

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН

УНИВЕРСИТЕТ ИМЕНИ СУЛЕЙМАНА ДЕМИРЕЛЯ

ИНЖЕНЕРНЫЙ ФАКУЛЬТЕТ



Кафедра «Компьютерная инженерия»

УДК 004.021

ГАЗИЗОВ ТИМУР

**РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ПОМОЩЬЮ
ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ**

Магистерская диссертация, представленная на соискание
академической степени магистра технических наук
по специальности

6М070400 – «Вычислительная техника и программное обеспечение»

Каскелен, 2013

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН

УНИВЕРСИТЕТ ИМЕНИ СУЛЕЙМАНА ДЕМИРЕЛЯ

ИНЖЕНЕРНЫЙ ФАКУЛЬТЕТ



Кафедра «Компьютерная инженерия»

«Допущен к защите»
Заведующий кафедрой
д.т.н., профессор Амиргалиев Е.Н.

« 12 » 2013 г.

Заведующий департаментом
послевузовского образования
к.ф.н. Айдогду Шабан

« 12 » 2013 г.

Магистерская диссертация

РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ПОМОЩЬЮ
ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

Специальность: 6М070400 – «Вычислительная техника и программное
обеспечение»

Магистрант Газизов Тимур Газизов Тимур

Научный руководитель Карлык Бекир Карлык Бекир
к.т.н., профессор

Каскелен, 2013

ABSTRACT

We should note the special role of differential equations in the solution of many problems in mathematics, physics and engineering, as it is not always possible to establish a functional relationship between the data and the variables, but it is often possible to derive a differential equation that allows you to accurately predict the course of a particular process under certain conditions. Differential equations have great practical importance, being a powerful tool for exploring the many problems of science and technology: they are widely used in mechanics, astronomy, physics, in many problems of chemistry and biology. This is because very often the laws that govern certain processes are recorded in the form of differential equations, and the equations themselves act as a mean of quantitative interpretation of thus laws.

To solve thus equations we take the most well suited networks belonging to a class of Hopfield neural networks. These networks have a way of transmitting output signals to the inputs, and the response of such networks is dynamic, i.e. after receive a new input the output is calculated and transmitting by feedback network modifies the input. Then the output is recalculated, and the process is repeated again. For the network, which can be considered as stable, the sequence of iterations lead to smaller changes in outputs, and at the end the output does not become permanent. There is also an unstable network, for which the process of selection of the output may never end. That's the essence of the network settings for gaining the desired result.

Of course, there is also a classical numerical methods. But there are situations where these methods may not lead to a solution, or it can be obtained for a very large number of iterations. The neural network is much more flexible in this respect, and generally, an algorithm based on them is more efficient.

РЕЗЮМЕ

Следует отметить особую роль дифференциальных уравнений при решении многих задач математики, физики и техники, поскольку не всегда удается установить функциональную зависимость между искомыми и данными переменными величинами, но зато часто удается вывести дифференциальное уравнение, позволяющее точно предсказать протекание определенного процесса при определенных условиях. Дифференциальные уравнения имеют огромное прикладное значение, являясь мощным орудием исследования многих задач естествознания и техники: они широко используются в механике, астрономии, физике, во многих задачах химии, биологии. Это объясняется тем, что весьма часто законы, которым подчиняются те или иные процессы, записываются в форме дифференциальных уравнений, а сами эти уравнения, таким образом, являются средством для количественного выражения этих законов.

Для решения уравнений наиболее удачно подходят сети, принадлежащие к классу нейронных сетей Хопфилда. Эти сети имеют пути, передающие сигналы от выходов к входам, фактически получается, что отклик таких сетей является динамическим, то есть после того, как поступит новый вход, вычисляется выход и, передаваясь по сети обратной связи, модифицирует вход. Затем выход повторно вычисляется, и процесс повторяется опять. Для сети, которую можно считать устойчивой, последовательности итераций приводят к все меньшим изменениям выхода, и в конце выход не становится постоянным. Бывает также и неустойчивые сети, для которых процесс подбора выхода никогда не может закончиться. В этом вся и суть настройки сети, чтобы она получала нужный результат.

Разумеется, существуют также классические численные методы решения. Но бывают такие ситуации, когда эти методы не могут привести к решению, либо оно получается за очень большое число итераций. Нейронные сети гораздо более гибки в этом плане, и, как правило, алгоритм на их основе оказывается более эффективным.

ТҮЙІН

Дифференциалдық теңдеулер үлкен қолданбалы мағынаға ие. Олар механикада, астрономияда, физикада кең ауқымда қолданылады. Негізінде айналамыздағы кез келген бір процессті (экономикада, физикада, биологияда ж.т.б.) дифференциалдық теңдеулердің формасында жазуға болады.

Теңдеулерді шешуде Хопфилдтың нейрондық желілері әлдеқайда сәйкес келеді. Бұл желілерде кіруден шығуға сигнал жіберетін жолдар бар. Қайтып келетін сигнал динамикалық болады. Өйткені шығу сигналы шықпағанша кіру сигналың жаңартып тұрады. Бірақ кейбір желілерде бұл цикл аяқталмайды. Ондай желілер тұрақсыз желілер деп аталады. Тұрақты желілерде итерациялар арқылы шығу сигналы аз өзгертіледі. Сонда шығу сигналы тұрақты болып қалады. Сонымен қажетті нәтиже шықпағанша желіні өзгерту керек.

Сонымен қатар теңдеулерді шешудің аналитикалық тәсілдері де бар. Кейбір жағдайда осы тәсілдер арқылы біз шешім ала алмаймыз, немесе шешім көп итерация арқылы шығады. Бірақ нейрондық желі аналитикалық тәсілдерге әлдеқайда қарапайым. Осының салдарынаң алгоритм әсерлірірек болады.

Нейрондық желілердің жұмыс жасау принциптері адамның миының жұмыс істеу процесі негізінде орындалады. Мысалы үшін, бір суретте не бейнеленгенін табуымыз керек: үстел немесе орындық. Классикалық бағдарламалау арқылы осы есепті шешу мүмкін емес. Бірақ кез келген бір бала осы есепті бір секундта үстел немесе орындық екенін таба алады. Нақты шешімді табудың жалғыз жолы адам миының моделін жасауымыз керек (әрине біз мидың толық өзін жасай алмаймыз).

СОДЕРЖАНИЕ

СОДЕРЖАНИЕ	6
ОБОЗНАЧЕНИЯ И СОКРАЩЕННЫЕ СЛОВ.....	8
ВВЕДЕНИЕ.....	9
1 ОСНОВЫ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ	10
1.1 БИОЛОГИЧЕСКИЙ ПРОТОТИП.....	10
1.2 ИСКУССТВЕННЫЙ НЕЙРОН.....	11
1.3 ОДНОСЛОЙНЫЕ ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ.....	14
1.4 МНОГОСЛОЙНЫЕ ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ.....	15
2 ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ.....	19
2.1 МОДЕЛЬ ЛОТКИ — ВОЛЬТЕРРЫ.....	19
2.2 РАВНОВЕСНАЯ ЦЕНА В МОДЕЛИ ВАЛЬРАСА.....	21
3 РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ПОМОЩЬЮ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ.....	23
3.1 ПРИБЛИЖЕНИЕ РЕШЕНИЯ ДРОБНЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ.....	23
3.1.1 ВВЕДЕНИЕ В ГЛАВУ.....	23
3.1.2 ИСПОЛЬЗОВАНИЕ МНОГОСЛОЙНОЙ НЕЙРОННОЙ СЕТИ.....	25
3.1.3 ПРИБЛИЖЕННАЯ РЕАЛИЗАЦИЯ НЕПРЕРЫВНЫХ ОТОБРАЖЕНИЙ НЕЙРОННЫХ СЕТЕЙ.....	26
3.1.4 ЧИСЛЕННЫЕ МЕТОДЫ.....	30
3.1.5 ЧИСЛЕННЫЙ ПРИМЕР.....	32
3.1.6 ЗАКЛЮЧЕНИЕ ГЛАВЫ.....	34
3.2 РЕШЕНИЕ НЕЛИНЕЙНЫХ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ПОМОЩЬЮ FFANN.....	34
3.2.1 ВВЕДЕНИЕ В ГЛАВУ.....	34
3.2.2 ПРИБЛИЖЕНИЕ ФУНКЦИИ.....	36
3.2.3 ТРАНСФОРМАЦИЯ КУСОЧНО-ЛИНЕЙНЫХ ФУНКЦИЙ АКТИВАЦИИ СПЛАЙНОВ ПО ОГРАНИЧЕНИЮ ВЕС.....	41
3.2.4 ОПРЕДЕЛЕНИЕ ВЫВОДА ВЕСА: МЕТОД ВЗВЕШЕННЫХ НЕВЯЗОК.....	45
3.2.5 ЧИСЛЕННЫЕ ПРИМЕРЫ.....	46
3.2.5.1 УРАВНЕНИЕ ФОЛКНЕР - СКАНА.....	46
3.2.5.1.1 ПРОЦЕДУРА РЕШЕНИЯ.....	52
3.2.5.1.2 ПОСЛЕДУЮЩАЯ ОБРАБОТКА РЕШЕНИЯ.....	54
3.2.5.1.3 РЕЗУЛЬТАТЫ.....	56
3.2.5.1.4 СХОДИМОСТЬ.....	60
3.2.5.2 СВЯЗАННЫЕ НЕЛИНЕЙНЫЕ ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ.....	63
3.2.5.2.1 ПРОЦЕДУРА РЕШЕНИЯ.....	67
3.2.5.2.2 ПОСЛЕДУЮЩАЯ ОБРАБОТКА РЕШЕНИЯ.....	69
3.2.5.2.3 РЕЗУЛЬТАТЫ.....	70
3.2.5.2.4 СХОДИМОСТЬ.....	74

ЗАКЛЮЧЕНИЕ.....	75
ЛИТЕРАТУРА.....	77

ОБОЗНАЧЕНИЯ И СОКРАЩЕННЫЕ СЛОВА

FFANN (feed forward artificial neural network) - искусственная нейронная сеть прямой связи

RBF (radial basis function) – радиальная базисная функция

ИНС (искусственная нейронная сеть)

RMS – значение среднеквадратичной разности

ВВЕДЕНИЕ

Нейронные сети представляют собой программную модель человеческого мозга. Необходимость копирования принципов работы мозга легко понять на примере: допустим, у нас есть картинка, на которой схематично нарисован какой-то предмет. Требуется определить, что там изображено, например стол или стул. Данную задачу решить невозможно классическими методами программирования, так как для этого необходимо сравнить картинку со всеми возможными видами столов и стульев, что весьма абсурдно. Но в то же время, даже мозг ребёнка в состоянии за доли секунды отличить стол от стула. Поэтому, единственное решение будет построить модель мозга (разумеется, не в полной степени) и тогда можно будет достаточно легко получить решение.

Рассмотрим теперь, что позволяет работать мозгу так эффективно, в чём особенности его архитектуры. Следует отметить, что эта тема ещё досконально не изучена, и буквально каждое очередное исследование позволяет узнать что-то новое.

1 ОСНОВЫ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

Искусственные нейронные сети чрезвычайно разнообразны по своим конфигурациям. Несмотря на такое разнообразие, сетевые парадигмы имеют много общего.

1.1 БИОЛОГИЧЕСКИЙ ПРОТОТИП

Развитие искусственных нейронных сетей вдохновляется биологией. То есть, рассматривая сетевые конфигурации и алгоритмы, исследователи мыслят их в терминах организации мозговой деятельности. Но на этом аналогия может и закончиться. Наши знания о работе мозга столь ограничены, что мало бы нашлось руководящих ориентиров для тех, кто стал бы ему подражать. Поэтому разработчикам сетей приходится выходить за пределы современных биологических знаний в поисках структур, способных выполнять полезные функции. Во многих случаях это приводит к необходимости отказа от биологического правдоподобия, мозг становится просто метафорой, и создаются сети, невозможные в живой материи или требующие неправдоподобно больших допущений об анатомии и функционировании мозга.

Нервная система человека, построенная из элементов, называемых нейронами, имеет ошеломляющую сложность. Около 10^{11} нейронов участвуют в примерно 10^{15} передающих связях, имеющих длину метр и более. Каждый нейрон обладает многими качествами, общими с другими элементами тела, но его уникальной способностью является прием, обработка и передача электрохимических сигналов по нервным путям, которые образуют коммуникационную систему мозга.

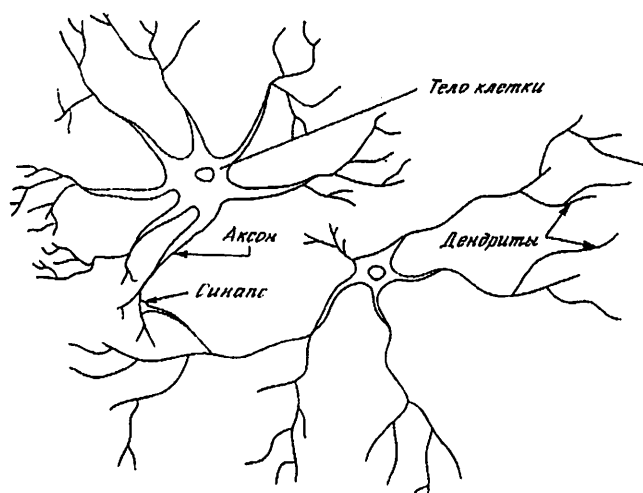


Рисунок 1 - Биологический нейрон

На рисунке 1 показана структура пары типичных биологических нейронов. Дендриты идут от тела нервной клетки к другим нейронам, где они принимают сигналы в точках соединения, называемых синапсами. Принятые синапсом входные сигналы подводятся к телу нейрона. Здесь они суммируются, причем одни входы стремятся возбудить нейрон, другие – воспрепятствовать его возбуждению. Когда суммарное возбуждение в теле нейрона превышает некоторый порог, нейрон возбуждается, посылая по аксону сигнал другим нейронам. У этой основной функциональной схемы много усложнений и исключений, тем не менее большинство искусственных нейронных сетей моделируют лишь эти простые свойства.

1.2 ИСКУССТВЕННЫЙ НЕЙРОН

Искусственный нейрон имитирует в первом приближении свойства биологического нейрона. На вход искусственного нейрона поступает некоторое множество сигналов, каждый из которых является выходом другого нейрона. Каждый вход умножается на соответствующий вес, аналогичный синаптической силе, и все произведения суммируются, определяя уровень активации нейрона. На рисунке 2 представлена модель, реализующая эту идею. Хотя сетевые парадигмы весьма разнообразны, в основе почти всех их лежит эта конфигурация. Здесь множество входных сигналов, обозначенных x_1, x_2, \dots, x_n , поступает на искусственный нейрон. Эти входные сигналы, в совокупности обозначаемые вектором X , соответствуют сигналам, приходящим в синапсы биологического нейрона. Каждый сигнал умножается на соответствующий вес w_1, w_2, \dots, w_n , и поступает на суммирующий блок, обозначенный Σ . Каждый вес соответствует «силе» одной биологической синаптической связи. (Множество весов в совокупности обозначается вектором W .) Суммирующий блок, соответствующий телу биологического элемента, складывает взвешенные входы алгебраически, создавая выход, который мы будем называть NET . В векторных обозначениях это может быть компактно записано следующим образом:

(1)

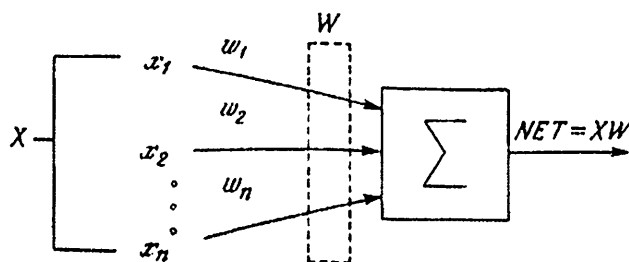


Рисунок 2 - Искусственный нейрон

Активационные функции

Сигнал далее, как правило, преобразуется активационной функцией и дает выходной нейронный сигнал. Активационная функция может быть обычной линейной функцией

(2)

где θ – постоянная, пороговой функции, если $NET > \theta$, в остальных случаях, где θ – некоторая постоянная пороговая величина, или же функцией, более точно моделирующей нелинейную передаточную характеристику биологического нейрона и представляющей нейронной сети большие возможности.

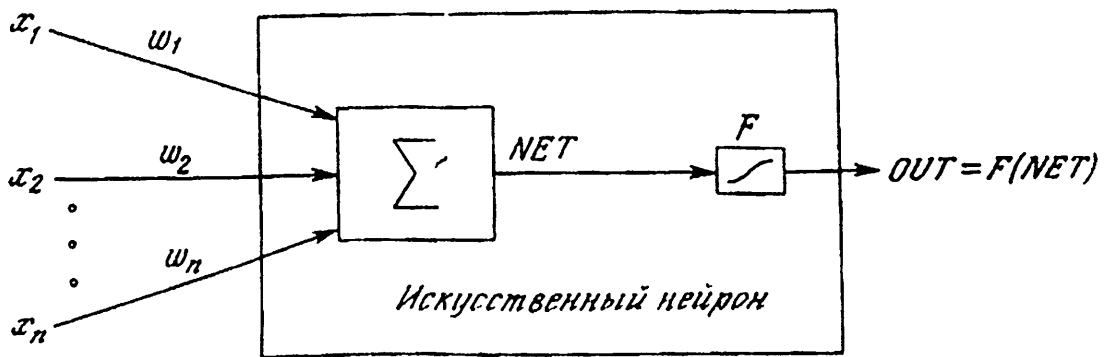


Рисунок 3 - Искусственный нейрон с активационной функцией

На рисунке 3 блок, обозначенный Σ , принимает сигнал NET и выдает сигнал $OUT = F(NET)$. Если блок сужает диапазон изменения величины NET так, что при любых значениях NET значения OUT принадлежат некоторому конечному интервалу, то F называется «сжимающей» функцией. В качестве «сжимающей» функции часто используется логистическая или «сигмоидальная» (S-образная) функция, показанная на рисунке 4. Эта функция математически выражается как

(3)

Таким образом,

$$OUT = \frac{1}{1 + e^{-NET}}$$

(4)



Рисунок 4 - Сигмоидальная логистическая функция

По аналогии с электронными системами активационную функцию можно считать нелинейной усилительной характеристикой искусственного нейрона. Коэффициент усиления вычисляется как отношение приращения величины к вызвавшему его небольшому приращению величины. Он выражается наклоном кривой при определенном уровне возбуждения и изменяется от малых значений при больших отрицательных возбуждениях (кривая почти горизонтальна) до максимального значения при нулевом возбуждении и снова уменьшается, когда возбуждение становится большим положительным. Гроссберг (1973) обнаружил, что подобная нелинейная характеристика решает поставленную им дилемму шумового насыщения. Каким образом одна и та же сеть может обрабатывать как слабые, так и сильные сигналы? Слабые сигналы нуждаются в большом сетевом усилении, чтобы дать пригодный к использованию выходной сигнал. Однако усилительные каскады с большими коэффициентами усиления могут привести к насыщению выхода шумами усилителей (случайными флуктуациями), которые присутствуют в любой физически реализованной сети. Сильные входные сигналы в свою очередь также будут приводить к насыщению усилительных каскадов, исключая возможность полезного использования выхода. Центральная область логистической функции, имеющая большой коэффициент усиления, решает проблему обработки слабых сигналов, в то время как области с падающим усилением на положительном и отрицательном концах подходят для больших возбуждений. Таким образом, нейрон функционирует с большим усилением в широком диапазоне уровня входного сигнала.

$$OUT = \frac{1}{1 + e^{-NET}} = F(NET)$$

(5)

Другой широко используемой активационной функцией является гиперболический тангенс. По форме она сходна с логистической функцией и часто используется биологами в качестве математической модели активации нервной клетки. В качестве активационной функции искусственной нейронной сети она записывается следующим образом:

$$OUT = th(x)$$

(6)

Подобно логистической функции гиперболический тангенс является S-образной функцией, но он симметричен относительно начала координат, и в точке $NET = 0$ значение выходного сигнала OUT равно нулю (см. рисунок 5). В отличие от логистической функции гиперболический тангенс принимает значения различных знаков, что оказывается выгодным для ряда сетей.

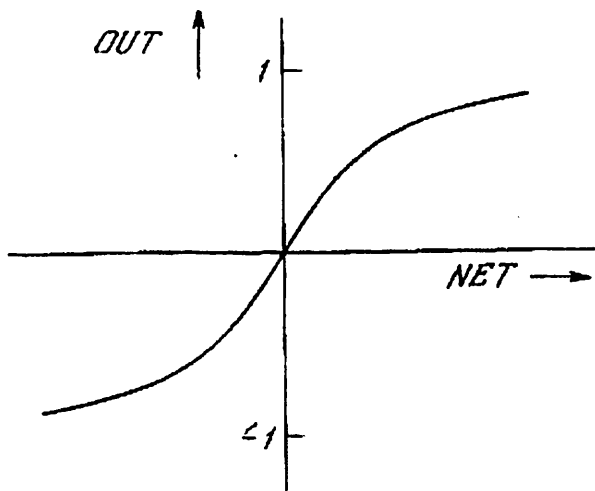


Рисунок 5 - Функция гиперболического тангенса

Рассмотренная простая модель искусственного нейрона игнорирует многие свойства своего биологического двойника. Например, она не принимает во внимание задержки во времени, которые воздействуют на динамику системы. Входные сигналы сразу же порождают выходной сигнал. И, что более важно, она не учитывает воздействий функции частотной модуляции или синхронизирующей функции биологического нейрона, которые ряд исследователей считают решающими.

Несмотря на эти ограничения, сети, построенные из этих нейронов, обнаруживают свойства, сильно напоминающие биологическую систему. Только время и исследования смогут ответить на вопрос, являются ли подобные совпадения случайными или следствием того, что в модели верно схвачены важнейшие черты биологического нейрона.

1.3 ОДНОСЛОЙНЫЕ ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

Хотя один нейрон и способен выполнять простейшие процедуры распознавания, сила нейронных вычислений проистекает от соединений нейронов в сетях. Простейшая сеть состоит из группы нейронов, образующих слой, как показано в правой части рисунка 6. Отметим, что вершины-круги слева служат

лишь для распределения входных сигналов. Они не выполняют каких-либо вычислений, и поэтому не будут считаться слоем. По этой причине они обозначены кругами, чтобы отличать их от вычисляющих нейронов, обозначенных квадратами. Каждый элемент из множества входов X отдельным весом соединен с каждым искусственным нейроном. А каждый нейрон выдает взвешенную сумму входов в сеть. В искусственных и биологических сетях многие соединения могут отсутствовать, все соединения показаны в целях общности. Могут иметь место также соединения между выходами и входами элементов в слое.

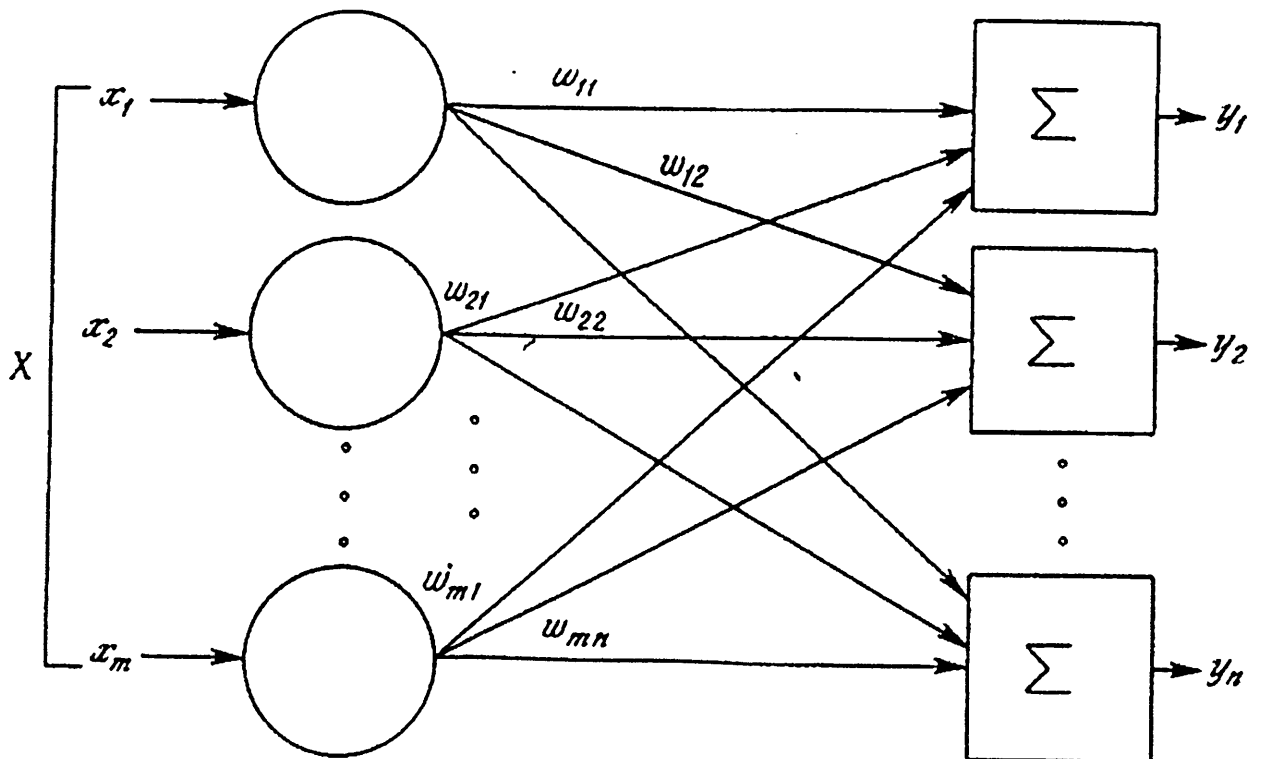


Рисунок 6 - Однослойная нейронная сеть

Удобно считать веса элементами матрицы W . Матрица имеет m строк и n столбцов, где m – число входов, а n – число нейронов. Например, $w_{2,3}$ – это вес, связывающий третий вход со вторым нейроном. Таким образом, вычисление выходного вектора N , компонентами которого являются выходы OUT нейронов, сводится к матричному умножению $N = XW$, где N и X – векторы-строки.

1.4 МНОГОСЛОЙНЫЕ ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

Более крупные и сложные нейронные сети обладают, как правило, и большими вычислительными возможностями. Хотя созданы сети всех конфигураций, какие только можно себе представить, послойная организация

нейронов копирует слоистые структуры определенных отделов мозга. Оказалось, что такие многослойные сети обладают большими возможностями, чем однослойные, и в последние годы были разработаны алгоритмы для их обучения.

Многослойные сети могут образовываться каскадами слоев. Выход одного слоя является входом для последующего слоя. Подобная сеть показана на рисунке 7 и снова изображена со всеми соединениями.

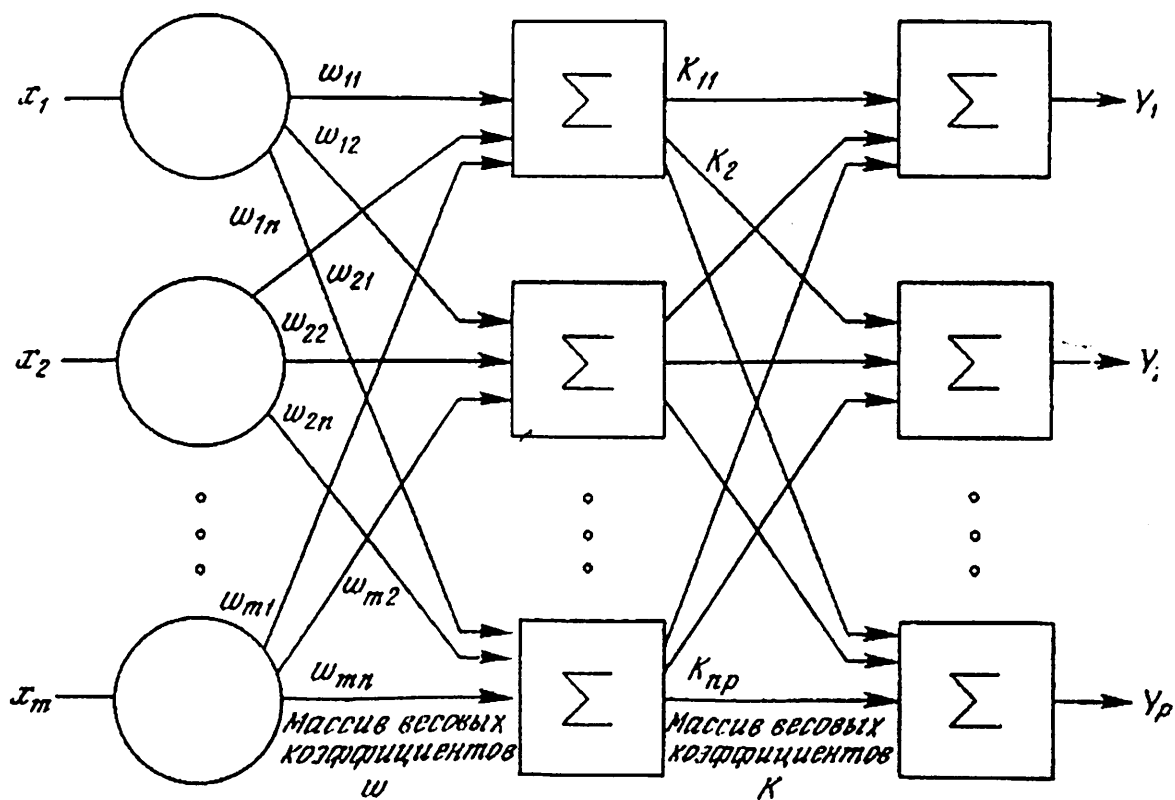


Рисунок 7 - Двухслойная нейронная сеть

Нелинейная активационная функция

Многослойные сети не могут привести к увеличению вычислительной мощности по сравнению с однослойной сетью лишь в том случае, если активационная функция между слоями будет нелинейной. Вычисление выхода слоя заключается в умножении входного вектора на первую весовую матрицу с последующим умножением (если отсутствует нелинейная активационная функция) результирующего вектора на вторую весовую матрицу.

$$(XW_1)W_2 \tag{7}$$

Так как умножение матриц ассоциативно, то

Это показывает, что двухслойная линейная сеть эквивалентна одному слою с весовой матрицей, равной произведению двух весовых матриц. Следовательно, любая многослойная линейная сеть может быть заменена эквивалентной однослойной сетью. Однослойные сети весьма ограничены по своим вычислительным возможностям. Таким образом, для расширения возможностей сетей по сравнению с однослойной сетью необходима нелинейная активационная функция.

Сети с обратными связями

У сетей, рассмотренных до сих пор, не было обратных связей, т. е. соединений, идущих от выходов некоторого слоя к входам этого же слоя или предшествующих слоев. Этот специальный класс сетей, называемых сетями без обратных связей или сетями прямого распространения, представляет интерес и широко используется. Сети более общего вида, имеющие соединения от выходов к входам, называются сетями с обратными связями. У сетей без обратных связей нет памяти, их выход полностью определяется текущими входами и значениями весов. В некоторых конфигурациях сетей с обратными связями предыдущие значения выходов возвращаются на входы; выход, следовательно, определяется как текущим входом, так и предыдущими выходами. По этой причине сети с обратными связями могут обладать свойствами, сходными с кратковременной человеческой памятью, сетевые выходы частично зависят от предыдущих входов.

Обучение искусственных нейронных сетей

Среди всех интересных свойств искусственных нейронных сетей ни одно не захватывает так воображения, как их способность к обучению. Их обучение до такой степени напоминает процесс интеллектуального развития человеческой личности что может показаться, что достигнуто глубокое понимание этого процесса. Возможности обучения искусственных нейронных сетей ограничены, и нужно решить много сложных задач, чтобы определить, на правильном ли пути мы находимся.

Цель обучения

Сеть обучается, чтобы для некоторого множества входов давать желаемое (или, по крайней мере, сообразное с ним) множество выходов. Каждое такое входное (или выходное) множество рассматривается как вектор. Обучение осуществляется путем последовательного предъявления входных векторов с одновременной подстройкой весов в соответствии с определенной процедурой. В

процессе обучения веса сети постепенно становятся такими, чтобы каждый входной вектор вырабатывал выходной вектор.

Обучение с учителем

Различают алгоритмы обучения с учителем и без учителя. Обучение с учителем предполагает, что для каждого входного вектора существует целевой вектор, представляющий собой требуемый выход. Вместе они называются обучающей парой. Обычно сеть обучается на некотором числе таких обучающих пар. Предъявляется выходной вектор, вычисляется выход сети и сравнивается с соответствующим целевым вектором, разность (ошибка) с помощью обратной связи подается в сеть и веса изменяются в соответствии с алгоритмом, стремящимся минимизировать ошибку. Векторы обучающего множества предъявляются последовательно, вычисляются ошибки и веса подстраиваются для каждого вектора до тех пор, пока ошибка по всему обучающему массиву не достигнет приемлемо низкого уровня.

Обучение без учителя

Трудно вообразить обучающий механизм в мозге, который бы сравнивал желаемые и действительные значения выходов, выполняя коррекцию с помощью обратной связи. Если допустить подобный механизм в мозге, то откуда тогда возникают желаемые выходы? Обучение без учителя является намного более правдоподобной моделью обучения в биологической системе. Развитая Кохоненом [3] и многими другими, она не нуждается в целевом векторе для выходов и, следовательно, не требует сравнения с predetermined идеальными ответами. Обучающее множество состоит лишь из входных векторов. Обучающий алгоритм подстраивает веса сети так, чтобы получались согласованные выходные векторы, т. е. чтобы предъявление достаточно близких входных векторов давало одинаковые выходы. Процесс обучения, следовательно, выделяет статистические свойства обучающего множества и группирует сходные векторы в классы. Предъявление на вход вектора из данного класса даст определенный выходной вектор, но до обучения невозможно предсказать, какой выход будет производиться данным классом входных векторов. Следовательно, выходы подобной сети должны трансформироваться в некоторую понятную форму, обусловленную процессом обучения. Это не является серьезной проблемой. Обычно не сложно идентифицировать связь между входом и выходом, установленную сетью.

2 ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

Далее основной задачей стал подбор более используемых или важных дифференциальных уравнений. Поэтому выбор пал на несколько функций.

2.1 МОДЕЛЬ ЛОТКИ — ВОЛЬТЕРРЫ

Первую математическую модель, которую мы рассмотрим, описывает биологические сообщества и называется *модель Лотки — Вольтерры*. Она описывает популяцию, состоящую из двух взаимодействующих видов. Первый из них, именуемый *хищниками*, при отсутствии второго вымирает по закону

$$x' = -ax \quad (a > 0), \quad (8)$$

а второй — *жертвы* — при отсутствии хищников неограниченно размножается в соответствии с *законом Мальтуса*

$$x' = ax \quad (a > 0). \quad (9)$$

Взаимодействие двух этих видов моделируется так. Жертвы вымирают со скоростью, равной числу встреч хищников и жертв, которое в данной модели предполагается пропорциональным численности обеих популяций, т. е. равной $dxу$ ($d > 0$). Поэтому

$$y' = by - dxу. \quad (10)$$

Хищники же размножаются со скоростью, пропорциональной числу съеденных жертв:

$$x' = -ax + cxy, \quad (c > 0). \quad (11)$$

Система уравнений:

$$x' = -ax + cxy, \quad (12)$$

$$y' = by - dxу, \quad (13)$$

Описывающая такую популяцию хищник — жертва и называется *системой* (или *моделью*) *Лотки — Вольтерры*.

Модель Лотки-Вольтерры имеет один недостаток. Она не учитывает совокупность всех факторов действующих на популяцию.

Этого недостатка лишена модель Холлинга — Тэннера, учитывающая большее число реальных факторов. В этой модели скорость изменения популяции хищников задается выражением $ax - bx^2/y = x(a - bx/y)$. Оно выбрано из следующих соображений. Когда пищи (жертв) много ($y \approx +\infty$), популяция хищников растет по правилу Мальтуса с показателем a . С уменьшением числа жертв скорость роста популяции хищников падает и при $y < bx/a$ становится отрицательной. Скорость изменения популяции жертв состоит из трех компонент. Первый член cy соответствует закону Мальтуса, второй $-dy^2$ описывает внутривидовую конкуренцию и вызван ограниченностью ресурсов экологической ниши, занимаемой популяцией жертв. При отсутствии хищников жертвы подчиняются уравнению

$$y' = y(c - dy) \quad (14)$$

Наконец, третий компонент скорости изменения популяции жертв в модели Холлинга — Тэннера описывает ее взаимодействие с хищниками и имеет вид —

$$pxy/(q + y) \quad (p, q > 0) \quad (15)$$

$$x' = (a - bx/y)x \quad (16)$$

$$y' = [c - dy - px/(q + y)]y \quad (17)$$

Процесс передачи инфекции и эпидемии.

Дана изолированная популяция. Особи популяции делятся на три класса. Инфицированный класс численностью $x(t)$ (t - время) состоит из инфицированных (заболевших) особей, каждая из этих особей заразна (предполагается, что инкубационный период заболевания пренебрежимо короток). Второй класс численностью $y(t)$ составляют восприимчивые особи, т. е. особи, которые могут заразиться при контакте с инфицированными особями. Третий класс состоит из невосприимчивых особей (приобретших иммунитет или погибших в результате заболевания). Его численность обозначается $z(t)$. Предполагается также, что общая численность популяции n постоянна (т. е. не учитываются рождения, естественные смерти и миграция). Две гипотезы, лежащие в основе модели таковы:

1) заболеваемость в момент времени t равна $x(t)y(t)$ (эта гипотеза основывается на правдоподобном предположении, что число заболевающих пропорционально числу встреч между больными и восприимчивыми особями, которое в свою очередь в первом приближении пропорционально $x(t)y(t)$); таким образом численность класса x растет, а численность класса y убывает со скоростью

$$ax(t)y(t) \quad (a > 0) \quad (18)$$

2) численность становящихся невосприимчивыми особей (приобретших иммунитет или погибших) растет со скоростью, пропорциональной численности заболевших, т. е. со скоростью $bx(t)$ ($b > 0$). В результате мы получаем систему уравнения

$$\begin{aligned}x' &= axy - bx, \\y' &= -axy \\z' &= bx.\end{aligned}\tag{19}$$

2.2 РАВНОВЕСНАЯ ЦЕНА В МОДЕЛИ ВАЛЬРАСА

Рассмотрим процесс, в ходе которого предложение и спрос уравниваются. Предполагается, что при любой цене, если спрос превышает предложение, то цена будет расти, если предложение превышает спрос, то цена будет падать. Обозначим цену через p ; величины спроса и предложения, соответственно, через $D(p, \alpha), S(p)$; α — параметр, соответствующий экзогенным (внешним) факторам. Тогда изменение цены при изменении времени t можно описать следующим соотношением:

$$\frac{dp}{dt} = D(p, \alpha) - S(p)\tag{20}$$

Что можно увидеть из этого соотношения? Например, поставим вопрос о том, существует ли равновесная цена, не изменяющаяся со временем. Её следует определять из условия

$$\frac{dp}{dt} = 0,$$

откуда

$$D(p, \alpha) - S(p) = 0.$$

Пусть p_0 есть решение данного уравнения. Разложим правую часть $D(p, \alpha) - S(p)$ в окрестности значения $p = p_0$, тогда изменение цены p при малом отклонении от равновесной цены $p = p_0$ будет определяться уравнением

$$\frac{dp}{dt} = (D_p - S_p)(p - p_0),\tag{21}$$

где опущены слагаемые выше первой степени величины $p - p_0$ и обозначено

$$D_p = \partial D(p, \alpha) / \partial p|_{p=p_0}, S_p = \partial S(p, \alpha) / \partial p|_{p=p_0}\tag{22}$$

Решение последнего уравнения даётся выражением

$$p(t) = p_0 + [p(0) - p_0] \exp[(D_p - S_p) t] \quad (23)$$

Пусть $D_p - S_p < 0$. Тогда, если начальная цена $p(0)$ отличалась от равновесной p_0 , то при $t \rightarrow \infty$ имеет место предел $p(t) \rightarrow p_0$, т.е. в процессе обращения товара на рынке его цена будет приближаться к равновесной. В этом случае можно сказать, что равновесная цена устойчива. Очевидно, что с увеличением цены падает спрос ($D_p < 0$) и увеличивается предложение, т.е. $S_p > 0$. В этих условиях существует равновесная цена. Данная модель предсказывает рост предложения при росте спроса на данный товар, выраженного в виде роста цены на товар.

Если $D_p - S_p > 0$, то цена $p(t)$ будет отклоняться от равновесной в ту или другую сторону. В этом случае равновесная цена неустойчива. В соответствии с положениями экономической теории, при неустойчивой цене данная модель не даёт разумных предсказаний.

3 РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ПОМОЩЬЮ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

3.1 ПРИБЛИЖЕНИЕ РЕШЕНИЯ ДРОБНЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ

3.1.1 ВВЕДЕНИЕ В ГЛАВУ

Нейронные сети с радиальными базисными функциями используются для решения начально краевых дробных уравнений в частных производных с переменными коэффициентами на конечной области. В итоге получается случай, когда левая или правая сторона дробной пространственной производной может присутствовать в дифференциальных уравнениях. Сходимости этого метода будут рассмотрены. Численный пример использования нейронных сетей способом RBF для двусторонних дробных частных дифференциальных уравнений также будут представлены и по сравнению с другими методами.

В этой статье я буду использовать метод нейронных сетей для решения частных дробных дифференциальных уравнений (ЧДДУ) в виде:

$$\frac{\partial u(x, t)}{\partial t} = c_+(x, t) \frac{\partial^\alpha u(x, t)}{\partial_+ x^\alpha} + c_-(x, t) \frac{\partial^\alpha u(x, t)}{\partial_- x^\alpha} + s(x, t). \quad (24)$$

На конечной области $L < x < R, 0 \leq t \leq T$. Здесь, я рассматриваю случай $1 \leq \alpha \leq 2$, где параметр α является дробным порядком пространственной производной. Функция $s(x, t)$ является источником / поглотителем первого предела. Функций $c_+(x, t) \geq 0$ и $c_-(x, t) \geq 0$ можно интерпретировать как транспортные коэффициенты. Мы также предполагаем начальное условие $u(x, t = 0) = F(x)$ для $L < x < R$ и нулевыми граничными условиями Дирихле. Для случая $1 < \alpha \leq 2$, добавление адвективно классической формы $-v(x, t) \partial u(x, t) / \partial x$ на правую сторону уравнения (24) не влияет на анализ, проведенный в данном документе и было опущено для упрощения обозначений.

(+) на правой стороне и (-) на левой стороне в уравнении (24) – дробные производные порядка α Риманна-Люовилля [15] определяются так:

$$\begin{aligned}
(D_{L+}^{\alpha} f)(x) &= \frac{d^{\alpha} f(x)}{d_{+} x^{\alpha}} = \frac{1}{\Gamma(n-\alpha)} \frac{d^n}{dx^n} \int_L^x \frac{f(\xi)}{(x-\xi)^{\alpha+1-n}} d\xi, \\
(D_{R-}^{\alpha} f)(x) &= \frac{d^{\alpha} f(x)}{d_{-} x^{\alpha}} = \frac{(-1)^n}{\Gamma(n-\alpha)} \frac{d^n}{dx^n} \int_x^R \frac{f(\xi)}{(x-\xi)^{\alpha+1-n}} d\xi,
\end{aligned} \tag{25}$$

где n целое число $n-1 < \alpha < n$. Если $\alpha = m$ целое число, тогда верхнее определение дает стандартные целые производные:

$$(D_{L+}^m f)(x) = \frac{d^m f(x)}{dx^m}, \quad (D_{R-}^m f)(x) = (-1)^m \frac{d^m f(x)}{d(-x)^m}. \tag{26}$$

Когда $\alpha = 2$ и подстановка $c(x, t) = c_+(x, t) + c_-(x, t)$ в уравнение (24) имеет место, то получается классическое параболическое частичное дифференциальное уравнение:

$$\frac{\partial u(x, t)}{\partial t} = c(x, t) \frac{\partial^2 u(x, t)}{\partial x^2} + s(x, t). \tag{27}$$

Точно так же, если $\alpha = 1$ и $c(x, t) = c_+(x, t) + c_-(x, t)$ подставляем в уравнение (24), то получится гиперболическое частичное дифференциальное уравнение:

$$\frac{\partial u(x, t)}{\partial t} = c(x, t) \frac{\partial u(x, t)}{\partial x} + s(x, t). \tag{28}$$

Случай $1 < \alpha < 2$ представляет собой супердиффузивный процесс, в котором частицы диффундируют быстрее, чем предсказывает классическая модель (27).

Следует отметить, что левая сторона дробной производной $f(x)$ в точке x зависит от всех значений функции слева от точки x , то есть, эта производная нагружена средним таких значений функции. Кроме того, правая сторона дробной производной $f(x)$ в точке x зависит от всех значений функции справа от этой точки. В общем, левые и правые производные не равны, если α не четное целое число, и в этом случае эти производные становятся локализованными и равными. При α , равным нечетному целому числу, эти производные становятся локализованными и противоположны по знаку. Для более подробной информации о дробной производной понятий и определений, см. [11, 13, 15, 16]. Ссылка [17] обеспечивает более детальное рассмотрение правой стороны дробных производных, а также существенное преобразование левых сторон дробных производных.

Другой способ решения дробного уравнения в частных производных уравнения(24) раскрыт в [14] ссылке. Они превращают это уравнение в частных производных в систему обыкновенных дифференциальных уравнений (методом прямых), которая затем решается с помощью формул дифференцирования назад. В другой работе [18] развивается метод конечных элементов для двухточечной краевой задачи, и [11], находит численное решение уравнения (24), методом конечных разностей.

3.1.2 ИСПОЛЬЗОВАНИЕ МНОГОСЛОЙНОЙ НЕЙРОННОЙ СЕТИ

Многослойная сеть Румельхарт-Хинтон-Уильяма [19], является опережающим типом сети со связями только между соседними слоями. Сети обычно обладают скрытыми слоями между входным и выходным слоями. Каждый слой состоит из вычислительных блоков. Отношение ввода-вывода каждой единицы представлены входами x_i и выходами y , вес соединения w_i , порог θ и дифференциальной функции φ следующим образом:

$$y(x_1, \dots, x_n) = \sum_{i=1}^m v_i \varphi(\|W^i - x\| + \theta_i), \quad \text{where } x_i \in R, W^i \in R^n, v_i, \theta_i \in R \quad (29)$$

Это сеть обладает обучением с обратным распространением [19], которое представляет собой алгоритм, который использует способ градиентного спуска для изменения веса и порога, как ошибку между требуемым результатом и выходным сигналом сети, которая сводится к минимуму. Я обычно использую ограниченные и монотонно возрастающие дифференцируемые функции, которые называются радиальными базисными функциями, для каждой функции выходного блока.

Если многослойная сеть имеет n входных узлов и m выходных, то отношение ввод-вывод определяет непрерывное отображение из n -мерного евклидова пространства в m -мерное евклидовое пространство. Я называю это отображение - отображение ввода-вывода сети. Я изучаю проблемы сетевых возможностей с точки зрения отображения ввода-вывода. По меркам изучения отображений многослойной сети становится ясным, что достаточно рассмотреть сети, чьи функции $\varphi(x)$ и выходные функции входных и выходных слоев являются линейными.

3.1.3 ПРИБЛИЖЕННАЯ РЕАЛИЗАЦИЯ НЕПРЕРЫВНЫХ ОТОБРАЖЕНИЙ НЕЙРОННЫХ СЕТЕЙ

Ссылка [20] рассматривает возможность представления непрерывных отображений с помощью нейронных сетей, у которых выходная функции в скрытых слоях является сигмовидной функцией, например, $\varphi = \frac{1}{1+e^{-x}}$. Общие непрерывные отображения не могут быть точно представлены сетями Румельхарт-Хинтон-Уильяма. Например, если используется действительная аналитическая функция вывода, такая как сигмовидная функция $\varphi = \frac{1}{1+e^{-x}}$, то отображение входа-выхода этой сети является аналитическим и вообще не может представлять все непрерывные отображения. Пусть точки n -мерного Евклидова пространства R^n обозначаются через $x = (x_1, \dots, x_n)$ и норма x определяется как

$$\|x\| = \left(\sum_{i=0}^n x_i^2\right).$$

Определение 3.1 (см. [21]): Пусть $H = R^n$ линейное пространство. Функция $F: H \rightarrow R$ называется радиальной базисной функцией, которая может быть представлена в виде $F = g \circ \varphi$, где $\varphi: R \rightarrow R$, $g: R^n \rightarrow R$, и $g(x) = \|x - w^i\|^2$, $x, w^i \in R^n$.

В данной работе я изучаю приближение U с помощью u , которая является представлением модели нейронных сетей посредством анализа предыдущих теоретических исследований. Кроме того, приведены исследования сходящегося решения U при приближенном решении u , где $\varphi = (w^i, x, \theta)$ является радиальной базисной функцией.

Мы также даем некоторые теоремы для получения условий сходимости приближительного решения для уравнения (24) с помощью нейронных сетей.

Теорема 3.2 (Ири-Мияки, [16]): Пусть $\psi(x) \in L^1 \times R$, то есть, пусть $\psi(x)$ будет абсолютно интегрируемой и $u(x_1, \dots, x_n) \in L^2(R^n)$. Пусть $\Psi(\xi)$ интегрируемая и $U(x_1, \dots, x_n)$ будет преобразованием функции Фурье $\psi(x)$ и $u(x_1, \dots, x_n) \in L^2(R^n)$ соответственно. Если $\Psi(1) \neq 0$, тогда:

$$u(x_1, \dots, x_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \psi\left(\sum_{i=1}^n x_i w_i - w_0\right) \frac{1}{(2\pi)^n \Psi(1)} U(w_1, \dots, w_n) \times \exp(iw_0) dw_0 dw_1 \dots dw_n. \quad (30)$$

Теорема 3.3. Пусть $\varphi(x)$ - радиальная базисная функция, непостоянная, ограниченная и монотонно возрастающая непрерывная функция. Пусть K - компактное подмножество R^n и $u(x_1, \dots, x_n)$ реальное значение непрерывной функции на K . Тогда для любого $\varepsilon > 0$ существует интеграл N и действительные константы v_j, θ_j, w_{ij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$) такие, что $y = (x_1, \dots, x_n) = \sum_{j=1}^m v_j \varphi[\sum_{i=1}^n (w_{ij} - x_i)^2]^{1/2} + \theta_j$

удовлетворяет

$$\max_{x \in K} |y(x_1, \dots, x_n) - u(x_1, \dots, x_n)| < \varepsilon.$$

Другими словами, для любого $\varepsilon > 0$ существует трехслойная сеть, выходная функция для скрытого слоя которой равна $\varphi(x)$, для вводного и выводного слоя которой являются линейными, и которая имеет входную-выходную функцию $y(x_1, \dots, x_n)$ такую, что

$$\max_{x \in K} |y(x_1, \dots, x_n) - u(x_1, \dots, x_n)| < \varepsilon$$

Доказательство первое

Так как $u(x)$ ($x = (x_1, \dots, x_n)$) непрерывная функция на компактном подмножестве K из R^n , то $u(x)$ можно расширить до непрерывной функции на R^n с компактным носителем.

Если я использую осреднение $\rho\alpha^*$ на $u(x)$, то $\rho\alpha^*u(x)$ является функцией C^∞ с компактным носителем. Кроме того,

$$\rho\alpha^*u(x) \rightarrow u(x) \rightarrow (\alpha \rightarrow +0) \text{ равномерно на } R^n.$$

Таким образом, мы можем предположить, что $u(x)$ - C^∞ функция с компактным носителем, для доказательства теоремы 3.3. По теореме Винера-Пэли [20], преобразование Фурье $U(w)$ ($w = (w_1, \dots, w_n)$) для $u(x)$ действительно аналитическая и для любого целого N , существует постоянная C_N так, что

$$|U(w)| \leq C_N(1 + |w|^{-N}) \tag{32}$$

В частности $U(w) \in L^1 \cap L^2(R^n)$

Я определяю $I_A(x_1, \dots, x_n), I_{A,\infty}(x_1, \dots, x_n), J_A(x_1, \dots, x_n)$ следующим образом:

$$I_A(x_1, \dots, x_n) = \int_{-A}^A \dots \int_{-A}^A \psi \left([\sum_{i=1}^n (x_i - w_i)^2]^{\frac{1}{2}} - w_0 \right) \times \frac{1}{(2\pi)^n \psi(1)} \times U(w_1, \dots, w_n) \times \exp(iw_0) dw_0 \dots dw_n,$$

$$I_{A,\infty}(x_1, \dots, x_n) = \int_{-A}^A \dots \int_{-A}^A \left[\int_{-\infty}^{\infty} \psi \left([\sum_{i=1}^n (x_i - w_i)^2]^{\frac{1}{2}} - w_0 \right) \times \frac{1}{(2\pi)^n \psi(1)} \times U(w_1, \dots, w_n) \times \exp(iw_0) \right] dw_0 \dots dw_n,$$

$$J_A(x_1, \dots, x_n) = \frac{1}{(2\pi)^n} \int_{-A}^A \dots \int_{-A}^A U(w_1, \dots, w_n) \exp[\sum_{i=1}^n (x_i - w_i)^2]^{\frac{1}{2}} dw_1 \dots dw_n, \quad (33)$$

где $\psi(x) \in L^1$ определена как

$$\psi(x) = \varphi(x/\delta + \alpha) - \varphi(x/\delta - \alpha), \delta, \alpha > 0.$$

Существенной частью доказательства интегральной формулы Ири-Мияке [11] является равенство $I_{A,\infty}(x_1, \dots, x_n) = J_A(x_1, \dots, x_n)$ и это происходит от

$$\int_{-A}^A \psi \left([\sum_{i=1}^n (x_i - w_i)^2]^{\frac{1}{2}} - w_0 \right) \exp(iw_0) dw_0 = \exp \left([\sum_{i=1}^n (x_i - w_i)^2]^{\frac{1}{2}} \right) \psi(1) \quad (34)$$

в нашей дискуссии, используя оценку $U(w)$, легко доказать, что

$\lim_{A \rightarrow \infty} I_{A,\infty}(x_1, \dots, x_n) = u(x_1, \dots, x_n)$ равномерно на R^n . Поэтому, $\lim_{A \rightarrow \infty} I_A(x_1, \dots, x_n) = u(x_1, \dots, x_n)$ равномерно на R^n . Я могу утверждать, что для любого $\varepsilon > 0$ существует $A > 0$ такое, что

$$\max_{x \in R^n} |I_{A,\infty}(x_1, \dots, x_n) - u(x_1, \dots, x_n)| < \frac{\varepsilon}{2} \quad (35)$$

Доказательство второе

Я приближу $I_{\infty,A}$ конечными интегралами на K . Для $\varepsilon > 0$, возьмем A , которое удовлетворяет (35). Для $A' > 0$, устанавливаем

$$I_{A',A}(x_1, \dots, x_n) = \int_{-A}^A \dots \int_{-A}^A \left[\int_{-A}^A \psi \left([\sum_{i=1}^n (x_i - w_i)^2]^{\frac{1}{2}} - w_0 \right) \times \frac{1}{(2\pi)^n \psi(1)} \times U(w_1, \dots, w_n) \times \exp(iw_0) dw_0 \right] dw_1 \dots dw_n \quad (36)$$

Я покажу, что для $\varepsilon > 0$, я могу взять $A' > 0$, так что

$$\max_{x \in K} |I_{A',A}(x_1, \dots, x_n) - I_{\infty,A}(x_1, \dots, x_n)| < \frac{\varepsilon}{2} \quad (37)$$

С помощью следующего уравнения:

$$\int_{-A'}^{A'} \psi \left(\left[\sum_{i=1}^n (x_i - w_i)^2 \right]^{\frac{1}{2}} - w_0 \right) \exp(iw_0) dw_0 = \int_{\sum_{i=1}^n x_i w_i - A'}^{\sum_{i=1}^n x_i w_i + A'} \psi(t) \exp(-it) dt \times \exp \left[\sum_{i=1}^n (x_i - w_i)^2 \right]^{\frac{1}{2}}, \quad (38)$$

Исходя из того, что $U(x) \in L^1$, и компактности $[-A, A]^n \times K$, можно взять A' так, чтобы

$$\left| \int_{-A'}^{A'} \psi \left(\left[\sum_{i=1}^n (x_i - w_i)^2 \right]^{\frac{1}{2}} - w_0 \right) \exp(iw_0) dw_0 - \int_{-\infty}^{\infty} \psi \left(\left[\sum_{i=1}^n (x_i - w_i)^2 \right]^{\frac{1}{2}} - w_0 \right) \exp(iw_0) dw_0 \right| \leq \frac{\varepsilon (2\pi)^2 |\Psi(1)|}{\left(2 \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |U(x)| dx + 1 \right)} \times \int_{-A'}^{A'} \dots \int_{-A'}^{A'} |U(x)| dx \quad \text{on } K \quad (39)$$

Следовательно,

$$\max_{x \in K} \left| I_{A', A}(x_1, \dots, x_n) - I_{\infty, A}(x_1, \dots, x_n) \right| \leq \frac{\varepsilon}{\left(2 \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |U(x)| dx + 1 \right)} \times \int_{-A'}^{A'} \dots \int_{-A'}^{A'} |U(x)| dx < \frac{\varepsilon}{2} \quad (40)$$

Доказательство третье

Исходя из (35) и (37), можно сказать, что для любого $\varepsilon > 0$ существуют $A, A' > 0$ такие, что

$$\max_{x \in K} \left| u(x_1, \dots, x_n) - I_{A', A}(x_1, \dots, x_n) \right| < \varepsilon \quad (41)$$

$u(x)$ может быть приближена с помощью конечного интеграла $I_{A', A}(X)$ на K . Интеграл $I_{A', A}(X)$ непрерывен на $[-A', A'] \times \dots \times [-A, A] \times K$ и может быть заменен, так как $I_{A', A}(X)$ может быть приближен с помощью сумм Римана на K .

Так как

$$\psi \left(\left[\sum_{i=1}^n (x_i - w_i)^2 \right]^{\frac{1}{2}} - w_0 \right) = \varphi \left(\frac{\left[\sum_{i=1}^n (x_i - w_i)^2 \right]^{\frac{1}{2}}}{\delta} - w_0 + \alpha \right) - \varphi \left(\frac{\left[\sum_{i=1}^n (x_i - w_i)^2 \right]^{\frac{1}{2}}}{\delta} - w_0 - \alpha \right) \quad (42)$$

сумма Римана может быть представлена в трехслойной сети. Поэтому $u(x)$ может быть представлена на трехслойной сети.

Теорема 3.4. Пусть H линейное пространство. Множество радиальных базисных функций обозначается $\Phi = \{\exp o\phi : \phi \in H^*\}$ фундаментальна в $C(H)$.

Доказательство. Пусть T обозначает линейную оболочку множества, Φ как я докажу T алгебра в $C(H)$, для $x \in H$ и $\phi, \psi \in H^*$ (где H^* является двойственным пространством H) [21]

$$[(\exp o\phi)(\exp o\psi)](x) = (\exp o\phi)(x) \cdot (\exp o\psi)(x) = \exp(\phi(x)) \exp(\psi(x)) = \exp(\phi(x) + \psi(x)) = \exp(\phi + \psi)(x) \quad (43)$$

Пусть ϕ_0 будет нулевой функционал в H^* , то $(\exp o\phi_0) = 1$ для всех $x \in H$. Следовательно, T содержит константы.

Теперь, если $x, y \in H$ и $x \neq y$, то существует $\phi \in H^*$ такой, что $\phi(x - y) \neq 0$.

$$\exp[\phi(x - y)] \neq \exp(0) = 1$$

$$\exp[\phi(x) - \phi(y)] \neq 1$$

$$\exp[\phi(x)] \cdot \exp(-\phi(y)) \neq 1$$

$$\frac{\exp[\phi(x)]}{\exp(-\phi(y))} \neq 1$$

$$\exp \phi(x) \neq \exp(\phi(y)) \quad (44)$$

Таким образом, T отделяет точку H .

Теперь, я докажу, что каждая базовая окрестность неподвижной точки в $C(H)$ пересекает T ; K компактное множество, такое, что $K \subset H$, основные окрестности u соответствуют K и имеют вид $B_{K,\varepsilon} = \{g \in C(H) : \|u - g\|_K < \varepsilon\}$, после этого ограничим u и все члены от T до K , потом $\{h/K : h \in T\}$ до сих пор на алгебре с содержанием распространяющих констант для K , следовательно, $\{h/K : h \in T\}$ плотно в $C(K)$, и, следовательно, T пересекает $B_{K,\varepsilon}$, что и требовалось. Таким образом, T является основополагающим в $C(H)$.

3.1.4 ЧИСЛЕННЫЕ МЕТОДЫ

Отказавшись от сетей с однослойным скрытым слоем и с выходным слоем, содержащим одну единицу, чтобы упростить изложение без потери общности.

Сеть предназначена для выполнения отображения нелинейной формы от входного пространства к скрытому пространству с последующим линейным отображением из скрытого пространства на выходное пространство.

Учитывая Множество N различных точек $\{x_i \in R^n, i = 1, 2, \dots, n\}$ и соответствующее множество N действительных чисел $\{u_i \in R, i = 1, 2, \dots, n\}$, найдем функцию $F: R^n \rightarrow R$, удовлетворяющую условию интерполяции.

$$f(x_i) = u_i, i = 1, 2, \dots, n \quad (45)$$

Для строгой интерполяции интерполяционная поверхность (то есть; функция $f(x)$) является сдерживающей на пути через все учебные точки данных, начальная $w^{(0)} \in R^{n+m}$, $\theta^{(0)}, d^{(0)} \in R$, $v^{(0)} \in R^n$, мы будем использовать алгоритм обратного распространения нейронной сети [21], чтобы найти лучший вес W, V, d и θ , вставляя условия интерполяции из (24) в (29).

Получится следующая система линейных уравнений для неизвестных коэффициентов разложения

$$\begin{bmatrix} \phi_{11} & \dots & \phi_{1n} \\ \vdots & \ddots & \vdots \\ \phi_{n1} & \dots & \phi_{nn} \end{bmatrix} \begin{bmatrix} v_1 \\ \dots \\ v_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix} \quad (46)$$

где

$$\phi_{ij} = \phi(\|W_j^i - X_i\|), (i, j) = 1, 2, \dots, n.$$

Вычислить ошибки

$$E = \left(\frac{1}{2}\right) \sum_{i=0}^N (y_i - u_i)^2,$$

и

$$\delta_i = (y_i - u_i) * \phi'(y_i).$$

Рассчитать

$$\Delta v_i = \alpha \delta_i * Z_i = \alpha \delta_i * \|X_i - W_j^i\| \quad (\alpha \text{ является обучение } 0 < \alpha < 1),$$

$$\Delta d_i = \alpha \delta_j, \rho_i = \sum_{j=1}^N \delta_j V_j, \beta_i = \rho_i \alpha'(\|X_i - W_j^i\|), \Delta W_{ij} = \alpha \beta_i x_i$$

$$\Delta\theta_i = \alpha\beta_i, V_i^{new} = V_j^{old} + \Delta V_j, W_{ij}^{new} = W_{ij}^{old} + \Delta W_{ij}, \theta_i^{new} = \theta_j^{old} + \theta_i, d_i^{new} = d_j^{old} + \Delta d_i, (47)$$

После этого лучшие веса W, V, θ, d найдены. Учитывая замену в (29), приближенное решение будет найдено.

Описанный выше метод должен работать в этой ситуации. В таком случае, значения отсутствия точного решения на независимых переменных будут заменены в граничных условиях для того, чтобы получить точные значения. Эти значения могут быть использованы на этапе подготовки рассматриваемого обратного распространения нейронной сети. Через этот подход можно получить значение хороших весов в (29). Радиальная базисная функция в теоремах 3.3 и 3.4 была использована для получения сходимости обучения нейронной сети обратного распространения.

3.1.5 ЧИСЛЕННЫЙ ПРИМЕР

Следующее двустороннее дробное уравнение в частных производных

$$\frac{\partial u(x,t)}{\partial u} = c_+(x,t) \frac{\partial^{1.8} u(x,t)}{\partial_+ x^{1.8}} + c_-(x,t) \frac{\partial^{1.8} u(x,t)}{\partial_- x^{1.8}} + s(x,t) \quad (48)$$

были рассмотрены на конечной области $0 < x < 2$ и $t > 0$ с коэффициентом функции $c_+(x,t) = \Gamma(25)x^{1.8}$ и вынужденной функции

$$s(x,t) = -32e^{-t} \left[x^2 + (2-x)^2 - 2.5(x^3 + (2-x)^3) + \frac{25}{22}(x^4 + (2-x)^4) \right] \quad (49)$$

Таблица 1: подготовка данных по значениям граничных условий.

x	t	$U(x,t)$	$Y(x,t) \times e^{-5}$
0.1	0	0.1444	0.1678
0.2	0	1.0404	1.5493
0.5	0	0.2500	0.9654
1.3	0	3.3124	6.7412
1.5	0	2.25	4.7759
1.7	0	1.0404	5.7418
2	0.1	0	0
2	0.4	0	0
2	0.6	0	0

Таблица 2: сравнение точного решения $u(x_i, t_i)$ и приближенного решения контрольных точек $y(x_i, y_i)$ при $\Delta x_i = 0.2, \Delta t_i = 0.1$

x	t	$U(x, t)$	$Y(x, t)$
0.2	0.1	0.4691	0.4246
0.4	0.2	1.3414	1.3752
0.6	0.3	0.2490	0.2831
0.8	0.4	2.0906	2.0597
1.0	0.5	2.4711	2.4244
1.2	0.6	2.4261	2.4997
1.4	0.7	2.0231	2.0488
1.6	0.8	0.7362	0.7651
1.8	0.9	0.2107	0.2239

Максимальная погрешность 0,0736

Начальное условие $u(x, 0) = 4x^2(2 - x)^2$ и граничные условия $u(0, t) = u(2, t)$. Это дробное уравнение в частных производных имеет точное решение $u(x, t) = 4e^{-t}x^2(2 - x)^2$, которые могут быть проверены применением дробной формулы

$$D_{L+}^{\alpha}(x - L)^{\rho} = \frac{\Gamma(\rho+1)}{\Gamma(\rho+1-\alpha)}(x - L)^{\rho-\alpha}, D_{R-}^{\alpha}(R - x)^{\rho} = \frac{\Gamma(\rho+1)}{\Gamma(\rho+1-\alpha)}(R - x)^{\rho-\alpha} \quad (50)$$

см.[11]

Таблица 1 показывает обучающие данные по значениям граничных условий.

Таблица 2 показывает сравнение точного решения и $u(x_i, t_i)$ и приближенного решения контрольных точках $y(x_i, y_i)$ при $\Delta x_i = 0.2, \Delta t_i = 0.1$.

Таблица 3 сравнивает максимальную погрешность с приближенным решением методом искусственных нейронных сетей и метода конечных разностей численного метода [11].

Таблица 3: Сравнение между максимальной погрешностью и приближенным решением.

Δx_i	Δt_i	FDM Maximum error	NNS Maximum error
0.2	0.1	0.1417	0.0736
0.1	0.05	0.0571	0.0541
0.05	0.025	0.0249	0.0217
0.025	0.0125	0.0113	0.0082

3.1.6 ЗАКЛЮЧЕНИЕ ГЛАВЫ

Возвращаясь к таблицам 1-3, ясно, что приближение дробных уравнений в частных производных с помощью нейронной сети с RBF позволяет хорошо приблизить решение. Кроме того, обсуждается способность метода для решения дробных уравнений в частных производных с более чем двумя переменными, где многие другие методы не срабатывают. Предлагаемый метод дает общее приближенное решение к интервалу определения и зависит от граничных условий.

3.2 РЕШЕНИЕ НЕЛИНЕЙНЫХ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ПОМОЩЬЮ FFANN

3.2.1 ВВЕДЕНИЕ В ГЛАВУ

Быстрое развитие в области коннекционизма связано с появлением параллельных, распределенных и адаптивных систем обработки информации, которые включают в себя такие инструменты, как генетические системы обучения, имитации отжига систем, ассоциативные воспоминания и нечеткие системы обучения. Однако основным инструментом, который представляет интерес, являются искусственные нейронные сети.

Термин искусственно нейронной сети относится к любой системе обработки информации, которые имеют биологическое начало. Обычно они принимают форму структур направленных графов [22], вершинами которых являются данные, подлежащие обработке через математические операции. Информация представлена в них с помощью численного веса, связанного между вершинами системы. Математические операции, выполняемые этими вершинами, способ их соединения с другими узлами и поток информации сквозь структуру расположены и работают на принципе биологической нервной системы. Большая часть терминологии, связанная с этими системами, также имеет отношение к биологическим структурам, поэтому сети часто называют тренированными, а не программированными, и обученными, а не моделированными.

ИНС зарекомендовали себя как универсальный инструмент для достижения того, что можно было бы назвать задачами более высокого класса, такие как датчик синтеза, распознавания образов, классификации и визуальной обработки. Все эти приложения представляют большой интерес в области инжиниринга, однако, настоящие приложения ИНС создали мнение, что сети не подходят для работы в задачах, требующих аккуратности и точности, например,

математического моделирования или физического анализа инженерных систем. Конечно, биологические основы нейро-сетевой концепции предполагают, что сети работали бы лучше на задачах, на которых биологические системы превосходят.

В противоположность этому мнению авторов считают, что дальнейшие исследования в приближении возможности сетей покажет, что нейронные сети могут быть как численно точной и предсказуемой системой исчисления, как и обычные вычислительные методы. Считается также, что при рассмотрении ИНС в виде инструмента числовых исчислений, можно заметить определенные особенности в обучении и анализе существующих сканирования архитектур. В более непосредственном смысле преимущества этого подхода позволят нейро-сетевой парадигме со всем своим богатством поведения и способностью к адаптации, которая в паре с более чисто вычислительными парадигмами математически ориентированного программирования, используется инженерами в области моделирования и анализа.

В настоящее время самым популярным приложением ИНС в науке и технике является эмуляция физических процессов по архитектуре искусственных нейронных сетей прямой связи (FFANN) с использованием алгоритма обучения. Потому что эта парадигма требует воздействия многочисленных множеств ввода-вывода, это может стать большим и трудоемким процессом, а также занять большой объем памяти. Значительные усилия можно сохранить, если математическую модель физического процесса можно будет включить непосредственно и точно в архитектуру FFANN без необходимости примеров, таким образом, сокращая и устраняя фазу обучения.

В предыдущей работе рассматривались математическое моделирование или коннекционистское решение уравнений, как новая область, в которой применяются обычные коннекционистские парадигмы. Например, Такеда и Гудман [23] и Барнард и Casasent [24] используют парадигму оптимизации сети, особенно вариации на сетях Hopeld, для решения систем линейных уравнений. Линейные системы, которые должны быть решены, выражены в виде минимизации квадратичной функции, и веса сети Hopeld обучаются, пока функция не сводится к минимуму. Ли и Кан [25] расширяют это использование сетей Hopeld для решения дифференциальных уравнений. Wang и Mendel [26] используют более или менее стандартные FFANN архитектуры с задним алгоритмом распространения для решения задач линейной алгебры, подвергая

сеть неоднократно решению с соединением структуры ограничения характера решаемой задачи.

В этих статьях авторы используют подход для исследования возможностей проведения параллелей между сетью и определенными математическими объектами из теории функций приближения. Таким образом, будет показано, что FFANNs могут быть построены для моделирования заданных нелинейных обыкновенных дифференциальных уравнений без необходимости обучения или создания множеств данных. Веса этих сетей могут быть определены путем решения системы линейных и нелинейных алгебраических уравнений. Полученные сети архитектурно идентичны тем, которые используют традиционные методы обучения, но они обладают монотонно убывающей ошибкой увеличения числа процессорных элементов.

Прогресс в этом подходе должен также представлять интерес для коннекционистского сообщества, так как такой подход может быть использован в качестве относительно прямого метода для изучения влияния различных параметров, регулирующих FFANN. Кроме того, применение FFANN к решению дифференциального уравнения может эффективно отделить влияние качества выборок данных, сетевой архитектуры и транспортных функций для приближения сети.

Простые FFANN для численных примеров с одним входным и несколькими выходными нейронами, с одним скрытым слоем обработки элементов и с использованием кусочно-линейного отображения в виде активационной функции используются для точного приближения решения двух тестовых задач из области механики жидкости и газа. Первая проблема возникает при тестировании нелинейного обыкновенного дифференциального уравнения третьего порядка, известного как уравнение Фолкнер-Скана. Вторая проблема заключается в системе трех связанных нелинейных обыкновенных дифференциальных уравнений второго и третьего порядка, а также уравнений из области механики жидкости. Кроме того, будет показано, как можно устранить погрешности в построенной FFANN.

3.2.2 ПРИБЛИЖЕНИЕ ФУНКЦИИ

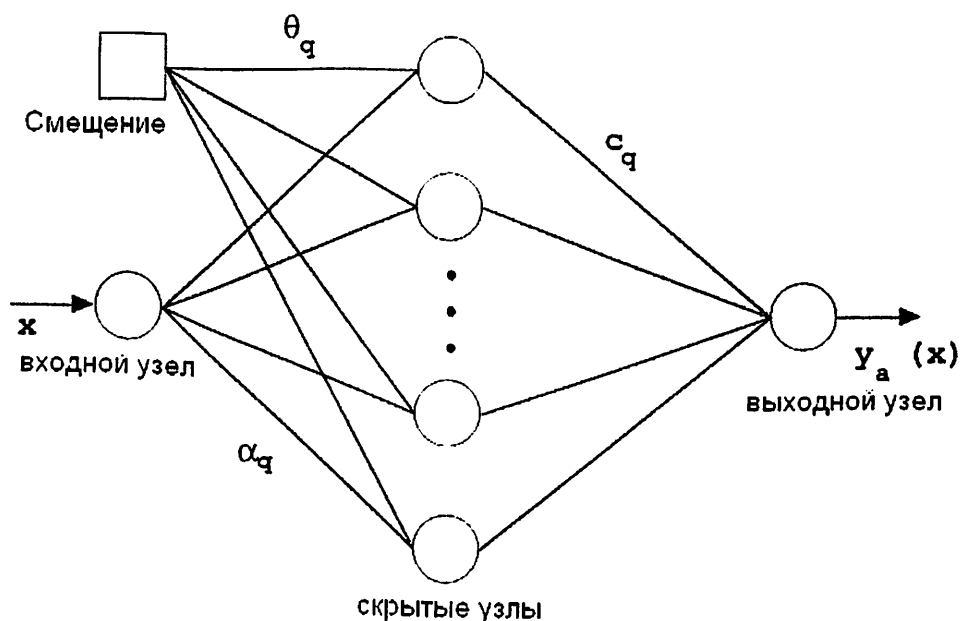
Теория приближения функций, которая решает задачи приближения или интерполяции непрерывных функций с множеством переменных. Этот метод

используется другими исследователями [27] - [29] для объяснения обучения с учителем и поведения ИНС в целом. Имея это в виду, считают, что наиболее распространенный метод приближения функций является функциональное расширение базиса. Функциональное расширение базиса представляет произвольную функцию $y(x)$ в виде взвешенных комбинаций линейно независимых базисных функций (Y), которые являются функциями независимой переменной x ,

$$y(x) = \sum_{q=1}^T Y_q(x)c_q \tag{51}$$

где T - общее число базисных функций и c_q - базисные коэффициенты. Типичным примером функционального расширения базиса является расширение рядом Фурье, которое использует функции с тригонометрическим базисом. Другие базисные функции можно использовать с тем же успехом, но значения коэффициентов c_q , варьируются в зависимости от выбранных базисных функций.

Исследование уравнения (51) показывает, что это скалярное произведение базисных функций и разложение коэффициентов. Кроме того, математические операции, выполняемые простой FFANN (рисунок 8) можно интерпретировать как скалярное произведение передаточной функции выходных значений и выходных весов.



θ_q : Смещенный вес для q -того скрытого узла

α_q : Входной вес для q -того скрытого узла

c_q : Выходной вес для q -того скрытого узла

Рисунок 8 - Стандартное построение нейронной сети

Апеллируя к теории приближения функций, можно разумно выбрать передаточные функции и распределение веса, которые оптимизируют работу FFANN в представляющих функциональных связях. Кроме того, мы можем назначать роли различных параметров в сети. Например, выходные веса могут выступать в качестве коэффициентов расширения для базиса и, в принципе, должно быть вычислимым при помощи обычных методов. Похожие замечания находятся в Cybenko [30].

Мы выбрали линейное отображение по частям [31] на рисунке 9 в качестве активационной функции для скрытого слоя, так как оно является одной из простейших функций для введения как на цифровых компьютерах так и на электронных аппаратах [32]. Кусочно-линейная функция активации может быть смоделирована с помощью следующих уравнений:

$$Y_q = -1.0 \text{ для } \xi_q < -1.0$$

$$Y_q = \xi_q(x) \text{ для } -1.0 < \xi_q < 1.0$$

$$Y_q = +1.0 \text{ для } \xi_q > 1.0 \tag{52}$$

где проверка рисунка 8 показывает, что передаточная функция действует на сумму взвешенных входов и смещение значений ($\xi_q = \alpha_q x + \theta_q$), т.е. линейного преобразования независимой переменной x . Эта нотация похожа на не масштабированное смещение вращений Ито [33].

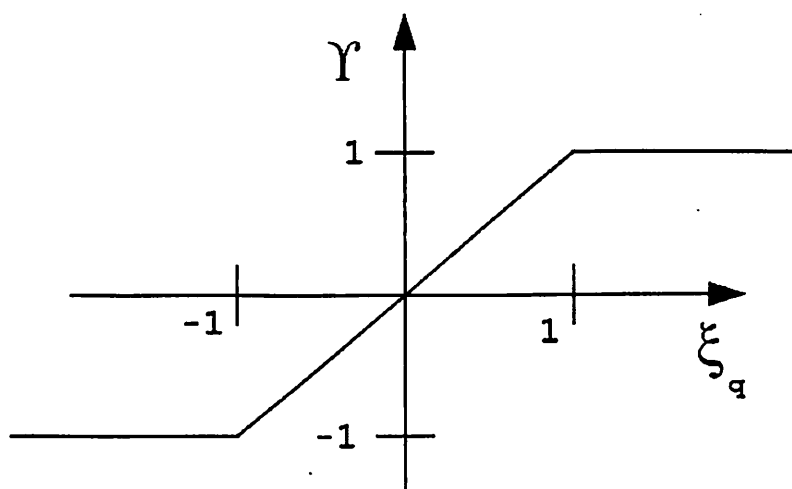


Рисунок 9 - Кусочно-линейное отображение, используемое как активационная функция.

Базисные функции в литературе о приближении функции описывается как локальная и глобальная как говорят, быть либо местными (экспонирование значения, отличного от нуля в ограниченной части области определения) или глобальным (ненулевой в большей части в области определения). Кусочно-линейную функцию активации можно рассматривать как глобальный полиномиальный базис линейной функции, определенные частичным образом с помощью превращения ξ_q . Все глобальные базисные функции имеют аналогичные недостатки, в том числе и жаберное кондиционирование при решении разложения коэффициентов и плохое приближение между точками дискретизации (узлами) [34].

Можно предположить, что многие из распространенных проблем, связанных с представлением FFANN могут возникнуть просто от использования математически плохо подходящих для приближения базиса. В теории приближения функций недостатки глобальных базисных функций можно избежать с помощью частично многочленных сплайнов.

Наиболее распространенным полиномиальным сплайном является Чарпау или шляпная функция, как показано на рисунке 10

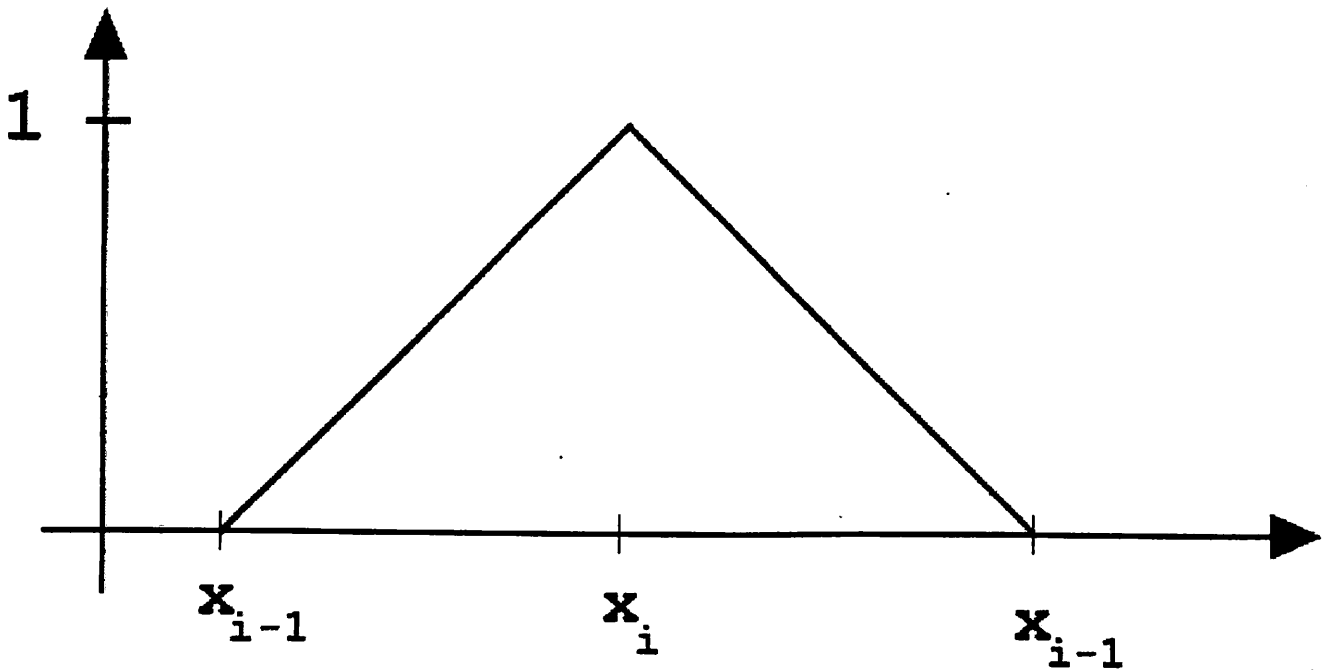


Рисунок 10 - Функция Шароу или Шляпная функция.

Она определяется так

$$\Phi_i = \frac{(x - x_{i-1})}{(x_i - x_{i-1})} \quad \text{для } x_{i-1} < x < x_i$$

$$\Phi_i = \frac{(x_{i+1} - x)}{(x_{i+1} - x_i)} \quad \text{для } x_i < x < x_{i+1}$$

$$\Phi_i = 0 \quad \text{для } x < x_{i-1} \quad \text{или} \quad x > x_{i+1} \quad (53)$$

Отметим, что многочлен варьируется в значениях от 0 до 1.

Рисунок 10 показывает, как шляпная функция должна быть распределена для обеспечения линейной интерполяцией между узлами в проблемной области $[x_1, x_N]$, дискретизированной с $N = 3$ узлами (x_1, x_2 и x_3). Алгебраическая система, разработанная для оценки разложения коэффициентов для частичного полиномиального сплайна, очень редка и в хорошем состоянии. Поскольку максимальное значение сплайнов равно 1, значения расширения коэффициентов являются идентичными значениям приближенных функций в узлах. Недостатком является то, что приближение разрывным после первой производной.

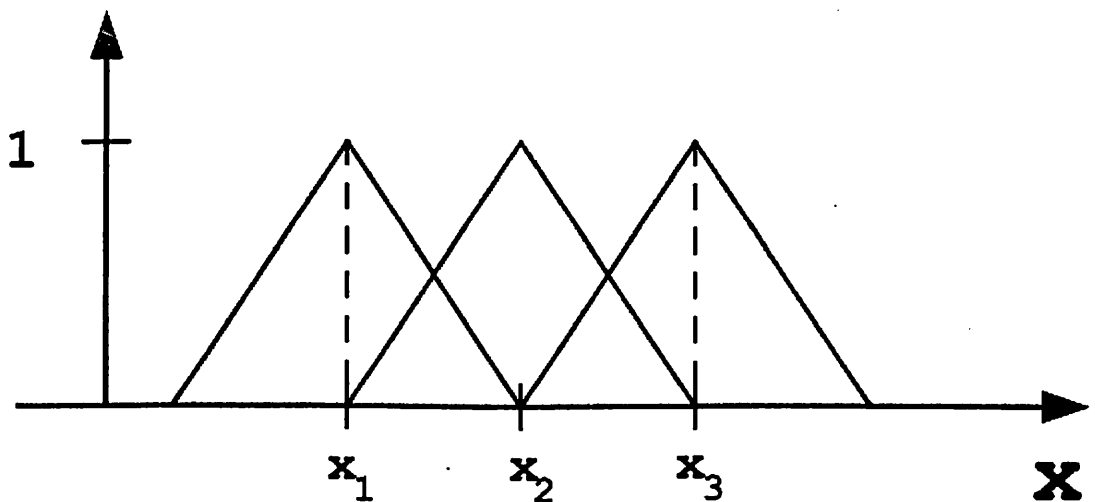


Рисунок 11 - Правильное распределение шляпной функции на проблемной области.

3.2.3 ТРАНСФОРМАЦИЯ КУСОЧНО-ЛИНЕЙНЫХ ФУНКЦИЙ АКТИВАЦИИ СПЛАЙНОВ ПО ОГРАНИЧЕНИЮ ВЕСА

Все математические преимущества шляпной функции предполагают, что ее преобразования были бы выгодны при представлении кусочно-линейного отображения FFANN в шляпную функцию. К счастью выбор кусочно-линейного отображения в виде функции активации упрощает эти моменты.

Рассмотрим одномерную область как на рисунке 12. Функция расположена между x_i и x_{i+1} , затем она умножается на -1 и прибавляется ко второй неизменной функции (рисунок 13), и в результате получается шляпная функция с максимальным значением, равным 2. Превращение кусочно-линейной функции активации в шляпную функцию может быть описано следующим уравнением:

$$\gamma_i^A - \gamma_i^B = 2\Phi_i, \quad (i = 1, 2, \dots, N) \quad (54)$$

где супер скрипты активационной функции A и B относятся к прилегающим интервалам $x_{i-1} \leq x \leq x_i$ и $x_i \leq x \leq x_{i+1}$ соответственно. Функции γ_i^A и γ_i^B определяются как ноль в серединах соответствующих интервалах. Следствием этой формулировки является то, что число активационных функций может быть связано с необходимым числом шляпных функций. Наш подход требует в два раза больше кусочно-линейных функций, чем шляпных функций ($T = 2N$).

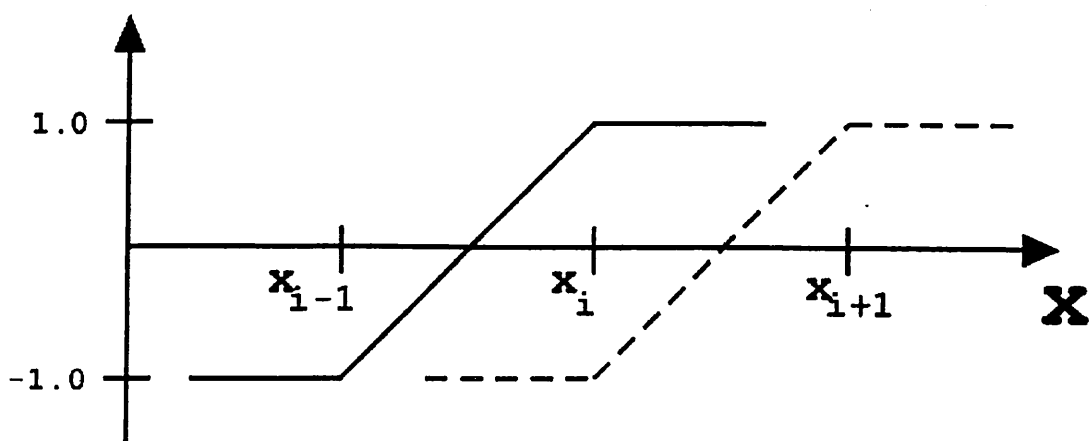


Рисунок 12 - Ограниченное распределение кусочно-линейной активационной функции по оси Ox , используя входные и смещенные веса.

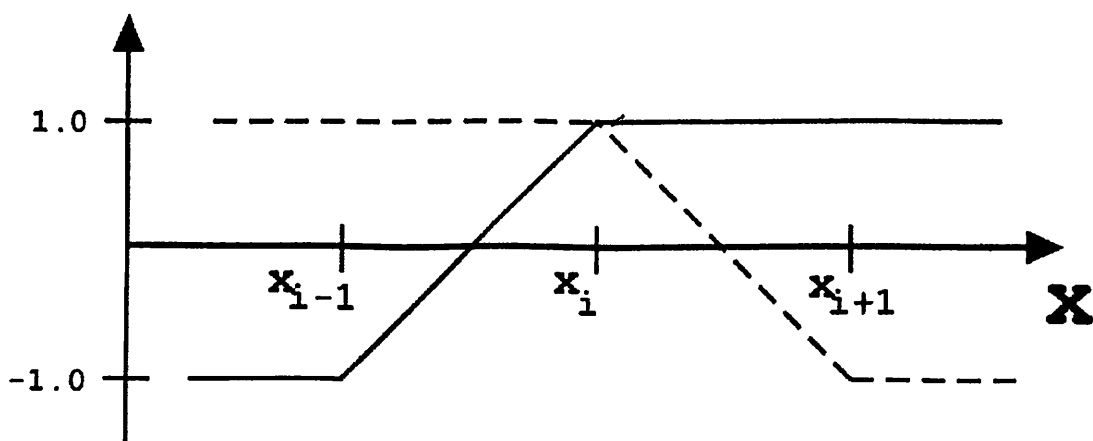


Рисунок 13: Правосторонняя активационная функция (пунктир), повернутая по изменению знака выходного веса. В сумме две активационные функции дают шляпную функцию. (см. рисунок 10)

Теперь определим оставшиеся ограничения, необходимые для уравновешения представления активационной функции (уравнение (51)) и шляпной функции. Давайте дискретизируем проблемную область $[a, b]$ с помощью узлов x_i следующим образом:

$$a = x_1 < x_2 < \dots < x_{N-1} < x_N = b$$

Мы обозначим эту дискретизацию как проблемная сетка. Для создания соответствующих шляпных функций на границах необходимо наличие двух вспомогательных узлов x_0 и x_{N+1} такие, что $x_0 < a$ и $x_{N+1} > b$. Используя ограничение $T = 2N$, мы можем переписать уравнение (51) так

$$\sum_{q=1}^{T=2N} Y_q(x) c_q = \sum_{q=1}^N Y_q(x) c_q + \sum_{q=N+1}^{2N} Y_q(x) c_q, \quad (55)$$

где слагаемые правой стороны можно переписать без потери общности, как

$$\sum_{q=1}^N Y_q c_q = \sum_{i=1}^N Y_i^A u_i \text{ и } \sum_{q=N+1}^{2N} Y_q c_q = \sum_{i=1}^N Y_i^B v_i. \quad (56)$$

Поэтому

$$\sum_{q=1}^{T=2N} Y_q(x) c_q = \sum_{i=1}^N Y_i^A u_i + \sum_{i=1}^N Y_i^B v_i. \quad (57)$$

Тогда можно показать, что основное разложение по кусочно-линейной активационной функции может быть преобразовано в основе разложения по шляпным функциям, выбрав следующие значения для FFANN весов на рисунке 8 [35]. Для входных весов и смещения веса:

$$\alpha_i^A = \frac{2}{(x_i - x_{i-1})}, \alpha_i^B = \frac{2}{(x_{i+1} - x_i)}, \theta_i^A = -\frac{2x_{i-1}}{(x_i - x_{i-1})} - 1, \theta_i^B = -\frac{2x_i}{(x_{i+1} - x_i)} - 1, \quad (58)$$

где α_i^A , α_i^B , θ_i^A , θ_i^B соответствуют активационным функциям уравнения (54). Для выходного веса:

$$c_i = u_i = \frac{w_i}{2}, (i = 1, 2, \dots, N)$$

$$c_{i+N} = v_i = \frac{-w_i}{2}, (i = 1, 2, \dots, N) \quad (59)$$

где нумерация веса по отношению к фактической сетевой архитектуре может быть в любом порядке.

Применение уравнений (54) - (59) позволяет сделать следующие эквивалентности:

$$y_a(x) = \sum_{q=1}^T Y_q(x) c_q = \sum_{i=1}^N \Phi_i(x) w_i \quad (60)$$

Шляпные функции Φ_i определяются как имеющий значение 1 в узле x , как и в формуле (53), и w_i базисные коэффициенты, связанные с шляпными функциями.

В то время как веса были ограничены в этом подходе, архитектура сети была сохранена. Рисунок 14 показывает результирующую архитектуру для обработки шести элементов, используемых для формирования трех шляпных функций с тремя узлами и с двумя вспомогательными узлами, как на рисунке 11.

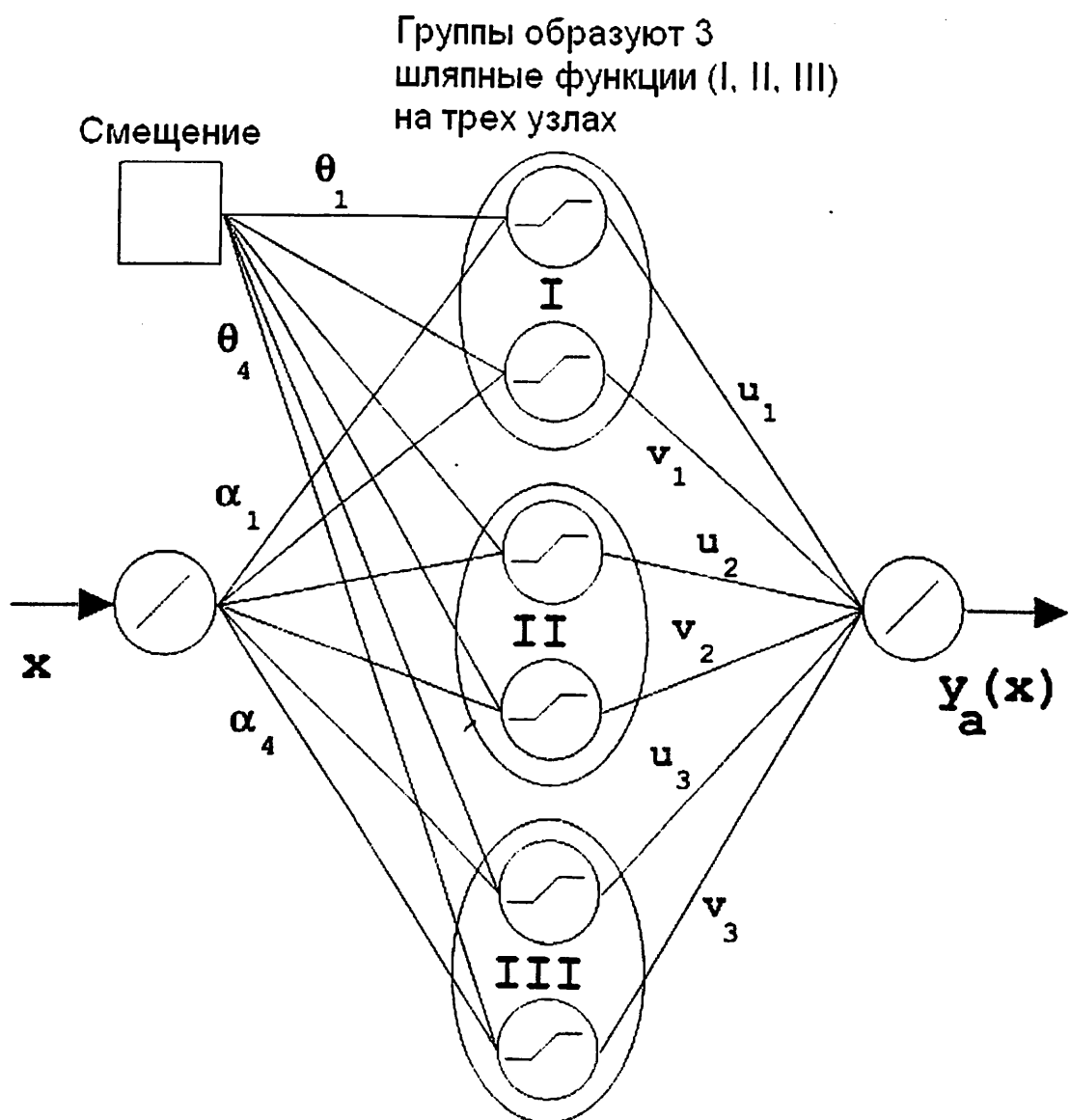


Рисунок 14 - Сеть с одним скрытым слоем и ограниченными весами ($T = 6, N = 3$)

Следует отметить, что нет никакой тренировки сети в этом подходе. Значения входов и смещения веса фиксированы для заданного числа узлов в проблемной области и эффективно отсоединяется от определения выходного веса. Построение сети было уменьшено для нахождения соответствующих значений коэффициентов w_i , и затем с помощью w_i определяется значение пары выходных весов u_i и v_i . Вычисление разложения коэффициентов w_i могут быть получены с помощью различных численных методов. Мы применим процедуру, вытекающую из метода взвешенных невязок [36].

3.2.4 ОПРЕДЕЛЕНИЕ ВЫВОДА ВЕСА: МЕТОД ВЗВЕШЕННЫХ НЕВЯЗОК

Дифференциальное уравнение, представленное в виде $L(y) = G(x)$, где $L(y)$ является некоторым дифференциальным оператором, действующим на зависимой переменной y , приближительное расширение y заменяется на y_α , где

$$y_\alpha(x) = \sum_{i=0}^N \Phi_i(x) w_i \quad (61)$$

и где $\Phi_i(x)$ базисная функция. Поэтому $L(y_\alpha) - g(x) = R(w_1, \dots, w_N, x)$, где - некоторая функция ненулевого значения, которое может быть описано как ошибки или дифференциальное уравнение остаточной.

В методе взвешенных невязок, остаточная умножается или нагружается функцией $F(x)$ (тестовая функция), интегрированной по проблемной области $[a, b]$, и равна 0,

$$\int_a^b F R dx = (F, R) = 0, \quad (62)$$

где (F, R) является внутренним произведением функций F и R . Заметим, что даже при том, что интеграл равен нулю, дифференциальное уравнение остаточные R стремится к нулю только в среднем и только в том случае, если $x_{i+1} - x_i$ стремится к нулю. В коннекционистской схеме это равносильно стремлению количества процессорных элементов (T) в скрытом слое к бесконечности. Уравнение (12) эквивалентно

$$(F, L(y_\alpha) - g(x)) = 0, \text{ или } (F, L(y_\alpha)) = (F, g) \quad (63)$$

Многие из наиболее популярных методов приближения функций, такие как конечные разности, субдоменный метод, конечные элементы, расстановка и спектральный метод могут быть получены из уравнения (63) при соответствующем выборе функции $F(x)$ [37]. Мы будем использовать технику Петрова-Галеркина [37], где $F(x)$ является модификацией базисных функций, использованной для описания приближения y_α

$$F(x) = F_k(x) = a(x)\Phi_k(x), k = 1, 2, \dots, N \quad (64)$$

где $a(x)$ выбирается, например, чтобы удовлетворить граничные условия задачи. Техника Петрова-Галеркина была выбрана для определения выходного веса, поскольку она облегчает дробление высоко нелинейной или прерывистых

проблем. Следует отметить, что техника Петрова-Галеркина работает только тогда, когда используемые базисные функции линейно независимы, как шляпная функция.

После того, как $F(x)$ определена, вычисления интегралов приводит к системе алгебраических уравнений. Обычно это система линейных уравнений, если приближение дифференциального уравнения является линейным. В противном случае алгебраические уравнения являются нелинейными. Для получения однозначного решения эти уравнения могут быть оценены по стандартным методикам [38].

3.2.5 ЧИСЛЕННЫЕ ПРИМЕРЫ

3.2.5.1 УРАВНЕНИЕ ФОЛКНЕР-СКАНА

В первом численном примере мы взяли один вход, множественный выход FFANN для приближения решения, первые и вторые производные решения для следующего нелинейного обыкновенного дифференциального уравнения:

$$\frac{d^3 f}{d\eta^3} + f \frac{d^2 f}{d\eta^2} + \beta \left[1 - \left(\frac{df}{d\eta} \right)^2 \right] = 0 \quad (65)$$

с соответствующими граничными условиями

$$f = \frac{df}{d\eta} = 0 \text{ при } \eta = 0; \frac{df}{d\eta} \rightarrow 1 \text{ при } \eta \rightarrow \infty \quad (66)$$

Кроме того, по сравнению с предыдущим граничные условия

$$\frac{df}{d\eta} \rightarrow 1 \text{ при } \eta \rightarrow \infty \text{ затем } \frac{d^2 f}{d\eta^2} \rightarrow 0 \text{ при } \eta \rightarrow \infty \quad (67)$$

Уравнение (65), известное как уравнение Фолкнер-Скана, происходит от преобразования подобия системы уравнений в частных производных при устойчивой модели с двумерным и несжимаемым импульсом в граничном слое. Центральное место в преобразовании подобия занимают следующие ограничения

$$x \geq 0, y \geq 0, \bar{u}(x, y) = \bar{U}(X) \frac{df}{d\eta}(\eta), \bar{U}(x) = K x^{\beta/(2-\beta)}, \eta = y \sqrt{\frac{\bar{u}}{(2-\beta)\nu x}} \quad (68)$$

где x и y пространственные измерения в продольных направлениях нормалей к поверхности соответственно. Переменная η является пространственной

скоростью граничного слоя. $\bar{U}(x)$ является пространственной скоростью внешнего слоя, K - это константа пропорциональности со значением $+1$, ν является кинематической вязкостью жидкости, а также $\eta(x, y)$ является подобием переменной. Детальное обсуждение происхождения и свойств уравнения (65) можно найти в работе [39].

Решение уравнения (65) демонстрирует многие явления пограничного слоя, таких как сильные благоприятные градиенты, отрыв пограничного слоя, превышение скорости и неоднозначность. Параметр β является измерением давления градиента и может быть непосредственно связано с пограничным слоем около клиньев и расширения поворота углов θ , где

$$\theta = \frac{\beta\pi}{2}$$

Если мы ограничим наше внимание на случаи, в которых уравнение (65) может физически моделировать подходящий поток жидкости и когда $\bar{u}(x, y)$ монотонно возрастает по y , то $-0,19884 \leq \beta \leq +\infty$. Как следствие уравнения (68) и $\bar{u}(x, y)$, $df/d\eta$ монотонно возрастает по η для ограниченного диапазона β . Следует отметить, что для $-0,19884 \leq \beta \leq +\infty$ в работе [41] показано, что существует по крайней мере два решения уравнения (65) для любой заданной β , но только одно из решений дает монотонно возрастающей $\bar{u}(x, y)$). Для примера этой проблемы мы будем исследовать три значения β

$\beta = -0,15$ обтекание расширения поворота угла $\theta = 13.5^\circ$

$\beta = 0$ обтекания вдоль на пластины

$\beta = +0,5$ обтекание клина половинного угла $\theta = 45^\circ$

В нашем подходе к решению уравнения Фолкнер-Скана отметим, что независимой переменной η в уравнении (65) полубесконечные, и само уравнение имеет конечные условия, применяемые на бесконечности. Тем не менее, для нашего ограниченного значения β простой асимптотический анализ уравнения (65), используя ограничение монотонно возрастающей $df/d\eta$, определяет, что зависимая переменная f , а также вторая и третья ее производная, изменяются монотонно по η . Таким образом, должно быть возможно преобразовать уравнение (65) в форму, в которой зависимая переменная или ее производные будут действовать в качестве независимой переменной. Как упоминалось ранее,

шляпная функция использует базисы при приближенном решении линейных уравнений, что означает, что вторая, третья и более высокие производные базисов, тождественно равны нулю. Для того, чтобы получить нетривиальное решение мы не только меняем независимую переменную η на u , но и с тем же успехом меняем зависимую переменную f на τ , где

$$u = \frac{df}{d\eta}, 0 \leq u \leq 1 \text{ и } \tau = \frac{d^2f}{d\eta^2} \quad (69)$$

С применением уравнения (69) и цепного правила, уравнение Фолкнер-Скана сводится к

$$\frac{d\tau}{du} + f + \beta \frac{(1-u^2)}{\tau} = 0 \quad (70)$$

Хотя производная большого порядка в уравнении была уменьшена, уравнение (70) все еще содержит f в явном виде. Для устранения f возьмем производную уравнения (70) относительно u и применим цепное правило, что приведет к изменению уравнения Фолкнер-Скана на уравнение:

$$\frac{d^2\tau}{du^2} + (1 - 2\beta) \frac{u}{\tau} + \beta(1 - u^2) \frac{d}{du} \left(\frac{1}{\tau} \right) = 0 \text{ для } 0 \leq u \leq 1. \quad (71)$$

Обратите внимание, что нелинейное обыкновенное дифференциальное уравнение третьего порядка на неограниченной проблемной области было преобразовано путем замены переменных на более управляемые дифференциальные уравнения второго порядка на единичном домене. Граничные условия в уравнениях (66) и (67) переписаны с ссылкой на преобразованные переменные:

$$u = 0 \text{ на } \eta = 0 \text{ и } u \rightarrow \infty \text{ на } \eta \rightarrow \infty$$

$$\tau = 0 \text{ на } u = 1$$

Применяя уравнение (62) к уравнению (71), получается слабая форма,

$$\left(F_k, \frac{d^2\tau}{du^2} \right) = -(1 - 2\beta) \left(F_k, \frac{u}{\tau} \right) - \beta \left(F_k, (1 - u^2) \frac{d}{du} \left(\frac{1}{\tau} \right) \right), k = 1, 2, \dots, N, \quad (72)$$

где, например

$$\left(F_k, \frac{d^2\tau}{du^2} \right) = \int_0^1 F_k \frac{d^2\tau}{du^2} du$$

Для понижения порядка производной к левой части уравнения (72) необходимо применить интегрирование по частям. Это приведет к:

$$\left(\frac{dF_k}{du}, \frac{d\tau}{du}\right) = (1 - 2\beta) \left(F_k, \frac{u}{\tau}\right) + \beta \left(F_k, (1 - u^2) \frac{d}{du} \left(\frac{1}{\tau}\right)\right) + \left[F_k \frac{d\tau}{du}\right]_0^1 \quad (73)$$

Введем приближенное решение τ_α . Как упоминалось в разделе 3.1, проблемная область для действующей независимой переменной и дискретизация в N узлов,

$$0 = u_1 < u_2 < \dots < u_N = 1$$

с двумя вспомогательными узлами за пределами проблемной области. С помощью входа и смещения веса кусочно-линейная активационная функция FFANN не способна сформировать шляпную функцию, направленную на каждый узел в проблемной области. Однако, вместо приближения τ с простым расширением базиса, как в уравнении (61), мы используем

$$\tau_\alpha = (1 - u) \sum_{i=1}^N \Phi_i b1_i \quad (74)$$

Следует отметить, что приближение τ равно нулю при $u = 1$, таким образом, преобразованное граничное условие выполняется автоматически. Вместо того, чтобы использовать обратную формулу (74) для приближения $1/\tau$ мы будем использовать метод приближения произведения [40] и введем отдельное приближение

$$\frac{1}{\tau_\alpha} = \frac{1}{(1-u)} \sum_{i=1}^N \Phi_i b2_i \quad (75)$$

Базисное расширение $1/\tau$ автоматически неограниченно при $u = 1$ и, как уравнение (74), базисные коэффициенты не требуют удовлетворения граничному условию.

Связь между множеством базисных коэффициентов может быть определена путем приравнивания $1/\tau_\alpha$ и обратная τ_α в каждом узле u_j для $j = 1, \dots, N$:

$$\frac{1}{\tau_\alpha(u_j)} = \frac{1}{(1-u_j)} \sum_{i=1}^N \Phi_i u_j b2_i = \frac{1}{(1-u_j) \sum_{i=1}^N \Phi_i(u_j) b1_i} \quad (76)$$

Для шляпной функции

$$\Phi_i(u_j) = 1, j = i \text{ и } \Phi_i(u_j) = 0, j \neq i \quad (77)$$

Таким образом, после замены фиктивной подписке j на i в уравнении (76) имеем

$$b2_i = \frac{1}{b1_i} \text{ для } i = 1, \dots, N \quad (78)$$

который обеспечивает дополнительное множество необходимых уравнений.

Метод приближения произведения применим также к проблемам с более чем одной зависимой переменной, назначая каждой мультипликативной группе переменных отдельные базисные разложения. Применение метода приближения произведения не оказывает вредного влияния на точность решения задачи, но увеличивает вычислительную эффективность, в то время как значительно упрощает выполнение нелинейных задач с помощью взвешенного остаточного метода [37].

Подставляя приближение и их производные в слабые формулировки результатов, получается

$$\begin{aligned} \sum_{i=1}^N \left(\frac{dF_k}{du}, (1-u) \frac{d\Phi_i}{du} - \Phi_i \right) b1_i &= (1-2\beta) \sum_{i=1}^N \left(F_k, \frac{u}{1-u} \Phi_i \right) b2_i + \\ + \beta \sum_{i=1}^N \left(F_k, \frac{(1-u^2)}{(1-u)^2} \left((1-u) \frac{d\Phi_i}{du} - \Phi_i \right) \right) b2_i &+ \left[F_k \frac{d\tau}{du} \right]_0^1 \end{aligned} \quad (79)$$

Отметим, что поскольку $\frac{1}{(1-u)}$ присутствует, это вызовет трудности в оценке внутреннего произведения, если его не устранить. Это может быть сделано путем применения техники Петрова-Галеркина для уравнения (79). Возьмем

$$F_k = (1-u)\Phi_k$$

Поэтому уравнение (79) переходит в

$$\begin{aligned} \sum_{i=1}^N \left((1-u) \frac{d\Phi_k}{du} - \Phi_k, (1-u) \frac{d\Phi_i}{du} - \Phi_i \right) b1_i &= (1-2\beta) \sum_{i=1}^N (\Phi_k, u\Phi_i) b2_i + \\ + \beta \sum_{i=1}^N \left(\Phi_k, (1+u) \left((1-u) \frac{d\Phi_i}{du} + \Phi_i \right) \right) b2_i &+ \left[(1-u)\Phi_k \frac{d\tau}{du} \right]_0^1 \end{aligned} \quad (80)$$

Если мы определим следующие матрицы

$$\begin{aligned}
M1 &= M1_{ki} \left((1-u) \frac{d\Phi_k}{du} - \Phi_k, (1-u) \frac{d\Phi_i}{du} - \Phi_i \right) \\
M2 &= M2_{ki} = (\Phi_k, u\Phi_i) \\
M3 &= M3_{ki} = \left(\Phi_k, (1+u) \left((1-u) \frac{d\Phi_i}{du} + \Phi_i \right) \right)
\end{aligned} \tag{81}$$

Затем алгебраическую систему сформированную с помощью уравнения (80) можно переписать так

$$\sum_{i=1}^N M1_{ki} b1_i = (1-2\beta) \sum_{i=1}^N M2_{ki} b2_i + \beta \sum_{i=1}^N M3_{ki} b2_i + \left[(1-u) \Phi_k \frac{d\tau}{du} \right]_0^1 \tag{82}$$

для

$$k = 1, \dots, N \text{ и } b2_i = \frac{1}{b1_i} \text{ для } i = 1, \dots, N \tag{83}$$

Характер шляпной функции по базису показывает, что матрицы **M1**, **M2**, **M3** положительно определены и тридиагональны, и поэтому преобретают порядок N ($O(N)$) [37]. Рассмотрим теперь термин

$$\left[(1-u) \Phi_k \frac{d\tau}{du} \right]_0^1 = \left[(1-u) \Phi_k \frac{d\tau}{du} \right]_1 - \left[(1-u) \Phi_k \frac{d\tau}{du} \right]_0$$

Так как $u = 0 = u_1$ и $u = 1 = u_N$, применение уравнения (77) показывает,

$$\left[(1-u) \Phi_k \frac{d\tau}{du} \right]_0 = \Phi_k(0) \frac{d\tau}{du}(0) = 0 \text{ для } k = 2, \dots, N \tag{84}$$

$$\left[(1-u) \Phi_k \frac{d\tau}{du} \right]_0 = \frac{d\tau}{du}(0) \text{ для } k = 1 \tag{85}$$

$$\left[(1-u) \Phi_k \frac{d\tau}{du} \right]_1 = 0 \text{ для } k = 1, \dots, N \tag{86}$$

Определение уравнения (70) при $u = 0$ и, учитывая, что $f(0) = 0$,

$$\frac{d\tau}{du}(0) = - \left[f + \beta \frac{(1-u^2)}{\tau} \right]_{u=0} = - \frac{\beta}{\tau(0)}$$

Таким образом, термин щ можно записать в виде

$$\left[(1-u)\Phi_k \frac{d\tau}{du} \right]_0^1 = \frac{\beta}{\tau(0)} = \beta \sum_{i=1}^N \Phi_i(0) b_{2i} = \beta b_{2i} \text{ для } k = 1 \quad (87)$$

Определим матрицу **M4**

$$M4_{ki} = (1 - 2\beta)M2_{ki} + \beta M3_{ki} + \beta \text{ для } k = 1 \text{ и } i = 1, \dots, N$$

и

$$M4_{ki} = (1 - 2\beta)M2_{ki} + \beta M3_{ki} \text{ для } k = 2, \dots, N \text{ и } i = 1, \dots, N \quad (88)$$

где **M4** положительно определена и тридиагональна. С помощью уравнений (82), (83), (87) и (88) получается алгебраическая система:

$$\sum_{i=1}^N M1_{ki} b_{1i} - \sum_{i=1}^N M4_{ki} \frac{1}{b_{1i}} = 0 \text{ для } k = 1, \dots, N \quad (89)$$

Приближение зависимой переменной τ будет в комплекте с решением уравнения (89).

3.2.5.1.1 ПРОЦЕДУРА РЕШЕНИЯ

Система, описанная уравнением (89), может быть решена путем любого метода, способного решить стандартное нелинейное уравнение. Мы прибегаем к методу исчисления, в частности, методу Пауэлла. Это было реализовано с помощью рутинных *DNEQNJ* из численных библиотек *IMSL* [42].

Алгоритм минимизирует скалярную функцию:

$$\epsilon(b_{1_1}, b_{1_2}, \dots, b_{1_N}) = \sum_{k=1}^N (\epsilon_k)^2,$$

где

$$\epsilon_k = \sum_{i=1}^N M1_{ki} b_{1i} - \sum_{i=1}^N M4_{ki} \frac{1}{b_{1i}}.$$

Как и в любом многомерном алгоритме поиска, метод Пауэлла подразумевает системы матриц Якоби. В этом примере проблемы якобиана **J** в алгоритм была поставлена аналитически для увеличения вычислительной эффективности. Якобиан рассчитывается следующим образом:

$$J = J_{kj} = \frac{d\epsilon_k}{db_{1j}} = M1_{kj} + M4_{kj} \left(\frac{1}{b_{1j}} \right)^2 \quad (90)$$

Обратите внимание, что, как матрицы **M1** и **M4**, так и якобиан также положительно определен и тридиагонален.

Процедура решения могут быть поставлена следующим образом:

1. Численные вычисления интегралов уравнения (81), чтобы сформировать матрицы **M1**, **M2** и **M3**.
2. Выбор значений для β и вычисления компонентов матрицы **M4**.
3. Выбор начальных значений для $b1_i$. В этом примере $b1_i = 1$ для $i = 1, \dots, N$.
4. Применим метод алгоритма Пауэлла, чтобы минимизировать оценки J_{kj} и ϵ_k в каждой итерации. Конвергенция достигается, когда изменение ϵ_k между последовательными итерациями меньше некоторого заданного значения.

3.2.5.1.2 ПОСЛЕДУЮЩАЯ ОБРАБОТКА РЕШЕНИЯ

С известными значениями $b1_i$ расширение для данного τ в уравнение (74) завершено. Построение разложения базиса для неизвестных f, η и $d^2/d\eta^2$, используя τ и u , является относительно устремленным вперед.

Предполагая, что приближение f может быть записано как

$$f_\alpha(\eta) = f_\alpha(u) = \sum_{k=1}^N \Phi_k(u) w1_k \quad (91)$$

где N и Φ такие же, как и те, которые используются для приближения τ , и используя преобразования координат уравнения (69), мы показываем:

$$u = \frac{df}{d\eta}; df = u d\eta \quad \text{и} \quad \tau = \frac{du}{d\eta}; d\eta = \frac{1}{\tau} du$$

Поэтому мы можем написать

$$df = \left(\frac{u}{\tau}\right) du \quad (92)$$

Используя узлы, можно интегрировать уравнение (92)

$$\int_{u_1}^{u_k} df = f(u_k) - f(u_1) = \int_{u_1}^{u_k} \left(\frac{u}{\tau}\right) du = f_\alpha(u_k) - f_\alpha(u_1) \quad (93)$$

Из граничных условий $f(u_1 = 0) = 0 = f_\alpha(0)$ и из свойств шляпной функции $f_\alpha(u_k) = w1_k$ уравнение (93) переходит в

$$w1_k = \int_0^{u_k} \left(\frac{u}{\tau}\right) du \quad (94)$$

После подстановки соответствующего расширения получается

$$w1_k = \int_0^{u_k} \frac{u}{(1-u)} \sum_{i=1}^N \Phi_i b2_i du = \sum_{i=1}^N \left(\int_0^{u_k} \frac{u}{(1-u)} \Phi_i du \right) b2_i = \sum_{i=1}^N G1_{ki} b2_i$$

$$\text{для } k = 1, 2, \dots, N-1 \quad (95)$$

где

$$G1_{ki} = \int_0^{u_k} \frac{u}{(1-u)} \Phi_i du \quad \text{для } k = 1, 2, \dots, N-1 \quad (96)$$

Если при $k = N$ мы используем узел $u_N = 1$, то значение $w1_N$ станет бесконечным.

Таким образом, в нашем примере мы устанавливаем

$$w1_N = \sum_{i=1}^N \left(\int_0^{0.99} \frac{u}{(1-u)} \Phi_i du \right) b2_i = \sum_{i=1}^N G1_{Ni} b2_i \quad (97)$$

С оценкой $w1_k$ расширение базиса для f закончено.

Для приближения оригинальной независимой переменной η в виде переменной u запишем

$$\eta_\alpha = \eta_\alpha(u) = \sum_{i=1}^N \Phi_k(u) w2_k \quad (98)$$

Вспомним

$$\tau = \frac{du}{d\eta} d\eta = \left(\frac{1}{\tau}\right) du$$

Интеграция $d\eta$ и признавая, что $\eta(0) = 0 = \eta_\alpha(0)$, имеем

$$\eta_\alpha(u_k) = \int_0^{u_k} \left(\frac{1}{\tau}\right) du = w2_k$$

Так что

$$w2_k = \sum_{i=1}^N G2_{ki} b2_i \quad \text{для } k = 1, 2, \dots, N \quad (99)$$

где

$$G2_{ki} = \int_0^{u_k} \frac{1}{(1-u)} \Phi_i du \text{ для } k = 1, 2, \dots, N-1 \text{ и } G2_{Ni} = \int_0^{0.99} \frac{1}{(1-u)} \Phi_i du \quad (100)$$

Остается приблизить вторую производную f с учетом η действующей независимой переменной u . Запишем

$$\frac{d^2 f_\alpha}{d\eta^2}(u) = \sum_{k=1}^N \Phi_k(u) w3_k \quad (101)$$

Чтобы определить $w3_k$, нужно приравнять приближение τ с приближением $d^2 f/d\eta^2$ на узлах с проблемной областью.

$$\tau_\alpha(u_k) = (1 - u_k) \sum_{i=1}^N \Phi_i(u_k) b1_i = \frac{d^2 f_\alpha}{d\eta^2}(u_k) = \sum_{k=1}^N \Phi_i(u_k) w3_k$$

Это приводит к

$$w3_k = (1 - u_k) b1_k \quad (102)$$

Поскольку дискретизация является похожей для всех базисных расширений, применение уравнения (59) к $w1, w2, w3$ разрешает использование одного входа, множественного выхода FFANN (рисунок 15). Для данного значения $df/d\eta$, сеть будет выводить значения для η, f, F и $d^2 f/d\eta^2$.

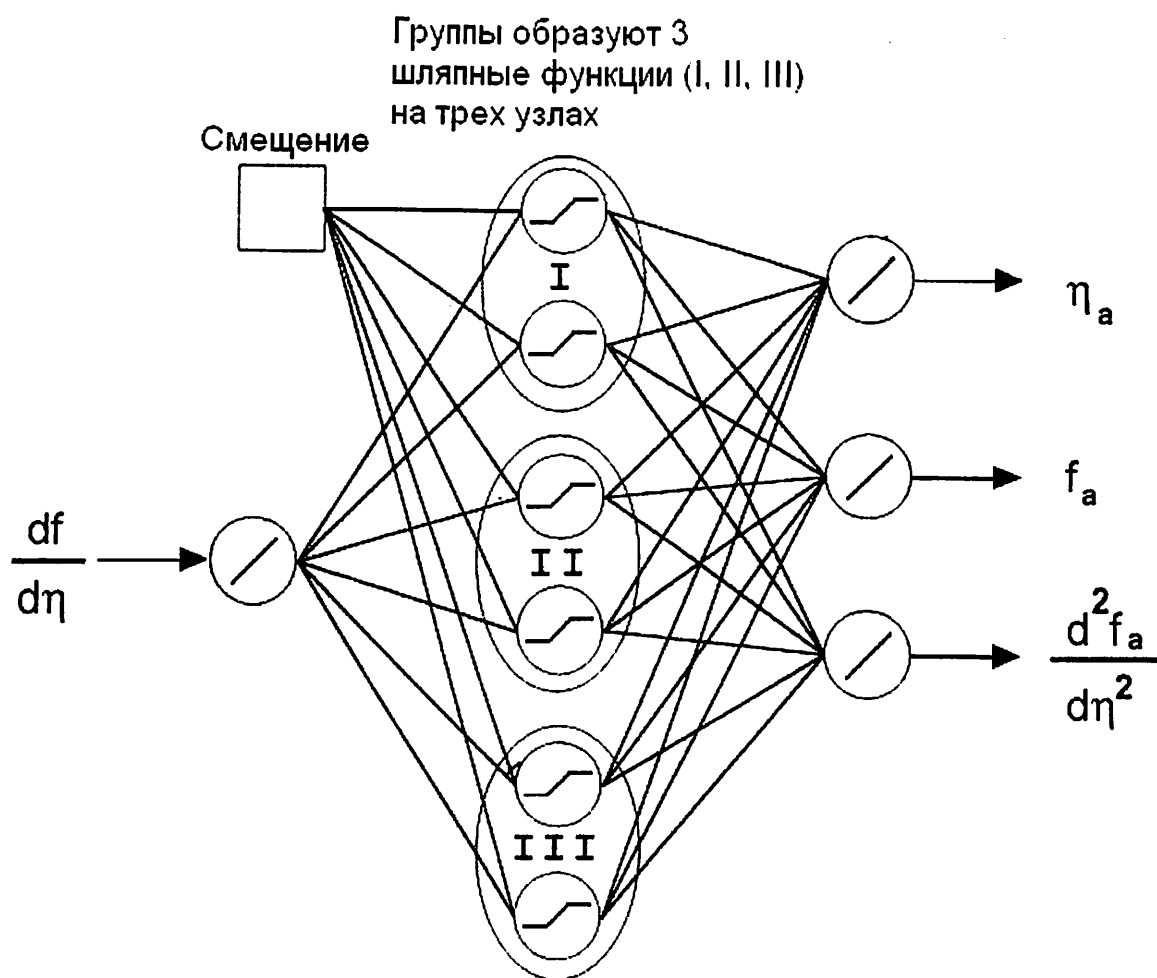


Рисунок 15 - Сеть с одним входом и множественным выходом для Примера 1 ($T = 6, N = 3$).

3.2.5.1.3 РЕЗУЛЬТАТЫ

Для примера уравнения Фолкнер-Скана входные данные, смещения и выходные веса для сети были рассчитаны с помощью обычного Fortran кода. Все расчеты были сделаны с двойной точностью. Как упоминалось ранее, для решения системы нелинейных алгебраических уравнений была использована подпрограмма *IMSL - DNEQNJ*. Также была использована подпрограмма *DQDAG* для численного интегрирования компонентов коэффициентов матрицы.

Хотя уравнение Фолкнер - Скана относительно простое по внешнему виду, но оно не имеет аналитического решения в настоящее время. Однако уравнение (65) решается численными методами. Для того, чтобы дать некоторое представление о точности, построенной FFANN, решение для нелинейного

дифференциального уравнения было рассмотрено через метод Рунге-Кутта с переменным шагом четвертого-пятого порядка и итерационный метод подбора.

FFANN была построена с использованием одного скрытого слоя из 42 процессорных элементов для случая $\beta = -0,15$ и 22 скрытых элементов обработки для случая $\beta = 0$ и $\beta = 0,5$. Это соответствует одинаковым расстояниям в пределах проблемной области, 21 и 11 узлов, соответственно. Размер переменного шага Рунге-Кутта проходит от $\eta = 0$ до $\eta = 5$, в 89 шагов при $\beta = -0,15$, 103 шага при $\beta = 0$ и 128 шагов $\beta = 0,5$. Решение по Рунге-Кутта удовлетворяет граничным условиям с погрешностью не более $1,0 \times 10^{-6}$ для всех случаев.

В рисунках с 16 по 18 сравниваются выходные данные FFANN с решением по Рунге-Кутта (сплошная линия) при 10 равномерно разнесенных точках образца ($S = 10$) для $\beta = 0,15$, $\beta = 0$, $\beta = 0,5$. Штрихи означают дифференцирование зависимых переменных по η . Число итераций (I), необходимое для решения базисных коэффициентов были: $I = 9$ при $\beta = -0,15$ с $\epsilon = 2,336 \times 10^{-13}$, $I = 7$ для $\beta = 0$ с $\epsilon = 4,611 \times 10^{-18}$ и $I = 8$ при $\beta = 0,5$ с $\epsilon = 8,596 \times 10^{-16}$. Значения среднеквадратичной разности (RMS) между выходными данными решения Рунге-Кутта и FFANN приведены в виде отдельной фигуры.

Программа запускалась на стандартной *SUN Microsystems Sparcstation 2* с выполнением порядка двух секунд для испытания (рисунки с 16 по 18).

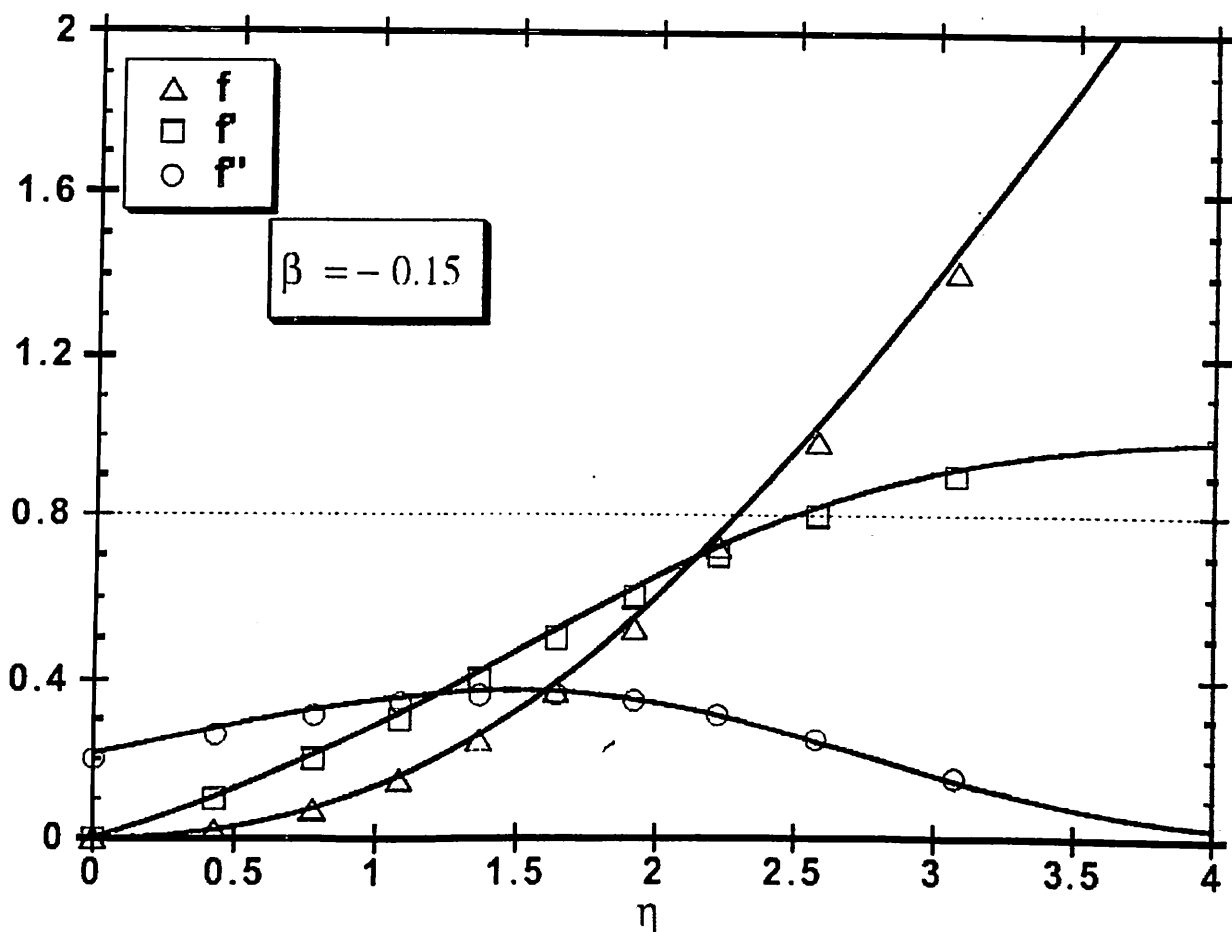


Рисунок 16 - $\beta = -0.15$: Сравнение решений Рунге-Кутты и выходных данных сети при количестве обрабатывающих элементов $T = 42$. $f: RMS = 3.033 \times 10^{-2}$; $\frac{df}{d\eta}: RMS = 7.200 \times 10^{-2}$; $\frac{d^2f}{d\eta^2}: RMS = 1.067 \times 10^{-2}$

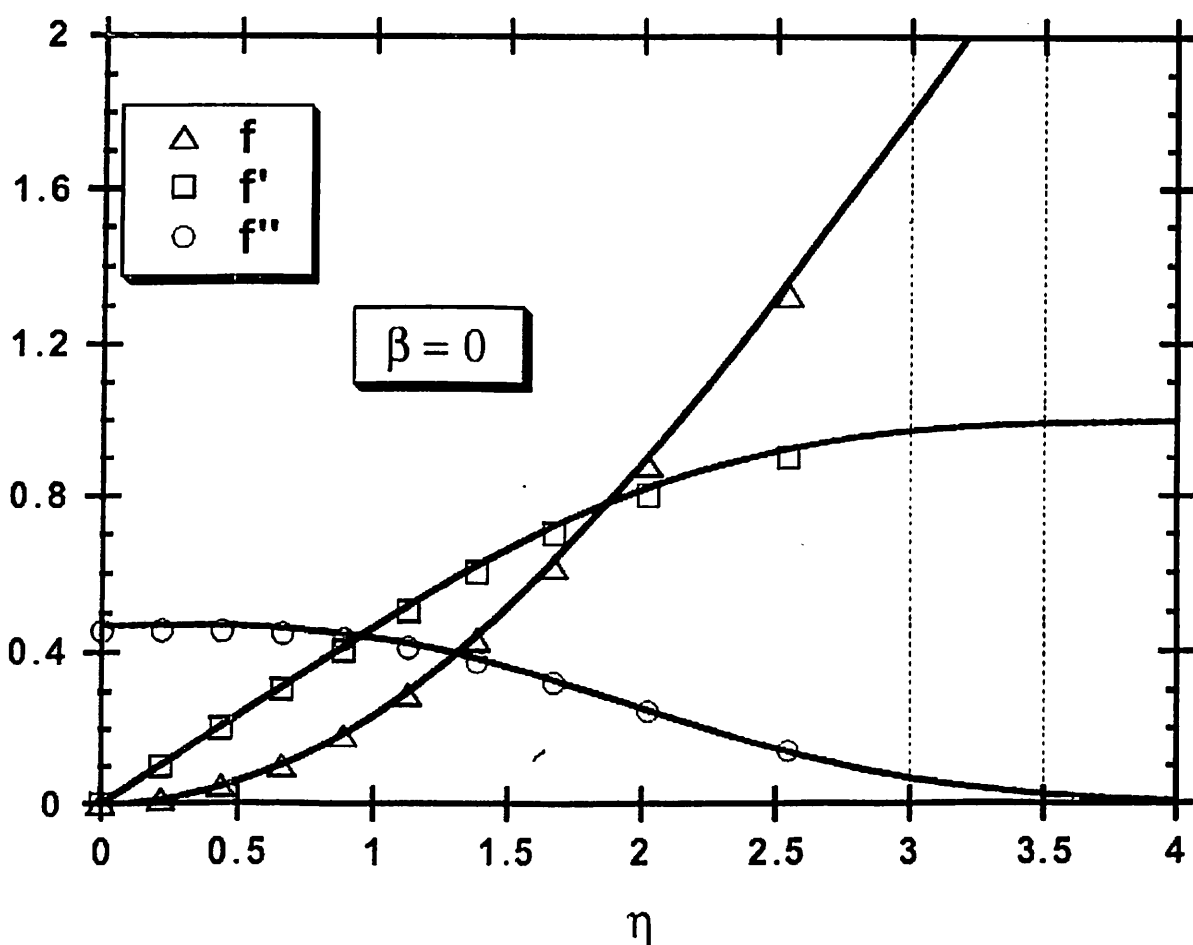


Рисунок 17 - $\beta = -0$: Сравнение решений Рунге-Кутта и выходных данных сети при количестве обрабатывающих элементов $T = 22$. $f: RMS = 3.563 \times 10^{-2}$; $\frac{df}{d\eta}: RMS = 6.215 \times 10^{-2}$; $\frac{d^2f}{d\eta^2}: RMS = 1.788 \times 10^{-2}$

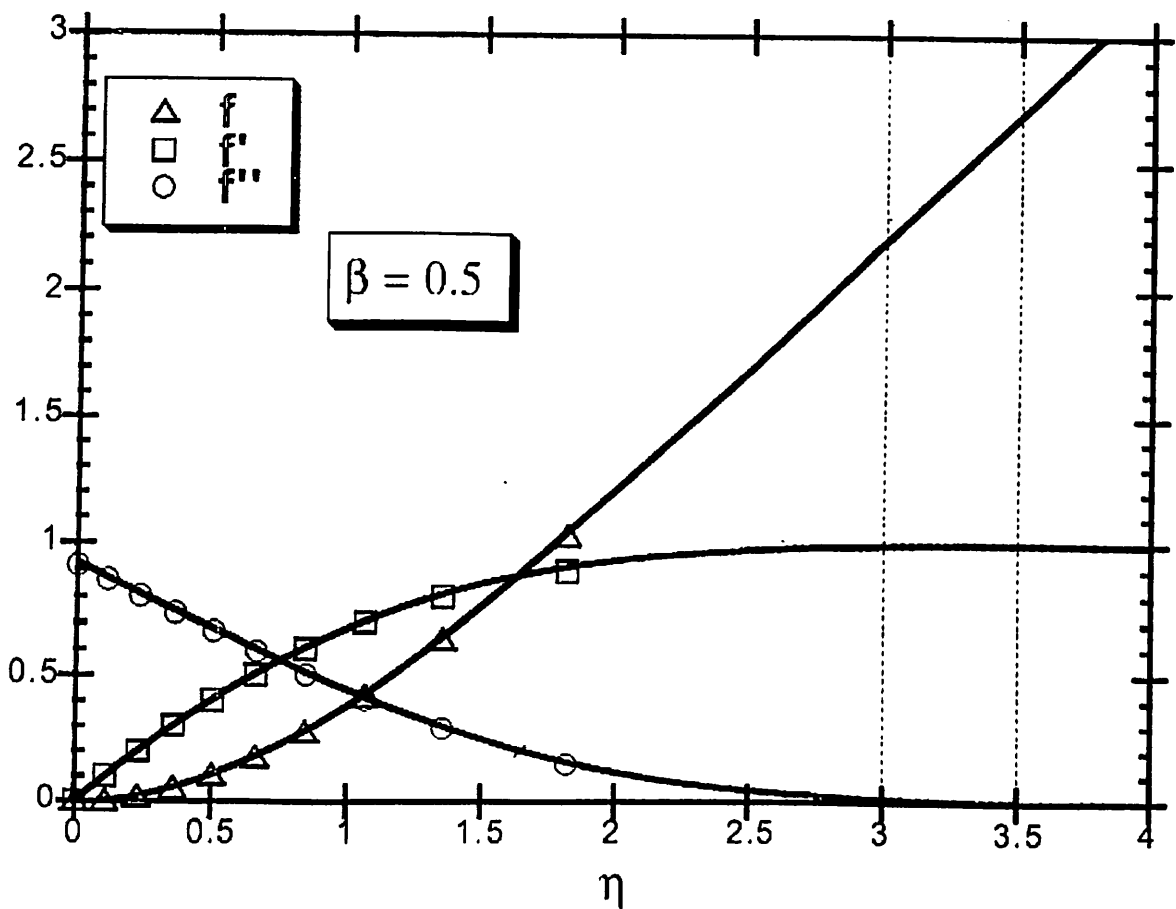


Рисунок 18 - $\beta = 0.5$: Сравнение решений Рунге-Кутты и выходных данных сети при количестве обрабатывающих элементов $T = 22$. f : $RMS = 2.554 \times 10^{-2}$; $\frac{df}{d\eta}$: $RMS = 3.534 \times 10^{-2}$; $\frac{d^2f}{d\eta^2}$: $RMS = 1.741 \times 10^{-2}$

3.2.5.1.4 СХОДИМОСТЬ

Определение полученных погрешностей для метода Петрова-Галеркина позволяют ожидать, что норма L_2 об ошибке будет ограничиваться квадратичным членом в расстояние между узлами (сетки) при приближении линейных обыкновенных дифференциальных уравнений [43]. Этот стандартный результат численного анализа был подтвержден в работе [35] для сетевого решения линейных обыкновенных дифференциальных уравнений. При приближении нелинейных дифференциальных уравнений квадратная скорость сходимости деградирует и не может быть легко предсказано.

Для равномерного шага сетки

$$u_{i+1} - u_i = u_i - u_{i-1} = h = \frac{1}{(N-1)}$$

Таким образом, для ошибки E в переменной $d^2f/d\eta^2$ норма L_2 от E определяется так

$$\|E\| = \sqrt{\int (E)^2 du} = \sqrt{\int \left(\frac{d^2f}{d\eta^2} - \frac{d^2f_\alpha}{d\eta^2} \right)^2 du} \quad (103)$$

На практике дискретная норма L_2 погрешности используется таким образом, чтобы избежать выполнения фактической интеграции в уравнении (53):

$$\|E\| = \sqrt{h \sum_{i=1}^N \left(\frac{d^2f}{d\eta^2}(u_i) - \frac{d^2f_\alpha}{d\eta^2}(u_i) \right)^2},$$

где интеграл была приближен масштабированным суммированием. Будем считать, что норма L_2 погрешности прямо пропорциональна до некоторого порядка A из сетки h

$$\|E\| \approx h^A \text{ или } \|E\| \leq Ch^A, \quad (104)$$

где C - постоянная и где A некоторое действительное число. Показатель A должен быть ≤ 2 для нелинейных дифференциальных уравнений. Переменная $d^2f/d\eta^2$ выбрана для исследования, поскольку сходимость численных экспериментов показали, что это функция наиболее подходящая и ее скорость сходимости должна обеспечивать верхний предел сходимости других переменных.

Поскольку нет точного решения уравнения Фолкнер - Скана, скорость сходимости была протестирована через последовательное уточнение сетки через следующие шаги:

1. Выбрать начальный интервал h_0 и множество $k = 0$.
2. Решить уравнение на сетке $h_k = h_0/2^k$ и $h_{k+1} = h_0/2^{(k+1)}$.
3. Найти дискретную норму L_2 разности между решениями на основных узлах.
4. Увеличиваем k и возвращаемся к шагу 2.

Целью этой процедуры заключается в измерении скорости, с которой приближается некоторые точные решения в дискретных точках. Если взять логарифм от $\|E\|$ в уравнении (104), тогда получится:

$$\log(\|E\|) \leq \log(C) + A \log(h) \quad (105)$$

Судя по форме уравнения (105), было бы разумно ожидать, что логарифмический участок нормы L_2 погрешности в зависимости от логарифма сетки даст прямую линию наклонной A .

Подходящий участок сходимости этого типа показан на рисунке 19 для трех значений β . Пунктирная линия представляет чисто линейную скорость сходимости. Масштабы проблемы сетки варьировались от $h = 0,1$ до $h = 0,0125$. Значение показателя A было определено для $\beta = 0,15, 0$ и $0,5$ и равно $0,892, 0,679$ и $0,598$ соответственно. Обратите внимание на высокую скорость сходимости, соответствующую отрицательному значению β .

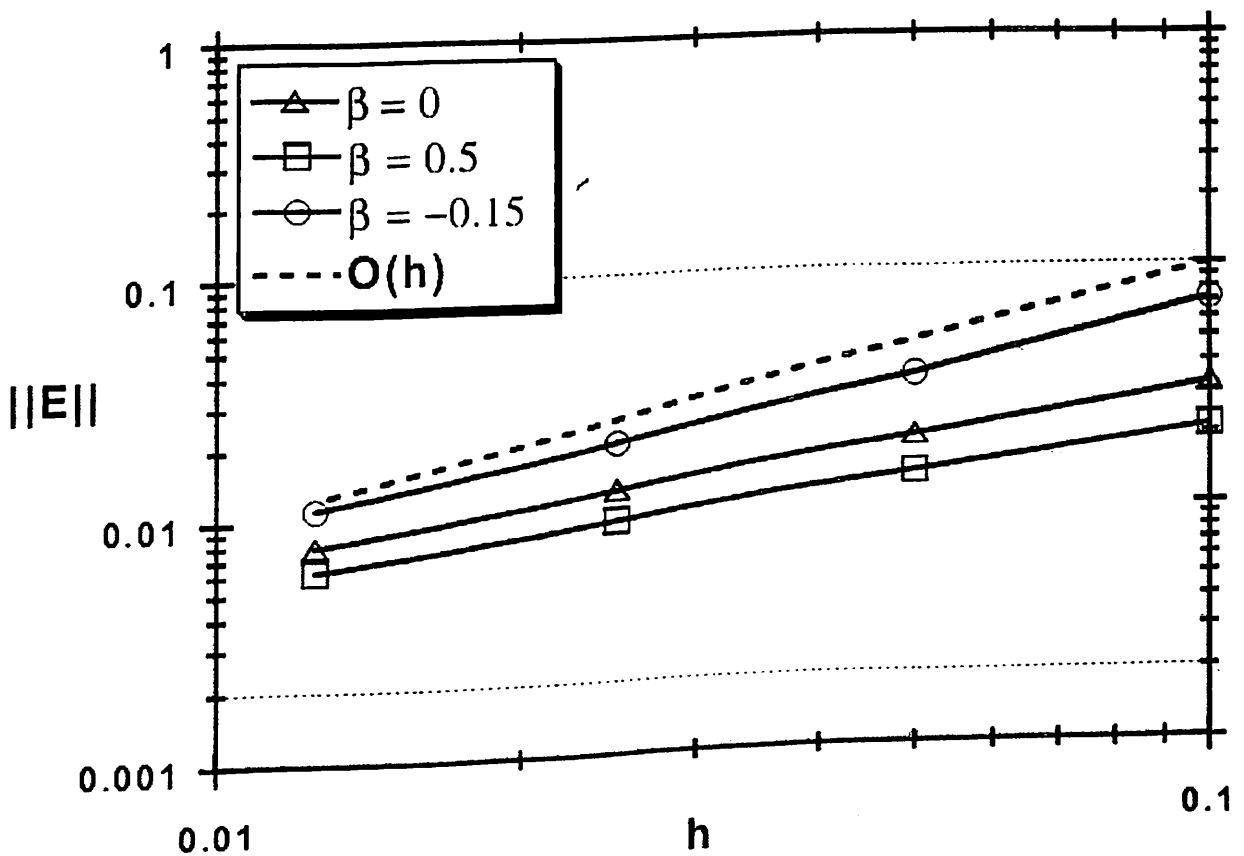


Рисунок 19 - Участок логарифмической сходимости для $d^2 f / d\eta^2$.

$\beta = -0.15$: slope = 0.892; $\beta = 0$: slope = 0.679; $\beta = +0.5$: slope = 0.598.

Величина ошибки не может быть оценена, пока не получено решение уравнения (89). Кроме того рисунок 19 показывает, что ошибка может

управляться путем изменения количества обрабатывающих элементов в скрытом слое сети.

3.2.5.2 СВЯЗАННЫЕ НЕЛИНЕЙНЫЕ ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

Для второго числового примера мы построили один вход, множественный выход FFANN для приближения решения и производных из связанной системы нелинейных обыкновенных дифференциальных уравнений второго и третьего порядка:

$$\frac{d^3 f}{d\eta^3} - (A_1 f + A_2 g) \frac{d^2 f}{d\eta^2} = 0 \quad (106)$$

$$\frac{d^3 g}{d\eta^3} - (A_1 f + A_2 g) \frac{d^2 g}{d\eta^2} - A_3 \frac{df}{d\eta} \frac{dg}{d\eta} - A_4 \left(\frac{dg}{d\eta}\right)^2 - A_5 \left(\frac{df}{d\eta}\right)^2 - A_6 \lambda - A_7 = 0 \quad (107)$$

$$\frac{d^2 g}{d\eta^2} - Pr(A_1 f + A_2 g) \frac{d\lambda}{d\eta} - A_8 \left(\frac{d^2 g}{d\eta^2}\right)^2 - A_8(A_1 f + A_2 g) \frac{df}{d\eta} \frac{d^2 f}{d\eta^2} = 0, \quad (108)$$

где $\eta \geq 0$. Коэффициенты A_1, \dots, A_7 константы и представляют собой комбинации различных физических параметров и коэффициент A_8 представляет собой функцию параметра Pr . Связанные граничные условия для уравнений (106) - (108)

$$f = 0, \frac{df}{d\eta} = 0, g = 0, \frac{dg}{d\eta} = 0, \lambda = 0 \text{ при } \eta = 0 \text{ и } \frac{df}{d\eta} \rightarrow 1 \text{ при } \eta \rightarrow \infty \quad (109)$$

Кроме того из граничных условий, так как

$$\frac{df}{d\eta} \rightarrow 1 \text{ при } \eta \rightarrow \infty \text{ тогда } \frac{d^2 f}{d\eta^2} \rightarrow 0 \text{ при } \eta \rightarrow \infty \quad (110)$$

Как и в предыдущем примере, уравнения (106) - (108) представляют практический интерес в области механики жидкости. Обыкновенные дифференциальные уравнения получаются от преобразования системы уравнений в частных производных. Эта модель трехмерного импульса и теплового пограничного слоя относительно наклонной прямого кругового конуса в сверхзвуковом потоке [44] - [46]. Если уравнения (106) - (108) используются для моделирования физически значимых ситуаций, $df/d\eta$ должно монотонно возрастать по η . Это может быть достигнуто путем ограничения значений физических параметров, что эквивалентно ограничению $A_1 - A_8$. Для конуса с половинным углом $^\circ$, в воздухе со свободным потоком число Маха равное 7 и в

свободном потоке число Рейнольдса равно 5×10^5 , мы имеем следующие численные значения для A_1 до A_7 для угла атаки 10° :

$$\begin{aligned} A_1 &= -0,960428; A_2 = -0,395063; A_3 = 0,640285; A_4 = 0,395063; \\ A_5 &= 0,946661; A_6 = -0,665229; A_7 = -0,665229 \end{aligned} \quad (111)$$

Мы должны также использовать ограничение $0,7 \leq Pr \leq 1,0$, что соответствует $0 \leq A_8 \leq 0,86$.

Подробное обсуждение вывода уравнений (106) - (108), необходимость в монотонно возрастающей $df/d\eta$ и оценки коэффициентов $A_1 - A_8$, можно найти в ссылке [47].

Для построения FFANN, которая моделирует связанные системы, мы должны трансформировать основные уравнения в более управляемые формы. Наш подход к решению уравнений (106) - (108) будет очень похож на подход, используемый для решения уравнения Фолкнера - Скана. Поскольку $df/d\eta$ монотонно изменяется по η для вышеуказанных ограничений $A_1 - A_8$, производная может быть использована в качестве действующей независимой переменной. Введем следующую замену переменных:

$$u = \frac{df}{d\eta} \quad 0 \leq u \leq 1, \tau = \frac{d^2f}{d\eta^2}, r = \frac{dg}{d\eta}, s = \frac{d\lambda}{d\eta}, p = \frac{d\tau}{d\eta}, q = \frac{dr}{d\eta} \quad (112)$$

В нашем преобразовании связанных уравнений мы должны устранить явную зависимость f от g . Перегруппировка уравнения (106) дает

$$\frac{d^4f}{d\eta^4} = \left(A_1 \frac{df}{d\eta} + A_2 \frac{dg}{d\eta} \right) \frac{d^2f}{d\eta^2} + \frac{d^3f}{d\eta^3} \left(\frac{d^2f}{d\eta^2} \right)^{-1} \frac{d^3f}{d\eta^3} \quad (114)$$

$$\frac{d^3g}{d\eta^3} = \frac{d^3f}{d\eta^3} \left(\frac{d^2f}{d\eta^2} \right)^{-1} \frac{d^2g}{d\eta^2} + A_3 \frac{df}{d\eta} \frac{dg}{d\eta} + A_4 \left(\frac{dg}{d\eta} \right)^2 + A_5 \left(\frac{df}{d\eta} \right)^2 + A_6 \lambda + A_7 \quad (115)$$

$$\frac{d^2\lambda}{d\eta^2} = Pr \frac{d^3f}{d\eta^3} \left(\frac{d^2f}{d\eta^2} \right)^{-1} \frac{d\lambda}{d\eta} + A_8 \left(\frac{d^2f}{d\eta^2} \right)^2 + A_8 \frac{d^3f}{d\eta^3} \frac{df}{d\eta} \quad (116)$$

При подстановке преобразованных зависимых переменных и использование цепного правила, мы имеем для уравнений (114) - (116)

$$\frac{dp}{du} = A_1 u + A_2 r + \left(\frac{p}{\tau} \right)^2 \quad (117)$$

$$\frac{dp}{du} = \left(\frac{p}{\tau}\right) \left(\frac{q}{\tau}\right) + A_3 u \left(\frac{r}{\tau}\right) + A_4 \left(\frac{r^2}{\tau}\right) + A_5 u^2 \left(\frac{1}{\tau}\right) + A_6 \left(\frac{\lambda}{\tau}\right) + A_7 \left(\frac{1}{\tau}\right) \quad (118)$$

$$\frac{ds}{du} = \text{Pr} \left(\frac{p}{\tau}\right) \left(\frac{q}{\tau}\right) + A_8 \tau + A_8 u \left(\frac{p}{\tau}\right) \quad (119)$$

Использование цепного правила на переменные уравнения (112) завершает преобразование путем уменьшения первоначальной системы трех связанных нелинейных смешанных дифференциальных уравнений второго и третьего порядка к системе шести уравнений первого порядка:

$$\frac{dp}{du} - A_1 u - A_2 r - \left(\frac{p}{\tau}\right)^2 = 0$$

$$\frac{dp}{du} - \left(\frac{p}{\tau}\right) \left(\frac{q}{\tau}\right) - A_3 u \left(\frac{r}{\tau}\right) - A_4 \left(\frac{r^2}{\tau}\right) - A_5 u^2 \left(\frac{1}{\tau}\right) - A_6 \left(\frac{\lambda}{\tau}\right) - A_7 \left(\frac{1}{\tau}\right) = 0$$

$$\frac{ds}{du} - \text{Pr} \left(\frac{p}{\tau}\right) \left(\frac{q}{\tau}\right) - A_8 \tau - A_8 u \left(\frac{p}{\tau}\right) = 0$$

$$\frac{d\tau}{du} - \left(\frac{p}{\tau}\right) = 0, \frac{dr}{du} - \left(\frac{q}{\tau}\right) = 0, \frac{d\lambda}{du} - \left(\frac{s}{\tau}\right) = 0 \quad (120)$$

Теперь осталось найти новые граничные условия для преобразованной системы. На основе определения уравнений (106) - (108) при $\eta = 0$ и $\eta \rightarrow \infty$ в сочетании с уравнением (110) и уравнением (112) становится очевидно, что

$$p = r = \lambda = 0 \text{ при } u = 0; \tau = q = s = 0 \text{ при } u = 1 \quad (121)$$

Теперь мы можем использовать базисные разложения для приближения преобразованных зависимых переменных. Используя метод приближения произведения, показанного в уравнения Фолкнер - Скана, получают следующие приближения, которые можно записать для преобразования зависимых переменных

$$\tau_\alpha = (1 - u) \sum_{i=1}^N \Phi_i b_{1i}, r_\alpha = \sum_{i=1}^N \Phi_i b_{2i}, \lambda_\alpha = \sum_{i=1}^N \Phi_i b_{3i}, p_\alpha = \sum_{i=1}^N \Phi_i b_{4i},$$

$$q_\alpha = (1 - u) \sum_{i=1}^N \Phi_i b_{5i}, s_\alpha = (1 - u) \sum_{i=1}^N \Phi_i b_{6i},$$

и для мультипликативных групп

$$\begin{aligned}
\frac{p_\alpha}{\tau_\alpha} &= \frac{1}{(1-u)} \Phi_i b7_i, \quad \frac{q_\alpha}{\tau_\alpha} = \frac{1}{(1-u)} \Phi_i b8_i, \quad \frac{s_\alpha}{\tau_\alpha} = \frac{1}{(1-u)} \Phi_i b9_i, \\
\left(\frac{p_\alpha}{\tau_\alpha}\right)^2 &= \frac{1}{(1-u)^2} \Phi_i b10_i, \quad \left(\frac{p_\alpha}{\tau_\alpha}\right) \left(\frac{q_\alpha}{\tau_\alpha}\right) = \frac{1}{(1-u)^2} \Phi_i b11_i, \quad \frac{r^2_\alpha}{\tau_\alpha} = \frac{1}{(1-u)} \Phi_i b12_i, \\
\frac{1}{\tau_\alpha} &= \frac{1}{(1-u)} \Phi_i b13_i, \quad \frac{\lambda_\alpha}{\tau_\alpha} = \frac{1}{(1-u)} \Phi_i b14_i, \quad \left(\frac{p_\alpha}{\tau_\alpha}\right) \left(\frac{s_\alpha}{\tau_\alpha}\right) = \frac{1}{(1-u)^2} \Phi_i b15_i, \\
\frac{r_\alpha}{\tau_\alpha} &= \frac{1}{(1-u)} \Phi_i b16_i,
\end{aligned} \tag{122}$$

где N число узлов в проблемной области. Обратите внимание, что умножение и деление базисных разложений на $(1-u)$ помогает в обеспечении граничных условий.

Из базисных коэффициентов только от $b1_i$ до $b6_i$ линейно независимые. Соотношение между множеством остальных базисных коэффициентов может быть определено путем использования одного и того же аргумента эквивалентности, сделанного в примере 1 (уравнения (76) - (78)):

$$\begin{aligned}
b7_i &= b4_i b13_i, \quad b8_i = b5_i b13_i, \quad b9_i = b6_i b13_i, \\
b10_i &= (b4_i b13_i)^2, \quad b11_i = b4_i b5_i (b13_i)^2, \quad b12_i = (b2_i)^2 b13_i \\
b13_i &= \frac{1}{b1_i}, \quad b14_i = b3_i b13_i, \quad b15_i = b4_i b6_i (b13_i)^2, \quad b16_i = b2_i b13_i
\end{aligned} \tag{123}$$

для $i = 1, \dots, N$.

Чтобы сформулировать алгебраические системы, решение которых даст коэффициенты разложения преобразованного уравнения (120), мы применяем метод Петрова-Галеркина с $F_k(u) = (1-u)\Phi_k(u)$:

$$\sum_{i=1}^N M1_{ki} b4_i - A_1 v_k - A_2 \sum_{i=1}^N M3_{ki} b2_i - \sum_{i=1}^N M4_{ki} b10_i = 0$$

$$\begin{aligned}
& \sum_{i=1}^N M1_{ki} b4_i - \sum_{i=1}^N M2_{ki} (A_4 b12_i + A_7 b13_i + A_6 b14_i) - \sum_{i=1}^N M4_{ki} b11_i \\
& \quad - A_3 \sum_{i=1}^N M5_{ki} b16_i - A_5 \sum_{i=1}^N M6_{ki} b13_i = 0 \\
& \sum_{i=1}^N M7_{ki} b6_i - Pr \sum_{i=1}^N M4_{ki} b15_i - A_8 \left(\sum_{i=1}^N M8_{ki} b1_i - \sum_{i=1}^N M5_{ki} b7_i \right) = 0 \\
& \sum_{i=1}^N M7_{ki} b1_i - M2_{ki} b7_i = 0 \\
& \sum_{i=1}^N M1_{ki} b2_i - M2_{ki} b8_i = 0; \sum_{i=1}^N M1_{ki} b3_i - M2_{ki} b9_i = 0, \tag{124}
\end{aligned}$$

где матрицы $M1, \dots, M8$ и вектор v определяются как

$$\begin{aligned}
M1_{ki} &= \left((1-u)^2 \Phi_k, \frac{d\Phi_i}{du} \right), & M2_{ki} &= ((1-u)\Phi_k, \Phi_i), \\
M3_{ki} &= ((1-u)^2 \Phi_k, \Phi_i) \\
M4_{ki} &= (\Phi_k, \Phi_i), & M5_{ki} &= (u(1-u)\Phi_k, \Phi_i), & M6_{ki} &= (u^2(1-u)\Phi_k, \Phi_i) \\
M7_{ki} &= \left((1-u)^2 \Phi_k, (1-u) \frac{d\Phi_i}{du} - \Phi_i \right), & M8_{ki} &= ((1-u)^3 \Phi_k, \Phi_i) \\
v_k &= ((1-u)^2 \Phi_k, u) \tag{125}
\end{aligned}$$

Граничные условия уравнения (121) дают

$$b2_1 = b3_1 = b4_1 = 0 (u = 0); b1_N = b5_N = b6_N = 1 (u = 1) \tag{126}$$

Как и в примере уравнения Фолкнера - Скана, эти ограничения включены в систему алгебраических уравнений путем замены соответствующих строк матриц.

3.2.5.2.1 ПРОЦЕДУРА РЕШЕНИЯ

Шесть систем уравнений в уравнении (124) должны быть решены для коэффициентов $b1_i, b2_i, \dots, b6_i$. Как и в уравнения Фолкнера - Скана, снова прибегаем к методу Пауэлла для решения нелинейных уравнений. В этом случае алгоритм минимизирует следующую скалярную функцию:

$$\epsilon(b_{1_1}, \dots, b_{1_N}, b_{2_1}, \dots, b_{6_1}, \dots, b_{6_N}) = \sum_{k=1}^N \left((\epsilon_k^{(1)})^2 + (\epsilon_k^{(2)})^2 + \dots + (\epsilon_k^{(6)})^2 \right),$$

где

$$\epsilon_k^{(1)} = \sum_{i=1}^N M1_{ki} b4_i - A_1 v_k - A_2 \sum_{i=1}^N M3_{ki} b2_i - \sum_{i=1}^N M4_{ki} b10_i$$

$$\begin{aligned} \epsilon_k^{(2)} = & \sum_{i=1}^N M7_{ki} b5_i - \sum_{i=1}^N M2_{ki} (A_4 b12_i + A_7 b13_i + A_6 b14_i) - \sum_{i=1}^N M4_{ki} b11_i \\ & - A_3 \sum_{i=1}^N M5_{ki} b16_i - A_5 \sum_{i=1}^N M6_{ki} b13_i \end{aligned}$$

$$\epsilon_k^{(4)} = \sum_{i=1}^N M7_{ki} b1_i - M2_{ki} b7_i$$

$$\epsilon_k^{(5)} = \sum_{i=1}^N M1_{ki} b2_i - M2_{ki} b8_i$$

$$\epsilon_k^{(6)} = \sum_{i=1}^N M1_{ki} b3_i - M2_{ki} b9_i \tag{127}$$

Как и в предыдущем примере, якобиан (J) вычисляется аналитически для повышения эффективности вычислений. В этом случае Якобиан имеет размер $(6N) \times (6N)$. Он состоит из 36 разделов, каждый из которых может рассматриваться, как суб-якобиана размерностью $N \times N$. Например, если раздел определяется как

$$J^{(AB)} = \frac{d\epsilon_k^{(A)}}{dbB_j}, \text{ где } A, B = 1, \dots, 6$$

Тогда

$$J = \begin{bmatrix} J^{(11)} & \dots & J^{(16)} \\ \vdots & \ddots & \vdots \\ J^{(61)} & \dots & J^{(66)} \end{bmatrix}$$

Каждый раздел якобиана матрицы тридиагонален, вследствие свойств шляпной функции. В результате полной Якобиан матрицы является редким.

3.2.5.2.2 ПОСЛЕДУЮЩАЯ ОБРАБОТКА РЕШЕНИЯ

С помощью известных коэффициентов $b_{1_i}, b_{2_i}, \dots, b_{6_i}$ можно вычислить $b_{7_i}, b_{8_i}, \dots, b_{16_i}$. Приближение зависимых переменных преобразования являются полными. После использование $u = df/d\eta$ в качестве входных переменных к сети, остается определить выход весов для исходных переменных $\eta, f, \frac{d^2f}{d\eta^2}, g, \frac{dg}{d\eta}, \frac{d^2g}{d\eta^2}, \lambda, \frac{d\lambda}{d\eta}$.

Запишем базисное разложение для приближения следующим образом:

$$\begin{aligned} \eta_\alpha &= \sum_{i=1}^N \Phi_i(u) w_{1_i}, & f_\alpha &= \sum_{i=1}^N \Phi_i(u) w_{2_i}, & \frac{d^2 f_\alpha}{d\eta^2} &= \sum_{i=1}^N \Phi_i(u) w_{3_i} \\ g_\alpha &= \sum_{i=1}^N \Phi_i(u) w_{4_i}, & \frac{dg}{d\eta} &= \sum_{i=1}^N \Phi_i(u) w_{5_i}, & \frac{d^2 g_\alpha}{d\eta^2} &= \sum_{i=1}^N \Phi_i(u) w_{6_i} \\ \lambda_\alpha &= \sum_{i=1}^N \Phi_i(u) w_{7_i}, & \frac{d\lambda_\alpha}{d\eta} &= \sum_{i=1}^N \Phi_i(u) w_{8_i} \end{aligned} \quad (128)$$

Наш подход к определению базисных коэффициентов w_{1_i}, \dots, w_{8_i} будет идентичным тому подходу, который используется в примере уравнения Фолкнер-Скана. Поэтому из уравнений (92) - (102) мы можем сразу написать:

$$\begin{aligned} w_{1_k} &= \sum_{i=1}^N G_{2_{ki}} b_{13_i}, & w_{2_k} &= \sum_{i=1}^N G_{1_{ki}} b_{13_i}, & w_{2_k} &= (1 - u_k) b_{1_k} \\ w_{4_k} &= \sum_{i=1}^N G_{2_{ki}} b_{16_i}, & w_{5_k} &= b_{2_k}, & w_{6_k} &= (1 - u_k) b_{5_k} \\ w_{7_k} &= \sum_{i=1}^N G_{2_{ki}} b_{9_i}, & w_{8_k} &= (1 - u_k) b_{6_k} \end{aligned} \quad (129)$$

где $k = 1, \dots, N$. При определении выходного веса от w_{1_k} до w_{8_k} , а также применение уравнения (9), модель FFANN уравнений (106) - (108) будет завершена.

3.2.5.2.3 РЕЗУЛЬТАТЫ

Как и в примере уравнения Фолкнер - Скана, эта система была решена с помощью итерационной техники подбора Рунге-Кутта с переменным шагом в течение двух шагов, соответствующих $Pr = 1.0$ ($A_8 = 0.0$) и $Pr = 0.7$ ($A_8 = 0.853837$), а также результаты, сравненные с выходными данными из FFANN, построенной с помощью техники Петрова-Галеркина (рисунки с 20 по 23). Построенная сеть использует 162 процессорных элемента в скрытом слое, соответствующие равномерному расположению 81 узлов и два вспомогательных узлов. Результаты сравнивались в 10 пробных точках ($S = 10$), и только пять точек выборки показаны на чертежах для ясности. Размер переменного шага Рунге-Кутта был от $\eta = 0$ до $\eta = 10$ в 40 шагов для $Pr = 1.0$ и 46 шагов для $Pr = 0.7$, погрешность менее чем 1.0×10^{-6}

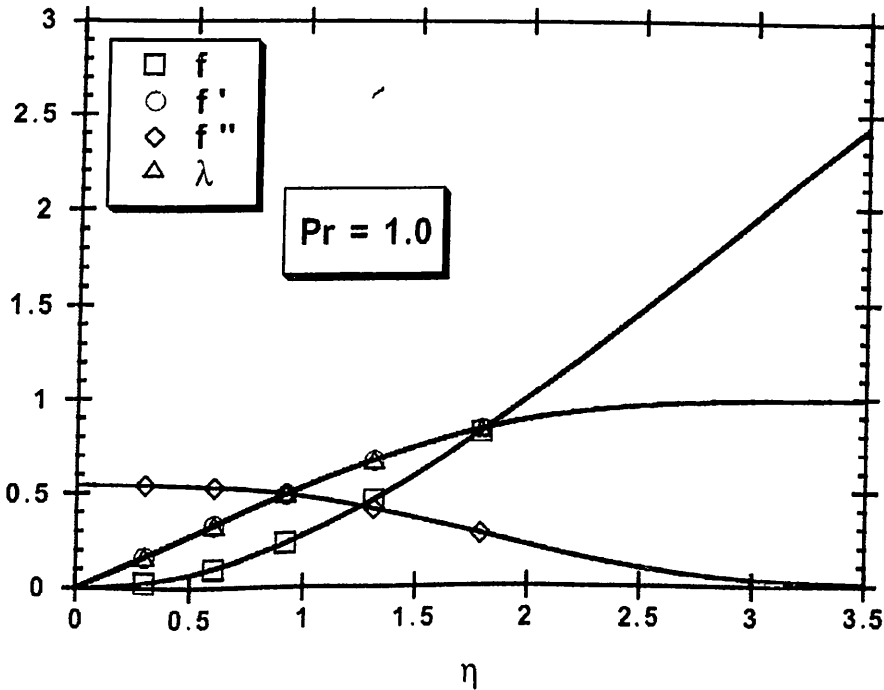


Рисунок 20 - $Pr = 1.0$: Сравнение решений Рунге-Кутта и выходных данных сети при количестве обрабатывающих элементов $T = 162$ (показаны 5 точек).

$$f: RMS = 2.740 \times 10^{-2}; \frac{df}{d\eta}: RMS = 1.907 \times 10^{-2}; \frac{d^2f}{d\eta^2}: RMS = 2.740 \times 10^{-3}; \lambda: RMS = 1.223 \times 10^{-6}$$

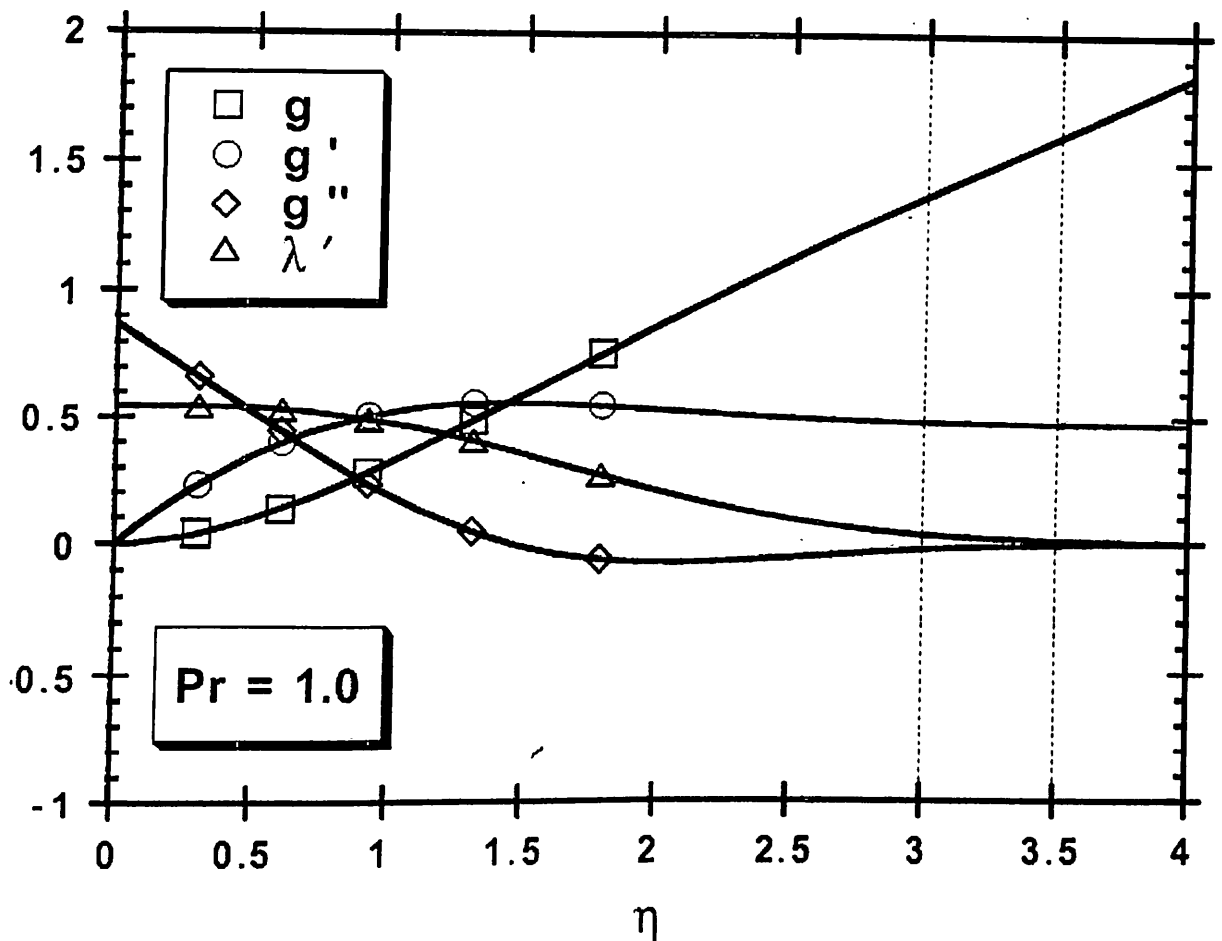


Рисунок 21 - $Pr = 1.0$: Сравнение решений Рунге-Кутты и выходных данных сети при количестве обрабатывающих элементов $T = 162$ (показаны 5 точек).

$$g: RMS = 1.178 \times 10^{-2}; \frac{dg}{d\eta}: RMS = 4.153 \times 10^{-3}; \frac{d^2g}{d\eta^2}: RMS = 1.524 \times 10^{-3}; \frac{d\lambda}{d\eta}: RMS = 2.739 \times 10^{-3}$$

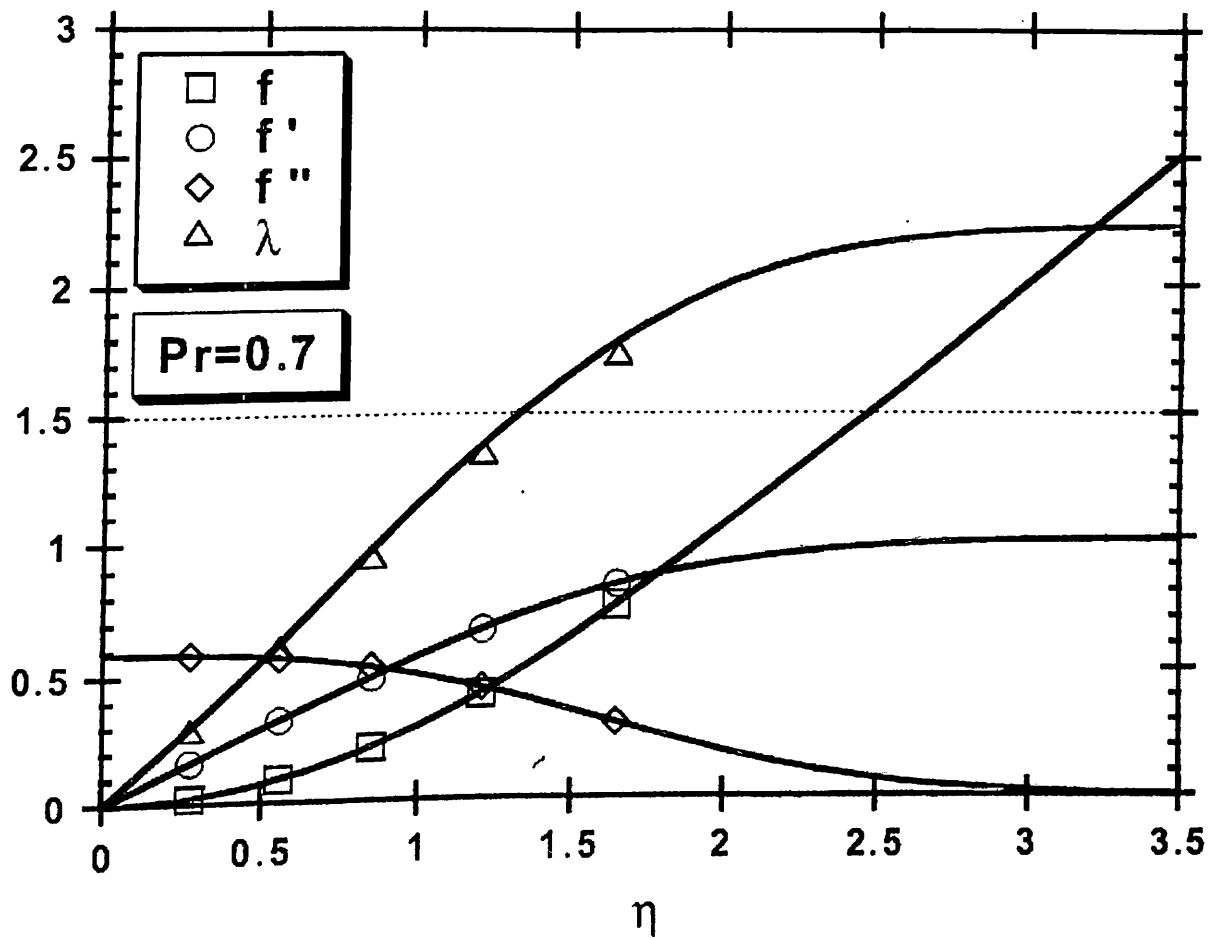


Рисунок 22 - $Pr = 0.7$: Сравнение решений Рунге-Кутты и выходных данных сети при количестве обрабатывающих элементов $T = 162$ (показаны 5 точек).

$$f: RMS = 1.628 \times 10^{-3}; \frac{df}{d\eta}: RMS = 2.854 \times 10^{-3}; \frac{d^2f}{d\eta^2}: RMS = 1.102 \times 10^{-3}; \lambda: RMS = 3.050 \times 10^{-2}$$

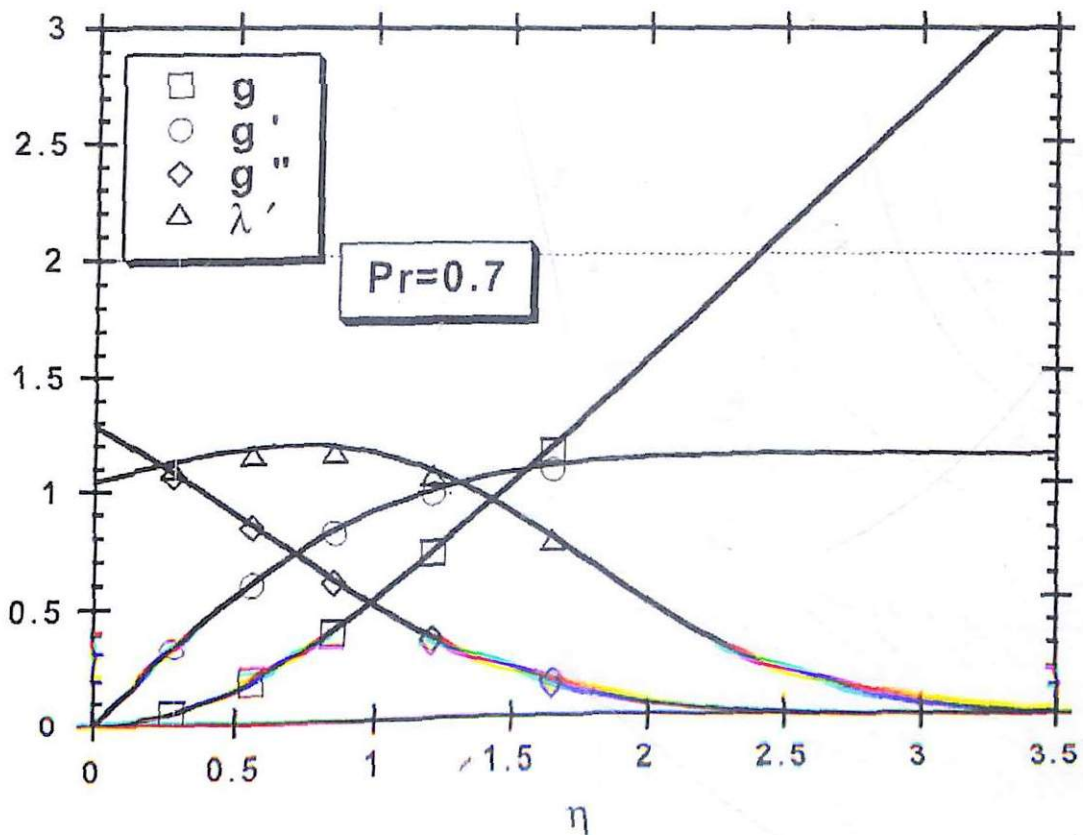


Рисунок 23 - $Pr = 0.7$: Сравнение решений Рунге-Кутты и выходных данных сети при количестве обрабатываемых элементов $T = 162$ (показаны 5 точек).

$$g: RMS = 1.274 \times 10^{-2}; \frac{dg}{d\eta}: RMS = 1.325 \times 10^{-2}; \frac{d^2g}{d\eta^2}: RMS = 1.385 \times 10^{-2}; \frac{d\lambda}{d\eta}: RMS = 2.706 \times 10^{-2}$$

Физически $Pr = 1.0$, что импульс и тепловой пограничный слой идентичны. Математически с $Pr = 1$ только в уравнении (106) и уравнении (107) должны быть линейно независимы, как это можно показать, что в этой ситуации мы должны иметь $\lambda = df/d\eta$, и также $s = \tau$.

Количество итераций (I), необходимое для решения базисных коэффициентов, было: $I = 19$ для $Pr = 1.0$ с $\epsilon = 1.834 \times 10^{-11}$ и $I = 36$ для $Pr = 0.7$ с $\epsilon = 3.711 \times 10^{-10}$. Значения среднеквадратичных различий между методом Рунге-Кутты и выходными данными сети приведены в отдельном графике. Программа работает на SUN Microsystems Sparcstation 2, с временем выполнения порядка 200 секунд для CPU с $Pr = 1.0$ и 700 секунд для $Pr = 0.7$, на 81 узле.

3.2.5.2.4 СХОДИМОСТЬ

Логарифмическая сходимость переменной $d^2f/d\eta^2$ (рисунок 24) получали путем последовательного уточнения сетки. Значения показателя для $Pr = 1.0$ составляет около 0,899. Сходимость кривой для $Pr = 0.7$ не имеет постоянный наклон, хотя его средняя, казалось бы, такая же, как для $Pr = 1.0$.

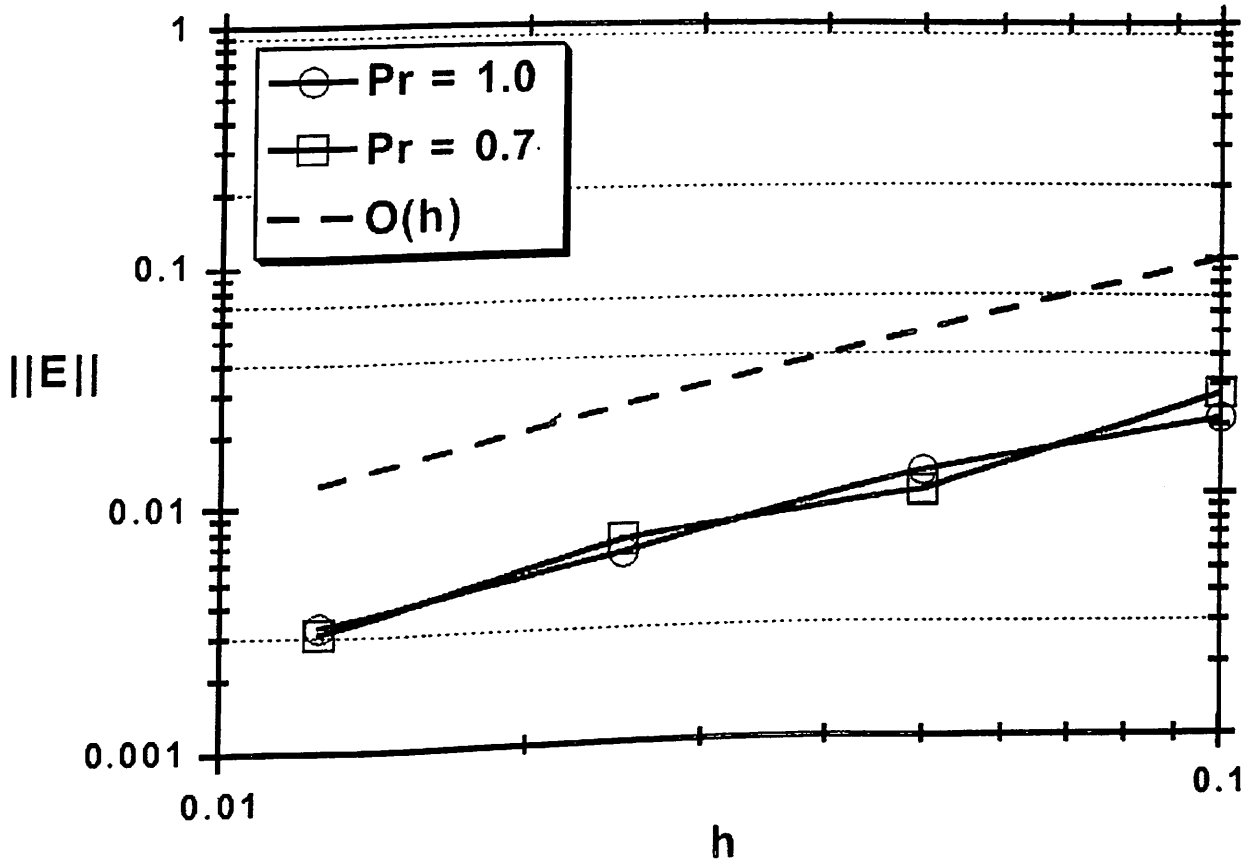


Рисунок 24 - Участок логарифмической сходимости для $d^2f/d\eta^2$.

$Pr = 1.0$: $slope = 0.899$; $Pr = 0.7$: $avg. slope \approx 0.899$

ЗАКЛЮЧЕНИЕ

Можно построить FFANN с кусочно-линейными функциями активации для моделирования решений нелинейных обыкновенных дифференциальных уравнений без необходимости обучения в обычном коннекционистском смысле. Построение требует введения определенных ограничений на значения входных данных, смещения и выходные веса, принадлежности некоторых ролей каждого из этих параметров, а также использование устоявшихся математических методов из теории приближения функций. Такой подход должен быть применим и к использованию гиперболических тангенсов, сигмоида и радиальных базисных функций.

Явные формулы для входа и смещения веса были сформулированы, и метод взвешенных невязок (в частности, одна из его вариаций техника Петрова-Галеркина) был введен как математический алгоритм для определения значений выходного веса FFANN. Два примера взяты из механики жидкости были приближены и решены такими методами, и результаты используются для построения нейронных сетей. Выход сети сравнивали с решениями, полученными с помощью итерационной техники подбора с использованием пошагового алгоритма Рунге-Кутты.

Как и методы обучения, алгоритмы, используемые в данной работе, имели итерационный характер. Тем не менее, алгебраические системы уравнений хорошо обусловлены и при возрастании числа элементов ошибка в сети выходных данных монотонно уменьшается. Кроме того, алгоритмы довольно обычные, не требующие никаких специальных допущений и работают на обычных аппаратах за разумное количество времени.

Поскольку качество базисных функций влияет на интерполяции, хранения, скорость и свойства сходимости метода построения FFANN, работа идет на создание сплайнов более высокого порядка с использованием гиперболического тангенса, сигмовидного и радиального базиса функций. Сплайны высшего порядка исследуются методами Лагранжа, Эрмита и В-сплайнов. Подход, изложенный в этой статье, позволит исследовать более подходящие алгоритмы для построения FFANNs, Sigma-Pi, и рекуррентные сети для приближения линейных и нелинейных уравнений в частных производных.

Весь прогресс от исследований в этой сфере ведется цели построения FFANN для обучения с учителем. В настоящее время разрабатывается не итерационный алгоритм обучения, который позволит решить данную проблему напрямую через необходимые веса и количества скрытых слоев. Алгоритм потребует только единственное решение в системе линейных алгебраических уравнений.

ЛИТЕРАТУРА

1. Grossberg S. 1973. Contour enhancement, short-term memory, and consistencies in reverberating neural networks. *Studies in Applied Mathematics* 52: 217, 257.
2. Hebb D. O. 1961. *Organization of behavior*. New York: Science Edition.
3. Kohonen T. 1984. *Self-organization and associative memory*. Series in Information Sciences, vol. 8. Berlin: Springer Verlag.
4. Rosenblatt F. 1962. *Principles of neurodynamics*. New York: Spartan Books. (Русский перевод: Розенблатт Ф. Принципы нейродинамики. – М.: Мир., 1965.)
5. Widrow B. 1959. Adaptive sampled-data systems, a statistical theory of adaptation. 1959 IRE WESCON Convention Record, part 4, pp. 88-91. New York: Institute of Radio Engineers.
6. Widrow B., Hoff M. 1960. Adaptive switching circuits. 1960 IRE WESCON Convention Record, pp. 96-104. New York: Institute of Radio Engineers.
7. Гроссман С. Тернер Дж. Математика для биологов: перевод с англ./предисл и коммент. Ю.М. Свирежева. – М. Высшая школа, 1983 – 383 с., ил.
8. Биология. В 2 кн. Кн. 2: Учебник для медиц. Спец. Вузов/ В.Н.Ярыгин, В.И.Васильева, И.Н. Волков, В.В.Синельщикова; под ред. В.Н.Ярыгина.- 5-е изд., испр. И доп. – М.: Высш. шк., 2003. – 334с.,ил.
9. Н. Бейли Математика в биологии и медицине. Издательство «МИР» Москва, 1970
10. <http://www-sbras.nsc.ru> Ахмеров Р.Р. Очерки по теории обыкновенных дифференциальных уравнений.
11. M. M. Meerschaert and C. Tadjeran, “Finite difference approximations for two-sided space-fractional partial differential equations,” *NSFgrant DMC*, pp. 563–573, 2004.
12. A.S. Chaves, “A fractional diffusion equation to describe Levy flights”, *Physics Letters A*, vol. 239, no.1-2, pp. 13-16, 1998.
13. D.A. Benson, S.W. Wheatcraft, and M.M. Meerschaert, “The fractional-order governing equation of Levy motion”, *Water Resources Research*, vol. 36, no. 6, pp. 1413-1423, 2000.

14. F. Liu, V. Aon, and I Turner, "Numerical solution of the fractional advection-dispersion equation", 2002, <http://academic.research.microsoft.com/Publication/3471879>.
15. K.S. Miller and B. Ross, *An introduction to the fractional calculus and fractional differential equations*, A Willey-Interscience Publication, John Wiley & Sons, New York, 1993.
16. R. Gorenflo, F. Mainardi, E. Scalas, and M. Raberto, "Fractional calculus and continuous-time finance. III. The diffusion limit", in *Mathematical Finance*, Trends Math., pp. 171-180, Birkha user, Basel, Switzerland, 2001.
17. S.G. Samko, A.A. Kilbas, and O.I. Marichev, *Fractional Integrals and Derivatives: Theory of Applications*, Gordon and Breach Science Publishers, Yverdon, Switzerland, 1993.
18. G.J. Fix and J.P. Roop, "Least squares finite-element solution of a fractional order two-point boundary value problem", *Computers & Mathematics with Applications*, vol. 48, no. 7-8, pp. 1017-1033, 2004.
19. S. Haykin, *Neural Networks*, Printice-Hall, 2006.
20. K.-I. Funahashi, "On the approximate realization of continuous mappings by neural networks", *Neural Networks*, vol. 2, no. 3, pp. 183-192, 1989.
21. A. Al-Marashi and K. Al-Wagih, "Approximation solution of boundary values of partial differential equations by using neural networks", *Thamar University Journal*, no. 6, pp. 121-136, 2007.
22. J. Freeman and D. Skapura, *Neural Networks: Algorithms, Applications, and Programming Techniques*, Addison Wesley, New York, (1991).
23. M. Takeda and J. Goodman, Neural Networks for Computation: number representation and programming complexity, *Applied Optics*, **25** (18), 3033, (1986).
24. E. Barnard and D. Casasent, New Optical Neural System Architectures and Applications, *Optical Computing* **88**, **963**, 537 Toulon, France, (1988).

25. H. Lee and I. Kang, Neural Algorithms for Solving Differential Equations, *Journal of Computational Physics*, **91**, 110, (1990).
26. L. Wang and J.M. Mendel, Structured Trainable Networks for Matrix Algebra, *IEEE International Joint Conference on Neural Networks*, **2**, 125 San Diego, June (1990).
27. S. Omohundro, Efficient Algorithms with Neural Network Behaviour, *Complex Systems*, **1**, 237, (1987).
28. T. Poggio and F. Girosi, A Theory for Networks for Approximation and Learning, *Artificial Intelligence Laboratory, Massachusetts Institute of Technology*, A.I. Memo No. 1140, July (1989).
29. F. Girosi and T. Poggio, Networks for learning: a view from the theory of approximation of functions, *Proceedings of the Genoa Summer School on Neural Networks and Their Applications*, Prentice Hall, (1989).
30. G. Cybenko, Approximation by Superposition of a Sigmoidal Function, *Math. Control Signals Systems*, **2**, 303, (1989).
31. A. Maren, C. Harston and R. Pap, *Handbook of Neural Computing Applications*, Academic Press, New York, (1990).
32. L.O. Chua, C.A. Desoer and E.S. Kuh, *Linear and Nonlinear Circuits*, McGraw Hill, New York, (1987).
33. Y. Ito, Approximation of Functions on a Compact Set by Finite Sums of a Sigmoid Function Without Scaling, *Neural Networks*, **4**, 817, (1991).
34. P.M. Prenter, *Splines and Variational Methods*, Wiley, New York, (1989).
35. A.J. Meade, Jr. and A.A. Fernandez, The Numerical Solution of Linear Ordinary Differential Equations by Feed forward Neural Networks, *Mathematical and Computer Modelling*, **19** (12), 1, (1994).
36. Finlayson, B. A. and Scriven, L.E., The Method of Weighted Residuals – A Review, *Applied Mechanics Review*, **19** (9), 735, (1966).
37. C.A.J. Fletcher, *Computational Galerkin Methods*, Springer-Verlag, New York, (1984).

38. G. Strang, *Linear Algebra and Its Applications*, Second Edition, Academic Press, New York, (1980).
39. Frank M. White, *Viscous Fluid Flow*, McGraw-Hill Book Company, New York, (1974).
40. I. Christie, D.F. Griffiths, A.R. Mitchell, and J.M. Sanz-Serna, Product Approximation for Non-linear Problems in the Finite Element Method, *Journal of Numerical Analysis* **1**, 253, (1981).
41. K. Stewartson, Further solutions of the Falkner-Skan equation, *Proc. Camb. Phil. Soc. Journal of Numerical Analysis* **50**, 454, (1954).
42. *IMSL User's Manual, MATH/LIBRARY*, Version 2, (1991).
43. C. Johnson, *Numerical Solution of Partial Differential Equations by The Finite Element Method*, Cambridge University Press, Cambridge, (1990).
44. F.K. Moore, Three-Dimensional Compressible Laminar Boundary-Layer Flow, NACA TN 2279, (1951).
45. F.K. Moore, Laminar Boundary Layer on Cone in Supersonic Flow at Large Angle of Attack, NACA TR 1132, (1952).
46. E. Reshotko, Laminar Boundary Layer with Heat Transfer on a Cone at Angle of Attack in a Supersonic Stream, NACA TN 4152, (1957).
47. A.J. Meade Jr., Semi-Discrete Galerkin Modelling of Compressible Viscous Flow Past a Circular Cone at Incidence, Ph.D Thesis, University of California, Berkeley, (1989).