

П  
МЗЧ

П  
МЗЧ



Н. Н. Евтихийев    Б. Н. Оныкий    В. В. Перепелица  
И. Б. Щербаков

004-94

**МАТЕМАТИЧЕСКИЕ МОДЕЛИ  
И ОПТИЧЕСКИЕ РЕАЛИЗАЦИИ  
МНОГОСЛОЙНЫХ И ПОЛИНОМИАЛЬНЫХ  
НЕЙРОННЫХ СЕТЕЙ**

Москва 1994

ГОСУДАРСТВЕННЫЙ КОМИТЕТ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ПО ВЫСШЕМУ ОБРАЗОВАНИЮ

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ  
ИНЖЕНЕРНО-ФИЗИЧЕСКИЙ ИНСТИТУТ  
(ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ)

Н. Н. Евтихийсв Б. Н. Оныкий В. В. Перепелица  
И. Б. Щербаков

**МАТЕМАТИЧЕСКИЕ МОДЕЛИ И ОПТИЧЕСКИЕ  
РЕАЛИЗАЦИИ МНОГОСЛОЙНЫХ И  
ПОЛИНОМИАЛЬНЫХ НЕЙРОННЫХ СЕТЕЙ**

Препринт 004-94

*Утверждено  
редсоветом института*

Москва 1994

УДК 681.3

Евтихий Н.Н., Оныкий Б.Н., Перепелица В.В., Щербаков И.Б.  
Математические модели и оптические реализации многослойных и  
полиномиальных нейронных сетей. М.: Препринт/ МИФИ, 004-94,  
1994. - 32 с.

Излагаются результаты исследования математических моделей и  
оптических реализаций нейронных сетей. Представлен вывод  
обобщенного градиентного алгоритма обучения многослойных  
нейронных сетей первого порядка с последовательными связями.  
Предлагаются реализации сетей на основе многоканального  
многочастотного акустооптического модулятора. Изложены основ-  
ные проблемы сопряжения математических моделей нейронных  
сетей и их оптических реализаций, намечены пути разрешения этих  
проблем.

© Московский государственный  
инженерно-физический институт  
(технический университет), 1994.

ISBN 5-72-62-0028-4

## ВВЕДЕНИЕ

Нейронная сеть - это совокупность простых вычислительных элементов (нейронов), соединенных между собой. В простейшем случае нейрон выполняет скалярное произведение вектора входных сигналов на вектор синаптических весов, к произведению добавляется смещение. Над суммой выполняется нелинейное преобразование (передаточная функция), результат которого есть выходной сигнал нейрона [6]. Такие нейроны называются аддитивными. Сеть, состоящая из аддитивных нейронов, называется нейронной сетью первого порядка.

Мультипликативными называются нейроны, выполняющие перемножение тензоров. Сети из мультипликативных нейронов называются сетями высших порядков; порядок сети определяется размерностью перемножаемых тензоров. Если все нейроны в сети имеют одинаковые передаточные функции, то сеть называется однородной.

В многослойной нейронной сети нейроны делятся на группы с общим входным сигналом - слои. Связи от выходов нейронов некоторого слоя  $m$  к входам нейронов следующего слоя  $(m+1)$  называются последовательными.

Нейрокомпьютер - это вычислитель, архитектура которого ориентирована на выполнение операций, заданных структурой нейронной сети [1]. Нейрокомпьютеры принято относить к шестому поколению вычислительных систем [2].

Уже доказано, что нейронный базис универсален. Для любого алгоритма можно построить нейронную сеть, которая его реализует. Но в настоящее время основное внимание уделяется разработке специализированных нейрокомпьютеров, ориентированных на решение задач, с которыми плохо справляются вычислительные системы прошлых поколений.

Нейрокомпьютеры обладают целым рядом свойств, привлекательных с точки зрения их практического использования :

- сверхвысокое быстродействие за счет использования массового параллелизма обработки информации;
- толерантность к ошибкам : работоспособность сохраняется при повреждении значительного числа нейронов;
- способность к обучению; программирование вычислительной системы заменяется обучением;

- способность к распознаванию в условиях сильных помех и искажений.

Эти свойства позволяют с высокой эффективностью решать задачи распознавания образов и адаптивного управления [11, 19].

Нейронная сеть для решения некоторой задачи может быть эмулирована программно либо реализована аппаратно. Нейронная сеть в виде программного продукта работает медленно, поскольку параллельные операции, заданные структурой нейронной сети, выполняются последовательно. Последовательного выполнения параллельных операций не удастся избежать даже на многопроцессорных вычислительных системах. Компьютерное моделирование нейронных сетей целесообразно проводить лишь в процессе разработки аппаратных реализаций нейронных сетей.

Аппаратная реализация нейронной сети может быть выполнена на нейрочипах (микросхемах, содержащих фрагменты нейронных сетей) [17], на СБИС-пластинах или оптоэлектронным способом [10]. Основными недостатками электронных нейронных сетей являются высокое энергопотребление и трудности в реализации большого числа связей [9].

Наиболее перспективными являются оптоэлектронные нейронные сети. Они сравнительно недороги, обладают повышенной помехоустойчивостью, надежностью, низким энергопотреблением, хорошими массогабаритными показателями [4,9]. Вообще, оптоэлектронные вычислители наиболее полно удовлетворяют требованиям, предъявляемым к бортовым вычислительным системам [5]. Их отличительной чертой является сверхвысокое быстродействие. Скорость вектор-матричных операций в оптическом вектор-матричном перемножителе превышает скорость суперкомпьютера CRAY-2 ( $10^9$  опер./с) тем больше, чем больше размерность матрицы (N) [12]. Начиная с  $N=50$  это преимущество растет как  $N^3$ .

Построение специализированного нейрокомпьютера разбивается на четыре основных этапа :

1. Построение математической модели нейронной сети для решения задачи (выбор модели нейронов и архитектуры связей в сети, определение значений параметров нейронной сети).
2. Разработка аппаратной реализации нейронной сети.
3. Сопряжение модели нейронной сети и аппаратной реализации.
4. Макетирование специализированного нейрокомпьютера.

В работе представлено описание математических моделей и

оптоэлектронных реализаций многослойных (разделы 1, 2) и полносвязных (раздел 3) нейронных сетей. В разделе 4 описаны алгоритмы сопряжения математических моделей нейронных сетей и их оптических реализаций.

## 1. МНОГОСЛОЙНАЯ НЕЙРОННАЯ СЕТЬ

Многие задачи распознавания образов и адаптивного управления можно свести к следующей обобщенной математической постановке: в соответствии с входным сигналом  $X$  сформировать требуемый выходной сигнал  $Y$ . Отображение  $X \rightarrow Y$  часто представляют множеством пар входных-выходных векторов  $\{(X^k, Y^k), k=1 \dots K\}$ .

В соответствии со следствием из теоремы Колмогорова-Арнольда для любого множества пар  $(X^k, Y^k)$  существует двухслойная однородная нейронная сеть первого порядка, которая выполняет отображение  $X \rightarrow Y$ , выдавая на каждый входной сигнал  $X^k$  правильный выходной сигнал  $Y^k$ . Нейроны в такой двухслойной нейронной сети должны иметь сигмоидные передаточные функции [19].

Иногда целесообразно использовать сети с большим числом слоев. Такие многослойные нейронные сети могут иметь меньшее количество нейронов и более простые передаточные функции, чем двухслойные сети, реализующие то же самое отображение. Таким образом, для реализации отображения будет использоваться многослойная нейронная сеть.

Схема сети представлена на рис. 1.1. Синаптические веса обозначены буквой  $w$ , смещения -  $b$ , выходы сумматоров -  $s$ , передаточные функции -  $f$ , выходы нейронов -  $y$  (внешний входной сигнал считается выходом нейронов слоя с номером 0). Индексы:  $i$  - номер нейрона ( $i=1 \dots N$ ,  $N$  - количество нейронов в слое),  $m$  - номер слоя ( $m=1 \dots L$ ,  $L$  - количество слоев в сети).

Формулы, определяющие функционирование многослойной нейронной сети первого порядка с последовательными связями:

$$s_{i_m}^m = \sum_{i_{m-1}=1}^{N_{m-1}} w_{i_m i_{m-1}}^m \times y_{i_{m-1}}^{m-1} - b_{i_m}^m, \quad (1.1)$$
$$i_m = 1 \dots N_m, \quad m=1 \dots L,$$

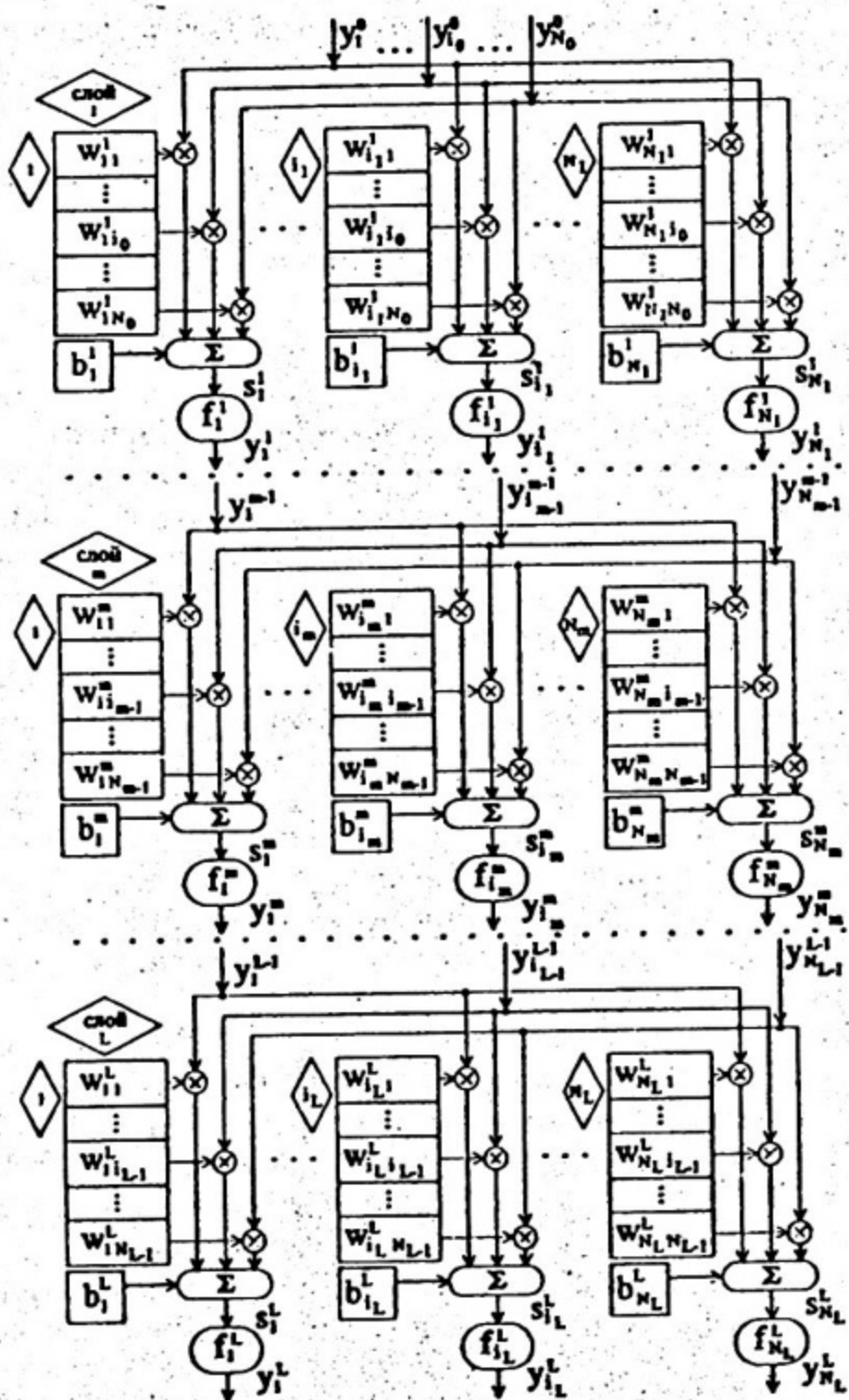


Рис. 1.1. Схема многослойной нейронной сети первого порядка.

$$y_{i_m}^m = f_{i_m}^m(s_{i_m}^m), \quad i_m = 1 \dots N_m, \quad m = 1 \dots L. \quad (1.2)$$

В процессе функционирования нейронной сети сигналы последовательно передаются от первого слоя нейронов ко второму, от второго - к третьему и т.д. В формулах (1.1) и (1.2) переменная  $m$  последовательно принимает значения 1, 2, 3 и так далее.

Для определения значений параметров нейронной сети используется алгоритм обучения [7]. Задача обучения определяется совокупностью пяти элементов:  $\langle X, Y, r, G, D \rangle$ , где  $X$  и  $Y$  - вход и выход соответственно;  $r$  - функция  $r: X \rightarrow Y$  (определяет желаемый результат обучения);  $G$  - множество функций  $g: X \rightarrow Y$  для всех  $g \in G$ ;  $D$  - функционал качества, показывающий для каждой функции  $g \in G$  степень близости к  $r$ . Обучение состоит в синтезе функции  $g \in G$ , оптимальной по  $D$ . В задаче обучения по примерам функция  $r$  задается обучающей выборкой - множеством пар входных-выходных данных. Архитектура связей нейронной сети считается заданной до начала обучения, она определяет множество функций  $G$ .

Цель обучения: построить функцию  $g \in G$  так, чтобы функционал качества обращался в оптимум для всех примеров обучающей выборки. Каждый элемент выборки (обучающий пример) имеет вид:

$$\langle \sigma_1^e, \sigma_2^e, \dots, \sigma_{i_0}^e, \dots, \sigma_{N_0}^e; \zeta_1^e, \zeta_2^e, \dots, \zeta_{i_L}^e, \dots, \zeta_{N_L}^e \rangle,$$

где  $\sigma_{i_0}^e$  - элемент входного сигнала,  $i_0 = 1 \dots N_0$ ;

$\zeta_{i_L}^e$  - элемент выходного сигнала,  $i_L = 1 \dots N_L$ ;

$e$  - номер обучающего примера в выборке ( $e = 1 \dots P$ );

$P$  - количество обучающих примеров в выборке.

На каждый входной сигнал  $\sigma^e$  сеть должна ответить выходным сигналом  $\zeta^e$ ,  $e = 1 \dots P$ .

Если в необученную нейронную сеть ввести входной сигнал одного из примеров обучающей выборки:

$$y_{i_0}^{0,e} = \sigma_{i_0}^e, \quad i_0 = 1 \dots N_0, \quad (1.3)$$

$$s_{i_m}^{m,e} = \sum_{i_{m-1}=1}^{N_{m-1}} w_{i_m i_{m-1}}^m \times y_{i_{m-1}}^{m-1,e} - b_{i_m}^m, \quad (1.4)$$

$$i_m = 1 \dots N_m, \quad m = 1 \dots L,$$

(5.1)

$$y_{i_m}^{m,e} = f_{i_m}^m(s_{i_m}^{m,e}), \quad i_m = 1 \dots N_m, \quad m = 1 \dots L, \quad (1.5)$$

то выходы нейронной сети будут отличаться от требуемых выходов, которые определены в обучающем примере.

Функция ошибки определяет степень близости выходов нейронной сети к требуемым выходам при решении всей совокупности примеров обучающей выборки :

$$D = \frac{1}{2} \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} (\zeta_{i_L}^e - y_{i_L}^{L,e})^2. \quad (1.6)$$

Обучение - это итерационная процедура. На каждой итерации происходит уменьшение функции ошибки.

Градиентный алгоритм обучения многослойных нейронных сетей первого порядка с последовательными связями может быть описан следующими формулами :

$$w_{i_m i_{m-1}}^m(t) = w_{i_m i_{m-1}}^m(t-1) + \varepsilon \times \sum_{e=1}^P \Delta_{i_m}^{m,e}(t-1) \times y_{i_{m-1}}^{m-1,e}(t-1), \quad (1.7)$$

$$i_m = 1, \dots, N_m, \quad i_{m-1} = 1 \dots N_{m-1},$$

$$m = L, (L-1) \dots 2;$$

$$w_{i_1 i_0}^1(t) = w_{i_1 i_0}^1(t-1) +$$

$$+ \varepsilon \times \sum_{e=1}^P \Delta_{i_1}^{1,e}(t-1) \times \sigma_{i_0}^e, \quad (1.8)$$

$$i_1 = 1 \dots N_1, \quad i_0 = 1 \dots N_0;$$

$$\Delta_{i_L}^{L,e} = [\zeta_{i_L}^e - f_{i_L}^L(s_{i_L}^{L,e})] \times (f_{i_L}^L(s_{i_L}^{L,e}))', \quad (1.9)$$

$$i_L = 1 \dots N_L;$$

$$\Delta_{i_m}^{m,e} = \sum_{i_{m+1}=1}^{N_{m+1}} \Delta_{i_{m+1}}^{m+1,e} \times w_{i_{m+1}i_m}^{m+1} \times (f_{i_m}^m(s_{i_m}^{m,e}))', \quad (1.10)$$

$$i_m = 1 \dots N_m, \quad m = L-1, \dots, 1;$$

$$b_{i_m}^m(t) = b_{i_m}^m(t-1) - \varepsilon \times \sum_{e=1}^P \Delta_{i_m}^{m,e}(t-1), \quad (1.11)$$

$$i_m = 1 \dots N_m, \quad m = L, \dots, 1.$$

В этих формулах :

$t$  - номер итерации алгоритма обучения,

$\Delta$  - двойственные переменные, которые вводятся для упрощения вычислений и ускорения процесса обучения.

Данный алгоритм является обобщением традиционного алгоритма обратного распространения ошибки [2, 19], он дает возможность обучать сети с произвольным числом слоев. Вывод алгоритма представлен в ПРИЛОЖЕНИИ.

## 2. ОПТИЧЕСКАЯ РЕАЛИЗАЦИЯ МНОГОСЛОЙНОЙ НЕЙРОННОЙ СЕТИ

Реализовывать многослойную нейронную сеть предлагается на основе оптического вектор-матричного перемножителя, использующего в качестве узла ввода многоканальный многочастотный акустооптический модулятор света (МАОМ). МАОМ обладает уникальным сочетанием достоинств: он является самым быстро-

действующим оптическим управляемым транспарантом (кадровая скорость в пределах  $1 \div 10$  мкс); обеспечивает абсолютный контраст, что упрощает регистрацию сигналов на выходе и увеличивает динамический диапазон оптического процессора в целом; позволяет вводить в оптический тракт сразу трехмерные массивы данных (третья координата появляется при применении частотного мультиплексирования).

Основной вычислительной нагрузкой в случае реализации многослойной нейронной сети является умножение вектора на матрицу:  $A * B = C$ . На рис. 2.1 показан оптический вектор-матричный перемножитель (ОВМП) на основе МАОМ, на рис. 2.2 - варианты заполнения апертуры МАОМ, на рис. 2.3 - схема лабораторного макета, реализующего оптическую нейронную сеть. Обозначения на рисунках: ЛЛ - линейка лазеров, ПЗС - приемник с зарядовой связью.

Перемножитель работает по хорошо известному алгоритму DMAC (digital multiplication via analog convolution), однако сам процесс вычисления существенно отличается от традиционного [13]. За счет использования возможности частотного мультиплексирования ОВМП позволяет перемножать  $2^D$  и  $3^D$  потоки данных.

Вектор в качестве  $2^D$  потока вводится в линейку лазеров, а матрица в качестве  $3^D$  потока - в МАОМ (см. рис. 2.2,б). Для данной архитектуры вектор  $A$  имеет  $P$  компонент  $a_k^P$ , каждая из которых представляется восьмиразрядным двоичным числом (байтом)  $a_k^P = 1,0 : k=1$  - старший бит,  $k=K$  - младший. Матрица  $B$  имеет размерность  $M * L$ . Каждый элемент  $b_{ij}^s$  тоже представлен байтом:  $s=1$  - старший бит,  $s=S$  - младший. Таким образом, вектор  $A$  представлен двумерным двоичным массивом, а матрица  $B$  - трехмерным. В нашем случае  $K=M=L=S=8$ .

Элементы трехмерного массива  $b_{ij}^s$ , то есть матрицы  $B$ , вводятся в апертуру МАОМ так, что индекс  $i$  определяет месторасположения элемента в канале,  $j$  - номер канала,  $s$  - номер несущей частоты  $f_s$ . Когда апертура МАОМ заполнена, производится  $K$  (по числу битов в каждом элементе вектора) тактов засветки от ЛЛ, причем, как показано на рисунке, каждый элемент вектора вводится в свой лазер и первым передается младший бит  $a_k^P$ , а последним - старший  $a_1^P$ . Излучение каждого лазера равномерно распределяется по апертуре соответствующего канала. Дифрагировавший в МАОМ свет переносится оптической системой (ОС2) на линейный ПЗС,

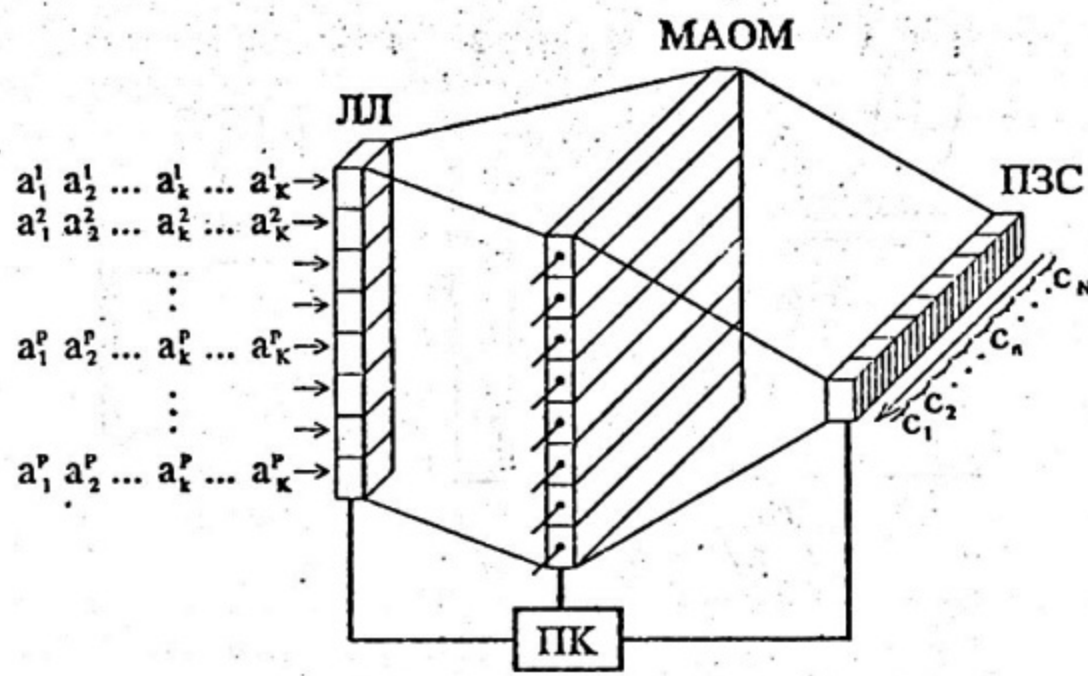
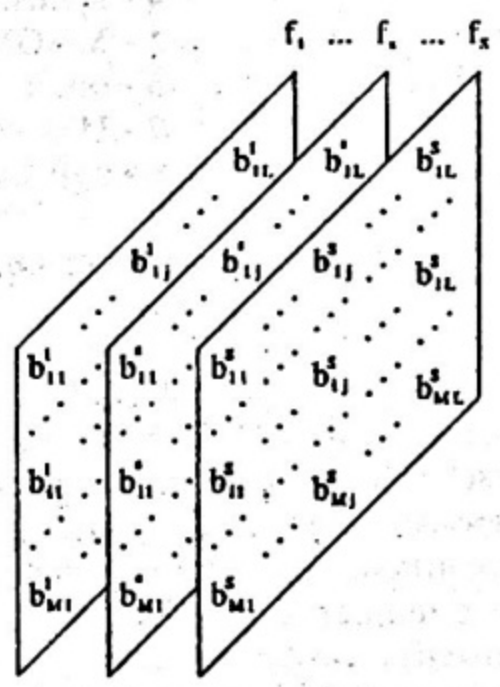


Рис.2.1. Вектор-матричный перемножитель.

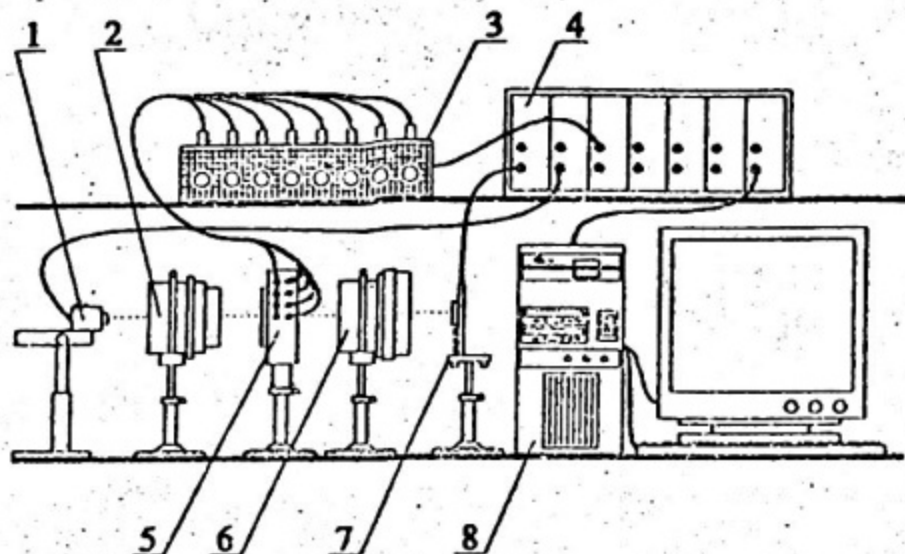
$b_{11}$	$b_{12}$	...	$b_{1j}$	...	$b_{1L}$
$b_{21}$	$b_{22}$	...	$b_{2j}$	...	$b_{2L}$
...	...	...	...	...	...
$b_{j1}$	$b_{j2}$	...	$b_{jj}$	...	$b_{jL}$
...	...	...	...	...	...
$b_{M1}$	$b_{M2}$	...	$b_{Mj}$	...	$b_{ML}$

а) заполнение апертуры по одной частоте



б) заполнение апертуры в многочастотном режиме

Рис. 2.2. Варианты заполнения апертуры МАОМ.



- 1 - линейка лазеров (ЛЛ)
- 2 - оптическая система 1 (ОС1)
- 3 - генератор-формирователь  
сигналов для управления МАОМ
- 4 - крейт КАМАК
- 5 - МАОМ
- 6 - оптическая система 2 (ОС2)
- 7 - ПЗС-фотоприемник
- 8 - персональный компьютер (ПК)

Рис. 2.3. Макет оптической нейронной сети.

на котором образуется  $M$  (по числу "окон" в канале) групп точек. В каждой группе  $S$  (по числу частот, активных в каждом окне) точек. ПЗС работает в режиме накопления со сдвигом, который синхронизирован с тактами засветки от ЛЛ таким образом, что заряд, накопленный в одной точке группы при первой экспозиции, перемещается по ПЗС к месту расположения соседней точки к моменту начала второй экспозиции. После окончания тактов засветки ПЗС содержит  $N$  групп зарядов. В нашем случае  $N=8$ , и в каждой группе содержится 15 ( $8+8-1$ ) точек свертки, которые представляют собой элементы  $c_n$  выходного вектора  $C$  в смешанном коде. После этого производится вывод из ПЗС, отцифровка и перевод в двоичное или десятичное представление элементов  $c_n$ .

Синхронизация всей системы, ввод данных в ЛЛ и МАОМ и отцифровка результатов производится ЭВМ.

Исходя из предельных параметров МАОМ, а именно: 32 канала, 32 "окна", 32 частоты, 10 мкс временная апертура, требования к остальным элементам ОП следующие: интервал следования импульсов ЛЛ около 6 нс, частота переноса в ПЗС около 120 МГц [8], динамический диапазон ПЗС не менее 50 дБ. Такой МАОМ, имея временную апертуру 10 мкс, обеспечивает скорость ввода на уровне  $10^8$  бит/с [20].

### 3. ПОЛИНОМИАЛЬНАЯ НЕЙРОННАЯ СЕТЬ

Алгоритм Хопфилда [3] в полиномиальных нейронных сетях может быть описан следующим образом. В памяти хранятся  $M$  бинарных векторов  $V^m$ ,  $m$  - номер вектора,  $V_k^m = \pm 1$ . Сила связи между нейронами определяется семейством тензоров  $b_{i...z}$ , каждый из которых однозначно определяется векторами памяти  $V^m$ :

$$b_{i...k...z} = \sum_{m=1}^M V_i^m \dots V_k^m \dots V_z^m. \quad (3.1)$$

Состояние нейронной сети в каждый отдельный момент времени определяется  $N$ -мерным ( $N$  - число нейронов в сети) вектором состояния  $a$ , каждый элемент которого  $a_k$  есть состояние  $i$ -го нейрона.

Вычислительный процесс в нейронной сети организован так, что при возбуждении сети неким входным вектором  $a$  состояние сети начинает меняться: она формирует выходной вектор

$$a'_i = \text{Tr} \left\{ \sum_{j=1}^N b_{ij} a_j + \sum_{j,k=1}^N b_{ijk} a_j a_k + \right. \\ \left. + \sum_{j,k,l=1}^N b_{ijkl} a_j a_k a_l + \dots \right\}, \quad (3.2)$$

где  $\text{Tr}$  - пороговая функция.

Вектор  $a'$  является новым вектором состояния, поэтому он вновь подается на вход и так далее до достижения некоторого устойчивого

состояния, когда выходной вектор станет равным входному. Вычисление выражения (3.2) является задачей специализированного процессора.

Рассмотрим подробно организацию вычислительного процесса в предлагаемом оптическом процессоре (рис. 3.1). Обозначения на рисунке: СИД - светоизлучающие диоды, ФД - фотодиоды.

Сначала рассмотрим случай, когда в (3.2) присутствует только второй член, то есть сеть является квадратичной:

$$a'_i = Tr \left\{ \sum_{j,k=1}^N b_{ijk} a_j a_k \right\}. \quad (3.3)$$

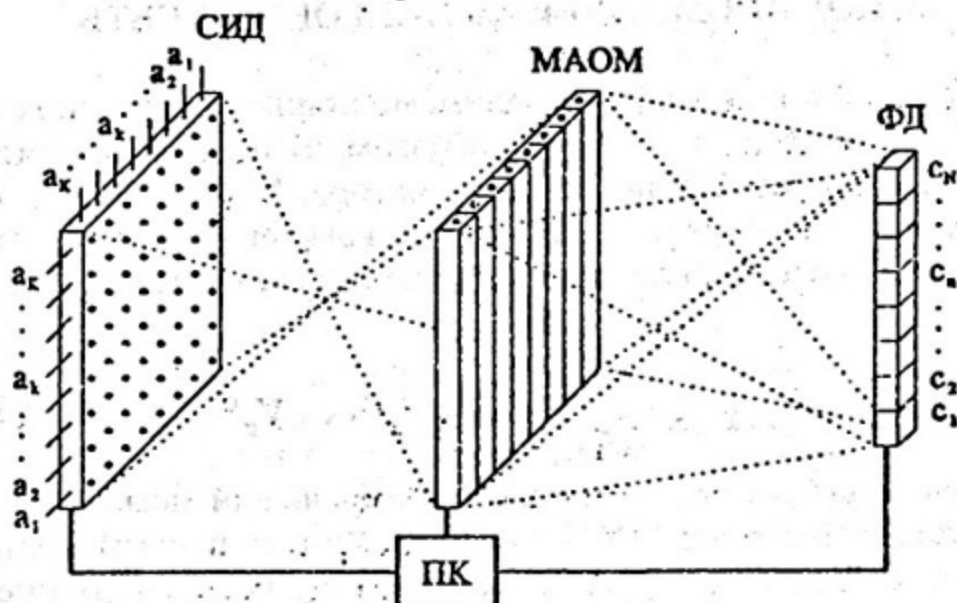


Рис. 3.1. Матрично-тензорный перемножитель.

Для реализации выражения (3.3) необходимо сформировать двумерный сигнал - матрицу внешнего произведения векторов  $a_j a_k$  - и получить  $N$  ее внутренних произведений (сверток) с  $N$  "срезами" тензора весов, то есть двумерными матрицами с элементами  $b_{ij}^1, b_{ij}^2, \dots, b_{ij}^s$ . Полученные  $N$  чисел будут представлять собой новый вектор состояния нейронной сети до пороговой операции. Тензор весов вводится в МАОМ как трехмерный массив таким образом, что положение элемента  $b_{ij}^s$  определяется значением его индексов:  $j$  - номер канала,  $i$  - положение элемента в канале к моменту экспозиции,  $s$  - номер несущей звуковой частоты  $f_s$ . Когда апертура МАОМ полностью заполнена, производится ее экспозиция двумерно

кодированным световым пучком от матрицы СИД. СИД представляет собой двумерную решетку полупроводниковых светодиодов [14] с матричной адресацией: аноды СИД в отдельно взятой строке соединены общей шиной, и соответственно катоды СИД в отдельном столбце тоже соединены. Возможно также использование матрицы светодиодов, которая существенно проще и дешевле в изготовлении и которая легко может быть совмещена с управляющей электроникой и слоем фотодиодов [18]. Распределение интенсивности света в поперечном сечении пучка является матрицей  $a_{jk}$  высшего векторного произведения вводимого в нейронную сеть вектора  $a_j$  на самого себя, что достигается подачей вектора  $a_j$  на X- и на Y-входы МЛ. Оптическая система (на рисунке не показана) отображает матрицу  $a_{jk}$  на апертуру МАОМ. Световой пучок  $a_{jk}$  дифрагирует на  $N$  частотах в  $N * N$  "окнах" МАОМ, формируя  $N$  расходящихся под углами дифракции параллельных световых пучков, суммарная интенсивность каждого из которых представляет собой элемент результирующего вектора  $a'_1$  до применения к нему пороговой операции  $T_{\theta}$ . Оптическая система (на рисунке не показана) фокусирует эти  $N$  пучков на линейку фотодиодов (ФД). Выходные сигналы ФД проходят пороговую операцию и вновь подаются на МЛ для следующей итерации. Таким образом, математическое вычисление выражения (3.3), то есть одна итерация, осуществляется в предлагаемой архитектуре за один такт.

Для непрерывной работы нейронной сети данной архитектуры тензор весов "заключивается" в МАОМ, поэтому интервал между итерациями не может быть меньше временной апертуры МАОМ (10 нс). Работа всей архитектуры организуется ЭВМ, которая осуществляет пороговую операцию, определяет степень отличия полученного вектора от предыдущего, управляет памятью МАОМ и синхронизирует работу всех узлов. При использовании в качестве фотоприемника и излучателя слоистой структуры, состоящей из слоя фотоприемников, слоя электронных пороговых и усиливающих устройств и слоя фотодиодов, можно заменить электронные цепи обратной связи на оптический тракт с обратным направлением распространения света, что снимет значительную часть вычислительной нагрузки с ЭВМ.

В том случае, когда необходимо реализовать в данной архитектуре полиномиальную сеть высокого порядка, число тактов, требующихся на одну итерацию, растет как  $N^{M-2}$ , где  $N$  - размерность

вектора, а  $M$  - порядок сети. Если в квадратичной сети происходит свертка двумерного  $2^D$  и трехмерного  $3^D$  массивов и получается одномерный:  $2^D * 3^D \rightarrow 1^D$ , то для  $M$ -степенной сети необходимо реализовать свертку  $(M-1)^D * M^D \rightarrow 1^D$ . Например, для  $M=4$  в выражении (3.2) появляется член  $b_{ijklm} a_j a_k a_l a_m$ . Сначала за  $N^2$  тактов получаем  $N^3$  произведений  $a_l a_m$  на  $N^3$  "срезов" тензора весов. Эти  $N^3$  чисел образуют трехмерный массив  $H_{ijk}$ :

$$H_{ijk} = \sum_{l,m} b_{ijklm} a_l a_m, \quad (3.4)$$

который за один такт перемножается на  $a_j a_k$  и получается выходной вектор:

$$a'_i = \sum_{j,k} H_{ijk} a_j a_k = \sum_{j,k,l,m} b_{ijklm} a_j a_k a_l a_m. \quad (3.5)$$

Таким образом, потребуется  $N^2+1$  такт для свертки  $5^D$  тензора. Сюда следует добавить число тактов для свертки  $2^D$ ,  $3^D$  и  $4^D$  тензоров. Свертка  $2^D$  тензора (обычное вектор-матричное перемножение) потребует одного такта. В этом случае на  $X$ -входы надо подать вектор с единицей в первом разряде и нулями в остальных, на  $Y$ -входы - вектор состояния  $a_s$ , в МАОМ ввести матрицу  $b_{ij}^s$  так, чтобы индекс  $j$  определял номер канала, а индекс  $s$  - номер частоты  $f_s$ . Временное кодирование по месту расположения элемента в канале должно быть таким, чтобы сигнал присутствовал только в верхней строке. Свертка  $3^D$  тензора потребует одного такта (квадратичная нейронная сеть), а для свертки  $4^D$  тензора потребуется  $N+1$  такт. Итого, для одной итерации в полной полиномиальной сети четвертого порядка потребуется  $N^2+N+4$  тактов.

#### 4. АЛГОРИТМЫ СОПРЯЖЕНИЯ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ И АППАРАТНЫХ РЕАЛИЗАЦИЙ НЕЙРОННЫХ СЕТЕЙ

При выполнении вычислений, соответствующих структуре нейронной сети, на вектор-матричном перемножителе возникают проблемы, связанные с ограниченной размерностью перемножителя, конечной точностью представления чисел, возможностью

оперировать только положительными числами. Ниже определены пути разрешения этих проблем.

Ввод ограничений на параметры сети. В процессе обучения учитываются ограничения на значения весовых коэффициентов и смещения. В литературе описываются два способа введения ограничений [2]. Первый состоит в том, что в функционал качества вводятся штрафные функции за выход из допустимого диапазона. Этот способ усложняет процесс обучения и не гарантирует того, что ограничения будут точно соблюдены. Второй способ заключается в проверке выхода из допустимого диапазона после каждой итерации обучения. Если значение параметра вышло из допустимого диапазона, то оно принудительно возвращается на границу диапазона.

Контрастирование сети. Контрастирование - это операция сведения непрерывного ряда значений параметров сети к фиксированному набору значений [2]. Эта операция не сильно ухудшает рабочие характеристики обученной нейронной сети, так как сеть проявляет устойчивость к небольшим изменениям значений синаптических коэффициентов и смещений.

Компьютерные эксперименты показали, что если контрастировать сеть после каждого шага обучения, то обучение замедляется. Другой путь - контрастирование после обучения - был более результативен.

Формирование матриц из положительных и отрицательных синаптических весов. Оптоэлектронный вектор-матричный перемножитель на базе МАОМ способен оперировать только положительными числами. Для того чтобы выполнить умножение вектора на матрицу, содержащую как положительные, так и отрицательные элементы, необходимо сформировать две матрицы. Первая из них содержит только положительные элементы исходной матрицы (отрицательные элементы заменяются нулями), вторая - только отрицательные. Последовательно умножая вектор на матрицу положительных элементов и матрицу отрицательных элементов и вычитая из первого результата второй, получаем требуемый результат. Вычитание выполняется на персональном компьютере или электронным способом.

Алгоритм разделения матриц. Основная операция, выполняемая нейронной сетью - умножение вектора на матрицу. Векторно-матричные перемножители, используемые для аппаратных

реализаций нейронных сетей, имеют фиксированную (часто небольшую) размерность. Для перемножения вектора на матрицу произвольной (сколь угодно большой) размерности может быть использован следующий алгоритм разделения матриц [15].

1. Пусть вектор-матричный перемножитель выполняет умножение матрицы размерностью  $K * L$  на вектор размерностью  $L$ . Необходимо перемножить матрицу  $W = \|w_{ij}\|_{MN}$  и вектор  $X = \|x_j\|_{NL}$ :  $W * X = S$ .

2. Строится новая матрица  $W'$  путем добавления к исходной матрице  $W$   $\beta$  нулевых столбцов и  $\alpha$  нулевых строк так, чтобы  $N + \beta = H * K$ ,  $H \geq 1$ ,  $M + \alpha = E * L$ ,  $E \geq 1$ .

3. Матрица  $W'$  делится на  $(H * E)$  матриц  $W^{eh}$  размерностью  $K * L$ , где каждая матрица  $W^{eh}$  ( $e = 1 \dots E$ ,  $h = 1 \dots H$ ) содержит элементы матрицы  $W'$ :  $W^{eh} = \|W'_{(e-1)*K+k, (h-1)*L+q}\|$ ,  $k = 1 \dots K$ ,  $q = 1 \dots L$ .

4. Аналогично строится вектор  $X'$ :

$X' = (X^1, X^2, \dots, X^h, \dots, X^H)^T$ ,  $X^h = \|X'_{(h-1)*L+q}\|$ ,  $q = 1 \dots L$ .

5. Результат умножения  $W'$  на  $X'$  - вектор  $S'$ :

$S' = (S^1, S^2, \dots, S^e, \dots, S^E)^T$ ,  $S^e = \|S'_{(e-1)*K+k}\|$ ,  $k = 1 \dots K$ .

6. Векторы  $S^{eh}$  получаются в результате умножения матриц  $W^{eh}$  на векторы  $X^h$ :  $W^{eh} * X^h = S^{eh}$ ,  $e = 1 \dots E$ ,  $h = 1 \dots H$ .

7. Векторы  $S^e$  вычисляются по формуле

$$S^e = \sum_{h=1}^H S^{eh}.$$

8. Первые  $M$  компонент вектора  $S'$  - это искомый вектор  $S$ .

Цифровое разделение. Для повышения точности представления значений, которые вводятся в вектор-матричный перемножитель, используется алгоритм цифрового разделения [16]. Пусть элементы вектора и матрицы представлены двухбайтными числами, а перемножитель способен оперировать только однобайтными числами. В этом случае строятся два вектора и две матрицы. Элементы первого вектора представляют собой младшие байты исходного вектора, элементы второго - старшие байты. Аналогично строятся матрицы. Парно умножая однобайтные векторы и матрицы и складывая со сдвигом получившиеся произведения, получаем требуемый результат.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Галушкин А.И., Судариков В.А., Шабанов Е.В. Методика решения задач на нейрокомпьютерах // *Нейрокомпьютер*. No. 1. 1992. С. 24-28.
2. Горбань А.Н. Обучение нейронных сетей. М.: СП ПараГраф. 1990.
3. Евтихийев Н.Н., Лазарев В.Н., Перепелица В.В. Оптическая реализация квадратичной нейронной сети // *Оптическая память и нейронные сети: Международный симпозиум*. Звенигород. 1990.
4. Евтихийев Н.Н., Оныкий Б.Н., Перепелица В.В., Щербаков И.Б. Разработка специализированного оптоэлектронного нейрокомпьютера // *Всероссийская научно-практическая конференция "Высшая школа России и конверсия"* : Тезисы докладов. М., 1993. С.149-151.
5. Морозов В.Н. Оптоэлектронные матричные процессоры. М.: Радио и связь. 1986.
6. Оныкий Б.Н., Сипицын С.В., Щербаков И.Б. Перспективы использования нейрокомпьютеров в бортовых интеллектуальных системах // *Вычислительные системы на базе транспьютеров и параллельные вычисления* : Материалы семинара. М., 1992. С.28-41.
7. Оныкий Б.Н., Щербаков И.Б. Перспективы использования нейронных сетей для построения экспертных систем, обучаемых на примерах // *Экспертные системы. Базы знаний и данных* : Материалы семинара. М., 1992. С. 149-153.
8. Приемники оптического излучения : Справочник / Под ред. М.Д.Аксененко, М.Л.Бараночкикова. М.: Радио и связь, 1987.
9. Уоссермен Ф. Нейрокомпьютерная техника : Теория и практика. М.: Мир, 1992.
10. Широков Ф.В. На пути к нейрокомпьютерам. // *Итоги науки и техники. Серия : Физические и математические модели нейронных сетей*. Т.1. М.: ВИНТИ. 1990.
11. Щербаков И.Б. Использование сети Хопфилда для решения задачи распознавания зашумленных образов // *Всесоюзный семинар "Распределенная обработка информации. - IV"* : Тезисы докладов. Новосибирск. 1991. С. 83.
12. M.Abushagur, H.Caulfield. Proc.SPIE. V.639. 1986. P.63.
13. P.R.Beaudet, A.P.Goutzoulis et al. // *Appl. Opt.* V.25. No.18. 1986. P.3097.
14. N.N.Evtihiev, A.Korolev-Korotkov, A.P.Dolgiy, N.A.Escpkina, A.P.Lavrov SAR System Optoelectronic Compact Processor With LED Array // *Proc. SPIE*. V.1704. 1992. P.115-124.

15. N.N. Evtihiev, V.N. Lazarev, B.N. Onyky, V.V. Perepelitsa, I.B. Scherbakov. Optoelectronic neurocomputer based on digital optical vector-matrix multiplier // International conference on optical information processing. St.- Peterburg. 2-7 Aug. 1993.
16. Gary C.K. Matrix-vector multiplication using digital partitioning for more accurate optical computing // Applied Optics. Vol. 31. No. 29. 1992. P. 6205-6211.
17. Graph H.P. et al. VLSI implementation of a neural network model // IEEE Computer. V.21. No.3. March 1988.
18. J.Javel, Y.Lee, N.Olsson. Two-dimensional Array Microlaser For Photonic Switching // International Conf. "Photonic Switching": Techn. Digest. Japan, Kobe. 1990. P. 14-19.
19. Muller B., Reinhardt J. Neural Networks. Springer-Verlag. 1990.
20. A. Pugovkin, A.Briganets, V.Dmitriev, L.Serebrennikov. Functional units of multichannel acoustooptic processors // School-seminar "Acoustooptics. Researches and developments". Leningrad. June 27 - July 1 1990.

ВЫВОД ГРАДИЕНТНОГО АЛГОРИТМА ОБУЧЕНИЯ  
МНОГОСЛОЙНЫХ ОДНОРОДНЫХ НЕЙРОННЫХ СЕТЕЙ  
С ПОСЛЕДОВАТЕЛЬНЫМИ СВЯЗЯМИ

Пусть  $t$  - номер шага обучения нейронной сети. В момент времени  $t=0$  весовым коэффициентам нейронов сети случайным образом присваиваются значения из интервала  $(w_{\min}, w_{\max})$ :

$$w_{i_m i_{m-1}} = (w_{\max} - w_{\min}) \times \rho + w_{\min}, \quad i_m = 1 \dots N_m, \quad m=1 \dots L,$$

где  $\rho$  - случайная величина, равномерно распределенная в интервале  $(0,1)$ .

Смещениям нейронов сети случайным образом присваиваются значения из интервала  $(b_{\min}, b_{\max})$ :

$$b_{i_m} = (b_{\max} - b_{\min}) \times \rho + b_{\min}, \quad i_m = 1 \dots N_m, \quad m=1 \dots L.$$

Как правило, абсолютные значения  $w_{\min}$  и  $w_{\max}$ ,  $b_{\min}$  и  $b_{\max}$  - некоторые небольшие числа (например:  $w_{\min} = -0.1$ ,  $w_{\max} = 0.1$ ,  $b_{\min} = -0.1$ ,  $b_{\max} = 0.1$ ). Значения  $w_{\min}$  и  $w_{\max}$  не должны выходить за пределы допустимых значений весовых коэффициентов, значения  $b_{\min}$  и  $b_{\max}$  - за пределы допустимых значений смещений. Пределы допустимых значений определяются до начала обучения сети.

На шаге обучения  $t$  параметрам нейронной сети присваиваются новые значения:

$$w_{i_m i_{m-1}}^m(t) = w_{i_m i_{m-1}}^m(t-1) + \delta w_{i_m i_{m-1}}^m(t-1),$$

$$b_{i_m}^m(t) = b_{i_m}^m(t-1) + \delta b_{i_m}^m(t-1),$$

$$i_m = 1 \dots N_m, \quad m=1 \dots L.$$

Новые значения параметров сети определяются в соответствии с методом скорейшего спуска. На каждом шаге обучения вычисляется градиент функции ошибки в адаптивном пространстве сети и делается шаг по адаптивному рельефу в направлении, противоположном градиенту.

Адаптивное пространство сети имеет размерность, которая на единицу превосходит число параметров нейронной сети, настраиваемых в процессе обучения:

$$dim = \sum_{m=1}^L N_m \times N_{m-1} + \sum_{m=1}^L N_m + 1.$$

По осям координат в адаптивном пространстве откладываются значения синаптических весов и смещений нейронов сети, значения функции ошибки.

Адаптивный рельеф - это поверхность в адаптивном пространстве, каждая точка которой определяется следующим набором переменных:

$$\left( \|w_{i_1 i_0}^1\|, \|b_{i_1}^1\|, \|w_{i_2 i_1}^2\|, \|b_{i_2}^2\|, \dots, \|w_{i_m i_{m-1}}^m\|, \|b_{i_m}^m\|, \dots \right. \\ \left. \dots, \|w_{i_L i_{L-1}}^L\|, \|b_{i_L}^L\|, D \right),$$

где

$\|w_{i_m i_{m-1}}^m\|$  - матрица весовых коэффициентов  $m$ -го слоя,

$\|b_{i_m}^m\|$  - столбец смещений  $m$ -го слоя ( $m = 1 \dots L$ ),

$D$  - значение функции ошибки, соответствующее значениям весовых коэффициентов и смещений нейронной сети.

Задача минимизации функции ошибки сводится к поиску минимума на адаптивном рельефе.

Для того чтобы определить значения  $\delta w$  и  $\delta b$ , необходимо вычислить частные производные функции ошибки по каждому из параметров нейронной сети, настраиваемых в процессе обучения.

Синаптические веса последнего слоя:

$$\delta w_{j_L j_{L-1}}^L = -\epsilon \times \frac{\partial D}{\partial w_{j_L j_{L-1}}^L} = \\ = -\epsilon \times \frac{\partial}{\partial w_{j_L j_{L-1}}^L} \left\{ \frac{1}{2} \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - y_{i_L}^{L,e}]^2 \right\} = \\ = -\epsilon \times \frac{\partial}{\partial w_{j_L j_{L-1}}^L} \left\{ \frac{1}{2} \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - f_{i_L}^L(s_{i_L}^{L,e})]^2 \right\} =$$

$$= \varepsilon \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - f_{i_L}^L(s_{i_L}^{L,e})] \times (f_{i_L}^L(s_{i_L}^{L,e}))' \times \frac{\partial s_{i_L}^{L,e}}{\partial w_{j_L j_{L-1}}^L};$$

$$\frac{\partial s_{i_L}^{L,e}}{\partial w_{j_L j_{L-1}}^L} = \frac{\partial}{\partial w_{j_L j_{L-1}}^L} \left\{ \sum_{i_{L-1}=1}^{N_{L-1}} w_{i_L i_{L-1}}^L \times y_{i_{L-1}}^{L-1,e} - b_{i_L}^L \right\} =$$

$$= \begin{cases} y_{j_{L-1}}^{L-1,e}, & i_L = j_L; \\ 0, & i_L \neq j_L; \end{cases}$$

$$\delta w_{j_L j_{L-1}}^L = \varepsilon \times \sum_{e=1}^P [\zeta_{j_L}^e - f_{j_L}^L(s_{j_L}^{L,e})] \times (f_{j_L}^L(s_{j_L}^{L,e}))' \times y_{j_{L-1}}^{L-1,e};$$

$$\Delta_{j_L}^{L,e} = [\zeta_{j_L}^e - f_{j_L}^L(s_{j_L}^{L,e})] \times (f_{j_L}^L(s_{j_L}^{L,e}))';$$

$$\delta w_{j_L j_{L-1}}^L = \varepsilon \times \sum_{e=1}^P \Delta_{j_L}^{L,e} \times y_{j_{L-1}}^{L-1,e}.$$

Смещения последнего слоя :

$$\delta b_{j_L}^L = -\varepsilon \times \frac{\partial D}{\partial b_{j_L}^L} = -\varepsilon \times \frac{\partial}{\partial b_{j_L}^L} \left\{ \frac{1}{2} \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - y_{i_L}^{L,e}]^2 \right\} =$$

$$= -\varepsilon \times \frac{\partial}{\partial b_{j_L}^L} \left\{ \frac{1}{2} \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - f_{i_L}^L(s_{i_L}^{L,e})]^2 \right\} =$$

$$= \varepsilon \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - f_{i_L}^L(s_{i_L}^{L,e})] \times (f_{i_L}^L(s_{i_L}^{L,e}))' \times \frac{\partial s_{i_L}^{L,e}}{\partial b_{j_L}^L};$$

$$\frac{\partial s_{i_L}^{L,e}}{\partial b_{j_L}^L} = \frac{\partial}{\partial b_{j_L}^L} \left\{ \sum_{i_{L-1}=1}^{N_{L-1}} w_{i_L i_{L-1}}^L \times y_{i_{L-1}}^{L-1,e} - b_{i_L}^L \right\} = \begin{cases} -1, & i_L = j_L; \\ 0, & i_L \neq j_L; \end{cases}$$

$$\begin{aligned} \delta b_{j_L}^L &= -\varepsilon \times \sum_{e=1}^P [\zeta_{j_L}^e - f_{j_L}^L(s_{j_L}^{L,e})] \times (f_{j_L}^L(s_{j_L}^{L,e}))' = \\ &= -\varepsilon \times \sum_{e=1}^P \Delta_{j_L}^{L,e}. \end{aligned}$$

Синаптические веса предпоследнего слоя :

$$\begin{aligned} \delta w_{j_{L-1}j_{L-2}}^{L-1} &= -\varepsilon \times \frac{\partial D}{\partial w_{j_{L-1}j_{L-2}}^{L-1}} = \\ &= -\varepsilon \times \frac{\partial}{\partial w_{j_{L-1}j_{L-2}}^{L-1}} \left\{ \frac{1}{2} \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - y_{i_L}^{L,e}]^2 \right\} = \\ &= -\varepsilon \times \frac{\partial}{\partial w_{j_{L-1}j_{L-2}}^{L-1}} \left\{ \frac{1}{2} \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - f_{i_L}^L(s_{i_L}^{L,e})]^2 \right\} = \\ &= \varepsilon \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - f_{i_L}^L(s_{i_L}^{L,e})] \times (f_{i_L}^L(s_{i_L}^{L,e}))' \times \frac{\partial s_{i_L}^{L,e}}{\partial w_{j_{L-1}j_{L-2}}^{L-1}} ; \\ \frac{\partial s_{i_L}^{L,e}}{\partial w_{j_{L-1}j_{L-2}}^{L-1}} &= \frac{\partial}{\partial w_{j_{L-1}j_{L-2}}^{L-1}} \left\{ \sum_{i_{L-1}=1}^{N_{L-1}} w_{i_L i_{L-1}}^L \times y_{i_{L-1}}^{L-1,e} - b_{i_L}^L \right\} = \\ &= \sum_{i_{L-1}=1}^{N_{L-1}} w_{i_L i_{L-1}}^L \times (f_{i_{L-1}}^{L-1}(s_{i_{L-1}}^{L-1,e}))' \times \frac{\partial s_{i_{L-1}}^{L-1,e}}{\partial w_{j_{L-1}j_{L-2}}^{L-1}} ; \\ \frac{\partial s_{i_{L-1}}^{L-1,e}}{\partial w_{j_{L-1}j_{L-2}}^{L-1}} &= \frac{\partial}{\partial w_{j_{L-1}j_{L-2}}^{L-1}} \left\{ \sum_{i_{L-2}=1}^{N_{L-2}} w_{i_{L-1} i_{L-2}}^{L-1} \times y_{i_{L-2}}^{L-2,e} - b_{i_{L-1}}^{L-1} \right\} = \\ &= \begin{cases} y_{j_{L-2}}^{L-2,e}, & i_{L-1} = j_{L-1}, \\ 0, & i_{L-1} \neq j_{L-1} \end{cases} \end{aligned}$$

$$\delta W_{j_{L-1}j_{L-2}}^{L-1} = \varepsilon \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - f_{i_L}^L(s_{i_L}^{L,e})] \times (f_{i_L}^L(s_{i_L}^{L,e}))' \times$$

$$\times W_{i_L j_{L-1}}^L \times (f_{j_{L-1}}^{L-1}(s_{j_{L-1}}^{L-1,e}))' \times y_{j_{L-2}}^{L-2,e} =$$

$$= \varepsilon \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} \Delta_{i_L}^{L,e} \times W_{i_L j_{L-1}}^L \times (f_{j_{L-1}}^{L-1}(s_{j_{L-1}}^{L-1,e}))' \times y_{j_{L-2}}^{L-2,e};$$

$$\Delta_{j_{L-1}}^{L-1,e} = \sum_{i_L=1}^{N_L} \Delta_{i_L}^{L,e} \times W_{i_L j_{L-1}}^L \times (f_{j_{L-1}}^{L-1}(s_{j_{L-1}}^{L-1,e}))';$$

$$\delta W_{j_{L-1}j_{L-2}}^{L-1} = \varepsilon \times \sum_{e=1}^P \Delta_{j_{L-1}}^{L-1,e} \times y_{j_{L-2}}^{L-2,e}.$$

Смещения предпоследнего слоя :

$$\delta b_{j_{L-1}}^{L-1} = -\varepsilon \times \frac{\partial D}{\partial b_{j_{L-1}}^{L-1}} =$$

$$= -\varepsilon \times \frac{\partial}{\partial b_{j_{L-1}}^{L-1}} \left\{ \frac{1}{2} \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - y_{i_L}^{L,e}]^2 \right\} =$$

$$= -\varepsilon \times \frac{\partial}{\partial b_{j_{L-1}}^{L-1}} \left\{ \frac{1}{2} \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - f_{i_L}^L(s_{i_L}^{L,e})]^2 \right\} =$$

$$= \varepsilon \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - f_{i_L}^L(s_{i_L}^{L,e})] \times (f_{i_L}^L(s_{i_L}^{L,e}))' \times \frac{\partial s_{i_L}^{L,e}}{\partial b_{j_{L-1}}^{L-1}};$$

$$\frac{\partial s_{i_L}^{L,e}}{\partial b_{j_{L-1}}^{L-1}} = \frac{\partial}{\partial b_{j_{L-1}}^{L-1}} \left\{ \sum_{i_{L-1}=1}^{N_{L-1}} W_{i_L i_{L-1}}^L \times y_{i_{L-1}}^{L-1,e} - b_{i_L}^L \right\} =$$

$$= \sum_{i_{L-1}=1}^{N_{L-1}} w_{i_{L-1}i_{L-1}}^L \times (f_{i_{L-1}}^{L-1}(S_{i_{L-1}}^{L-1, \theta}))' \times \frac{\partial S_{i_{L-1}}^{L-1, \theta}}{\partial b_{j_{L-1}}^{L-1}};$$

$$\frac{\partial S_{i_{L-1}}^{L-1, \theta}}{\partial b_{j_{L-1}}^{L-1}} = \frac{\partial}{\partial b_{j_{L-1}}^{L-1}} \left\{ \sum_{i_{L-2}=1}^{N_{L-2}} w_{i_{L-1}i_{L-2}}^{L-1} \times y_{i_{L-2}}^{L-2, \theta} - b_{i_{L-1}}^{L-1} \right\} =$$

$$= \begin{cases} -1, & i_{L-1} = j_{L-1}; \\ 0, & i_{L-1} \neq j_{L-1}; \end{cases}$$

$$\delta b_{j_{L-1}}^{L-1} = -\varepsilon \times \sum_{\theta=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^{\theta} - f_{i_L}^L(S_{i_L}^{L, \theta})] \times (f_{i_L}^L(S_{i_L}^{L, \theta}))' \times$$

$$\times w_{i_L j_{L-1}}^L (f_{j_{L-1}}^{L-1}(S_{j_{L-1}}^{L-1, \theta}))' = -\varepsilon \sum_{\theta=1}^P \sum_{i_L=1}^{N_L} \Delta_{i_L}^{L, \theta} \times w_{i_L j_{L-1}}^L \times$$

$$\times (f_{j_{L-1}}^{L-1}(S_{j_{L-1}}^{L-1, \theta}))' = -\varepsilon \times \sum_{\theta=1}^P \Delta_{j_{L-1}}^{L-1, \theta}.$$

• • • • •

Синаптические веса m-го слоя :

$$\delta w_{j_m j_{m-1}}^m = -\varepsilon \times \frac{\partial D}{\partial w_{j_m j_{m-1}}^m} =$$

$$= -\varepsilon \times \frac{\partial}{\partial w_{j_m j_{m-1}}^m} \left\{ \frac{1}{2} \times \sum_{\theta=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^{\theta} - y_{i_L}^{L, \theta}]^2 \right\} =$$

$$= -\varepsilon \times \frac{\partial}{\partial w_{j_m j_{m-1}}^m} \left\{ \frac{1}{2} \times \sum_{\theta=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^{\theta} - f_{i_L}^L(S_{i_L}^{L, \theta})]^2 \right\} =$$

$$= \varepsilon \times \sum_{\sigma=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^{\sigma} - f_{i_L}^L(s_{i_L}^{L,\sigma})] \times (f_{i_L}^L(s_{i_L}^{L,\sigma}))' \times \frac{\partial s_{i_L}^{L,\sigma}}{\partial w_{j_m j_{m-1}}^m}$$

Для некоторого  $i_n$  ( $n = L, L-1, \dots, m$ ):

$$\frac{\partial s_{i_n}^{n,\sigma}}{\partial w_{j_m j_{m-1}}^m} = \frac{\partial}{\partial w_{j_m j_{m-1}}^m} \left\{ \sum_{i_{n-1}=1}^{N_{n-1}} w_{i_n i_{n-1}}^n \times y_{i_{n-1}}^{n-1,\sigma} - b_{i_n}^n \right\} =$$

$$= \begin{cases} \begin{cases} y_{j_{m-1}}^{m-1,\sigma}, & i_n = j_m, \quad n=m; \\ 0, & i_n > j_m \end{cases}, & n=m; \\ \sum_{i_{n-1}=1}^{N_{n-1}} w_{i_n i_{n-1}}^n \times (f_{i_{n-1}}^{n-1}(s_{i_{n-1}}^{n-1,\sigma}))' \times \frac{\partial s_{i_{n-1}}^{n-1,\sigma}}{\partial w_{j_m j_{m-1}}^m}, & n > m. \end{cases}$$

Введем обозначения:

$$\Delta_{j_L}^{L,\sigma} = [\zeta_{j_L}^{\sigma} - f_{j_L}^L(s_{j_L}^{L,\sigma})] \times (f_{j_L}^L(s_{j_L}^{L,\sigma}))';$$

$$\Delta_{j_n}^{n,\sigma} = \sum_{i_{n+1}=1}^{N_{n+1}} \Delta_{i_{n+1}}^{n+1,\sigma} \times w_{i_{n+1} j_n}^{n+1} \times (f_{j_n}^n(s_{j_n}^{n,\sigma}))', \quad n=1 \dots L-1.$$

$$\delta w_{j_m j_{m-1}}^m = \varepsilon \times \sum_{\sigma=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^{\sigma} - f_{i_L}^L(s_{i_L}^{L,\sigma})] \times (f_{i_L}^L(s_{i_L}^{L,\sigma}))' \times$$

$$\times \sum_{i_{L-1}=1}^{N_{L-1}} w_{i_L i_{L-1}}^L \times (f_{i_{L-1}}^{L-1}(s_{i_{L-1}}^{L-1,\sigma}))' \times \sum_{i_{L-2}=1}^{N_{L-2}} w_{i_{L-1} i_{L-2}}^{L-1} \times$$

$$\times (f_{i_{L-2}}^{L-2}(s_{i_{L-2}}^{L-2,\sigma}))' \times \dots \times \sum_{i_{m+1}=1}^{N_{m+1}} w_{i_{m+2} i_{m+1}}^{m+2} \times (f_{i_{m+1}}^{m+1}(s_{i_{m+1}}^{m+1,\sigma}))' \times$$

$$\times w_{i_{m+1} j_m}^{m+1} \times (f_{j_m}^m(s_{j_m}^{m,\sigma}))' \times y_{j_{m-1}}^{m-1,\sigma} =$$

$$\begin{aligned}
&= \varepsilon \times \sum_{e=1}^P \sum_{i_{L-1}=1}^{N_{L-1}} \sum_{i_L=1}^{N_L} \Delta_{i_L}^{L,e} \times W_{i_L i_{L-1}}^L \times (f_{i_{L-1}}^{L-1}(S_{i_{L-1}}^{L-1,e}))' \times \\
&\times \sum_{i_{L-2}=1}^{N_{L-2}} W_{i_{L-1} i_{L-2}}^{L-1} \times (f_{i_{L-2}}^{L-2}(S_{i_{L-2}}^{L-2,e}))' \times \dots \times \sum_{i_{m+1}=1}^{N_{m+1}} W_{i_{m+2} i_{m+1}}^{m+2} \times \\
&\times (f_{i_{m+1}}^{m+1}(S_{i_{m+1}}^{m+1,e}))' \times W_{i_{m+1} j_m}^{m+1} \times (f_{j_m}^m(S_{j_m}^m,e))' \times y_{j_{m-1}}^{m-1,e} = \\
&= \varepsilon \times \sum_{e=1}^P \sum_{i_{L-2}=1}^{N_{L-2}} \sum_{i_{L-1}=1}^{N_{L-1}} \Delta_{i_{L-1}}^{L-1,e} \times W_{i_{L-1} i_{L-2}}^{L-1} \times (f_{i_{L-2}}^{L-2}(S_{i_{L-2}}^{L-2,e}))' \times \\
&\times \dots \times \sum_{i_{m+1}=1}^{N_{m+1}} W_{i_{m+2} i_{m+1}}^{m+2} \times (f_{i_{m+1}}^{m+1}(S_{i_{m+1}}^{m+1,e}))' \times W_{i_{m+1} j_m}^{m+1} \times \\
&\times (f_{j_m}^m(S_{j_m}^m,e))' \times y_{j_{m-1}}^{m-1,e} = \dots = \varepsilon \times \sum_{e=1}^P \sum_{i_{m+1}=1}^{N_{m+1}} \Delta_{i_{m+1}}^{m+1,e} \times W_{i_{m+1} j_m}^{m+1} \times \\
&\times (f_{j_m}^m(S_{j_m}^m,e))' \times y_{j_{m-1}}^{m-1,e} = \varepsilon \times \sum_{e=1}^P \Delta_{j_m}^{m,e} \times y_{j_{m-1}}^{m-1,e}.
\end{aligned}$$

Смещения  $m$ -го слоя :

$$\begin{aligned}
\delta b_{j_m}^m &= -\varepsilon \times \frac{\partial \mathcal{D}}{\partial b_{j_m}^m} = -\varepsilon \times \frac{\partial}{\partial b_{j_m}^m} \left\{ \frac{1}{2} \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - y_{i_L}^{L,e}]^2 \right\} = \\
&= -\varepsilon \times \frac{\partial}{\partial b_{j_m}^m} \left\{ \frac{1}{2} \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - f_{i_L}^L(S_{i_L}^{L,e})]^2 \right\} =
\end{aligned}$$

$$= \varepsilon \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - f_{i_L}^L(s_{i_L}^{L,e})] \times (f_{i_L}^L(s_{i_L}^{L,e}))' \times \frac{\partial s_{i_L}^{L,e}}{\partial b_{j_m}^m}$$

Для некоторого  $i_n$  ( $n = L, L-1, \dots, m$ ):

$$\frac{\partial s_{i_n}^{n,e}}{\partial b_{j_m}^m} = \frac{\partial}{\partial b_{j_m}^m} \left\{ \sum_{i_{n-1}=1}^{N_{n-1}} w_{i_n i_{n-1}}^n \times y_{i_{n-1}}^{n-1,e} - b_{i_n}^n \right\} =$$

$$= \begin{cases} \begin{cases} -1, & i_n = j_m, \\ 0, & i_n \neq j_m \end{cases}, & n = m; \\ \sum_{i_{n-1}=1}^{N_{n-1}} w_{i_n i_{n-1}}^n \times (f_{i_{n-1}}^{n-1}(s_{i_{n-1}}^{n-1,e}))' \times \frac{\partial s_{i_{n-1}}^{n-1,e}}{\partial b_{j_m}^m}, & n > m. \end{cases}$$

$$\delta b_{j_m}^m = -\varepsilon \times \sum_{e=1}^P \sum_{i_L=1}^{N_L} [\zeta_{i_L}^e - f_{i_L}^L(s_{i_L}^{L,e})] \times (f_{i_L}^L(s_{i_L}^{L,e}))' \times$$

$$\times \sum_{i_{L-1}=1}^{N_{L-1}} w_{i_L i_{L-1}}^L \times (f_{i_{L-1}}^{L-1}(s_{i_{L-1}}^{L-1,e}))' \times \sum_{i_{L-2}=1}^{N_{L-2}} w_{i_{L-1} i_{L-2}}^{L-1} \times$$

$$\times (f_{i_{L-2}}^{L-2}(s_{i_{L-2}}^{L-2,e}))' \times \dots \times \sum_{i_{m+1}=1}^{N_{m+1}} w_{i_{m+2} i_{m+1}}^{m+2} \times (f_{i_{m+1}}^{m+1}(s_{i_{m+1}}^{m+1,e}))' \times$$

$$\times w_{i_{m+1} j_m}^{m+1} \times (f_{j_m}^m(s_{j_m}^m))' = -\varepsilon \times \sum_{e=1}^P \sum_{i_{L-1}=1}^{N_{L-1}} \sum_{i_L=1}^{N_L} \Delta_{i_L}^{L,e} \times w_{i_L i_{L-1}}^L \times$$

$$\times (f_{i_{L-1}}^{L-1}(s_{i_{L-1}}^{L-1,e}))' \times \sum_{i_{L-2}=1}^{N_{L-2}} w_{i_{L-1} i_{L-2}}^{L-1} \times (f_{i_{L-2}}^{L-2}(s_{i_{L-2}}^{L-2,e}))' \times \dots \times$$

$$\times \sum_{i_{m+1}=1}^{N_{m+1}} w_{i_{m+2} i_{m+1}}^{m+2} \times (f_{i_{m+1}}^{m+1}(s_{i_{m+1}}^{m+1,e}))' \times w_{i_{m+1} j_m}^{m+1} \times (f_{j_m}^m(s_{j_m}^m))' =$$

$$\begin{aligned}
&= -\varepsilon \times \sum_{\theta=1}^P \sum_{i_{L-2}=1}^{N_{L-2}} \sum_{i_{L-1}=1}^{N_{L-1}} \Delta_{i_{L-1}}^{L-1, \theta} \times W_{i_{L-1} i_{L-2}}^{L-1} \times (f_{i_{L-2}}^{L-2} (S_{i_{L-2}}^{L-2, \theta}))' \times \\
&\times \dots \times \sum_{i_{m+1}=1}^{N_{m+1}} W_{i_{m+2} i_{m+1}}^{m+2} \times (f_{i_{m+1}}^{m+1} (S_{i_{m+1}}^{m+1, \theta}))' \times W_{i_{m+1} j_m}^{m+1} \times \\
&\times (f_{j_m}^m (S_{j_m}^{m, \theta}))' = \dots = -\varepsilon \times \sum_{\theta=1}^P \sum_{i_{m+1}=1}^{N_{m+1}} \Delta_{i_{m+1}}^{m+1, \theta} \times W_{i_{m+1} j_m}^{m+1} \times \\
&\times (f_{j_m}^m (S_{j_m}^{m, \theta}))' = -\varepsilon \times \sum_{\theta=1}^P \Delta_{j_m}^{m, \theta}.
\end{aligned}$$

• • • • •

Синаптические веса первого слоя :

$$\delta w_{j_1 j_0}^1 = \varepsilon \times \sum_{\theta=1}^P \Delta_{j_1}^{1, \theta} \times y_{j_0}^{0, \theta} = \varepsilon \times \sum_{\theta=1}^P \Delta_{j_1}^{1, \theta} \times \sigma_{j_0}^{\theta}.$$

Смещения первого слоя :

$$\delta b_{j_1}^1 = -\varepsilon \sum_{\theta=1}^P \Delta_{j_1}^{1, \theta}.$$

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	3
1. МНОГОСЛОЙНАЯ НЕЙРОННАЯ СЕТЬ.....	5
2. ОПТИЧЕСКАЯ РЕАЛИЗАЦИЯ МНОГОСЛОЙНОЙ НЕЙРОННОЙ СЕТИ.....	9
3. ПОЛИНОМИАЛЬНАЯ НЕЙРОННАЯ СЕТЬ.....	13
4. АЛГОРИТМЫ СОПРЯЖЕНИЯ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ И АППАРАТНЫХ РЕАЛИЗАЦИЙ НЕЙРОННЫХ СЕТЕЙ.....	16
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	19
ПРИЛОЖЕНИЕ. ВЫВОД ГРАДИЕНТНОГО АЛГОРИТМА ОБУЧЕНИЯ МНОГОСЛОЙНЫХ ОДНОРОДНЫХ НЕЙРОННЫХ СЕТЕЙ С ПОСЛЕДОВАТЕЛЬНЫМИ СВЯЗЯМИ.....	21

Николай Николаевич Евтихийев  
Борис Николаевич Оныкий  
Вадим Викторович Перспеклица  
Игорь Борисович Щербаков

**МАТЕМАТИЧЕСКИЕ МОДЕЛИ И ОПТИЧЕСКИЕ РЕАЛИЗАЦИИ  
МНОГОСЛОЙНЫХ И ПОЛИНОМИАЛЬНЫХ  
НЕЙРОННЫХ СЕТЕЙ**

Рукопись поступила в издательский отдел 31.03.94 г.

Редактор Н.Н. Антонова

Ответственный за выпуск И.Б. Щербаков

Лицензия ЛР No 020676 от 09.12.92г.

---

Подписано в печать 25.04.94	Формат 60x84 1/16
Печл. 2,0 Уч.-издл. 2,0 Тираж 100 экз.	Изд. No 004-94 Заказ 592

---

Московский государственный инженерно-физический институт  
(технический университет). Типография МИФИ.  
115409, Москва, Каширское ш., 31