





Известия высших учебных заведений. Прикладная нелинейная динамика. 2023. Т. 31, № 4
Izvestiya Vysshikh Uchebnykh Zavedeniy. Applied Nonlinear Dynamics. 2023;31(4)

Научная статья
УДК 004.032.26, 530.152.2

DOI: 10.18500/0869-6632-003052
EDN: XGRKMR

Влияние шума на рекуррентные нейронные сети с нелинейными нейронами

В. М. Москвитин, Н. И. Семенова 

Саратовский национальный исследовательский
государственный университет имени Н. Г. Чернышевского, Россия
E-mail: vmoskvitin@gmail.com, semenovani@sgu.ru

Поступила в редакцию 6.03.2023, принята к публикации 2.05.2023,
опубликована онлайн 12.07.2023, опубликована 31.07.2023

Аннотация. Цель настоящего исследования — установить особенности распространения и накопления шумов в рекуррентной нейронной сети на примере упрощенной эхо-сети. В данной работе исследовалось влияние типа функции активации искусственных нейронов и матриц связи между ними. *Методы.* В качестве источников шумов рассматриваются источники белого Гауссова шума. В зависимости от способа воздействия шума на искусственные нейроны использовались аддитивный, мультипликативный и смешанный шум. Оценка шумового воздействия проводилась с использованием дисперсии (вариации) выходного сигнала. *Результаты.* Показано, что функция активации играет существенную роль в накоплении шумов. Рассмотрены две нелинейные функции активации: гиперболический тангенс и сигмоида со значениями от 0 до 1. Показано, что для второй функции происходит подавление некоторых типов шумов. В результате рассмотрения влияния матриц связи было выявлено, что диагональные матрицы связи с большим коэффициентом рассеяния приводят к меньшему накоплению шума в резервуаре эхо-сети при увеличении влияния памяти резервуара. *Заключение.* Показано, что функции активации вида сигмоиды со значениями от 0 до 1 подходят для подавления мультипликативных и смешанных шумов. Было рассмотрено накопление шума в резервуаре для трех типов матриц связи внутри резервуара: однородная матрица, ленточная матрица с малым коэффициентом рассеяния и ленточная матрица с большим коэффициентом рассеяния. Было установлено, что эхо-сети с ленточной матрицей с большим коэффициентом рассеяния накапливают шумы меньше всех остальных. Это работает как для аддитивных, так и для мультипликативных шумов.

Ключевые слова: нейронные сети, рекуррентные нейронные сети, эхо-сети, шумовое воздействие, белый шум, нелинейная функция активации.

Благодарности. Работа была поддержана грантом Российского Научного Фонда 21-72-00002.

Для цитирования: Москвитин В. М., Семенова Н. И. Влияние шума на рекуррентные нейронные сети с нелинейными нейронами // Известия вузов. ПНД. 2023. Т. 31, № 4. С. 484–500. DOI: 10.18500/0869-6632-003052. EDN: XGRKMR

Статья опубликована на условиях Creative Commons Attribution License (CC-BY 4.0).

Noise influence on recurrent neural network with nonlinear neurons

V. M. Moskvitin, N. I. Semenova✉

Saratov State University, Russia

E-mail: vmmoskvitin@gmail.com, ✉semenovani@info.sgu.ru

Received 6.03.2023, accepted 2.05.2023, available online 12.07.2023,
published 31.07.2023

Abstract. The purpose of this study is to establish the features of noise propagation and accumulation in a recurrent neural network using a simplified echo network as an example. In this work, we studied the influence of activation function of artificial neurons and the connection matrices between them. *Methods.* We have considered white Gaussian noise sources. We used additive, multiplicative and mixed noise depending on how the noise is introduced into artificial neurons. The noise impact was estimated using the dispersion (variance) of the output signal. *Results.* It is shown that the activation function plays a significant role in noise accumulation. Two nonlinear activation functions have been considered: the hyperbolic tangent and the sigmoid function with range from 0 to 1. It is shown that some types of noise are suppressed in the case of the second function. As a result of considering the influence of coupling matrices, it was found that diagonal coupling matrices with a large blurring coefficient lead to less noise accumulation in the echo network reservoir with an increase in the reservoir memory influence. *Conclusion.* It is shown that activation functions of the form of sigmoid with range from 0 to 1 are suitable for suppressing multiplicative and mixed noise. The accumulation of noise in the reservoir was considered for three types of coupling matrices inside the reservoir: a uniform matrix, a band matrix with a small blurring coefficient, and a band matrix with a large blurring coefficient. It has been found that the band matrix echo networks with a high blurring coefficient accumulate the least noise. This holds for both additive and multiplicative noise.

Keywords: neural networks, recurrent neural networks, echo-state networks, noise influence, white noise, nonlinear activation function.

Acknowledgements. This work was supported by Russian Science Foundation (Project no. 21-72-00002).

For citation: Moskvitin VM, Semenova NI. Noise influence on recurrent neural network with nonlinear neurons. *Izvestiya VUZ. Applied Nonlinear Dynamics.* 2023;31(4):484–500. DOI: 10.18500/0869-6632-003052

This is an open access article distributed under the terms of Creative Commons Attribution License (CC-BY 4.0).

Введение

На данный момент искусственные нейронные сети (ИНС) уже успешно используются в многочисленных областях машинного обучения и решают задачи различной сложности [1]. К таким задачам можно отнести распознавание образов [2, 3], их классификация, улучшение звуковых записей, распознавание речи [4], предсказание климатических явлений [5] и многое другое.

В данной работе рассматриваются рекуррентные сети на примере упрощенной эхо-сети. Такие сети характеризуются тем, что часть их нейронов обладает запаздывающей обратной связью, что позволяет сети «запоминать» свои предыдущие состояния. Подобные сети часто используются для решения задач прогнозирования [6] или распознавания в реальном времени [7].

Усложнение топологий ИНС и поставленных задач может вскоре привести к кризису [8, 9], когда возможностей современных компьютеров и вычислительных кластеров уже не будет достаточно для удовлетворения растущих потребностей. Здесь на помощь приходит новое направление конструирования ИНС — аппаратные нейронные сети [10]. Согласно этому подходу, нейронные сети не создаются с помощью компьютера, а представляют собой реальное устройство, способное обучаться и решать задачи. Сами нейроны и связь между ними реализуются на физическом уровне, то есть система не моделируется на компьютере, а реализуется аппаратно согласно соответствующим физическим принципам. В последние годы наблюдается экспоненциальный рост работ с аппаратными реализациями ИНС. Наибольшую эффективность на данный момент

показывают ИНС, в основе которых лежат лазеры [11], мемристоры [12], спин-трансферные осцилляторы [13]. Связь между нейронами в оптических реализациях ИНС основывается на принципах голографии [14], дифракции [15, 16], интегрированных сетях модуляторов Маха–Цендера [17], спектрального уплотнения каналов [18], оптических связях, реализованных при помощи 3D принтера [19–21].

Основным принципом построения ИНС является распространение сигнала между нейронами при помощи связей с некоторыми коэффициентами (весами). При этом наибольшей эффективности и быстродействия можно добиться при помощи распараллеливания вычислений на высокопроизводительных вычислительных кластерах. Однако в этом случае «узким горлышком» служит скорость обращения к памяти и обработка данных. Максимальная производительность вычислений может быть достигнута только в том случае, если ИНС реализована полностью аппаратно. В этом случае исчезает проблема обращения к памяти и математических операций над большим количеством данных, так как каждый нейрон соответствует аппаратному нелинейному компоненту, а каждое соединение — физическому каналу связи. Физическая реализация ИНС принципиально меняет особенности влияния шумов. В случае цифровой компьютерной реализации ИНС шум может попадать в систему исключительно с входным сигналом, тогда как в аппаратной ИНС существует множество внутренних источников шума с различными свойствами. Настоящая статья направлена на исследование особенностей распространения внутренних шумов в рекуррентных ИНС, выявление способов подавления таких шумов и обоснование устойчивости сетей к некоторым типам шумов.

В наших предыдущих работах мы уже рассматривали влияние шумов на глубокие нейронные сети [22, 23] и предлагали стратегии подавления шума в них [24]. Данная статья является существенным усложнением поставленной задачи, так как рекуррентные сети обладают свойством памяти о своих предыдущих состояниях, следовательно накопление шумов в них должно происходить сложнее, чем в ИНС, не зависящих от временной реализации. В качестве примера рекуррентной сети будет рассмотрена упрощенная эхо-сеть.

Основным предметом исследования данной работы является влияние функции активации и свойства памяти эхо-сети на накопление шума. По этой причине рассматриваются упрощенные матрицы связи между нейронами. Конкретный вид матриц связи существенно зависит от задач, для которых была обучена ИНС, и их статистические особенности могут существенно влиять на накопление шума. Поэтому влияние особых матриц связи уже обученных ИНС на накопление шума будет предметом исследования в последующих работах.

1. Влияние шума на один нейрон

1.1. Основные уравнения. ИНС состоят из искусственных нейронов, чья роль заключается в линейном или нелинейном преобразовании входного сигнала. Это преобразование реализуется при помощи так называемой функции активации. Тип функции активации зависит от поставленной задачи и от эффективности обучения. Так, в рекуррентных нейронных сетях часто используются кусочно-линейные функции и семейство функций класса «сигмоиды». В данной статье будут рассмотрены две функции этого класса: гиперболический тангенс $f(x) = \tanh ax$ и сигмоида вида $f(x) = \frac{1}{1+e^{-ax}}$. В обеих функциях множитель a отвечает за «крутизну» нелинейности. Таким образом, выходной сигнал i -го искусственного нейрона без влияния шума определяется следующим образом:

$$x_i^{\text{out}}(t) = f(x_i^{\text{in}}(t)), \quad (1)$$

где $x_i^{\text{out}}(t)$ — выходной сигнал i -го нейрона в момент времени t , а $x_i^{\text{in}}(t)$ — его входной сигнал.

Для того чтобы указать, какие именно нейроны находятся под шумовым воздействием, мы будем использовать оператор шума \hat{N} , который будет применяться к выходному сигналу без шумового воздействия: $y_i^{\text{out}} = \hat{N}x_i^{\text{out}}$.

В данной статье рассматриваются аддитивный и мультипликативный шумы. Особенность аддитивного шума заключается в том, что он добавляется к выходному сигналу, а мультипликативный умножается на него:

$$y_i^{\text{out}}(t) = \hat{N}x_i^{\text{out}}(t) = x_i^{\text{out}}(t) \cdot (1 + \xi_M(t, i)) + \xi_A(t, i), \quad (2)$$

где ξ — это независимые источники аддитивного (индекс «А») и мультипликативного (индекс «М») белого Гауссова шума с нулевым средним и соответствующими дисперсиями σ_A^2 и σ_M^2 . Свойство независимости источников шума в этом разделе роли не играет, так как рассматривается всего один нейрон. Однако в следующих разделах, в которых изучаются особенности поведения ИНС, это свойство является важным и будет влиять на последующие выводы. Также стоит отметить, что изучение особенностей влияния коррелированного шума на ИНС и его накопления тоже представляет большой интерес, и будет предметом исследования в последующих работах.

1.2. Оценка шумового воздействия. В наших предыдущих работах [22–24] для оценки уровня шума использовалась характеристика «отношение сигнал–шум» (signal-to-noise ratio, SNR), которая вычислялась как отношение среднего сигнала к его стандартному отклонению. Использование такой интерпретации SNR имеет некоторые ограничения. В частности, она может быть применена только к положительным сигналам. Так как в этой статье одной из функций активации является гиперболический тангенс, выходные значения которого находятся в диапазоне от -1 до 1 , такую формулу расчета SNR уже нельзя использовать. По этой причине в данной статье используется более универсальная характеристика, дисперсия $\sigma^2[\cdot]$, которая является мерой разброса значений случайной величины относительно ее математического ожидания и вычисляется следующим образом:

$$\sigma^2[X] = \text{Var}[X] = E[(X - E[X])^2], \quad (3)$$

где X — некоторая дискретная случайная величина, $E[X]$ — ее математическое ожидание (среднее значение). Для оценки уровня дисперсии в качестве множества X берется множество выходных значений, которые являются откликом сети или одного отдельно взятого нейрона на один и тот же входной сигнал. Такая интерпретация позволяет оценить, какой диапазон значений наиболее подвержен разным шумовым воздействиям. Независимо от знака выходного сигнала, дисперсия будет всегда положительной, и чем ближе ее значение к нулю, тем меньше зашумленность выходного сигнала. Далее в тексте используются общепринятые обозначения математического ожидания $E[\cdot]$ и вариации $\text{Var}[\cdot]$ для оценки уровня шума.

Дисперсия выходного сигнала одного зашумленного нейрона для одного и того же входного сигнала может быть вычислена следующим образом:

$$\sigma^2[y^{\text{out}}] = \text{Var}[x^{\text{out}} \cdot (1 + \xi_M) + \xi_A] = \left(E[y^{\text{out}}]\right)^2 \cdot \sigma_M^2 + \sigma_A^2. \quad (4)$$

Как видно из этого уравнения, функция активации не влияет на зависимость дисперсии от среднего выходного сигнала. В случае аддитивного шума дисперсия не зависит от выходного сигнала и определяется только дисперсией источника аддитивного шума. В случае мультипликативного шума зависимость дисперсии выходного сигнала от его среднего является квадратичной. Для оценки справедливости этих выводов далее рассмотрены результаты численного моделирования для двух функций активации с различной крутизной нелинейности.

Для численного моделирования дисперсия выходного сигнала вычисляется следующим образом. Искусственный нейрон в каждый момент времени t получает входной сигнал $x_i^{\text{in}}(t)$ и выдает выходной сигнал $y_i^{\text{out}}(t)$, включающий в себя влияние функции активации и шума.

Для того чтобы оценить его статистические характеристики, каждый входной сигнал $x_i^{\text{in}}(t)$ повторяется $K = 1000$ раз. В результате для каждого t получается набор из K значений $y_i^{\text{out}}(k, t)$, по которому вычисляется среднее и дисперсия: $\mu[y_i^{\text{out}}(t)]$ и $\sigma^2[y_i^{\text{out}}(t)]$. В дальнейшем такая же техника будет использоваться для оценки зашумленности выходного сигнала эхо-сети.

1.3. Один нейрон с функцией активации \tanh . Рассмотрим особенности влияния шумового воздействия на один нейрон с функцией активации «гиперболический тангенс». В качестве входного сигнала используем $T = 200$ случайных значений из диапазона от -1 до 1 . На рис. 1, *a* приведен график функции $f(x) = \tanh \alpha x$ при $\alpha = 1$.

Для того чтобы оценить влияние шума на нейрон, на рис. 1, *b* приведены значения дисперсии в зависимости от среднего значения выходного сигнала. Зеленым цветом показаны зависимости, полученные для аддитивного шума с дисперсией $\sigma_A^2 = 10^{-2}$. Оранжевым цветом показаны зависимости, полученные для мультипликативного шума с $\sigma_M^2 = 10^{-2}$, а синим — смешанный

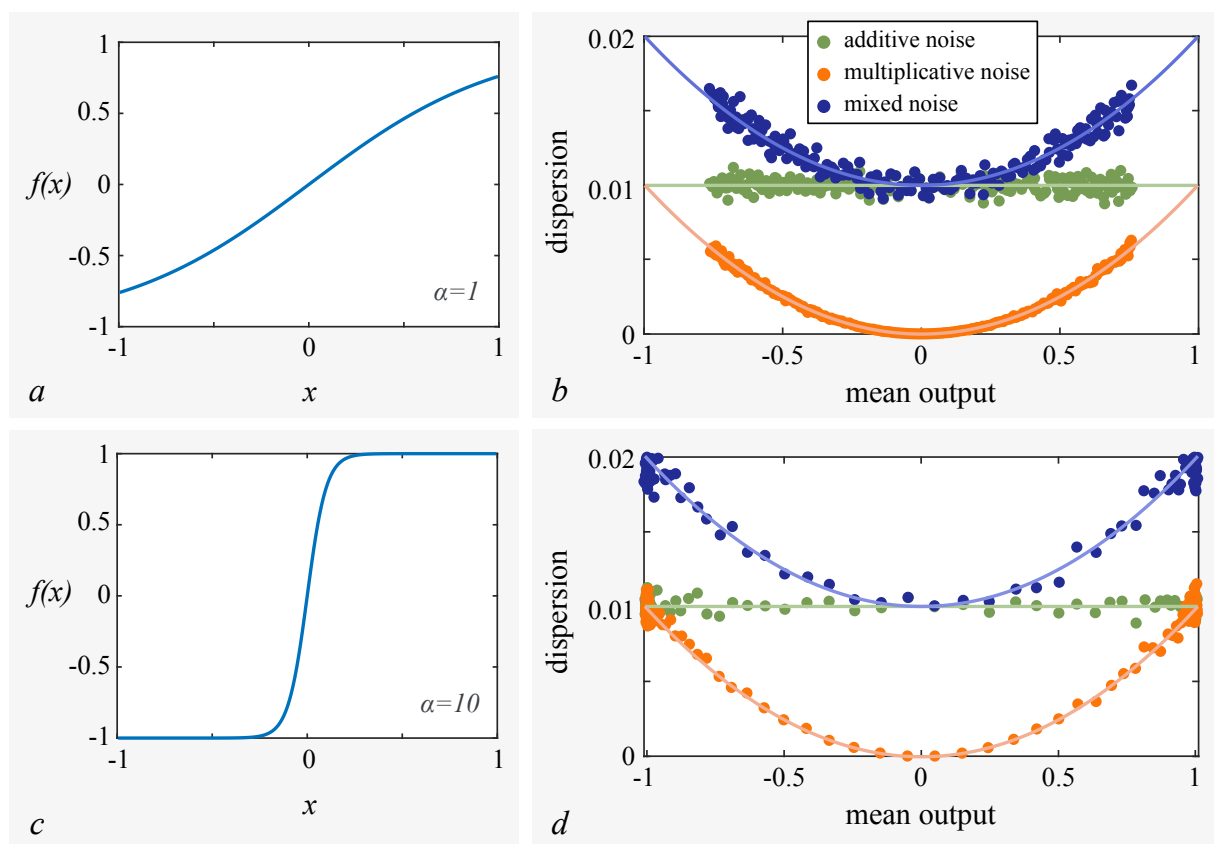


Рис. 1. Влияние шума на один нейрон с функцией активации $f(x) = \tanh \alpha x$. Входной сигнал нейрона — случайные числа из диапазона от -1 до 1 . Нелинейная функция с параметром $\alpha = 1$ (*a*) использовалась для получения зависимости дисперсии от среднего выходного сигнала (*b*) для аддитивного шума (зеленые точки), мультипликативного шума (оранжевые точки) и смешанного шума (синие точки) с дисперсиями $\sigma_A^2 = \sigma_M^2 = 10^{-2}$. Аналогично построены фрагменты *c, d* для параметра $\alpha = 10$. На фрагментах *b, d* при помощи линий соответствующих цветов показана теоретическая оценка уровня шума, полученная исходя из (4) (цвет онлайн)

Fig. 1. Noise impact on one neuron with activation function $f(x) = \tanh \alpha x$. Input signal contains random numbers from the range $[-1; 1]$. Nonlinear function with $\alpha = 1$ (*a*) was used to calculate the dependencies of dispersion on the mean output signal (*b*) for additive (green points), multiplicative (orange points) and mixed noise (blue points) with dispersions $\sigma_A^2 = \sigma_M^2 = 10^{-2}$. Panels *c, d* were prepared in the same manner for parameter $\alpha = 10$. Panels *b, d* contain analytical estimation of the noise level based on Eq. (4) shown by the lines of corresponding colors (color online)

шум, совмещающий аддитивный и мультипликативный шумы. Выбранные характеристики шума соответствуют шумам, наблюдаемым в ИНС, реализованным в фотонном эксперименте [15]. Линиями этих же цветов показаны зависимости, полученные исходя из уравнения (4).

Несмотря на то, что функция гиперболического тангенса может принимать значения из диапазона от -1 до 1 , на рис. 1, *b* диапазон средних выходных значений гораздо меньше. Это объясняется следующей особенностью. Если диапазон значений переменной x составляет от -1 до 1 при значении $\alpha = 1$, тогда этому соответствует диапазон значений $f(x)$ от -0.76 до 0.76 . Таким образом, из-за небольшого коэффициента крутизны α диапазон принимаемых значений может быть сильно уменьшен. В остальном сам вид зависимостей дисперсии, а также распределение точек по ним, очень похожи на то, что нами уже было получено для линейных нейронов [22]. Зависимости, полученные для мультипликативного шума, совмещают в себе уровни дисперсий, полученных для аддитивного и мультипликативного шума по отдельности.

Рассмотрим, что происходит с выходным сигналом зашумленного нейрона при увеличении крутизны нелинейности. На рис. 1, *c* приведен график гиперболического тангенса при $\alpha = 10$, что сопровождается существенным изменением крутизны нелинейности по сравнению с графиком на рис. 1, *a*. Как видно из графиков рис. 1, *d*, общий уровень дисперсии остается примерно тем же, как и для $\alpha = 1$, их небольшое изменение вызвано тем, что теперь $f(x)$ простирается уже на весь диапазон значений от -1 до 1 . Строго говоря, это должна быть та же самая зависимость, описываемая уравнением (4). Основной особенностью рис. 1, *d* является фокусирование точек. Если при $\alpha = 1$ точки были однородно распределены вдоль всего диапазона принимаемых значений, то теперь при $\alpha = 10$ точки сфокусированы около средних значений выходного сигнала -1 и 1 .

Свойство фокусирования точек можно объяснить следующим образом. Возвращаясь к графику рис. 1, *c*, можно отметить, что из-за ярко выраженной нелинейности и особенностей класса функций сигмоид большинство отрицательных значений x соответствуют значениям $f(x) \approx -1$, а большинство положительных значений x — значениям $f(x) \approx 1$. По этой причине для ярко выраженной нелинейности точки начинают фокусироваться около двух значений. Далее мы будем называть их *точками фокусирования*.

1.4. Один нейрон с функцией активацией «сигмоида». Рассмотрим, что происходит при другой нелинейности — сигмоиде. Соответствующие графики приведены на рис. 2. Так же, как и для гиперболического тангенса здесь рассматриваются два значения коэффициента крутизны нелинейности $\alpha = 1$ (рис. 2, *a, b*) и $\alpha = 10$ (рис. 2, *c, d*). Сравнивая сами функции активации на рис. 1 и 2, можно заметить, что в случае сигмоиды диапазон принимаемых значений изменяется до $(0; 1)$. Этим объясняется то, что дисперсия теперь имеет несколько другой вид. Однако, если сравнить эти зависимости с тем, что было получено для гиперболического тангенса, то видно, что в случае сигмоиды мы видим только правую часть зависимостей, полученных для гиперболического тангенса. Общий уровень шума остается примерно тем же, и при больших α снова появляются точки фокусирования. Если для гиперболического тангенса точки фокусирования были -1 и 1 , то для сигмоиды это 0 и 1 .

Стоит обратить внимание, что если одна из точек фокусирования равна нулю, то половина точек дисперсии будет сфокусирована около значения 0 . В случае мультипликативного шума (оранжевые точки) получается, что половина выходных значений лишается шумового воздействия из-за умножения на 0 . Таким образом, в случае такого шумового воздействия лучше использовать функцию активации типа сигмоиды, так как половина выходного сигнала перестает испытывать шумовое воздействие. Для аддитивного шума это не работает, а для мультипликативного и смешанного шумов такая стратегия может быть успешно применена. Это подтверждают графики рис. 2, *b, d*.

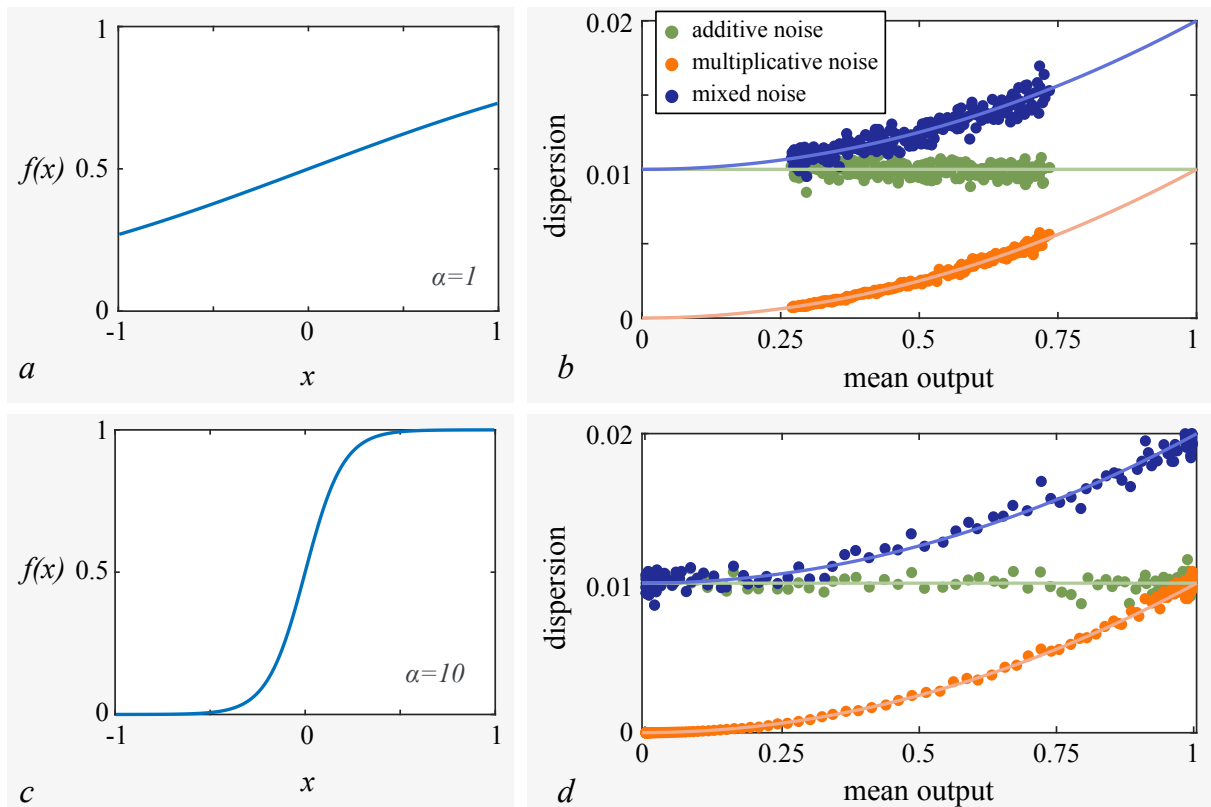


Рис. 2. Влияние шума на один нейрон с функцией активации $f(x) = 1/(1 + e^{-\alpha x})$. Входной сигнал нейрона — случайные числа из диапазона от -1 до 1 . Нелинейная функция с параметром $\alpha = 1$ (a) использовалась для получения зависимости дисперсии от среднего выходного сигнала (b) для аддитивного шума (зеленые точки), мультипликативного шума (оранжевые точки) и смешанного шума (синие точки) с дисперсиями $\sigma_A^2 = \sigma_M^2 = 10^{-2}$. Аналогично построены фрагменты c, d для параметра $\alpha = 10$. На фрагментах b, d при помощи линий соответствующих цветов показана теоретическая оценка уровня шума, полученная исходя из (4) (цвет онлайн)

Fig. 2. Noise impact on one neuron with activation function $f(x) = 1/(1 + e^{-\alpha x})$. Input signal contains random numbers from the range $[-1; 1]$. Nonlinear function with $\alpha = 1$ (a) was used to calculate the dependencies of dispersion on the mean output signal (b) for additive (green points), multiplicative (orange points) and mixed noise (blue points) with dispersions $\sigma_A^2 = \sigma_M^2 = 10^{-2}$. Panels c, d were prepared in the same manner for parameter $\alpha = 10$. Panels b, d contain analytical estimation of the noise level based on Eq. (4) shown by the lines of corresponding colors (color online)

Все перечисленные выше результаты подтверждаются аналитической оценкой зашумленности выходного сигнала (4).

Рассмотрим, как ведет себя дисперсия при изменении параметра крутизны нелинейности α и входного сигнала x^{in} (рис. 3). На рисунке цветом показана вариация в зависимости от параметра α и входного сигнала. В случае обеих функций активации уровень аддитивного шума не зависит ни от входного сигнала, ни от крутизны нелинейности (рис. 3 a, d).

Как видно из рис. 3, b, в случае гиперболического тангенса наименьшей дисперсии можно добиться при работе с малыми значениями $|\alpha|$ или малыми $|x^{\text{in}}|$. Однако при этом функция активации близка к линейной, и для некоторых задач это не применимо. Если увеличивать $|\alpha|$, то общий уровень дисперсии растет. Как видно из рис. 3, e, наиболее эффективной функцией активации с точки зрения подавления мультипликативного шума является нелинейность вида $f(x) = 1/(1 + e^{-\alpha x})$. Свойство уменьшения шума при малых $|\alpha|$ и $|x^{\text{in}}|$ справедливо и для этой функции активации, однако к этому еще добавляется свойство больших $|\alpha|$, которое заключается

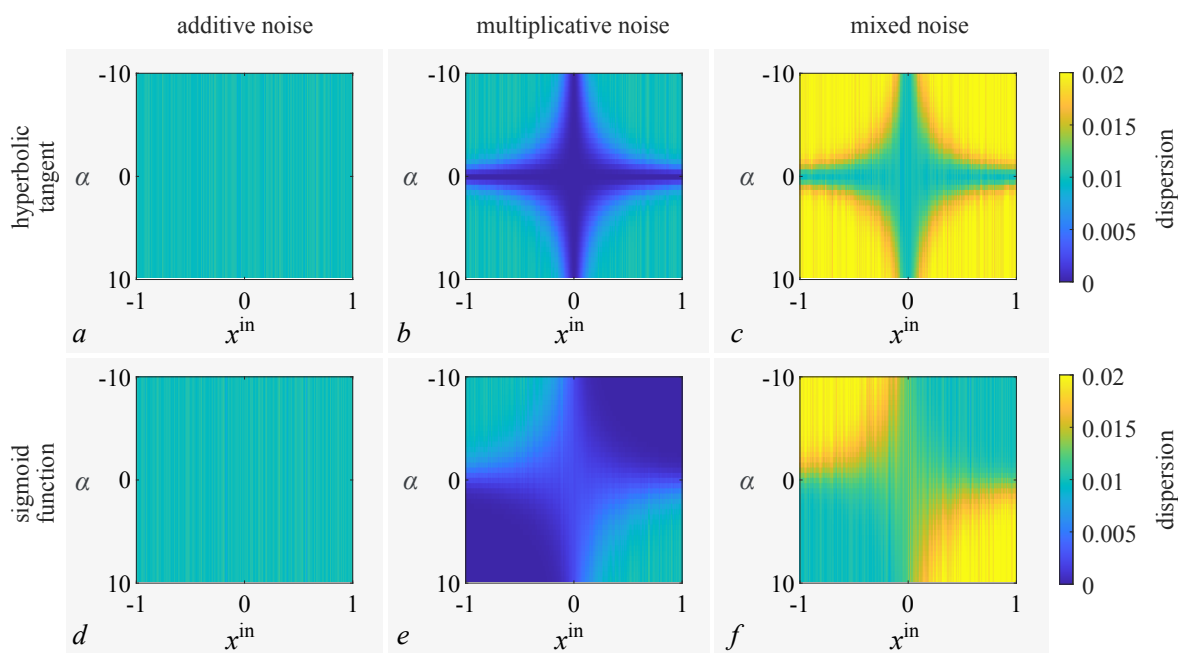


Рис. 3. Влияние входного сигнала x^{in} и параметра α на дисперсию выходного сигнала одного нейрона с функцией активации $f(x) = \tanh \alpha x$ (a-c) и функцией активации $f(x) = 1/(1 + e^{-\alpha x})$ (d-f). Типы шумов: a, d – аддитивный шум, b, e – мультипликативный шум, c, f – смешанный шум. Остальные параметры: $\sigma_A^2 = \sigma_M^2 = 10^{-2}$ (цвет онлайн)

Fig. 3. The influence of input signal x^{in} and parameter α on dispersion of the output signal of one neuron with activation function $f(x) = \tanh \alpha x$ (a-c) and activation function $f(x) = 1/(1 + e^{-\alpha x})$ (d-f). Noise types: a, d – additive noise, b, e – multiplicative noise, c, f – mixed noise. Other parameters: $\sigma_A^2 = \sigma_M^2 = 10^{-2}$ (color online)

в том, что при высокой крутизне нелинейности половина точек дисперсии фокусируется около 0, обнуляя при этом шумовое воздействие. Эта нелинейность оптимальна для мультипликативного и смешанного шума.

Все выводы, сделанные для мультипликативного шума, справедливы и для смешанного шума (рис. 3, c, f). Качественный вид дисперсии и его зависимость от входного сигнала и параметра α для обоих типов шумов одинаков. Количественно уровень дисперсии является суммой дисперсий, полученных для аддитивного и мультипликативного шума по отдельности.

2. Рекуррентные нейронные сети

2.1. Основные определения и составляющие сети. Существует большое количество разных типов нейронных сетей. Их топология и тип искусственных нейронов существенно зависит от решаемой задачи. Для решения задач прогнозирования или распознавания во времени, как правило, применяются рекуррентные сети. В данной работе в качестве примера таких сетей будет рассмотрена эхо-сеть, схема которой приведена на рис. 4.

Эхо-сеть характеризуется наличием входного слоя, отвечающего за получение и трансформацию входного сигнала. Далее этот сигнал передается в скрытый слой, который называется резервуаром. Резервуар состоит из N нейронов, которые связаны как с входными нейронами, которые передают входной сигнал в момент времени t , так и с тем же резервуаром, принимая сигнал резервуара в предыдущий момент времени ($t - 1$). Связь с входным слоем осуществляется при помощи матрицы связи \mathbf{W}^{in} . Если входной слой состоит из одного нейрона, то размер этой

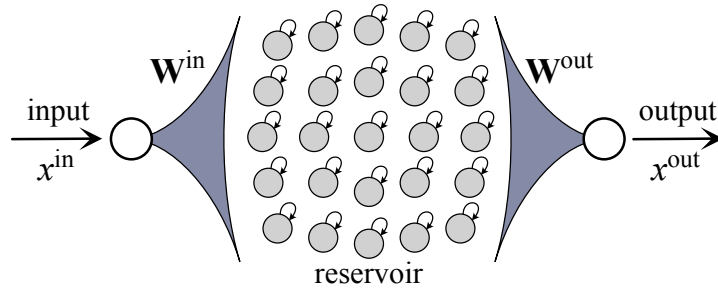


Рис. 4. Схема рассматриваемой эхо-сети. Белым цветом показаны входной и выходной нейроны без шумового воздействия. Серым показаны нейроны резервуара с нелинейной функцией активации, находящиеся под шумовым воздействием

Fig. 4. Schematic representation of considered echo state network (ESN). Input and output neurons without noise impact are shown by white color. The neurons of reservoir with nonlinear activation function receiving noise influence are colored in gray

матрицы составляет $(1 \times N)$. Матрица связи внутри резервуара \mathbf{W}^{res} имеет размер $(N \times N)$. Тогда состояние нейронов в момент времени t можно задать при помощи следующего уравнения:

$$\mathbf{x}_t^{\text{res}} = f\left(\beta x_t^{\text{in}} \cdot \mathbf{W}^{\text{in}} + \gamma \mathbf{y}_{t-1}^{\text{res}} \cdot \mathbf{W}^{\text{res}}\right); \quad \mathbf{y}_t^{\text{res}} = \hat{\mathbf{N}} \mathbf{x}_t^{\text{res}}, \quad (5)$$

где индексы t и $t - 1$ обозначают моменты времени, функция $f(\cdot)$ – функция активации. Жирным шрифтом обозначены \mathbf{x}^{res} и \mathbf{y}^{res} , которые являются векторами $(1 \times N)$, описывающими состояние N нейронов резервуара. В дальнейшем значение $N = 100$ не будет изменяться. Было показано, что описанная выше эхо-сеть позволяет предсказывать квазипериодическую и хаотическую динамику [25, 26].

Параметры β и γ отвечают за вклад входного сигнала и «памяти» резервуара, соответственно. В уже обученной сети нет возможности изменять матрицы связи \mathbf{W}^{in} и \mathbf{W}^{res} , и поэтому невозможно контролировать, что в большей степени влияет на состояние резервуара: входной сигнал или предыдущие состояния резервуара. В данной работе основной задачей является установление общих особенностей влияния шума, поэтому планируется рассмотреть, как сильно свойство памяти влияет на выходной сигнал путем введения дополнительного параметра γ . Однако, если изменять только параметр γ , то это приведет к увеличению или уменьшению амплитуды выходного сигнала. Для того чтобы сохранить диапазон выходных значений, был добавлен параметр β , управляющий вкладом входного сигнала, с условием $\beta + \gamma = 1$. Похожий подход использовался в работе [27].

В случае эхо-сети количество нейронов в выходном слое, как правило, совпадает с количеством нейронов во входном слое. Это делается для того, чтобы нейронную сеть можно было замкнуть саму на себя для решения задач прогнозирования, когда выход нейронной сети подается на ее же вход. Таким образом, как видно из рис. 4, выходной слой состоит из одного нейрона. Выходной сигнал этого нейрона, как и выходной сигнал всей сети задается как

$$\mathbf{x}_t^{\text{out}} = \mathbf{y}_t^{\text{res}} \cdot \mathbf{W}^{\text{out}}, \quad (6)$$

где матрица \mathbf{W}^{out} размером $(N \times 1)$ отвечает за связь резервуара с выходным слоем.

В данной статье нас в первую очередь интересует вопрос накопления шума внутри резервуара. Поэтому шумовое воздействие подается только в нейроны резервуара (см. (5)). Таким образом, входной и выходной нейрон являются линейными без шумового воздействия, поэтому на рис. 4 они обозначены белым цветом, а нейроны резервуара – серым. Входная матрица состоит из N элементов, равных единице, то есть каждый нейрон резервуара получает один и тот же входной сигнал. Для того чтобы исключить влияние матриц связи на статистику, выходная матрица будет состоять из значений $1/N$.

В эхо-сетях матрица связи внутри резервуара задается один раз и не меняется в процессе обучения. Как правило, в процессе обучения изменяется только матрица, связывающая резервуар с выходным слоем [28]. В следующем разделе рассмотрены и описаны основные типы матриц \mathbf{W}^{res} , которые часто используются в эхо-сетях.

2.2. Матрицы связи внутри резервуара. В данной статье рассмотрены два основных типа матриц связи \mathbf{W}^{res} .

Один из них — однородная матрица связи (uniform connection matrix), в которой элементы равны $1/N$. Как уже было показано в наших предыдущих работах [23, 24], с точки зрения распространения шумового воздействия это аналогично матрице, состоящей из случайных значений со средним значением $1/N$.

С другой стороны, в рекуррентных нейронных сетях матрицы связи внутри резервуара иногда имеют особый вид, когда все элементы задаются равными нулю, кроме элементов главной диагонали и некоторых параллельных им [15, 28]. Такие матрицы часто называют ленточными (band matrix). При такой конфигурации элементы главной диагонали оказываются больше остальных элементов, то есть наибольший вклад в состояние i -го нейрона резервуара вносит этот же i -й элемент в предыдущий момент времени. Для того чтобы воссоздать похожую ситуацию, введем дополнительный коэффициент рассеяния ζ , при помощи которого будем задавать ширину полосы, то есть сколько элементов слева и справа от главной диагонали не равны нулю. Для сравнения шумового воздействия необходимо сохранить общий диапазон значений $y_{t-1}^{\text{res}} \cdot \mathbf{W}^{\text{res}}$, то есть чтобы суммы элементов каждой строки и столбца были равны единице. Чтобы задать матрицу, удовлетворяющую описанным выше условиям, воспользуемся Гауссовой функцией. Тогда ненулевые элементы диагональной матрицы с коэффициентом рассеяния ζ будут задаваться следующим образом:

$$W_{k,i}^{\text{res}} = \frac{e^{-(k/\zeta^2)}}{\sum_{j=i-\zeta}^{i+\zeta} e^{-(j/\zeta^2)}}, \quad k \in [i - \zeta; i + \zeta]. \quad (7)$$

Примеры заданных таким образом матриц приведены на рис. 5 для коэффициентов рассеяния $\zeta = 2$ (a) и $\zeta = 20$ (b), соответствующих ширине полосы ленточной матрицы 4 и 40.

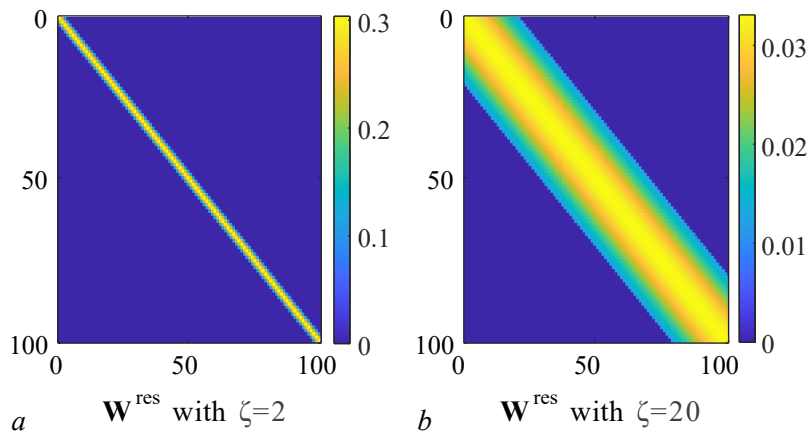


Рис. 5. Примеры рассматриваемых ленточных матриц, заданных при помощи (7) с коэффициентами рассеяния $\zeta = 2$ (a) и $\zeta = 20$ (b), которые соответствуют ширине полосы ленточной матрицы 4 и 40 (цвет онлайн)

Fig. 5. Examples of considered band matrices obtained using (7) with blurring coefficient $\zeta = 2$ (a) and $\zeta = 20$ (b). These values correspond to bandwidth of a band matrix equal to 4 and 40 (color online)

3. Влияние матрицы связи \mathbf{W}^{out}

Рассмотрим, как влияет шумовое воздействие внутри резервуара на статистические характеристики выходного сигнала эхо-сети. Ранее был рассмотрен один нейрон, и для оценки его выходного сигнала использовалась дисперсия.

Например, на рис. 6, *a* приведена зависимость дисперсии выходного сигнала эхо-сети x_t^{out} от среднего выходного значения. Одновременно по тем же данным был построен график зависимости дисперсии от времени (рис. 6, *b*) для эхо-сети, в которой входной сигнал меняется во времени согласно графику рис. 6, *c*. Эти результаты были получены для резервуара с нейронами, у которых в качестве функции активации выступает гиперболический тангенс. Аналогично были построены графики рис. 6, *d-f* для функции активации $f(x) = 1/(1 + e^{-\alpha x})$.

Для того чтобы понять, как свойство памяти резервуара влияет на конечный результат, сначала рассмотрим случай, когда $\gamma = 0$. Это соответствует ситуации, когда нейроны резервуара лишены свойства памяти, и получают только входной сигнал в момент времени t . Эти результаты приведены на рис. 6 для обеих функций активации. Сравнивая этот график с результатами, полученными для одного нейрона, можно заметить, что формы зависимостей совпадают, но количественно порядок дисперсии для сети изменяется с 10^{-2} до 10^{-4} . Падение общего уровня дисперсии до 10^{-4} вызвано выходной матрицей связи с элементами $1/N$. Выходной сигнал задается следующим образом:

$$x^{\text{out}} = \sum_{i=1}^N y_i^{\text{res}} \cdot W_i^{\text{out}} = \frac{1}{N} \sum_{i=1}^N y_i^{\text{res}}. \quad (8)$$

По правилам вычисления вариации, вариация суммы некоррелированных независимых случайных величин вычисляется как сумма вариаций этих величин: $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$.

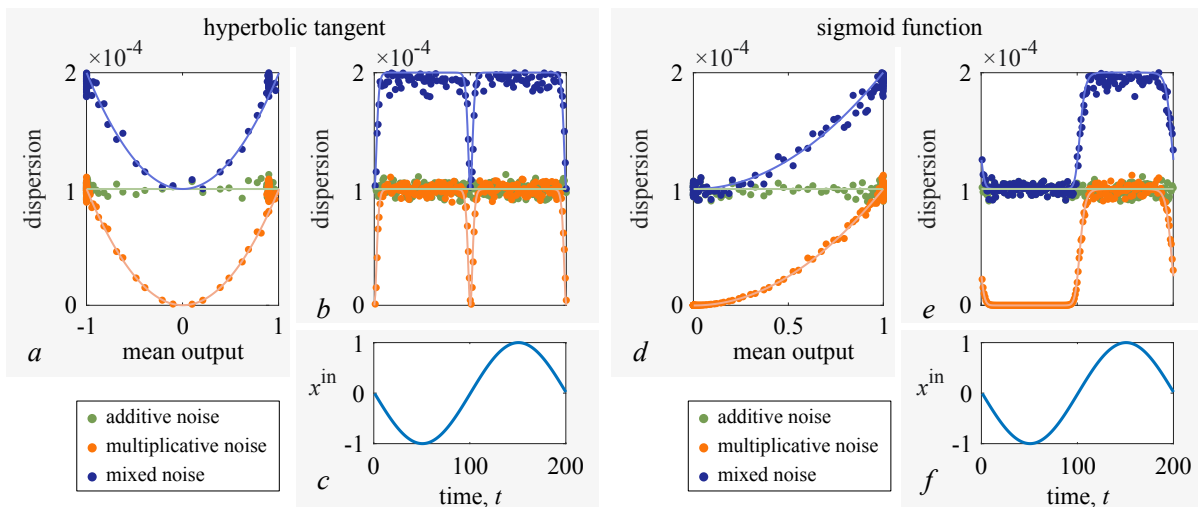


Рис. 6. Зависимость дисперсии выходного сигнала от среднего выходного сигнала (*a, d*), зависимость дисперсии (*b, e*) и входного (*c, f*) сигнала от времени для эхо-сети с однородной матрицей связи \mathbf{W}^{res} . Фрагменты *a-c* построены для сети с функцией активации $f(x) = \tanh \alpha x$, а фрагменты *d-f* — для функции $f(x) = 1/(1 + e^{-\alpha x})$. Параметры: $\gamma = 0$, $\alpha = 10$, $\sigma_A^2 = \sigma_M^2 = 10^{-2}$. На фрагментах *a, b, d, e* при помощи линий соответствующих цветов показана теоретическая оценка уровня шума, полученная исходя из (4) (цвет онлайн)

Fig. 6. Dependencies of dispersion of the output ESN signal on mean output signal (*a, d*), and how this dispersion (*b, e*) and input signal (*c, f*) changes in time in ESN with uniform connection matrix \mathbf{W}^{res} . Panels *a-c* were prepared for ESN with activation function $f(x) = \tanh \alpha x$, while panels *d-f* correspond to function $f(x) = 1/(1 + e^{-\alpha x})$. Parameters: $\gamma = 0$, $\alpha = 10$, $\sigma_A^2 = \sigma_M^2 = 10^{-2}$. Panels *a, b, d, e* contain analytical estimation of the noise level based on Eq. (4) shown by the lines of corresponding colors (color online)

Если случайная величина умножается на некоторую константу, то результирующая вариация вычисляется как произведение квадрата этой константы и вариации случайной величины: $\text{Var}[c \cdot X] = c^2 \cdot \text{Var}[X]$. Тогда если на резервуар действует аддитивный шум, вариация выходного сигнала вычисляется как

$$\text{Var}[x^{\text{out}}] = \frac{1}{N^2} \sum_{i=1}^N \text{Var}[y_i^{\text{res}}] = \frac{1}{N} \cdot \sigma_A^2. \quad (9)$$

Для одного изолированного нейрона вариация составляла σ_A^2 . Таким образом, вариация выходного сигнала уменьшается в $N = 100$ раз. Аналогичные выводы можно сделать и для мультипликативного шума.

Конкретный вид матрицы связи \mathbf{W}^{out} напрямую зависит от решаемой задачи и процесса обучения. Однако с точки зрения распространения шума эта матрица может быть рассмотрена как набор случайных значений с некоторым средним значением и средним квадратическим значением. Подобный подход несмотря на упрощение был нами успешно применен для глубокой сети [23] и показал свою эффективность для обученной сети. В общем случае множитель $1/N$ не может быть вынесен за пределы суммы в формуле (9), и тогда дисперсия выходного сигнала может быть вычислена следующим образом:

$$\begin{aligned} \text{Var}[x^{\text{out}}] &= \sum_{i=1}^N \text{Var}[y_i^{\text{res}} \cdot W_i^{\text{out}}] = \sum_{i=1}^N (W_i^{\text{out}})^2 \cdot \text{Var}[y_i^{\text{res}}] = \\ &= \sum_{i=1}^N (W_i^{\text{out}})^2 \cdot (\sigma_A^2 + \sigma_M^2 \cdot (E[y_i^{\text{res}}])^2) = \sigma_A^2 \cdot N \eta^2(\mathbf{W}^{\text{out}}) + \sigma_M^2 \cdot \sum_{i=1}^N (W_i^{\text{out}})^2 (E[y_i^{\text{res}}])^2, \end{aligned} \quad (10)$$

где $\eta^2(\mathbf{W}^{\text{out}})$ — среднее квадратическое значение матрицы \mathbf{W}^{out} , которое вычисляется как $\eta^2(\mathbf{W}^{\text{out}}) = \frac{1}{N} \sum_{i=1}^N (W_i^{\text{out}})^2$. Таким образом, если на резервуар действует только аддитивный шум, дисперсия выходного сигнала в общем случае выглядит как $N \cdot \eta^2(\mathbf{W}^{\text{out}}) \cdot \sigma_A^2$. Если матрица \mathbf{W}^{out} состоит из элементов $1/N$, то это значение может быть сведено к формуле (9).

В следующих разделах главным образом рассмотрено влияние матриц связи *внутри* резервуара на полученные результаты. Для того чтобы выходная матрица связи \mathbf{W}^{out} не искажала полученные результаты, мы будем задавать ее однородной и состоящей из элементов $1/N$.

4. Распространение шума в эхо-сети с однородной матрицей \mathbf{W}^{res}

В этом разделе рассматривается в том числе и влияние параметра γ на накопление шума. Параметр γ отвечает за вклад предыдущих состояний резервуара в новый выходной сигнал. По сути, этот параметр отвечает за «память» эхо-сети. Это говорит о том, что теперь недостаточно использовать в качестве входного сигнала случайную последовательность чисел, так как форма входного сигнала также может иметь большое значение. По этой причине теперь мы будем приводить зависимость дисперсии не только от среднего выходного сигнала, но и от времени t .

Рассмотрим, что происходит с дисперсией, если нейроны резервуара принимают на вход одновременно и сигнал от входного нейрона, и предыдущий сигнал резервуара. На рис. 7 приведены результаты расчета дисперсии для случая $\gamma = 0.5$ для обоих типов функции активации. Из сравнения рис. 6 и 7 видно, что общий вид зависимостей меняется.

В случае $\gamma = 0$ вид зависимости дисперсии от среднего выходного сигнала совпадает с соответствующими графиками для одного нейрона. Что же касается зависимостей дисперсии от времени, то их можно охарактеризовать следующим образом. Аддитивный шум приводит к постоянному уровню дисперсии независимо от времени для обеих функций активации, а влияние мультипликативного шума существенно зависит от входного сигнала. Дисперсия находится на том же уровне, что и для аддитивного шума, если входной сигнал соответствует выходному

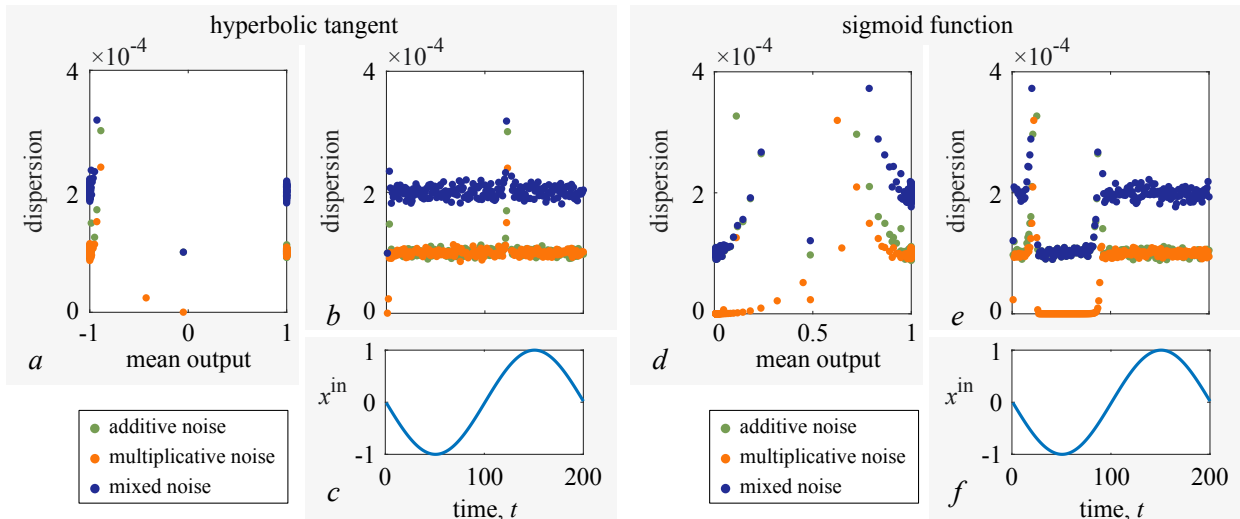


Рис. 7. Зависимость дисперсии выходного сигнала эхо-сети от его среднего и времени. Рисунок подготовлен аналогично рис. 6, но для параметра $\gamma = 0.5$ (цвет онлайн)

Fig. 7. Dispersion of the output ESN signal depending on its mean and time. This figure was prepared in the same manner as Fig. 6 but for parameter $\gamma = 0.5$ (color online)

сигналу, близкому к 1 или -1 . Например, в случае гиперболического тангенса (см. рис. 6, b) левая половина оранжевой зависимости соответствует отрицательному входному сигналу (см. рис. 6, c), а правая — положительному. Аналогично для сигмоиды: положительные значения входного сигнала приводят к тому же уровню дисперсии (см. рис. 6, e). Если же значения входного сигнала приводят к нулевому выходному сигналу, то уровень дисперсии мультипликативного шума падает до нулевого уровня. Для выбранной формы входного сигнала в сочетании с функцией активации \tanh (см. рис. 6, b) это приводит к резкому падению уровня дисперсии около значений $t = 0, 100, 200$, когда входной сигнал равен нулю. В случае сигмоиды нулевому выходному сигналу соответствует не одно значение входного сигнала, а целый диапазон отрицательных значений входного сигнала, поэтому дисперсия для мультипликативного шума в этом случае становится близкой к нулю для времени $t \in [0; 100]$ (см. рис. 6, e). Кроме этого, для обеих нелинейностей, как и для одного нейрона, характерно фокусирование точек.

В случае $\gamma = 0.5$ это свойство фокусирования проявляется сильнее (рис. 7, a, d). Увеличение «памяти» эхо-сети приводит к росту дисперсии промежуточных точек. По этой причине, несмотря на то, что большинство точек зависимостей находится на том же уровне, что и для $\gamma = 0$, средний уровень дисперсии растет с увеличением γ . В случае функции активации «сигмоида» количество точек, соответствующих нулевому уровню дисперсии, уменьшается при увеличении параметра γ , и при $\gamma \rightarrow 1$ все точки зависимости дисперсии становятся на один уровень, как это было для гиперболического тангенса. Падение дисперсии в случае мультипликативного шума почти до нуля объясняется тем, что для части отрицательного входного сигнала выходной сигнал равен нулю, как это было и для одного нейрона и для $\gamma = 0$.

5. Распространение шума в эхо-сети с ленточной матрицей \mathbf{W}^{res}

В случае, если рассматривается эхо-сеть с диагональной матрицей, зависимости дисперсии от среднего выходного сигнала и времени качественно остаются такими же, как для однородной матрицы. Существуют лишь незначительные количественные отличия.

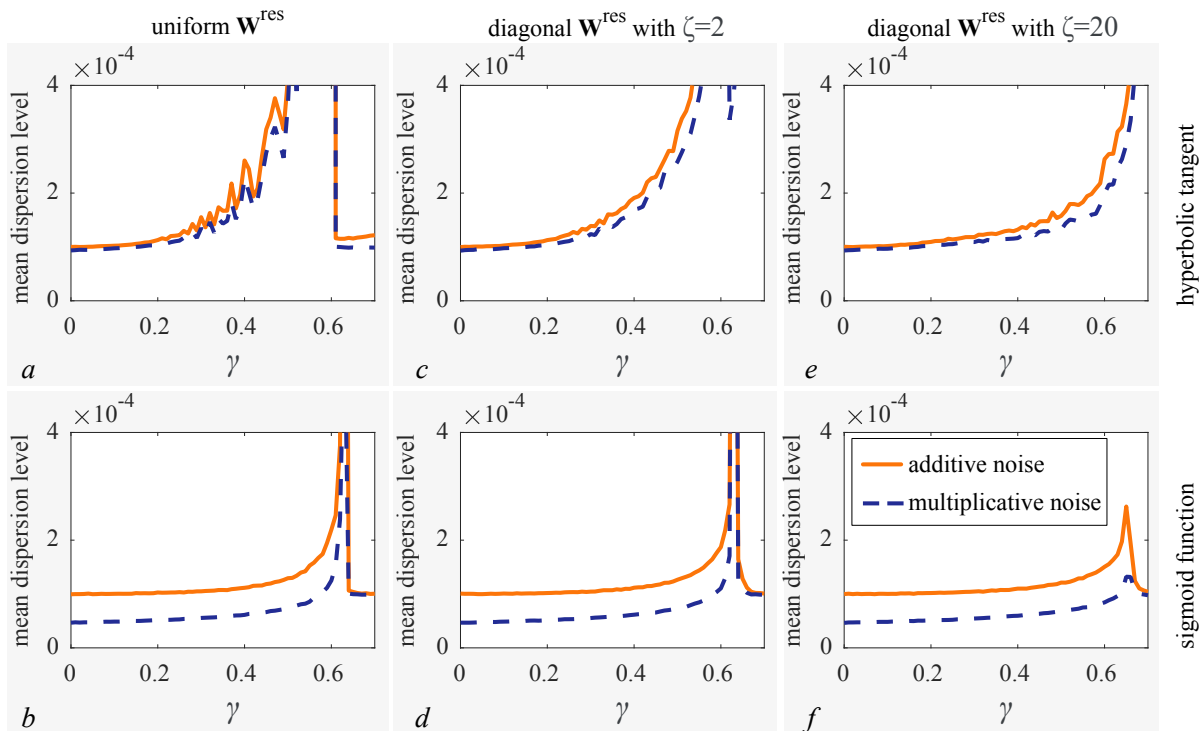


Рис. 8. Зависимость среднего уровня дисперсии от параметра γ . Дисперсия рассчитывалась по выходному сигналу эхо-сети с однородной матрицей \mathbf{W}^{res} (a, b), ленточной матрицей \mathbf{W}^{res} с $\zeta = 2$ (c, d), ленточной матрицей \mathbf{W}^{res} с $\zeta = 20$ (e, f). Верхние фрагменты подготовлены для функции активации $f(x) = \tanh \alpha x$, а нижние – для функции $f(x) = 1/(1 + e^{-\alpha x})$. Остальные параметры: $\alpha = 10$, $\sigma_A^2 = \sigma_M^2 = 10^{-2}$ (цвет онлайн)

Fig. 8. Dependencies of mean dispersion level on parameter γ . Dispersion is calculated by the output signal of ESN with uniform matrix \mathbf{W}^{res} (a, b), band matrix \mathbf{W}^{res} with $\zeta = 2$ (c, d), band matrix \mathbf{W}^{res} with $\zeta = 20$ (e, f). Top panels were prepared for activation function $f(x) = \tanh \alpha x$, while bottom panels correspond to function $f(x) = 1/(1 + e^{-\alpha x})$. Parameters: $\alpha = 10$, $\sigma_A^2 = \sigma_M^2 = 10^{-2}$ (color online)

В случае гиперболического тангенса дисперсия находится на примерно одном уровне 10^{-4} независимо от входного сигнала для аддитивного и мультипликативного шума. В случае сигмоиды и аддитивного шума дисперсия находится примерно на том же уровне. Если функция активации задается как сигмоида и на нейроны резервуара действует мультипликативный или смешанный шум, тогда для значений дисперсии характерно падение почти до нулевого уровня при отрицательном входном сигнале. При увеличении параметра γ количество таких точек уменьшается, и при $\gamma \rightarrow 1$ точки зависимости дисперсии тоже стремятся к одному уровню. Таким образом, все выводы про качественный вид зависимостей совпадают с выводами, описанными в разделе 4. Поэтому далее рассмотрим количественные аспекты.

Рассмотрим, что происходит со средним уровнем дисперсии при изменении параметра γ . На рис. 8 приведена зависимость дисперсии выходного сигнала от степени влияния запаздывающей обратной связи резервуара γ для трех матриц связи \mathbf{W}^{res} : однородная матрица связи, рассмотренная в предыдущем разделе (рис. 8, a, b), и диагональные матрицы с коэффициентами рассеяния $\zeta = 2$ (рис. 8, c, d) и $\zeta = 20$ (рис. 8, e, f). Эти зависимости были построены для обоих типов функции активации: гиперболический тангенс (верхние фрагменты) и сигмоида (нижние фрагменты).

Параметр γ отвечает за вклад памяти резервуара. При $\gamma > 0.5$ вклад входного сигнала уменьшается ($\beta < 0.5$), и больший вклад поступает со стороны резервуара с зашумленными нейронами в предыдущий момент времени. Так как источники шума независимые, это приводит

к существенному накоплению шума. По этой причине при $\gamma < 0.5$ общий уровень шума остается примерно на одном уровне, а при $\gamma > 0.5$ он начинает резко расти.

Как видно из рис. 8, при $\gamma \rightarrow 0.6$ на всех зависимостях наблюдается рост. Это связано с тем, что на зависимостях дисперсии от времени присутствуют некоторые выбросы (см. рис. 7). Их количество и амплитуда увеличивается с ростом γ , что сказывается на вычислении среднего уровня дисперсии. Строго говоря, независимо от параметра γ большинство точек зависимости дисперсии от времени принадлежит уровню 10^{-4} для аддитивного и мультипликативного шумов при функции активации, заданной как гиперболический тангенс. Такая же ситуация наблюдается, если функция активации задана как сигмоида и в резервуаре присутствуют только аддитивные шумы. Если в системе есть мультипликативные шумы, то правильнее использовать сигмоиду в качестве функции активации, так как средний уровень дисперсии выходного сигнала для нее ниже (синие графики на рис. 8, *b, d, f*). Сравнивая зависимости, изображенные на рис. 8, можно также сделать вывод, что накопление выбросов и увеличение среднего уровня дисперсии происходит медленнее, если матрица резервуара \mathbf{W}^{res} задана как диагональная матрица с большим коэффициентом рассеяния $\zeta = 20$, нежели при $\zeta = 2$ или для однородной матрицы.

Для некоторых фрагментов рис. 8 характерно резкое падение среднего уровня дисперсии при $\gamma > 0.6$. Из-за этого может сложиться неверное впечатление, что оптимальным будет выбор $\gamma > 0.6$. Однако этот эффект вызван тем, что в системе наблюдается почти полное замещение входного сигнала. В связи с этим такой коэффициент γ используется крайне редко.

Заключение

В данной работе рассмотрено влияние шума на рекуррентные сети на примере эхо-сети, а также влияние нелинейности функции активации и типов матриц связи на распространение и накопление аддитивного, мультипликативного и смешанного шумов.

На примере функции сигмоиды вида $f(x) = 1/(1 + e^{-\alpha x})$ с одной из точек фокусирования около нуля мы показали, что такие функции активации хорошо подходят для подавления мультипликативного и смешанного шумов. Влияние аддитивного шума и дисперсия выходного сигнала не зависят от входного сигнала и типа функции активации.

Кроме того, было рассмотрено накопление шума в резервуаре эхо-сети для трех типов матриц связи внутри резервуара: однородная матрица, ленточная матрица с малым коэффициентом рассеяния и ленточная матрица с большим коэффициентом рассеяния. Для того чтобы показать влияние шума в чистом виде, были рассмотрены упрощенные модели основных типов матриц связи внутри резервуара.

Рассмотренная однородная матрица соответствует матрице, заданной случайным образом с некоторым средним значением. Еще одним типом матриц, которые часто используются в рекуррентных нейронных сетях, являются ленточные матрицы, в которых все элементы кроме элементов главной диагонали и некоторых элементов параллельных им, равны нулю. Похожие матрицы были воссозданы в этой статье, когда элементы матрицы задавались при помощи Гауссовой функции с коэффициентом рассеяния ζ , задающим, сколько ненулевых элементов слева и справа от главной диагонали не равны нулю. Нами было установлено, что эхо-сети с диагональной матрицей связи с большим коэффициентом рассеяния меньше накапливают шумы с увеличением влияния памяти сети, контролируемого параметром γ .

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444. DOI: 10.1038/nature14539.
2. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun. ACM. 2017;60(6):84–90. DOI: 10.1145/3065386.

3. Maturana D, Scherer S. VoxNet: A 3D Convolutional Neural Network for real-time object recognition. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 28 September 2015–02 October 2015, Hamburg, Germany. New York: IEEE; 2015. P. 922–928. DOI: 10.1109/IROS.2015.7353481.
4. Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. 26–31 May 2013, Vancouver, BC, Canada. New York: IEEE; 2013. P. 6645–6649. DOI: 10.1109/ICASSP.2013.6638947.
5. Kar S, Moura JMF. Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise. *IEEE Transactions on Signal Processing*. 2009;57(1):355–369. DOI: 10.1109/TSP.2008.2007111.
6. Mandic DP, Chambers JA. *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*. New York: Wiley; 2001. 304 p. DOI: 10.1002/047084535X.
7. Bailador G, Roggen D, Tröster G, Triviño G. Real time gesture recognition using continuous time recurrent neural networks. In: 2nd International ICST Conference on Body Area Networks. 11th–13th Jun 2007, Florence, Italy. ICST; 2007. 8 p. DOI: 10.4108/bodynets.2007.149.
8. Hasler J, Marr H. Finding a roadmap to achieve large neuromorphic hardware systems. *Frontiers in Neuroscience*. 2013;7:118. DOI: 10.3389/fnins.2013.00118.
9. Gupta S, Agrawal A, Gopalakrishnan K, Narayanan P. Deep learning with limited numerical precision. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. 6-11 July 2015, Lille, France. JMLR; 2015. P. 1737–1746.
10. Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. *Nature Reviews Physics*. 2021;3(6):422–440. DOI: 10.1038/s42254-021-00314-5.
11. Brunner D, Soriano MC, Mirasso CR, Fischer I. Parallel photonic information processing at gigabyte per second data rates using transient states. *Nature Communications*. 2013;4(1):1364. DOI: 10.1038/ncomms2368.
12. Tuma T, Pantazi A, Le Gallo M, Sebastian A, Eleftheriou E. Stochastic phase-change neurons. *Nature Nanotechnology*. 2016;11(8):693–699. DOI: 10.1038/nnano.2016.70.
13. Torrejon J, Riou M, Araujo FA, Tsunegi S, Khalsa G, Querlioz D, Bortolotti P, Cros V, Yakushiji K, Fukushima A, Kubota H, Yuasa S, Stiles MD, Grollier J. Neuromorphic computing with nanoscale spintronic oscillators. *Nature*. 2017;547(7664):428–431. DOI: 10.1038/nature23011.
14. Psaltis D, Brady D, Gu XG, Lin S. Holography in artificial neural networks. *Nature*. 1990;343(6256):325–330. DOI: 10.1038/343325a0.
15. Bueno J, Maktoobi S, Froehly L, Fischer I, Jacquot M, Larger L, Brunner D. Reinforcement learning in a large-scale photonic recurrent neural network. *Optica*. 2018;5(6):756–760. DOI: 10.1364/OPTICA.5.000756.
16. Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y, Jarrahi M, Ozcan A. All-optical machine learning using diffractive deep neural networks. *Science*. 2018;361(6406):1004–1008. DOI: 10.1126/science.aat8084.
17. Shen Y, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M, Sun X, Zhao S, Larochelle H, Englund D, Soljačić M. Deep learning with coherent nanophotonic circuits. *Nature Photonics*. 2017;11(93):441–446. DOI: 10.1038/nphoton.2017.93.
18. Tait AN, de Lima TF, Zhou E, Wu AX, Nahmias MA, Shastri BJ, Prucnal PR. Neuromorphic photonic networks using silicon photonic weight banks. *Scientific Reports*. 2017;7(1):7430. DOI: 10.1038/s41598-017-07754-z.
19. Moughames J, Porte X, Thiel M, Ulliac G, Larger L, Jacquot M, Kadic M, Brunner D. Three-dimensional waveguide interconnects for scalable integration of photonic neural networks. *Optica*. 2020;7(6):640–646. DOI: 10.1364/OPTICA.388205.
20. Dinc NU, Psaltis D, Brunner D. Optical neural networks: The 3D connection. *Photonics*. 2020;(104):34–38. DOI: 10.1051/photon/202010434.

21. Moughames J, Porte X, Larger L, Jacquot M, Kadic M, Brunner D. 3D printed multimode-splitters for photonic interconnects. *Opt. Mater. Express*. 2020;10(11):2952–2961. DOI: 10.1364/OME.402974.
22. Semenova N, Porte X, Andreoli L, Jacquot M, Larger L, Brunner D. Fundamental aspects of noise in analog-hardware neural networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2019;29(10):103128. DOI: 10.1063/1.5120824.
23. Semenova N, Larger L, Brunner D. Understanding and mitigating noise in trained deep neural networks. *Neural Networks*. 2022;146:151–160. DOI: 10.1016/j.neunet.2021.11.008.
24. Semenova N, Brunner D. Noise-mitigation strategies in physical feedforward neural networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2022;32(6):061106. DOI: 10.1063/5.0096637.
25. Jaeger H. Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the “echo state network” approach. GMD-Report 159. Bonn: German National Research Center for Information Technology; 2002. 48 p.
26. Prokhorov D. Echo state networks: appeal and challenges. In: *Proceedings. 2005 IEEE International Joint Conference on Neural Networks*. Vol. 3. 31 July 2005 – 04 August 2005, Montreal, QC, Canada. New York: IEEE; 2005. P. 1463–1466. DOI: 10.1109/IJCNN.2005.1556091.
27. Cerina L, Santambrogio MD, Franco G, Gallicchio C, Micheli A. EchoBay: Design and optimization of echo state networks under memory and time constraints. *ACM Transactions on Architecture and Code Optimization*. 2020;17(3):1–24. DOI: 10.1145/3404993.
28. Lukoševičius M, Jaeger H. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*. 2009;3(3):127–149. DOI: 10.1016/j.cosrev.2009.03.005.



Москвитин Виктор Максимович — родился в Саратове (2004). С сентября 2022 года проходит обучение в Институте физики Саратовского национального исследовательского государственного университета имени Н. Г. Чернышевского по профилю «Информационные технологии и компьютерное моделирование в радиофизике». Научные интересы — компьютерное моделирование, объектно-ориентированное программирование, анализ данных, нейронные сети, машинное обучение.

Россия, 410012 Саратов, ул. Астраханская, 83
 Саратовский национальный исследовательский
 государственный университет имени Н. Г. Чернышевского
 E-mail: vmmoskvitin@gmail.com



Семенова Надежда Игоревна — родилась в Саратове (1992). Окончила с отличием физический факультет Саратовского государственного университета по специальности «Радиофизика и электроника» (2014). Защитила диссертацию на соискание учёной степени кандидата физико-математических наук по специальности «Радиофизика» (2017, СГУ). Получила PhD степень по специальности «Оптика» в университете Бургундии Франш-Конте (Франция, 2021). С 2021 года работает на кафедре «Радиофизика и нелинейная динамика» Саратовского государственного университета имени Н. Г. Чернышевского в должности доцента и старшего научного сотрудника. Научные интересы — нейронаука, математическое моделирование, нейронные сети, машинное обучение, влияние шумов, пространственно-временные режимы в ансамблях и сетях, химерные состояния, уединенные состояния, запаздывающая обратная связь. Опубликовала свыше 40 научных статей по указанным направлениям.

Россия, 410012 Саратов, ул. Астраханская, 83
 Саратовский национальный исследовательский
 государственный университет имени Н. Г. Чернышевского
 E-mail: semenovani@sgu.ru
 ORCID: 0000-0002-9180-3030
 AuthorID (eLibrary.Ru): 850757