

УДК 004.032.26(06)

## НАЧАЛЬНАЯ ИНИЦИАЛИЗАЦИЯ МНОГОСЛОЙНЫХ ПРОГНОСТИЧЕСКИХ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

*В.В. Бочкарёв, Ю.С. Масленникова*

### Аннотация

В работе предложен метод начальной инициализации искусственных нейронных сетей, предназначенных для решения задач прогнозирования временных рядов. Для выбора начальных значений весов нейронов используется аналогия с фильтром линейного предсказания. Рассматриваются различные варианты декомпозиции матрицы преобразования, соответствующего фильтру линейного предсказания, для наиболее эффективной инициализации многослойной нейронной сети. На примере задачи предсказания траекторий хаотической системы Лоренца показано, что применение данного метода позволяет значительно повысить точность нейросетевого прогноза.

**Ключевые слова:** искусственные нейронные сети, прогнозирование, начальная инициализация, фильтр линейного предсказания.

### Введение

Построение достоверных прогнозов в настоящее время является важной частью не только научных исследований, но и повседневной деятельности. На сегодняшний день для решения данной задачи наиболее часто используется метод линейного предсказания. Решение задачи построения прогноза в этом случае сводится к нахождению коэффициентов линейного предсказывающего фильтра  $p$ -го порядка, который наилучшим образом прогнозирует текущее значение вещественной последовательности  $\hat{x}(n)$  по ее предыдущим отсчетам:

$$\hat{x}(n) = -a(1)x(n-1) - a(2)x(n-2) - \dots - a(p)x(n-p), \quad (1)$$

где  $p$  задает порядок фильтра,  $a$  – его коэффициенты. Как правило, минимизируется сумма квадратов ошибок предсказания как функция коэффициентов фильтра [1]. Глобальный минимум такой функции определён однозначно, так как она имеет вид параболоида.

Метод линейного предсказания эффективен для стационарных систем, свойства которых не изменяются со временем, а распределение флуктуаций параметров близко к нормальному закону [1, 2]. Для более сложных случаев предпочтительнее использовать нелинейные методы, например искусственные нейронные сети [3, 4]. Однако данный метод также имеет недостатки. В случае нейронных сетей за счёт нелинейности функций активации нейронов функционал ошибок имеет сложный характер, и найденный минимум может быть не глобальным, а всего лишь локальным. Это может привести к тому, что результаты нейросетевого прогнозирования окажутся хуже, чем при использовании линейного прогнозирования, несмотря на большие потенциальные возможности нейронных сетей [5].

Как отмечается в [3], при использовании нейронных сетей многое определяется удачным выбором правила начальной инициализации весов нейронов. Алгоритмы

инициализации нейросетей предполагают выбор детерминированных либо случайных (например, распространённый алгоритм Нгуена – Уидроу [5]) начальных значений весов. В последнем случае обучение сети, как правило, выполняется много раз с тем, чтобы увеличить вероятность нахождения глобального минимума функции ошибки. Это приводит к многократному росту общего времени работы, притом что не всегда удаётся получить удовлетворительный результат.

В работе [6] предложен интересный алгоритм построения гибридной нейронной сети [4] для задач прогнозирования временных рядов. Процесс начинается с обучения сети, состоящей из одного линейного нейрона. Полученные коэффициенты используются для инициализации гибридной сети, также содержащей один нейрон. После этого сеть последовательно наращивается добавлением нейронов, пока ошибка прогнозирования не перестает уменьшаться. В настоящей работе мы рассмотрим другой подход, также основанный на предварительном решении задачи оптимального линейного прогноза.

Из формулы (1) видно, что фильтр линейного предсказания можно представить в виде однослойной линейной нейронной сети прямой передачи сигнала. Используя эту аналогию, за основу можно взять предварительно найденные коэффициенты линейного предсказания, включив их в соответствующие элементы нейронной сети. Идея состоит в том, чтобы построить нейронную сеть, которая до обучения преобразует данные полностью аналогично выражению (1). Для того чтобы наиболее полно использовать возможности нейросетевых алгоритмов по аппроксимации сложных зависимостей, сеть должна иметь 2 и более слоёв [3]. При этом возникает проблема, связанная с неоднозначностью декомпозиции исходного линейного преобразования на несколько последовательных преобразований, соответствующих слоям нейронной сети. Далее будут рассмотрены два возможных варианта решения этой проблемы.

### 1. Простейший алгоритм начальной инициализации нейронной сети на основе коэффициентов линейного предсказания

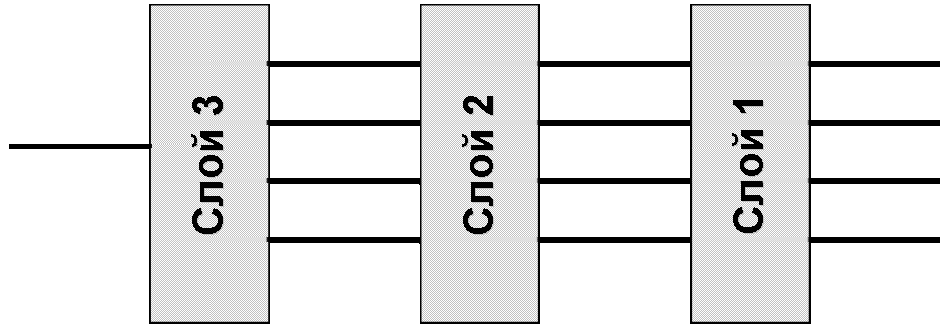
Рассмотрим простейший вариант решения поставленной задачи, когда функции, выполняемые различными слоями сети, чётко разделены. Погрешность, вносимая нелинейностью функций активации нейронов сети до её обучения, должна быть небольшой. Это накладывает определённые ограничения на выбор активационных функций для слоёв. Например, гиперболический тангенс (сигмоид) имеет область со слабой нелинейностью в начале координат. При использовании нейронов с сигмоидальными функциями активации, соответствующим образом масштабируя обучающую выборку, мы можем добиться того, чтобы преобразования сигнала нейронами сети были практически линейными. Разберём предлагаемый алгоритм инициализации на примере трёхслойной нейронной сети прямой передачи сигнала [4].

Предполагая, что значения на входах всех нейронов соответствуют линейным участкам функции активации, представим результирующую матрицу преобразований трёхслойной нейронной сети в виде:

$$A = A_3 A_2 A_1, \quad (2)$$

где  $A_1$ ,  $A_2$ ,  $A_3$  – матрицы преобразования 1-го, 2-го и 3-го слоёв соответственно (рис. 1). Входной слой сети масштабирует данные так, чтобы все значения ряда не выходили за пределы линейной области гиперболического тангенса.

Промежуточный слой осуществляет слабонелинейное преобразование с помощью весовой матрицы, содержащей в себе коэффициенты линейного прогнозирования  $a$ . Выходной слой осуществляет обратное масштабирование данных. Ошибка,



$$A = \begin{bmatrix} 0 & 0 & 0 & \dots & 1/\alpha \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ -a_1 & -a_2 & -a_3 & \dots & -a_p \end{bmatrix} \cdot \begin{bmatrix} \alpha & 0 & 0 & \dots & 0 \\ 0 & \alpha & 0 & \dots & 0 \\ 0 & 0 & \alpha & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & \alpha \end{bmatrix}$$

Рис. 1. Матрица преобразования нейронной сети после начальной инициализации

обусловленная нелинейностью функции активации, может быть сделана сколь угодно малой за счёт подходящего выбора коэффициента  $\alpha$ . В этом случае нейросеть с выбранными таким образом весами будет по поданным на входы  $p$  идущим подряд отсчётам сигнала  $x_{k-1}, x_{k-2}, \dots, x_{k-p}$  вычислять оценку следующего значения  $\hat{x}_k$  в соответствии с выражением (1).

Следует отметить, что результаты прогнозирования данной нейронной сети до обучения будут практически идентичны выходу фильтра линейного предсказания. Если мы используем для обучения алгоритм, при котором ошибка монотонно уменьшается (таковы градиентные, квазиньютоновские алгоритмы, алгоритмы, основанные на методе сопряжённых градиентов и др.), то на любом этапе обучения результат будет, как минимум, не хуже получаемого с помощью фильтра линейного предсказания. В настоящей работе использовался алгоритм Левенберга–Марквардта, относящийся к квазиньютоновским методам, обеспечивающим быструю сходимость [7].

## 2. Прогнозирование хаотической системы Лоренца

Проверка предложенного способа инициализации нейронной сети осуществлялась применительно к задаче прогнозирования траекторий системы Лоренца, которая является простейшим примером детерминированной хаотической системы [8]. Она впервые появилась в численных экспериментах Лоренца, исследовавшего поведение траекторий системы из трёх связанных квадратичных обыкновенных дифференциальных уравнений, описывающих три моды уравнений Обербека–Буссинеска для конвекции жидкости в двумерном слое, подогреваемом снизу. Упомянутые уравнения имеют вид:

$$\begin{cases} \dot{x} = \sigma(y - x), \\ \dot{y} = x(r - z) - y, \\ \dot{z} = xy - bz. \end{cases} \quad (3)$$

Как было установлено численным моделированием, решения системы при многих значениях параметров асимптотически приближаются к неустойчивым цик-

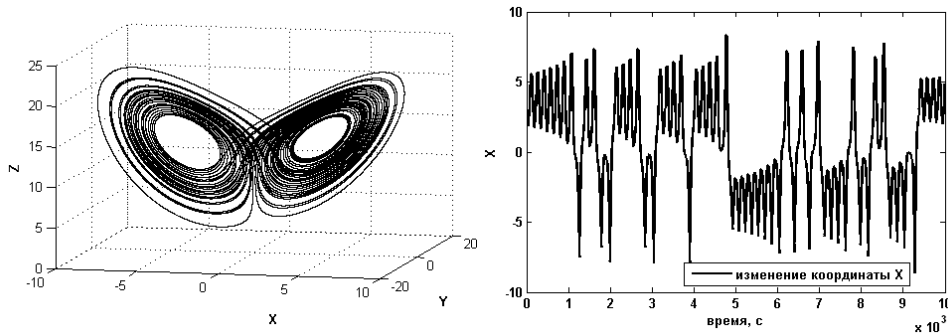


Рис. 2. Траектории аттрактора Лоренца при  $r = 28$  (слева); изменение координаты  $X$  (справа)

лам, но при этом имеют два ярко выраженных центра кластеризации, формируя тем самым так называемый странный аттрактор [8]. На рис. 2 приведён пример решения системы при  $\sigma = 10$ ,  $r = 28$ ,  $b = 8/3$ . Важной особенностью решений системы Лоренца также является существенная зависимость их от начальных условий – основная черта хаотической динамики. Таким образом, поведение системы на длительных промежутках времени становится слабопредсказуемым.

Выбор системы Лоренца в качестве тестовой задачи объясняется тем, что она считается достаточно адекватной моделью для многих реальных систем. В литературе отмечается, что этой системой можно приближенно описать колебания параметров в турбулентных потоках, вариации параметров геомагнитного поля [8, 9]. Подобная система уравнений возникает также при моделировании экономических процессов [10].

Для тестирования предлагаемого алгоритма инициализации были построены ряды, соответствующие решениям системы уравнений (3). Использовался метод Рунге–Кутты с автоматическим выбором шага. Затем полученные ряды были прорежены с тем, чтобы получить выборку приемлемого объёма, охватывающую в то же время достаточно большой интервал времени, при этом шаг по времени составил 0.01 с. Был также удалён начальный сегмент выборки, чтобы используемые для тестирования данные соответствовали движению вблизи аттрактора и результаты не зависели от выбора начальной точки.

Ниже представлены результаты прогнозирования поведения данной системы с помощью нейронной сети, начальные значения весов которой были основаны на коэффициентах фильтра линейного прогнозирования. Для начала были найдены коэффициенты линейного прогноза, на основании которых описанным выше способом затем была смоделирована нейронная сеть. Предсказание очередного значения осуществлялось по 5-ти идущим подряд предшествующим отсчётам. Первый и второй слой нейронной сети состояли из 5 нейронов каждый. Обучение нейронной сети производилось методом обратного распространения ошибки с минимизацией функционала ошибки по алгоритму Левенберга – Марквардта за 300 эпох обучения. Прогнозирование осуществлялось на интервал от 0.01 до 1 с вперёд.

На рис. 3 представлен фрагмент спрогнозированного ряда, для сравнения приведён также ряд, полученный по методу линейного предсказания. Данный пример является типичным: для всего диапазона изменения интервала предсказания при использовании предлагаемого алгоритма начальной инициализации наблюдается существенное улучшение по сравнению как с методом линейного предсказания, так и с результатами нейросетевого прогнозирования при использовании алгоритма начальной инициализации Нгуена – Уидроу. Среднеквадратические ошибки

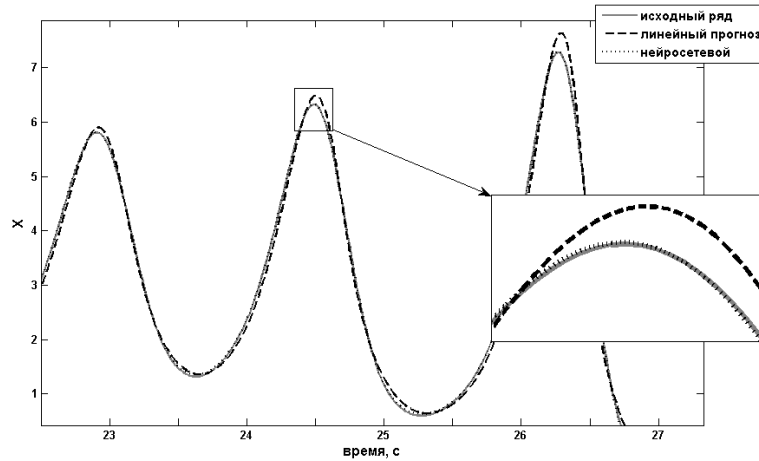


Рис. 3. Фрагмент предсказанного ряда изменения координаты  $X$  на 0.15 с вперёд, область внутри прямоугольника показывает фрагмент того же ряда в увеличенном масштабе

прогнозирования на различные временные интервалы в сравнении с результатами других методов приведены в табл. 1.

Полученные результаты показали эффективность предложенного метода начальной инициализации нейронной сети. Можно также утверждать, что применительно к рассмотренной задаче нейросетевой прогноз значительно эффективнее метода линейного предсказания.

### 3. Усовершенствованный метод начальной инициализации сети

Основным недостатком рассмотренного выше простейшего алгоритма следует считать то, что слои сети (по крайней мере, на начальной стадии обучения) в разной степени участвуют в преобразовании данных. Для того чтобы более полно использовать возможности многослойных сетей, желательно более равномерно распределить вычисления между слоями. Как уже указывалось, описанный способ начальной инициализации нейронной сети не единственный. Для рассмотренного выше примера трёхслойной нейронной сети в итоговую матрицу преобразования  $A$  (см. выражение (2) и рис. 1) можно включить действительные ортогональные матрицы  $U_1$  и  $U_2$ . Тогда  $U_1^T U_1 = E$  и  $U_2^T U_2 = E$ , и выполняется соотношение

$$A = A_3 U_2^T U_2 A_2 U_1^T U_1 A_1. \quad (4)$$

Таким образом, до обучения нейронная сеть, начальные значения весовых коэффициентов слоёв которой заданы матрицами  $U_1 A_1$ ,  $U_2 A_2 U_1^T$  и  $A_3 U_2^T$  соответственно, по-прежнему полностью будет соответствовать исходному фильтру линейного предсказания. Однако в процессе обучения за счет нелинейности передаточной функции нейронов можно ожидать улучшения качества прогноза. Для того чтобы подобрать оптимальные матрицы  $U_1$  и  $U_2$ , удобно представить их в экспоненциальном виде

$$U_i = e^{G_i}. \quad (5)$$

Как известно, в этом случае для того, чтобы матрицы  $U_i$  были ортогональными, матрицы  $G_i$  должны быть кососимметрическими

$$G_i^T = -G_i.$$

Пусть число нейронов в 1-м и 2-м слоях равно  $N$ . Выберем в линейном пространстве кососимметрических матриц размера  $N \times N$  (размерность его равна  $N(N-1)/2$ ) некоторый базис, например, следующим образом

$$G^{(k,m)}, \quad k < m, \quad G_{i,j}^{(k,m)} = \delta_{m,i} \delta_{k,j} - \delta_{k,i} \delta_{m,j}.$$

Тогда

$$G_i = \sum_{k,m} x_{k,m,i} G^{(k,m)}. \quad (6)$$

Например, для случая  $N = 3$  имеем представление

$$G_i = \begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix} = a_i \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + b_i \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} + c_i \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}.$$

При малом числе нейронов в слоях для поиска неизвестных в выражении (4) можно воспользоваться методом градиентного спуска [7]. Наиболее последовательным было бы при этом использовать в качестве целевой функции ошибку нейронной сети, инициализированной указанным способом и прошедшей затем полный цикл обучения. Ясно, однако, что данный подход требует большого объёма вычислений. Только для вычисления градиента по коэффициентам  $x_{k,m,i}$  из (6) в этом случае требуется выполнить  $\sim N^2$  циклов обучения сети.

Для решения этой проблемы можно воспользоваться алгоритмами стохастического поиска, либо задавать при инициализации случайные значения коэффициентов  $x_{k,m,i}$ . В этом случае вместо экспоненциальной параметризации ортогональных матриц (5) может быть также использована требующая меньшего объёма вычислений параметризация в виде

$$U_i = (I - G_i) (I + G_i)^{-1}.$$

Для поиска глобального минимума в этом случае потребуются многократное повторение обучения нейронной сети. Другой возможный подход, который и будет рассмотрен далее, основан на следующих соображениях. Постараемся выбрать такие матрицы  $U_1$  и  $U_2$ , чтобы выигрыш за счёт нелинейного преобразования данных в нейронной сети сказался наиболее быстро. Скорость обучения нейронной сети зависит от градиента, полученного дифференцированием функции ошибки по весам нейронов. Будем варьировать коэффициенты  $x_{k,m,i}$  в (6), стремясь максимизировать норму градиента. Отметим, что этот процесс полностью аналогичен процедуре обучения нейронной сети по методу обратного распространения ошибки, ведь, по сути, мы просто накладываем некоторые линейные ограничения на изменение весовых коэффициентов в соответствии с выражениями (4). В силу этого можно ожидать примерно двукратного увеличения общего времени работы по сравнению с изложенным выше простейшим алгоритмом инициализации. Реализация данного алгоритма требует минимальных добавлений к имеющимся библиотекам программ нейросетевых вычислений. Как вариант, в качестве целевой функции можно выбрать функцию ошибок сети после первой эпохи обучения. Для решения задачи оптимизации в обоих случаях можно использовать, например, метод градиентного спуска.

На тех же модельных данных и при той же архитектуре нейронной сети, что и в предыдущем случае, было проведено тестирование предлагаемого алгоритма. В качестве целевой функции из соображений простоты программирования использовалась ошибка нейронной сети после первой эпохи обучения. Выполнялось по

Табл. 1

Метод прогнозирования	$\sigma$ (0.01 с)	$\sigma$ (0.02 с)	$\sigma$ (0.05 с)	$\sigma$ (0.1 с)	$\sigma$ (1 с)
Линейный	$3.138 \cdot 10^{-4}$	$13 \cdot 10^{-4}$	0.0108	0.0646	3.4031
Нейросетевой	$1.184 \cdot 10^{-4}$	$9.441 \cdot 10^{-4}$	0.0062	0.0377	0.7285
Нейросетевой, основанный на линейном	$0.505 \cdot 10^{-4}$	$1.997 \cdot 10^{-4}$	0.0016	0.0097	0.0862
Нейросетевой, усовершенствованный	$0.275 \cdot 10^{-4}$	$0.985 \cdot 10^{-4}$	0.0009	0.0048	0.0564

10 итераций градиентного спуска. Для сети с найденными оптимальными параметрами начальных весов затем проводилось 300 эпох обучения, как и в предыдущем случае. При принятых параметрах общее время работы увеличилось менее чем в 2 раза по сравнению с рассмотренным выше простейшим алгоритмом инициализации.

В табл. 1 приведены среднеквадратические ошибки прогнозирования, полученные с применением 4-х методов (метод линейного предсказания, нейросетевое прогнозирование с инициализацией по алгоритму Нгуена–Уидроу, простейшему и улучшенному алгоритмам инициализации на основе коэффициентов линейного предсказания) в одинаковых условиях для интервалов предсказания 0.01, 0.02, 0.05, 0.1, 1 с. Можно видеть, что применение улучшенного алгоритма начальной инициализации во всех случаях приводит к дальнейшему значительному уменьшению ошибки прогнозирования.

Для сравнения также был проведён поиск значений коэффициентов  $x_{k,m,i}$  по критерию минимума ошибки на выходе полностью обученной сети. Ошибка прогнозирования на 0.05 с вперед в этом случае составила 0.00072 против 0.0009 при использовании критерия наибольшего градиента. Как видим, выигрыш в точности от перехода к методу, требующему гораздо большего объёма вычислений, не столь значителен.

### Заключение

Проведенное прогнозирование поведения хаотической системы Лоренца показало, что предложенный метод начальной инициализации нейронной сети позволяет существенно уменьшить ошибку прогноза по сравнению как с методом линейного предсказания, так и с нейросетевым прогнозированием при использовании универсальных алгоритмов начальной инициализации. Можно утверждать также, что применительно к данной задаче в равных условиях нейросетевой метод прогнозирования значительно эффективнее линейного. Усовершенствованный алгоритм начальной инициализации позволяет дополнительно улучшить качество прогнозирования ценой незначительного увеличения затрачиваемого на обучение времени. Таким образом, разработанный метод показал свою состоятельность применительно к прогнозированию траекторий хаотической системы Лоренца, что позволяет использовать его в задачах прогнозирования подобных сложных рядов.

### Summary

*V.V. Bochkarev, Yu.S. Maslennikova.* Initialization of Multilayered Forecasting Artificial Neural Networks.

In this paper a method of initializing a neural network solving the problem of forecasting a time series is offered. The analogy with a linear prediction filter is used to select an initial

weighting coefficient for neurons. Also the variants of decomposition of a transformation matrix corresponding to the linear prediction filter are offered to improve the initialization method quality. Through the neural nets forecasting of the Lorentz chaotic system's trajectory it is shown that the application of the given method allows significantly increasing the accuracy of the neural network prediction.

**Key words:** neurofuzzy modeling, forecasting, neural network initialization, linear prediction filter.

#### Литература

1. *Андерсон Т.* Статистический анализ временных рядов. – М.: Мир, 1976. – 757 с.
2. *Арато М.* Линейные стохастические системы с постоянными коэффициентами. Статистический подход. – М.: Наука, 1989. – 304 с.
3. *Галушкин А.И.* Теория нейронных сетей. – М.: ИПРЖР, 2000. – 416 с.
4. *Круглов В.В., Дли М.И., Голунов Р.Ю.* Нечёткая логика и искусственные нейронные сети. – М.: Изд-во физ.-матем. лит., 2001. – 224 с.
5. *Медведев В.С., Потёмкин В.Г.* Нейронные сети. МАТЛАВ 6. – М.: ДИАЛОГ-МИФИ, 2002. – 496 с.
6. *Gholipour A., Lucas C., Araabi B.N., Shafiee M.* Solar activity forecast: Spectral analysis and neurofuzzy prediction // J. Atmos. Solar-Terrest. Phys. – 2005. – V. 67. – P. 595–603.
7. *Грилл Ф., Мюррей У., Райт М.* Практическая оптимизация. – М.: Мир, 1985. – 509 с.
8. *Кроновер Р.М.* Фракталы и хаос в динамических системах. Основы теории. – М.: Постмаркет, 2000. – 352 с.
9. *Фрик П.Г.* Турбулентность: подходы и методы. – Москва-Ижевск: ИКИ, 2003. – 292 с.
10. *Пу Т.* Нелинейная экономическая динамика. – Москва-Ижевск: НИЦ «РХД», 2000. – 198 с.

Поступила в редакцию  
26.12.09

---

**Бочкарёв Владимир Владимирович** – ассистент кафедры радиофизики Казанского государственного университета.

E-mail: *Vladimir.Bochkarev@ksu.ru*

**Масленникова Юлия Сергеевна** – магистрант 2-го года обучения Казанского государственного университета.

E-mail: *YuliaMsl@gmail.com*