

# АНАЛИТИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ИНФОКОММУНИКАЦИОННЫХ ДАННЫХ

DOI: 10.36724/2072-8735-2025-19-8-29-37

Manuscript received 10 June 2025;  
Accepted 20 July 2025

**Толстых Виктор Николаевич,**  
Санкт-Петербургский государственный университет  
аэрокосмического приборостроения, Санкт-Петербург, Россия,  
[victor.n.tolstykh@gmail.com](mailto:victor.n.tolstykh@gmail.com)  
ORCID 0009-0003-0202-5500

*Работа выполнена при финансовой поддержке  
Министерства науки и высшего образования Российской  
Федерации, соглашение № FSRF-2023-0003,  
"Фундаментальные основы построения помехозащищенных  
систем космической и спутниковой связи, относительной  
навигации, технического зрения и аэрокосмического  
мониторинга"*

**Ершов Владимир Николаевич,**  
Санкт-Петербургский государственный университет  
аэрокосмического приборостроения, Санкт-Петербург, Россия

**Савинов Максим Сергеевич,**  
Санкт-Петербургский государственный университет  
аэрокосмического приборостроения, Санкт-Петербург, Россия

**Ключевые слова:** аппроксимация, экстраполяция,  
перцептрон, нейронные сети, монотонное  
преобразование, передаточная функция

В статье рассматривается вопрос о возможных расширениях семейств аппроксимирующих функций суперпозицией с некоторыми классами передаточных функций с целью улучшения качества аппроксимации непрерывных на отрезке и таблично заданных зависимостей. Как с точки зрения точности, так и с точки зрения скорости вычисления. Особое внимание уделяется возможности использования метода наименьших квадратов для таких расширений, поскольку этот метод даёт наиболее быстрое решение по сравнению с методами пошагового приближения. Исследуется вопрос о возможном изменении критерия сходимости функции ошибок для расширения возможности этого метода, что позволит добиться большей гибкости в решении этой задачи. Дополнительно, рассматриваются две практические задачи – это применимость метода наименьших квадратов к обучению искусственных нейронных сетей, а также к задачам экстраполяции и прогноза, к которым полиномы не очень пригодны. Для обоснования концепции передаточных функций формулируются теоремы с доказательством о линейной передаточной функции и полноте семейства обратных полиномов. Обсуждается вопрос применимости метода ортогональных многочленов, а также рациональных функций для этих целей. В качестве метода исследования предлагается дополнить аппроксимирующее семейство на примере полиномов строго монотонным преобразованием как передаточной функцией. Рассматриваются конкретные варианты таких передаточных функций, указывается для каких практических задач они могут быть полезны. Результатом является демонстрация явного преимущества полинома с передаточной функцией над полиномом без передаточной функции для аппроксимации конкретного вида зависимостей, часто фигурирующих в инфокоммуникационных задачах. Остальные результаты имеют теоретическое значение и могут быть использованы для задач прогноза и создания искусственных нейронных сетей новой архитектуры.

## Для цитирования:

Толстых В.Н., Ершов В.Н., Савинов М.С. Аналитическое представление инфокоммуникационных данных // Т-Comm: Телекоммуникации и транспорт. 2025. Том 19. №8. С. 29-37.

## For citation:

V.N. Tolstykh, V.N. Yershov, M.S. Savinov, "Analytical Representation of Infocommunication Data", T-Comm, 2025, vol. 19, no.8, pp. 29-37. (in Russian)

**Введение**

Аналитическое представление экспериментальных данных играет ключевую роль для исследования внутренних зависимостей в данных, а также для прогноза их поведения за границами известного. Наиболее часто для этих целей используют семейства полиномов<sup>1</sup>, а также ортогональные многочлены Чебышёва, Лежандра, Эрмита, Бернштейна и др. [1] со свободными коэффициентами в качестве параметров, в связи с тем, что их аппроксимирующие свойства хорошо изучены. А также в связи с тем, что эти коэффициенты могут быть определены достаточно просто с помощью хорошо изученного метода наименьших квадратов [2]. Не для всех непрерывных функций или дискретно заданных зависимостей такая аппроксимация подходит, поскольку теоретическая возможность и практическая реализация это не одно и то же.

В этой связи можно отметить требование на минимизацию необходимого количества коэффициентов и степень полинома для достижения заданной точности аппроксимации на обучающей выборке. Также имеют значение такие свойства как склонность моделей на основе полиномов к переобучению на дискретных данных и недостаточные или даже откровенно плохие экстраполирующие их свойства. Поэтому, иногда используют и другие аппроксимирующие семейства, в том числе такие, которые не совсем удовлетворяют теореме Вейерштрасса-Стоуна [3] об аппроксимации. То есть, подсемейства, плотные не на всём множестве непрерывных функций, а только на его подмножестве.

В работе предлагается расширить семейство классических полиномов или именных ортогональных полиномов, добавлением строго монотонных функций в качестве передаточных функций. Которые, с одной стороны, добавляют иногда полезную нелинейность в линейные формы, а, с другой стороны, позволяют использовать метод наименьших квадратов для более быстрой и качественной аппроксимации. Это даст возможность строить аналитические зависимости с новыми свойствами, что демонстрируется на примерах.

**Полиномы, ортогональные многочлены и другие аппроксимирующие семейства**

Пока рассмотрение применения передаточных функций ограничивалось полиномами в качестве аргумента. В то же время, как было сказано ранее, полиномы для практического применения – это не самое лучшее аппроксимирующее семейство. Математиками прошлого, частично указанными во введении, были разработаны более совершенные семейства, объединенные названием «ортогональные многочлены», которые компенсируют недостатки полиномов на отрезке. Ортогональными многочленами называется бесконечная последовательность частных многочленов возрастающей степени  $n$ , образующих ортогональный базис, по которому можно разложить любую непрерывную на отрезке  $(a, b)$  непрерывную зависимость. Пусть  $p^0(x), p^1(x), p^2(x), \dots, p^n(x), \dots$  – такая последовательность. На отрезке  $[-1, 1]$  она обладает свойством

$$\int_{-1}^1 w(x) p^i(x) p^j(x) dx = 0, \forall i, j \leq n,$$

которое позволяет говорить об ортогональности всего набора многочленов. Соответственно, эти многочлены образуют базис, по которому может быть разложена любая непрерывная на отрезке функция (или таблично заданная зависимость с возрастающей последовательностью аргументов  $\forall i, x_{i+1} > x_i$ )

$$f(x) = \sum_{k=1}^{\infty} a_k \cdot w(x) p^k(x), x \in (-1, 1).$$

Здесь  $a_k$  – это коэффициенты разложения,  $w(x)$  – финитная весовая функция, которая служит для компенсации больших значений на концах отрезка. Коэффициенты  $a_k$  находятся методом наименьших квадратов, поскольку входят в разложение линейно, образуя при этом линейную по коэффициентам форму. Таким образом, равномерное представление непрерывных функций ортогональными многочленами на отрезке возможно так же, как и полиномами. (Для простоты изложения, будем использовать термин полиномы, в отличие от многочленов, которые могут быть именованными). Аналогичным свойством обладают члены ряда Фурье и некоторые другие менее известные семейства. Обобщая, их можно записать в

виде линейной формы  $y = \sum_{k=1}^n a_k f_k(x)$ , где  $f_k \in F(-1, 1)$  –

одна из функций плотного в  $C^1[-1, 1]$  семейства. Иначе говоря, аппроксимирующего на данном отрезке семейства  $F$ . В исследованиях данной области практически обходят вниманием плотные семейства, которые удовлетворяют условию линейности относительно коэффициентов, но с некоторыми ограничениями. Примеры таких непрерывных зависимостей – это зависимости с асимптотическим ограничением значений или значений, непрерывных не на всех отрезках оси абсцисс. Вместе с тем, такие зависимости достаточно распространены. К примеру, инфокоммуникационные данные не принимают отрицательных значений и не заданы на отрицательной полуоси времени, а результаты работы искусственных нейронных сетей типа MLP<sup>2</sup>, как правило, не выходят за пределы значений ноль и единица, но асимптотически к ним приближаются.

В целом, надо заметить, что аппроксимирующие семейства не отличаются разнообразием и порой доставляют неудобства – обрезают максимумы-минимумы, характеризующихся всплесками, образуют волнообразные артефакты на краях интервала, и особенно неважно ведут себя в задачах экстраполяции, то есть прогноза, который зачастую является целью аппроксимации. Отход от классического представления задачи порождает нелинейность формы и, в результате, невозможность применения метода наименьших квадратов. Который в прежние времена был практически единственным способом аппроксимировать и экстраполировать зависимости. Ниже рассматриваются приемы, которые позволяют расширить семейства аппроксимирующих функций в сторону ограниченно нелинейных форм, которые поддаются классическому методу определения коэффициентов.

<sup>1</sup> То же, что многочлены в классическом виде Колмогорова-Габор

<sup>2</sup> Multilayer Perceptron – многослойная нейронная сеть

**Строго монотонное преобразование**

Задача для метода наименьших квадратов обычно формулируется для полиномов степени  $m$  и размерности  $n$ . Это связано с тем, что, согласно теореме Вейерштрасса-Стоуна об аппроксимации, любую непрерывную зависимость  $f(A \subset R^n)$ , определенную на компакте  $A$  можно приблизить полиномами со сколь угодно высокой точностью  $\varepsilon > 0$

$$\forall \varepsilon > 0 \exists p(\mathbf{x}, \mathbf{a}, m, n) \in P^{m,n}(A \subset R^n) \text{ такой, что}$$

$$\|f(A) - p(\mathbf{x}, \mathbf{a}, m, n)\| < \varepsilon. \quad (1)$$

Метод наименьших квадратов сводит задачу об аппроксимации к минимизации среднеквадратичной ошибки и далее к решению системы линейных уравнений. Это приводит к решению задачи аппроксимации за одну итерацию, что быстро и удобно, особенно в сравнении с довольно медленными и зачастую нестабильными итерационными методами пошагового приближения [4].

Расширением области применения метода наименьших квадратов в общем виде, с учетом многомерности, является зависимость в записи

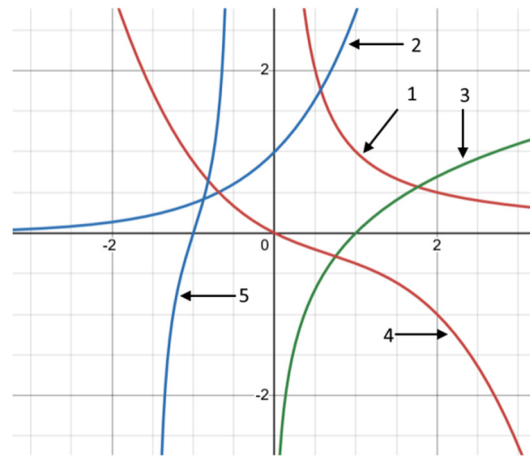
$$y(\mathbf{x}, \mathbf{a}, n) = \sum_{k=1}^K a_k f_k(\mathbf{x}), \quad (2)$$

в которой  $f_k(\mathbf{x})$  – частные функции произвольного вида в количестве  $K$  от нескольких переменных, не содержащие настраиваемых коэффициентов  $a_k$ . В таких условиях метод наименьших квадратов также применим, даже при том, что частные функции, такие как синусы или гауссианы<sup>3</sup> не являются полиномами. Как оказалось, дальнейшее расширение возможно с применением суперпозиции  $y = f \circ g = g(f(x))$ , где  $g$  – передаточная функция из класса строго монотонных функций. Пример улучшения качества аппроксимации для некоторого класса непрерывных функций приведены в работах [5, 6].

В данной работе предлагается дальнейшее расширение области применения метода наименьших квадратов добавлением к выражению (2) строго монотонной функции  $g$ . Напомним, что строго монотонной функцией называется биективная функция, у которой приращение аргумента  $\Delta x$  не меняет знак приращения функции  $\Delta y$ , и приращение функции не принимает значения ноль

$$y(\mathbf{x}, \mathbf{a}, n) = g\left(\sum_{k=1}^K a_k f_k(\mathbf{x})\right). \quad (3)$$

Примерами строго монотонной функции  $\sigma$  могут служить функции  $a + bt$ ,  $(a+bt)^{-1}$ ,  $\ln(t)$ ,  $\exp(t)$ ,  $\arctg(t)$ ,  $\text{sigma}(t)$  и др. Все они имеют обратную функцию (рис. 1)  $y = g(t) \rightarrow t = g^{-1}(y)$ .



**Рис. 1.** Строго монотонные функции, которые могут быть использованы в качестве передаточных:

- 1 – ветвь гиперболы; 2 – экспонента; 3 – логарифм;
- 4 – кубический полином; 5 – ветвь тангенса

Функция с добавленной передаточной функцией (3), строго говоря, не подходит для поиска коэффициентов методом наименьших квадратов, поскольку она включает нелинейное преобразование над коэффициентами (иначе нелинейность). Однако, это выражение может быть обращено относительно нелинейности и тогда выражение примет вид

$$\sum_{k=1}^K a_k f_k(\mathbf{x}) = g^{-1}(y(\mathbf{x}, \mathbf{a}, n)), \quad (4)$$

или

$$z(\mathbf{x}, \mathbf{a}, n) = g^{-1}(y(\mathbf{x}, \mathbf{a}, n)) = \sum_{k=1}^K a_k f_k(\mathbf{x}), \quad (5)$$

что превращает его в линейную по коэффициентам форму, аналогичную уравнению (2). В этой форме записи оно уже подходит для поиска коэффициентов методом наименьших квадратов. Однако, это будут не совсем те коэффициенты, которые получились бы прямым вычислением.

Напомним, что метод наименьших квадратов заключается в минимизации целевой функции (она же функция ошибок – критерий оптимизации) определенного вида – среднеквадратической ошибки

$$s(y, a) = \frac{1}{n} \sum_{i=1}^n (y(x_i, a, m, n) - y_i)^2 \xrightarrow{a} \min. \quad (6)$$

В данном случае для зависимости (5) возможно взять другой критерий оптимизации и получить результат

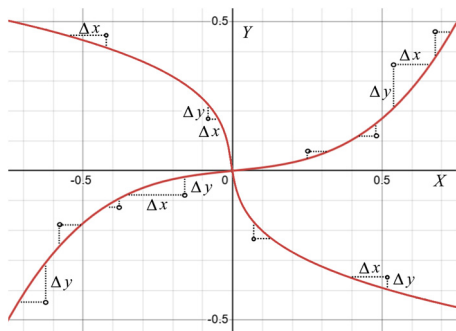
$$s(z, a) = \frac{1}{n} \sum_{i=1}^n (z(x_i, a, m, n) - z_i)^2 \xrightarrow{a} \min. \quad (7)$$

Результаты оптимизации при применении метода с двумя разными критериями будут разными. То есть, разными будут полученные двумя способами аппроксимирующие зависимости и их параметры. Сравнивать результаты между собой по критерию из теории полезности «лучше-хуже» в общем слу-

<sup>3</sup> Функции, использованные К.Вейерштрассом в доказательстве теоремы об аппроксимации

чае проблематично, поскольку первая зависимость (6) находится только методом пошаговой оптимизации, а вторая (7) как методом пошаговой оптимизации, так и методом наименьших квадратов. Результат действительно будет другим, вопрос только насколько это критично. Будет ли он хуже или лучше. Если хуже, то насколько хуже. Что перевесит – небольшое ухудшение результата или значительное ускорение получения результата за счет более быстрого алгоритма? Вычислительный эксперимент показывает половинчатый результат – в одних случаях лучше, в других хуже.

Метод наименьших квадратов имеет точное решение – он приводит к минимуму критерия оптимизации за один шаг. Но, критерии в формулах (6) и (7) разные, поэтому они могут приводить к разным решениям и оба будут точными, поскольку каждый вычисляется по своему критерию строгим аналитическим методом. Следует иметь в виду, что приближенное решение в общем смысле не обязательно реализуется через минимум среднеквадратической ошибки по значению – также его можно получить через взвешенное отклонение<sup>4</sup>, через абсолютную величину отклонения, через среднеквадратическую ошибку аргумента (рис. 1) или другими способами.



**Рис. 2.** Иллюстрация различия между аппроксимацией по двум критериям ошибки на примере двух кубических кривых  $y = f^3(x)$  и  $x = f^{-3}(y)$ . Показаны отклонения по координатам  $x$  и  $y$ . Влияние ошибок по  $x$  и по  $y$  диаметрально противоположное

На рисунке 2 показаны две кубические строго монотонные зависимости  $y = f^3(x)$  и  $x = f^{-3}(y)$ , и условные значения для обучающей выборки  $D = \{x_i, y_i\}$ . Можно заметить, что критерий ошибки по аргументу  $x$  более значим в области малых по абсолютной величине значений (горизонтальные пунктирные линии), чем по значению функции  $y$  (вертикальные пунктирные линии). Для другой строго монотонной зависимости ситуация обратная. В целом, выбранный критерий ошибки не является однозначно лучшим и иногда вместо функции  $y = f(x)$  под задачу может подойти обратная функция  $x = g^{-1}(y)$ . Такие вычисления с обратными функциями назовём «обратными» вычислениями.

### Линейная передаточная функция

Начнем с простых передаточных функций. Интуитивно очевидно, что линейная передаточная функция не должна оказывать какого-то влияния на результат аппроксимации. Докажем это.

**Теорема 1.** Линейная передаточная функция не влияет на одномерную полиномиальную аппроксимацию.

<sup>4</sup> Отклонение равно корню квадратному от ошибки

**Доказательство.** Пусть, для начала, аппроксимирующая функция имеет вид полинома первой степени  $y = a_0 + a_1x$ . И пусть, передаточная функция будет линейной  $\tilde{y} = (ct + d)$ , где  $t = y$ . Тогда

$$y = a_0 + a_1x \Rightarrow \tilde{y} = c(a_0 + a_1x) + d. \quad (8)$$

Для нахождения коэффициентов  $a_0$  и  $a_1$  воспользуемся методом наименьших квадратов. Примем в качестве критерия минимизации функции (8) среднеквадратическую ошибку

$$S(a) = \frac{1}{m} \sum_{k=1}^m ((c(a_0 + a_1x_k) + d) - y_k)^2 \xrightarrow{a} \min.$$

Для краткости изложения обозначим  $x = x_k$ ,  $y = y_k$  и

$$\Sigma(*) = \sum_{k=1}^m (*).$$

Найдём частные производные критерия  $S(a)$  по компонентам вектора коэффициентов  $a$

$$\begin{cases} S'_0(a) = \frac{2}{m} \Sigma((c(a_0 + a_1x) + d) - y)c = 0 \\ S'_1(a) = \frac{2}{m} \Sigma((c(a_0 + a_1x) + d) - y)cx = 0 \end{cases} \quad (9)$$

После сокращения коэффициентов, система линейных уравнений 9 запишется в виде

$$\begin{cases} a_0 \Sigma c + a_1 \Sigma cx = \Sigma(y - d) \\ a_0 \Sigma cx + a_1 \Sigma cx^2 = \Sigma(y - d)x. \end{cases}$$

Решим систему уравнений методом Крамера

$$\Delta = \begin{vmatrix} \Sigma c & \Sigma cx \\ \Sigma cx & \Sigma cx^2 \end{vmatrix} = \Sigma c \Sigma cx^2 - \Sigma cx \Sigma cx = c^2 (m \Sigma cx^2 - (\Sigma x)^2),$$

$$\Delta_0 = \begin{vmatrix} \Sigma(y - d) & \Sigma cx \\ \Sigma(y - d)x & \Sigma cx^2 \end{vmatrix} = \Sigma(y - d) \Sigma cx^2 - \Sigma(y - d)x \Sigma cx,$$

$$\Delta_1 = \begin{vmatrix} \Sigma c & \Sigma(y - d) \\ \Sigma cx & \Sigma(y - d)x \end{vmatrix} = \Sigma c \Sigma(y - d)x - \Sigma cx \Sigma(y - d).$$

$$\begin{aligned} \tilde{a}_0 &= \frac{\Delta_0}{\Delta} = \frac{\Sigma(y - d) \Sigma cx^2 - \Sigma(y - d)x \Sigma cx}{c^2 (m \Sigma cx^2 - (\Sigma x)^2)} = \\ &= \frac{\Sigma y \Sigma cx^2 - \Sigma yx \Sigma cx - d (\Sigma 1 \Sigma cx^2 - \Sigma x \Sigma cx)}{c^2 (m \Sigma cx^2 - (\Sigma x)^2)} = \\ &= \frac{\Sigma y \Sigma cx^2 - \Sigma yx \Sigma cx - dc (m \Sigma cx^2 - (\Sigma x)^2)}{c^2 (m \Sigma cx^2 - (\Sigma x)^2)} = \frac{a_0 - d}{c}. \\ \tilde{a}_1 &= \frac{\Delta_1}{\Delta} = \frac{\Sigma c \Sigma(y - d)x - \Sigma cx \Sigma(y - d)}{c^2 (m \Sigma cx^2 - (\Sigma x)^2)} = \\ &= \frac{\Sigma c \Sigma(yx) - \Sigma cx \Sigma y - d (\Sigma c \Sigma x - \Sigma cx \Sigma 1)}{c^2 (m \Sigma cx^2 - (\Sigma x)^2)} = \\ &= \frac{m \Sigma(yx) - \Sigma x \Sigma y - d (m \Sigma cx - m \Sigma x)}{c (m \Sigma cx^2 - (\Sigma x)^2)} = \frac{a_1}{c}. \end{aligned}$$

Подставим новые значения коэффициентов  $\tilde{a}_0$  и  $\tilde{a}_1$  в формулу (8) и вернем индексы на место

$$\tilde{y}_k = c(\tilde{a}_0 + \tilde{a}_1 x_k) + d = \frac{a_0 - d}{c} c + d + \frac{a_1}{c} c x_k = a_0 + a_1 x_k. \quad (10)$$

Таким образом, найденные значения коэффициентов с передаточной функцией приводят к тому же результату аппроксимации, что и без передаточной функции  $\tilde{y} = y = a_0 + a_1 x$ . Это доказывает, что линейная передаточная функция не влияет на аппроксимацию линейной функции. Чтобы распространить это доказательство на полиномы степени больше единицы, представим полином кусочно-линейной (приближающей) зависимостью. Каждый линейный кусок этой зависимости линейная передаточная функция оставляет на месте, следовательно, оставляет на месте и всю зависимость. Это доказывает теорему.

### Нелинейные передаточные функции

В отличие от линейной передаточной функции, нелинейная передаточная функция влияет на результат аппроксимации. При этом, как показывает рисунок 2b, некоторые передаточные функции могут существенно улучшить полиномиальную аппроксимацию [6].

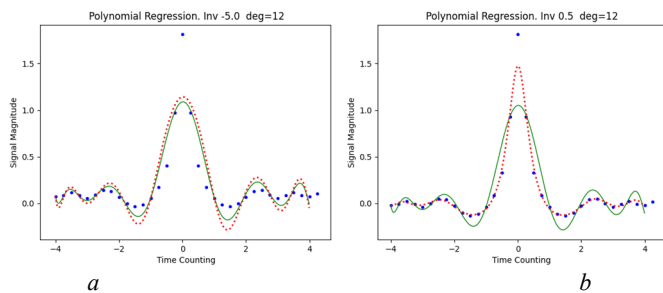


Рис. 3. Сравнение двух способов аппроксимации модельной зависимости (точки) полиномом двенадцатой степени (сплошная линия) и обратным полиномом (пунктирная линия) с параметром, принимающим значения а) -5,0 и б) 0,5

На рисунке 3 модельная зависимость  $y = f_{mod}(x)$  задана дискретным набором точек. Выбор зависимости определялся тем, что в реальных данных часто присутствуют «выбросы», с которыми полиномиальная аппроксимация плохо справляется. Сплошной линией обозначен результат аппроксимации полиномом двенадцатой степени  $n=12$  от одной переменной. Рисунок 3b демонстрирует явное преимущество полинома с передаточной функцией над полиномом той же степени, но без передаточной функции.

Пусть  $y = g\left(\sum_{i=0}^n a_i x^i\right)$  – полином с передаточной функцией  $g(t) = t^{-1}$ . Обратная функция – это инволюция<sup>5</sup>

<sup>5</sup> Оператор инволюции возвращает аргумент к прежнему значению при двукратном применении

$g^{-1}(t) = t^{-1}$ . Записанная в виде  $\tilde{y} = g^{-1}(y) = \sum_{i=0}^n a_i x^i$ , переда-

точная функция позволяет определить коэффициенты полинома методом наименьших квадратов.

Для использования обратной функции необходимо внести дополнительную переменную – параметр  $b$ , который переместит все значения  $y$  либо в положительную, либо в отрицательную область значений во избежание деления на ноль.

$y + b = \sum_{i=0}^n a_i x^i + b$ , что приводит к обратной задаче

$$\tilde{y} = \frac{1}{y + b} = \sum_{i=0}^n \tilde{a}_i x^i.$$

Волна над параметром  $a$  означает, что эти коэффициенты отличаются от тех, которые могут быть получены прямым вычислением, к примеру, итерационным методом Ньютона-Рафсона. Решать оптимизационную задачу будем методом наименьших квадратов

$$S(\tilde{a}) = \sum_{k=1}^m \left( \sum_{i=0}^n \tilde{a}_i x_k^i - \frac{1}{y_k + b} \right)^2 \xrightarrow{\tilde{a}} \min.$$

Здесь внешний знак суммы производит суммирование по  $m$  точкам обучающей выборки – датасета. Оптимум критерия ищется по стандартной схеме через вычисление нуля градиента  $a$ , которое ведёт к построению и решению системы линейных уравнений размером  $n$  на  $n$ . После чего новое приближение, с учетом сдвига, будет

$$\tilde{y} = \left( \sum_{i=0}^n \tilde{a}_i x_k^i \right)^{-1} - b.$$

На левом рисунке 3a параметр  $b$  равен минус 5, то есть, вся область значений  $y$  зависимости сдвинута в отрицательную область под линию  $y = 0$ . Пунктирная линия – это результат обратной аппроксимации. Видно, что обратный полином показал не лучший результат аппроксимации, чем обычный полином. На правом рисунке 3b параметр  $b$  равен плюс 1 и зависимость сдвинута в положительную область. При этом, качество аппроксимации обратными полиномами (пунктирная линия) принципиально улучшилось, что говорит о перспективности использования технологии обратных полиномов для такого типа исходных данных, характерных для области инфокоммуникационных технологий. Вычислительные эксперименты также показали, что при увеличении значения параметра  $b$  различие между этими двумя способами аппроксимации нивелируется.

Покажем, что обратное вычисление с инвертированием полинома позволяет аппроксимировать со сколь угодно высокой точностью любую функцию, непрерывную на ограниченном отрезке числовой оси. Для этого, чтобы не запутаться в терминологии, полином назовём  $p$ -функцией, а функцию, полученную с применением передаточной функции  $g(t) = 1/(t+b)$ , назовём  $q$ -функцией.

**Теорема 1:** Подмножество  $q$ -функций плотно<sup>6</sup> на множестве положительно определенных непрерывных функций на отрезке

$$q[x_1, x_2] \subset C_+^\infty[x_1, x_2].$$

**Доказательство:** пусть  $f$  – непрерывная функция на отрезке, принимающая только положительные значения. Покажем, что ее можно аппроксимировать  $q$ -функциями. По теореме Вейерштрасса существует полином  $p^m$  степени  $m$  такой, что  $\|f - p^m\| < \varepsilon$  для любого заданного  $\varepsilon > 0$ . Функции  $f$  и  $p^m$  не принимают нулевых значений, поэтому произведение их норм имеет положительное ненулевое значение. Тогда, для любого  $\varepsilon > 0$  существуют число  $\delta > 0$  и полином  $p^m$  такие, что  $\|f - p^m\| < \delta$ . Выберем  $\delta$  равным  $\varepsilon \|fp^m\|$ . Тогда

$$\left\| \frac{1}{f} - \frac{1}{p^m} \right\| = \frac{\|f - p^m\|}{\|fp^m\|} < \frac{\delta}{\|fp^m\|} = \varepsilon.$$

Из этого следует, что любые инвертированные полиномы (как просто непрерывные функции, а не только полу-рациональные функции) аппроксимируются  $q$ -функциями с любой наперед заданной точностью. А, поскольку из условия теоремы следует, что каждая непрерывная положительно определенная функция может быть обращена в инвертированную, то из этого следует, что каждая непрерывная положительно определенная функция может быть аппроксимирована  $q$ -функциями.

### Передаточные функции

Рассмотрим некоторые передаточные функции, отличные от обратных полиномов из тех, что упоминались выше.

#### 1. Экспоненциальная передаточная функция

$$y = \exp\left(\sum_{i=0}^n a_i x^i\right) - b.$$

Обращение данной зависимости также приводит к линейной форме относительно коэффициентов  $a$

$$\ln(y + b) = \sum_{i=0}^n a_i x^i.$$

После нахождения коэффициентов методом наименьших квадратов, решение представляется в виде

$$\tilde{y} = \exp\left(\sum_{i=0}^n \tilde{a}_i x^i\right) - b.$$

Полученные коэффициенты также будут отличаться от тех, которые могут быть получены итерационным методом. Но, будет ли результат всегда хуже? Моделирование не даёт на это однозначного ответа. Параметр  $b$  также нужен для поднятия значения аргумента  $y$  в область положительных значений, поскольку логарифм определён только на положитель-

ной полуоси. Из рисунка видно, что такой способ аппроксимации может конкурировать с полиномиальной аппроксимацией.

#### 2. Сигмоидная передаточная функция

Вычислительные эксперименты с сигмоидной функцией приведены в статье [7]. Это, так называемая активационная функция, используемая в нейронных сетях MLP – «много-слойный перцептрон»

$$y = \sigma(t) = \frac{1}{1 + \exp(-t)}.$$

Эта передаточная функция имеет обратную функцию  $t = \sigma^{-1}(y) = -\ln\left(\frac{1}{y} - 1\right)$  на открытом интервале  $(0, 1)$ . То есть, обладает аппроксимируемыми возможностями, ограниченными по оси ординат значениями 0 и 1

$$y = \sigma\left(b\left(\sum_{i=1}^n a_i x^i + \theta\right)\right).$$

Обратная функция может быть представлена уравнением, в правой части которого стоит численное выражение без настраиваемых параметров, а в правой части линейная по коэффициентам  $a$  форма

$$\sum_{i=1}^n a_i x^i + \theta = \frac{\sigma^{-1}(y)}{b}.$$

Критерий оптимизации для обратного перцептрона запишем в виде

$$S = \frac{1}{m} \sum_{k=1}^m \left( \sum_{i=1}^n \tilde{a}_i x_k^i + \theta - \frac{\sigma^{-1}(y_k)}{b} \right)^2 \xrightarrow{\tilde{a}} \min.$$

После определения набора весовых коэффициентов методом наименьших квадратов для нового критерия, получится функция

$$\tilde{y} = \sigma\left(b\left(\sum_{i=1}^n \tilde{a}_i x^i + \theta\right)\right).$$

Рисунок показывает, что аппроксимация с участием сигмоидной  $s$ -образной функции имеет право на применение в задачах аппроксимации, по крайней мере, на интервале аргумента от минус четырех до плюс четырех. Параметр  $b$  позволяет этот интервал расширить.

Моделирование при  $b = 2$  показало, что на интервале от минус до плюс восьми аппроксимация также вполне удовлетворительная, особенно если учесть, что для обучения перцептрона большой точности в передаточной функции не требуется. Это открывает новые возможности в использовании нейросетевых технологий для задач регрессии и прогноза [8].

<sup>6</sup> Подмножество  $B$  называется плотным в множестве  $A$  если любая окрестность точки  $a$  из  $A$  содержит точки из  $B$

### 3. Сигмоидная передаточная функция

Сигмоидная передаточная функция, а также её аналоги, особенно часто используется в искусственных нейронных сетях в качестве «активационной функции» [9]. При этом, элементарный вычислительный элемент нейронной сети – искусственный нейрон типа перцептрон имеет тот же вид, что и полином с передаточной функцией  $\sigma$

$$y = \sigma \left( \sum_{k=1}^n w_k x_k + \theta \right),$$

где  $x$  – координаты точки в  $n$ -мерном пространстве, а  $\theta$  – параметр смещения, который можно рассматривать как свободный член в многомерном полиноме первой степени. Обычный способ получения коэффициентов нейронной сети, как и коэффициентов отдельного перцептрона, состоит в «обучении» – методе пошагового градиентного спуска или, что фактически то же самое, методе «обратного распространения ошибки»<sup>7</sup> [8, 9].

Известно, что этот способ очень затратный, требует не только времени работы компьютера, но и вычислительной мощности, использования графических ускорителей, а также затрат электроэнергии на их охлаждение. Переход к идеологии обратного вычисления в ряде случаев позволит существенно ускорить и удешевить этот процесс, поскольку методом наименьших квадратов можно найти все коэффициенты за один шаг через решение системы линейных уравнений.

Для этой передаточной функции чем больше аргумент функции, тем меньше отклонение по ординате, в то время как отклонение по абсциссе возрастает. Если пользоваться обратными вычислениями, то большую чем в первом случае роль будут играть большие по модулю значения. Из этого можно сделать некоторые выводы:

- Первый: обучающая выборка должна быть максимально однородной, чтобы результаты расчетов по двум критериям были близки.

- Второй: в обучающей выборке не должно быть значений 0 и 1, а данные  $y_k$  в обучающей выборке следует нормировать с учётом характера их экспоненциального распределения.

- Третий: следует пересмотреть идеологию начальной инициализации весовых коэффициентов нейронной сети, поскольку в методе наименьших квадратов случайным числам места нет.

- Четвёртый: в полносвязных нейронных сетях сигмоидная активационная функция с обратным вычислением может стать эффективной заменой функции ReLU (Rectified Linear Unit) за счет повышения скорости вычисления на каждом нейроне.

- И последний: чем выше размерность пространства, в котором производится оптимизация сети, тем больше будет размер системы линейных уравнений при обратных вычислениях. Это может оказаться «ахиллесовой пятой» идеи обратных вычислений. Поэтому отказываться от классического способа обучения нейронных сетей в пользу нового подхода следует с осторожностью.

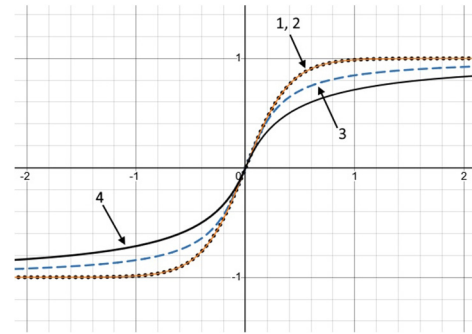


Рис. 4. Активационные функции с равными значениями и углами наклона в начале координат, для сравнения: 1,2) сдвинутая сигмоида и гиперболический тангенс; 3) а) тангенс; 4) сборка из двух гипербол

На рисунке 4 представлены четыре активационные функции: первые две – это сигмоида и гиперболический тангенс, подогнанные под общие значение и производную в начале координат

$$y = \frac{2}{1 + e^{-5x}} - 1,$$

$$y = \tanh(2.5x).$$

Обращает на себя внимание, что эти две активационные функции – сигмоида и гиперболический тангенс – взаимозаменяемы

$$2\sigma(2x) - 1 = \frac{2}{1 + e^{-2x}} - 1 = \frac{e^x(1 - e^{-2x})}{e^x(1 + e^{-2x})} = \tanh(x).$$

Третья функция – это обратный тангенс

$$y = \frac{2}{\pi} \arctan(4x).$$

Арктангенс отличается «толстым хвостом», то есть, медленным спаданием к нулю, отличным от экспоненциального при стремлении аргумента к минус бесконечности. Это может оказаться полезным при его использовании в рекуррентных нейронных сетях, в которых возникает проблема «затухания градиента». Ещё более толстым хвостом обладает функция номер 4, которая составлена из двух гипербол (без учета нормирующего коэффициента)

$$y = \begin{cases} \frac{1}{1-x} - 1, & x < 0 \\ 1 - \frac{1}{1+x}, & x \geq 0. \end{cases}$$

Привлекательным свойством этой передаточной функции является как простота её вычисления, так и простота вычисления её производной, что критически важно для вычисления градиента в задачах многомерной оптимизации

$$y' = xy^2.$$

<sup>7</sup> Авторство приписывают Д.Хинтону, хотя в СССР этот метод был опубликован раньше А.И.Галушковым

То же можно сказать и про обратную функцию, вычисление которой не сложнее. Это позволяет задуматься о возможности и перспективности её применения для обратных вычислений.

### Результаты моделирования

Численные эксперименты с упомянутыми передаточными функциями подтверждают теоретические выкладки о том, что прямые и обратные вычисления коэффициентов ведут к близким, но не идентичным результатам. Обратные полиномы показали интересные и полезные к практическому применению результаты. В том числе, продемонстрирована качественная аппроксимация одномерных зависимостей. Особенно это относится к инфокоммуникационным данным, которые характеризуются неотрицательными значениями, а также к моделированию одно и двумерных зависимостей, часто представленных в форме всплесков. Эксперименты с перцептроном показали перспективность применения обратных вычислений при некоторых ограничениях на значения из обучающей выборки.

### Задача экстраполяции, прогноза и рациональные функции

За пределами интервала назвать поведение ортогонального многочлена  $p$ , как обобщение полинома, экстраполяцией сложно. Во многом выручают обратные вычисления если убрать весовую функцию, значения которой на концах отрезка равны нулю. Но, в целом, результат будет неудовлетворительным. В работе [8] показано, что лучшим вариантом для задач экстраполяции является использование рациональных функций, являющихся отношением двух полиномов  $r(x) = p(x)/q(x)$ . Современные возможности вычислительной техники для определения коэффициентов рациональных функций пошаговыми методами оптимизации критерия ошибки, позволяют не особо беспокоиться о применимости более быстрого метода наименьших квадратов. И тем не менее, этот метод остаётся востребованным.

Пусть  $y(x) = r(a, b, x) = \frac{p(a, x)}{q(b, x)}$  – рациональная функция.

Тогда  $z(x) = y(x)q(b, x) - p(a, x) = 0$  – полином от двух переменных  $x$  и  $y$ , значение которого для всех пар  $(x_k, y_k)$  точек аргумента равны нулю  $z_k = 0 \forall k = 1, \dots, m$ , где  $m$  – число точек в обучающей выборке. Способом, определенном выше, здесь также может быть найдено решение методом наименьших квадратов

$$s(a, b) = \frac{1}{m} \sum_{k=1}^m (y_k q(b, x_k) - p(a, x_k))^2 \xrightarrow{a, b} \min.$$

В этом случае применяется способ вычисления координат кривой (графика) неявно заданной функции вида  $F(x, y) = 0$ . Для этого строим явно заданную функцию от двух координат  $z = F(x, y)$  и минимизируем её отклонение от нулевого значения. После чего находим частные производные по переменным  $a$  и  $b$ , приравниваем их нулю и решаем систему линейных уравнений. Строго говоря, в этом случае речь идет не о передаточных функциях, а о взвешенной сумме двух полиномов с весами  $y_k$  и  $-1$ . Поменяв местами эти веса, приходим к

обратной функции, которая уже является передаточной. Из этой комбинации пар полиномов и передаточной функции дополнительно следует, что использование просто передаточных функций – не единственный способ расширения семейств аппроксимирующих функций. К примеру, мы можем рассмотреть более общий случай  $F(p(x, y), q(x, y)) = 0$  функции от двух полиномов, в которой рациональные функции являются частным случаем. Это перспективное обобщение и тут есть ещё над чем работать.

Современные методы прогноза временных рядов многочисленны. Среди них выделяются фильтр ARIMA, фильтр Калмана, рекуррентный фильтр Хэмминга и нейросеть с ассоциативной памятью Дж.Хопфилда. Общая идея, которая их объединяет – это динамическая минимизация ошибки прогноза. Но, сам долгосрочный прогноз так или иначе основан на краткосрочном прогнозе с помощью полиномов. Которые, как известно, для прогноза временных рядов не очень подходят. Поэтому использование более широкого набора аппроксимирующих семейств актуально.

### Заключение

Расширение семейства полиномов, а также других аппроксимирующих семейств за счет использования строго монотонного преобразования в передаточных функциях имеет далеко идущие и пока ещё малоизученные перспективы. Из них наиболее интригующим является перспектива разработки новых архитектур малоразмерных нейронных сетей, ориентированных на задачи регрессии и прогноза. Их обучение, возможно, будет не столь затратно по времени и энергопотреблению как это происходит в современных нейронных сетях. Также новые семейства функций могут оказаться полезными для доработки существующих и разработки новых методов прогноза временных рядов.

### Литература

1. Сегё Г. Ортогональные многочлены. М.: гос. изд. ф-м литературы, 1962, 500 с.
2. Линник Ю.В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. 2-е изд. М., 1962.
3. Зорич В.А. Математический анализ. Часть вторая. М.: МЦНМО, 2002.
4. Иванов А.П. Практикум по численным методам. Метод Ньютона. Санкт-Петербург, СПбГУ, 12 с.
5. Толстых В.Н., Еришов В.Н. Аналитическое представление экспериментальных данных полурациональными функциями // Научные технологии в космических исследованиях Земли. Т. 16, № 6, 2024. С. 4-11.
6. Толстых В.Н., Еришов В.Н., Лавриненко К.А. Семейства параметрических функций для аппроксимации данных // Радиотехнические, оптические и биотехнические системы, 2024. С. 276-280.
7. Толстых В.Н., Швиндт М.О. Архитектура быстрообучающихся нейронных сетей // Радиотехнические, оптические и биотехнические системы, 2024. С. 289-293.
8. Толстых В.Н. Нейронные сети для экстраполяции временных рядов // Научные технологии в космических исследованиях Земли. Т.15, 2023. С. 4-12.
9. Галушкин А. И. Синтез многослойных систем распознавания образов. М.: Энергия, 1974.

## ANALYTICAL REPRESENTATION OF INFOCOMMUNICATION DATA

**Viktor N. Tolstykh**, Saint Petersburg State University of Aerospace Instrumentation, St. Petersburg, Russia,  
[victor.n.tolstykh@gmail.com](mailto:victor.n.tolstykh@gmail.com)

**Vladimir N. Yershov**, Saint Petersburg State University of Aerospace Instrumentation, St. Petersburg, Russia

**Maksim S. Savinov**, Saint Petersburg State University of Aerospace Instrumentation, St. Petersburg, Russia

**Abstract**

The article discusses the question of possible extensions of families of approximating functions by superposition with some classes of transfer functions in order to improve the quality of approximation of continuous on a segment and tabular dependencies. Both in terms of accuracy and speed of computation. Particular attention is paid to the possibility of using the least squares method for such extensions, since this method provides the fastest solution compared to step-by-step approximation methods. The question of a possible change in the convergence criterion of the error function to expand the possibility of this method, which will allow more flexibility in solving this problem, is being investigated. Additionally, two practical problems are considered - the applicability of the Least Squares method to the training of artificial neural networks, as well as to extrapolation and prediction problems, to which polynomials are not very suitable. To justify the concept of transfer functions, theorems are formulated with proof of the linear transfer function and the completeness of the family of inverse polynomials. The applicability of the method of orthogonal polynomials, as well as rational functions for these purposes, is discussed. As a research method, it is proposed to supplement the approximating family with an example of polynomials with a strictly monotone transformation as a transfer function. Specific variants of such transfer functions are considered, it is indicated for what practical tasks they can be useful. The result is a demonstration of the visible advantage of a polynomial supported with a transfer function over a polynomial without it to approximate a specific type of dependencies often featured in info-communication data. The rest of the results have theoretical significance and can be used for forecasting and develop a new type of artificial neural networks architecture.

**Keywords:** approximation, extrapolation, perceptron, neural networks, monotonic transformation, transfer function

**References**

- [1] G. Sego, "Orthogonal polynomials", Moscow, 1962, 500 p.
- [2] Yu. V. Linnik, "The least squares method and the foundations of the mathematical and statistical theory of observation processing", 2nd ed. Moscow, 1962.
- [3] V. A. Zorich, "Mathematical analysis", Part two. Moscow, 2002.
- [4] A. P. Ivanov, "Workshop on numerical methods. Newton's method", St. Petersburg, St. Petersburg State University, 12 p.
- [5] V. N. Tolstykh, V. N. Ershov, "Analytical presentation of experimental data by semirational functions", *H-ES Research*, vol. 16, no. 6, 2024, pp. 4-11.
- [6] V. N. Tolstykh, V. N. Ershov, K. A. Lavrinenko, "Families of parametric functions for approximation of data", *Radio, optical and biotechnical systems*, 2024, pp. 276-280.
- [7] V. N. Tolstykh, M. O. Schwindt, "Architecture of rapidly learning neural networks", *Radio, optical and biotechnical systems*, 2024, pp. 289-293
- [8] V. N. Tolstykh, "Neural networks for extrapolating time series", *H-ES Research*, vol. 15, 2023, pp. 4-12.
- [9] A. I. Galushkin, "Synthesis of multilayer pattern recognition systems", Moscow: Energy, 1974.