$ и нелинейной сложности $s' - 1$.*

Причем полученные оценки нелинейной сложности и глубины являются неулучшаемыми.

Доказательство. Отметим сначала, что в базисе B_2 можно выразить функцию $\max(x, y)$. Нетрудно убедиться, что данная функция выражается по формуле (31).

$$\max(x, y) = RELU(x - y) + y \tag{31}$$

Но имея функцию $\max(x,y)$, можно оптимальным образом [9] выразить функцию $\max(x_1,...,x_{s'})$, объединяя аргументы по сбалансированному бинарному дереву [10], [11], высота которого, как известно [11], равна $\lceil \log_2 s' \rceil$ и количество элементов $\max(x,y)$ в котором равно $s'-1$.

Далее, заменяя $\max(x,y)$ по формуле (31), получаем нейронную схему той же нелинейной сложности и глубины (так как вместо каждого максимума при замене возникает одна функция $RELU$ и какие-то линейные функции, то есть число нелинейных элементов при данном преобразовании не меняется).

При этом, очевидно, что при замене аргументов функции $\max(x_1,...,x_{s'})$ на соответствующие им значения линейных функций $f_{R^j}(\overline{x})$ в формуле (30) новые нелинейные элементы в конструируемую нейронную схему не добавляются.

Таким образом, заданную в условии кусочно-линейную функцию, можно представить в виде нейронной схемы нелинейной глубины $\lceil \log_2 s' \rceil$ и нелинейной сложности $s'-1$. □

Отметим, что существуют и вогнутые (выпуклые вверх) функции. Пусть дана вогнутая функция $f(\overline{x})$ класса CPL. Тогда $-f(\overline{x})$ является выпуклой функцией класса CPL. Но из теоремы 1 и следствия 1 следует, что всякая выпуклая CPL-функция представима в виде некоторой нейронной схемы $g_1(\overline{x})$ нелинейной сложности 1 над базисом B_1 и некоторой нейронной схемы $g_2(\overline{x})$ над базисом B_2 нелинейной сложности $s'-1$ и нелинейной глубины $\lceil \log_2 s' \rceil$.

Но тогда $-1 \cdot g_1(\overline{x})$ и $-1 \cdot g_2(\overline{x})$ являются нейронными схемами той же нелинейной сложности и глубины, что и схемы $g_1(\overline{x})$, $g_2(\overline{x})$, но при этом данные схемы задают уже функцию $f(\overline{x})$.

Из вышесказанного вытекает, что все доказанные утверждения верны и для любой вогнутой CPL-функции.

Отметим также, что класс CPL-функций состоит не только из выпуклых или вогнутых CPL-функций. Так, например, функция $f(x) = x - 2 \cdot RELU(x) + 2 \cdot RELU(x-1)$, очевидно, является CPL-функцией, но не является ни выпуклой, ни вогнутой функцией (см. рис. 6).

Автор выражает благодарность научному руководителю, к.ф.-м.н., научному сотруднику В.С. Половникову за постановку задачи и научное руководство, а также д.ф.-м.н., доценту А.А. Часовских за помощь в проверке полученных результатов.

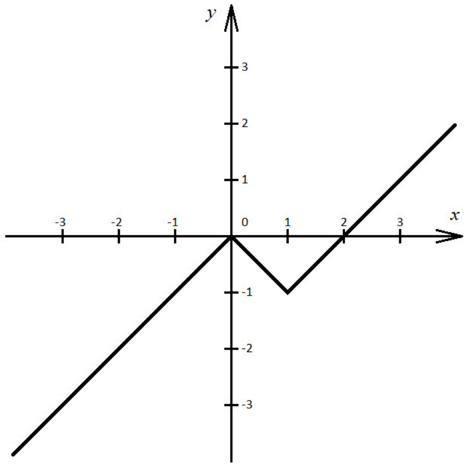

Рис. 6. CPL-функция, не являющаяся выпуклой или вогнутой

Список литературы

[1] Половников В.С., *Об оптимизации структурной реализации нейронных сетей*, Диссертация на соискание степени кандидата наук, Москва, 2007, 96 с.

[2] Яблонский С.В., *Введение в дискретную математику*, Наука. Гл. ред. физ.-мат. лит., Москва, 1986, 384 с.

[3] Simon Haykin, *Neural Networks. A Comprehensive Foundation*, Prentice Hall International, Inc., Canada, 1999, 1999 с.

[4] Шишляков В.Г., "О построении явной архитектуры нейронной сети, приближающей кусочно-линейные функции", *Интеллектуальные системы. Теория и приложения*, **26**:2 (2022), 42–60.

[5] McCulloch W.S., Pitts W., "A logical calculus of the ideas immanent nervous activity", *Bull. Of math. Biophysics*, **5** (1943), 115–133.

[6] Кан А.Н., "Вопросы полноты в классе кусочно-линейных непрерывных функций", *Интеллектуальные системы. Теория и приложения*, **21**:2 (2017), 46–56.

[7] Мищенко А.С., Фоменко А.Т., *Курс дифференциальной геометрии и топологии*, Факториал Пресс, Москва, 2000, 448 с.

[8] Виро О.Я., Иванов О.А., Нецветаев Н.Ю., Харламов В.М., *Задачи по топологии*, СПбГУ, Санкт-Петербург, 2000, 208 с.

[9] Абельхарисов А.В., “О существовании нейронных схем произвольной минимальной нелинейной глубины в базисе с функцией активации ReLU”, *Ломоносовские чтения 2021*, 2021.

[10] Кнут Д., *Искусство программирования. Том 1. Основные алгоритмы*, Вильямс, Москва, 2019, 720 с.

[11] Седжвик Р., *Фундаментальные алгоритмы на C++*, ДиаСофт, Санкт-Петербург, 2001, 688 с.

Convex CPL-functions recovering by neural networks on RELU-bases
Shishlyakov V.G.

The present paper considers a problem of functional classes obtained by using neural networks on max non-linearities bases. Firstly, some properties of CPL-functions and equivalence classes generating them are investigated. Proceeding from these properties a theorem is proved that neural networks built on the basis of linear and max non-linearity functions can exactly recover any convex CPL-function.

Secondly, RELU-basis, a special case of max non-linearities bases, is investigated, with a theorem similar to the previous one mentioned above proved. The question of estimating the number of neurons and layers in obtained architectures is also discussed.

All the mentioned theorems have a constructive proof, i.e. neural network architectures with mentioned features are built explicitly.

Keywords: Neural networks, architecture, functions recovery, functions expressibility, convex functions, particle-linear functions, ReLU function, max function.

References

[1] Polovnikov V.S., *On optimization of the structural implementation of neural networks*, Ph.D. Thesis … physical and mathematical sciences, MSU, Moscow, 2007 (In Russian)

[2] YAblonskij S.V., *Introduction to discrete mathematics*, eds. fiz.-mat.lit., «Science», Moscow, 1986 (In Russian)

[3] Simon Haykin, *Neural Networks. A Comprehensive Foundation*, Prentice Hall International, Inc., Canada, 1999, 1999 pp.

[4] Shishlyakov V.G., “On the construction of an explicit neural network architecture that approximates particle-linear functions”, *Intelligent systems. Theory and applications.*, **26**:2 (2022), 42–60 (In Russian)

[5] McCulloch W.S., Pitts W., “A logical calculus of the ideas immanent nervous activity”, *Bull. Of math. Biophysics*, **5** (1943), 115–133

[6] Kan A.N., “Questions of completeness in the class of particle-linear continuous functions”, *Intelligent systems. Theory and applications.*, **21**:2 (2017), 46–56 (In Russian)

[7] Mishchenko A.S., Fomenko A.T., *Course in Differential Geometry and Topology*, Factorial Press, Moscow, 2000 (In Russian)

[8] Viro O.Ya., Ivanov O.A., Netsvetaev N.Yu., Kharlamov V.M., *Topology tasks*, St. Petersburg State University, St. Petersburg, 2000 (In Russian)

[9] Abelkharisov A.V., “On the Existence of Neural Circuits of Arbitrary Minimum Nonlinear Depth in a Basis with ReLU Activation Function”, *Lomonosov Readings 2021*, 2021 (In Russian)

[10] Knut D., *The art of programming. Volume 1. Basic Algorithms*, Williams, Moscow, 2019 (In Russian)

[11] Sedgwick R., *Fundamental Algorithms in C++*, DiaSoft, St. Petersburg, 2001 (In Russian)