

# МОДИФИКАЦИЯ ФУНКЦИОНАЛА КАЧЕСТВА В ЗАДАЧАХ НЕЛИНЕЙНОЙ РЕГРЕССИИ ДЛЯ УЧЕТА НЕОДНОРОДНЫХ ПОГРЕШНОСТЕЙ ИЗМЕРЯЕМЫХ ДАННЫХ

Г. И. Рудой<sup>1</sup>

## Аннотация

Рассматривается проблема восстановления существенно нелинейной регрессионной зависимости по данным, имеющим погрешности измерения как зависимых, так и независимых переменных, при этом считается, что распределения погрешностей различных измерений имеют разную дисперсию. Предлагается модифицированный функционал среднеквадратичной ошибки, учитывающий погрешности определения независимых переменных и различные распределения погрешностей в разных точках, и выводимый из геометрических соображений. Приводятся результаты численного моделирования на данных, полученных в ходе эксперимента по измерению зависимости мощности лазера от прозрачности резонатора, показывающие преимущество предложенного функционала. Для этой же модели рассматривается сходимость вектора параметров, минимизирующего предлагаемый функционал качества, к оптимальному для классического функционала среднеквадратичной ошибки. Кроме того, фиксируется вектор параметров этой модели, принимаемый за «истинный», и исследуется сходимость оптимальных параметров для предлагаемого и для классического функционалов к «истинным» параметрам в зависимости от распределения погрешностей измерений.

**Ключевые слова:** гетероскедастичные ошибки, ошибки измерения независимых переменных, символьная регрессия, нелинейная регрессия.

## 1 Введение

В ряде приложений возникает задача нахождения оптимальных коэффициентов  $\omega$  некоторой регрессионной модели  $f$ , заданной в виде аналитической формулы, по набору экспериментальных данных. Для этого в предположении о нормальном распределении регрессионных остатков строится функционал  $\sum_i (y_i - f(x_i, \omega))^2$ , представляющий сумму квадратов отклонений экспериментальных точек  $y_i$  от регрессионной кривой, и находится вектор параметров  $\omega$ , его минимизирующий.

Однако, данный функционал корректен только для точно измеренных независимых переменных и гомоскедастичности ошибок измерения зависимой переменной. Иными словами, предполагается существование лишь ошибок измерения зависимой переменной, для которых дисперсия соответствующего распределения принимается одинаковой.

В большинстве естественнонаучных приложений это предположение не выполняется, особенно при измерении некоторой зависимости в достаточно широких диапазонах. Например, в задаче нахождения зависимости коэффициента преломления  $n$  прозрачного полимера от длины волны  $\lambda$  погрешности измерения каждого физического параметра в различных точках, вообще говоря, различны [1]. Так, если для измерения длины волны  $\lambda$  используется дифракционная решетка, то постоянной является относительная погрешность определения длины волны  $\frac{\sigma_{\lambda_i}}{\lambda_i} \approx \text{const}$ , и, следовательно, погрешность определения длины волны зависит от самой длины волны. Подобная ситуация фиксированной относительной (а не абсолютной) ошибки является типичной для физических экспериментов.

---

<sup>1</sup>Московский Физико-Технический Институт, georgii.rudoi@phystech.edu

Таким образом, возникает задача поиска оптимальных коэффициентов регрессионной формулы с учетом отличающихся погрешностей измерения в различных экспериментальных точках. Для некоторых частных случаев эта задача уже была решена.

Например, в [2] рассматривается модель Басса, описывающая принятие и распространение новых потребительских продуктов в популяции со временем, для которой вводится предположение о неравной точности измерений в различных экспериментальных точках, что описывается различными весовыми коэффициентами при соответствующих регрессионных остатках. При этом весовые коэффициенты имеют достаточно общий вид и вводятся произвольно в виде экспертно указанных значений.

Другим примером является [3], где рассматривается задача оценки коэффициентов трехпараметрического распределения Вейбулла по неточно измеренным зависимым и независимым данным. Для этого используется метод латентных переменных: к независимым переменным  $t_i$  добавляются «свободные» переменные  $\delta_i$ , предоставляющие степень свободы в пространстве независимых переменных, и минимизируется функционал вида

$$T(\alpha, \beta, \eta, \delta) = \sum_{i=0}^n w_i [f(t_i + \delta_i; \alpha, \beta, \eta) - y_i]^2 + \sum_{i=0}^n p_i \delta_i^2,$$

где  $\alpha, \beta, \eta$  — параметры распределения, а  $w_i$  и  $p_i$  являются некоторыми экспертно заданными весами, соответствующими относительной точности  $i$ -го измерения аналогично [2].

В настоящей работе рассмотрена более общая ситуация, в которой не только зависимые, но и независимые переменные определяются неточно, и каждая переменная имеет свою собственную погрешность измерения. Исследуется случай нелинейной регрессионной зависимости, в отличие от, например, [4], изучающей линейную модель.

Предложен модифицированный функционал качества, учитывающий погрешности как зависимых, так и независимых переменных в виде, достаточном для большинства практических приложений. Весовые коэффициенты при регрессионных остатках в настоящей работе выводятся из базовых предположений о распределении погрешностей измерения и о поведении регрессионной модели в окрестности каждой экспериментальной точки. На примере широко применяемого алгоритма Левенберга-Марквардта [5] рассматривается модификация имеющихся методов оптимизации, применяемых в подобных задачах (классический и модифицированный алгоритмы будем в дальнейшем обозначать АЛМ и мАЛМ соответственно). По построению показывается, что модифицированный алгоритм оптимизации имеет те же свойства (например, скорость сходимости), что и исходный алгоритм. Приводятся результаты анализа экспериментальных данных по измерению параметров усиливающей среды газового лазера. В рамках анализа, помимо непосредственного нахождения коэффициентов регрессионной модели, сравнивалась сходимость минимизирующих как предложенный, так и классический функционал параметров к некоторым «истинным» значениям. Показано, что в подавляющем большинстве рассмотренных случаев предложенный функционал дает лучшие приближения.

## 2 Постановка задачи

Дана обучающая выборка  $D$ :

$$D = \{\mathbf{x}_i, y_i\} | i \in \{1, \dots, \ell\}, \mathbf{x}_i \in \mathbb{R}^m, y_i \in \mathbb{R}. \quad (1)$$

Для каждой зависимой переменной  $y_i$  известно стандартное отклонение ошибки ее измерения  $\sigma_{y_i}$ , а для соответствующего вектора независимых переменных  $\mathbf{x}_i$  аналогично известны стандартные отклонения его компонент  $\sigma_{x_{ij}} | j \in \{1, \dots, m\}$ . Пусть, кроме того, выбрана некоторая регрессионная модель  $y = f(\mathbf{x}, \boldsymbol{\omega})$ , параметризованная вектором  $\boldsymbol{\omega}$ .

Для удобства введем вектор, составленный из ошибок измерений зависимых переменных  $\sigma_{y_i}$ :

$$\boldsymbol{\sigma}_y = \{\sigma_{y_1}, \dots, \sigma_{y_\ell}\}.$$

Аналогично введем матрицу, составленную из ошибок измерений независимых переменных  $\sigma_{x_{ij}}$ :

$$\Sigma_x = \|\sigma_{x_{ij}}\| | i \in \{1, \dots, \ell\}, j \in \{1, \dots, m\}.$$

Требуется построить функционал ошибки  $\check{S}(\boldsymbol{\omega})$  вектора параметров  $\boldsymbol{\omega}$  модели  $f$ , учитывающий ошибки измерений  $\boldsymbol{\sigma}_y$  и  $\Sigma_x$ :

$$\check{S}(\boldsymbol{\omega}) = \check{S}(\boldsymbol{\omega}, \boldsymbol{\sigma}_y, \Sigma_x), \quad (2)$$

и, кроме того, найти вектор параметров  $\boldsymbol{\omega}$ , минимизирующий функционал (2):

$$\hat{\boldsymbol{\omega}} = \arg \min_{\boldsymbol{\omega}} \check{S}(\boldsymbol{\omega}). \quad (3)$$

### 3 Модифицированный функционал качества

Воспользуемся следующим естественным качественным физическим соображением: чем больше погрешность определения зависимой или независимых переменных для некоторой экспериментальной точки, тем меньше соответствующий регрессионный остаток должен учитываться при оптимизации параметров модели. Кроме того, с физической точки зрения складывать можно только величины, имеющие одинаковую размерность, либо безразмерные величины, поэтому необходима соответствующая нормировка невязок по каждому из измерений.

Сначала рассмотрим случай одной независимой переменной:  $x \in \mathbb{R}$ . С учетом приведенных выше соображений введем следующее определение расстояния  $\rho(x, i)$  от точки  $(x_i, y_i)$  до некоторой точки  $(x, f(x, \boldsymbol{\omega}))$  на кривой, описываемой регрессионной моделью  $y = f(x, \boldsymbol{\omega})$ :

$$\rho^2(x, i) = \frac{(x_i - x)^2}{\sigma_{x_i}^2} + \frac{(y_i - y)^2}{\sigma_{y_i}^2}. \quad (4)$$

Непосредственное точное определение расстояния от экспериментальной точки до регрессионной кривой представляется отдельной сложной вычислительной задачей, однако в подавляющем большинстве практических приложений регрессионные зависимости достаточно гладкие, а погрешности измерения достаточно малы. При этих предположениях будем рассматривать расстояние не от точки до кривой, а от точки до линейаризованной в окрестности этой точки кривой. Так, на рис. 1 показаны различные варианты определения расстояния, при этом в иллюстративных целях размерности и погрешности определения  $x$  и  $y$  приняты одинаковыми.

Итак, линейаризуем  $f(x, \boldsymbol{\omega})$  в окрестности точки  $(x_i, f(x_i, \boldsymbol{\omega}))$ , обозначив оператор линейаризации в окрестности этой точки как  $\mathbb{L}_i$ :

$$f(x, \boldsymbol{\omega}) \approx \mathbb{L}_i[f](x, \boldsymbol{\omega}) = f(x_i, \boldsymbol{\omega}) + (x - x_i) \frac{\partial f}{\partial x}(x_i, \boldsymbol{\omega}), \quad (5)$$

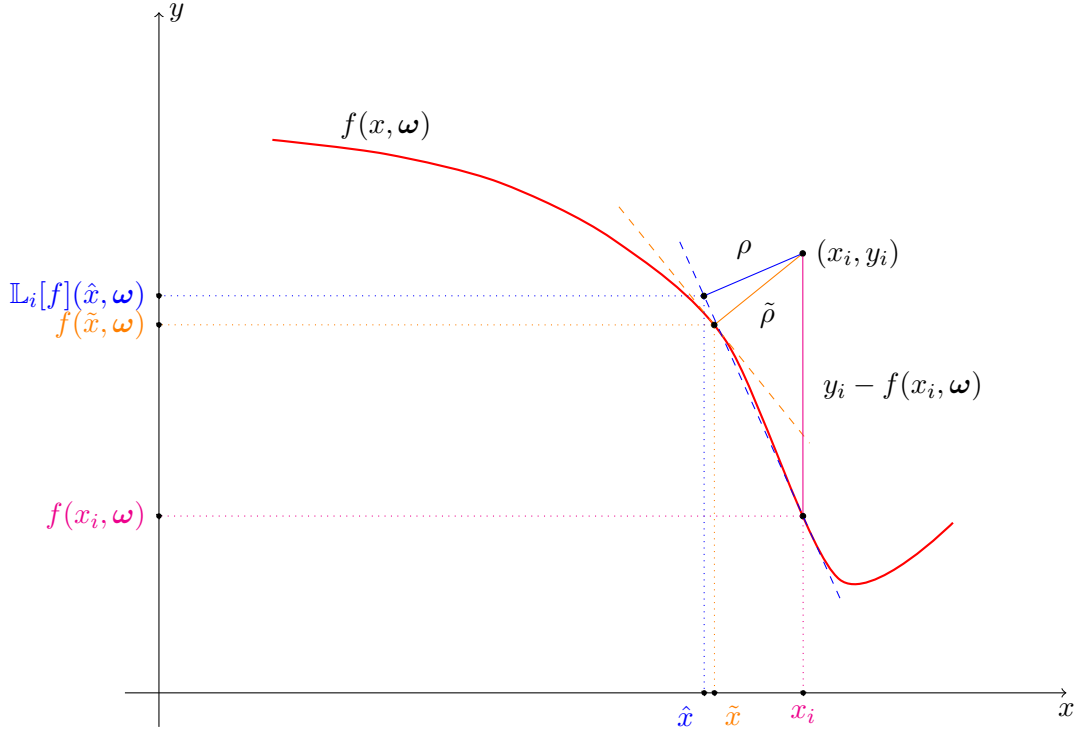


Рис. 1: Различные способы определения расстояния от точки до прямой:  $\tilde{\rho}$  — истинное расстояние как минимум расстояния от точки  $(x_i, y_i)$  до какой-либо точки на прямой,  $y_i - f(x_i, \omega)$  — расстояние в классическом функционале среднеквадратичной ошибки в предположении об отсутствии ошибок измерения независимой переменной,  $\rho$  — предлагаемое нами расстояние.

Расстояние (4) выражается для линеаризованной функции (5) следующим образом:

$$\rho^2(x, i) = \frac{(x_i - x)^2}{\sigma_{x_i}^2} + \frac{(y_i - f(x_i, \omega) - \frac{\partial f}{\partial x}(x_i, \omega)(x - x_i))^2}{\sigma_{y_i}^2}. \quad (6)$$

Минимизируя это выражение по  $x$ :

$$\hat{x} = \arg \min_x \rho^2(x, i),$$

находим расстояние от точки  $(x_i, y_i)$  из обучающей выборки до линеаризованной в ее окрестности регрессионной зависимости  $f$  при данном векторе параметров  $\omega$ :

$$\rho^2(f, \omega, i) = \rho^2(\hat{x}, i) = \frac{(y_i - f(x_i, \omega))^2}{\sigma_{y_i}^2 + \frac{\partial f}{\partial x}(x_i, \omega)^2 \sigma_{x_i}^2}. \quad (7)$$

Аналогично можно получить выражение для случая, когда объекты в обучающей выборке представлены  $m$  независимыми переменными ( $\mathbf{x} \in \mathbb{R}^m$ ). Таким образом, предла-

гаемый нами функционал, минимизирующий сумму введенных согласно (4) расстояний, для достаточно гладких функций выглядит следующим образом:

$$\check{S}(\omega) = \sum_{i=1}^{\ell} \frac{(y_i - f(\mathbf{x}_i, \omega))^2}{\sigma_{y_i}^2 + \sum_{j=1}^m (\frac{\partial f}{\partial x_j}(\mathbf{x}_i, \omega))^2 \sigma_{x_{ij}}^2}. \quad (8)$$

Отметим следующее:

- Функционал (8) соответствует классической сумме квадратов регрессионных остатков при условии нормировки квадрата каждого остатка на сумму квадратов погрешности определения зависимой величины  $\sigma_{y_i}$  и произведения частной производной регрессионной модели по  $j$ -ой компоненте вектора независимых величин на погрешность определения соответствующей компоненты  $\sigma_{x_{ij}}$ .
- При прочих равных условиях в выражении для расстояния (7) и, соответственно, в функционале (8) с большим весом учитываются те точки, в которых производная регрессионной модели  $\frac{\partial f}{\partial x_j}$  по соответствующей компоненте  $x_j$  больше, что соответствует соображениям здравого смысла: чем меньше наклон регрессионной зависимости в окрестности данной точки, тем меньше влияние неточного измерения соответствующей независимой переменной на значение регрессионной зависимости в этой точке.
- Если все независимые переменные измерены точно, то есть,  $\forall i, j : \sigma_{x_{ij}} = 0$ , то предложенный функционал переходит в рассмотренный в [2]. Если же, кроме того, все зависимые переменные имеют одну и ту же погрешность, то предложенный функционал переходит в известную сумму квадратов регрессионных остатков с точностью до некоторого множителя  $\sigma_y$ , не влияющего на положения минимумов функционала среднеквадратичной ошибки.

## 4 Метод оптимизации предложенного функционала

Для численной оптимизации функционала (8) представим его в виде суммы квадратов регрессионных остатков путем следующего переобозначения переменных. Вместо выборки (1) рассмотрим выборку

$$\tilde{D} = \{\tilde{\mathbf{x}}_i, \tilde{y}_i\} | i \in \{1, \dots, \ell\}, \tilde{\mathbf{x}}_i \in \mathbb{R}^{m+1}, y_i \in \mathbb{R},$$

где  $\tilde{y}_i \equiv 0$ , а  $\tilde{\mathbf{x}}_i = \{\mathbf{x}_i, y_i\}$  — исходный вектор  $\mathbf{x}_i$  с дополнительно приписанным к нему значением  $y_i$ . Кроме того, примем

$$\tilde{f}(\tilde{\mathbf{x}}_i, \omega) = \frac{f(\mathbf{x}_i, \omega) - y_i}{\sqrt{\sigma_{y_i}^2 + \sum_{j=1}^m (\frac{\partial f}{\partial x_j}(\mathbf{x}_i, \omega))^2 \sigma_{x_{ij}}^2}}.$$

Тогда минимизация функционала (8) возможна известными методами оптимизации, так как легко видеть, что (8) в этом случае эквивалентен

$$S(\omega) = \sum_{i=1}^{\ell} (\tilde{y}_i - \tilde{f}(\tilde{\mathbf{x}}_i, \omega))^2.$$

Как видно из построения, в качестве базового алгоритма оптимизации может быть использован любой метод решения задачи о наименьших квадратах, как, например, метод градиентного спуска или алгоритм Левенберга-Марквардта [6]. Кроме того, при соответствующих условиях гладкости частных производных функции  $f$  (что практически всегда выполняется в реальных физических приложениях) сохраняются все свойства исходного алгоритма.

Отметим, что предложенная идея введения весовых коэффициентов, отвечающих различным измерениям и зависящих от точности этих измерений, вообще говоря, применима и для прочих методов решения задачи восстановления регрессии, отличных от символьной регрессии. Подробное рассмотрение этих методов в совокупности с предлагаемым подходом выходит за рамки статьи, однако укажем, что при невозможности выполнить аналитическое дифференцирование функции  $f$  предлагается использовать следующий следующий итеративный алгоритм, предназначенный для использования с уже имеющимися реализациями методов оптимизации. Предполагается, что соответствующая реализация «принимает на вход» массив значений  $y_i$ , функцию вычисления значения  $f$  в точках  $\mathbf{x}_i$  с вектором параметров  $\boldsymbol{\omega}$ , и критерий останова в виде числа итераций.

Алгоритм выглядит следующим образом:

1. Выбирается некоторое начальное приближение вектора параметров  $\boldsymbol{\omega}$ .
2. Для каждой пары  $(\mathbf{x}_i, y_i)$  из обучающей выборки численно или аналитически рассчитывается значение частной производной  $\frac{\partial f}{\partial x}$  в точке  $(\mathbf{x}_i, \boldsymbol{\omega})$ .
3. Каждое значение зависимой переменной  $y_i$  и значение функции  $f(\mathbf{x}_i, \boldsymbol{\omega})$  нормируется на соответствующую величину

$$\sigma_{y_i}^2 + \sum_{j=1}^m \left( \frac{\partial f}{\partial x_j}(\mathbf{x}_i, \boldsymbol{\omega}) \right)^2 \sigma_{x_{ij}}^2.$$

4. Выполняется итерация классического алгоритма оптимизации для таким образом модифицированных значений функции  $f$  и зависимых переменных  $y_i$ , таким образом получается новое значение вектора  $\boldsymbol{\omega}$ .
5. Если критерий останова не достигнут, алгоритм продолжает выполнение с пункта 2.

Отметим следующее:

- Критерием останова могут служить обычные критерии, такие как достижение некоторого числа итераций, порог нормы изменения вектора  $\boldsymbol{\omega}$ , и т. п.
- Если известно, что производная  $\frac{\partial f}{\partial x}$  является достаточно гладкой в окрестности  $(\mathbf{x}_i, \boldsymbol{\omega}) \mid i \in \{1, \dots, \ell\}$ , на шаге 4 алгоритма представляется разумным выполнить сразу несколько итераций классического алгоритма во избежание потенциально ресурсоемкого пересчета производных и перенормировки значений  $y_i$  и  $f$ .

## 5 Вычислительный эксперимент

В вычислительном эксперименте рассматриваются данные, полученные в ходе измерения зависимости интенсивности излучения  $I$  лазера от прозрачности его резонатора. Изучался лазер высокого давления ( $\approx 3$  атм  $He$ ,  $\approx 60$  Торр  $Ne$ ,  $\approx 20$  Торр  $Ar$ ) на  $3p - 3s$  переходах неона (основной переход — 585 нм), возбуждаемый электронным пучком [7].

Пусть насыщающая переход интенсивность излучения —  $I_s$ , наблюдаемая интенсивность —  $I_l$ . В таком случае для безразмерной величины  $y = \frac{I_l}{I_s}$  с учетом однородного уширения линии усиления при высоком давлении газа и хорошей однородности возбуждения, обеспечиваемой электронным пучком, можно получить нелинейное уравнение [8]:

$$\alpha_0 L - \frac{1}{2} \ln R_0 = g_0 L \frac{1 + \sqrt{R_0}}{1 - \sqrt{R_0}} \frac{1}{y} \ln \left( 1 + \frac{y \frac{1 - \sqrt{R_0}}{1 + \sqrt{R_0}}}{1 + y \frac{2\sqrt{R_0}}{1 - R_0}} \right), \quad (9)$$

где  $\alpha_0$  — распределенные потери (например, на рассеяние света),  $g_0$  — коэффициент усиления слабого сигнала,  $R_0$  — коэффициент отражения выходного зеркала лазера. Однородность накачки означает, что  $g_0$  и  $\alpha_0$  одинаковы по всему объему с хорошей точностью.

Значение  $R_0$  является независимой переменной, изменяемой экспериментаторами, и в данном разделе также обозначается  $x$  сообразно остальной части работы.

Для достаточно больших  $R_0$ , близких к единице (фактически для  $R_0 \geq 0.6 \div 0.7$ ), можно упростить (9), заменив  $2\sqrt{R_0} \approx 1 + R_0$  и получив хорошо известное выражение:

$$y(R_0) = \gamma \frac{1 - R_0}{1 + R_0} \left( \frac{g_0}{\alpha_0 - \frac{1}{2} L \ln R_0} - 1 \right), \quad (10)$$

где  $\gamma$  — нормировочный коэффициент.

Длина активной среды  $L = 150$  см, точность определения мощности лазера  $y$  имеет относительную погрешность в 2%, точность определения прозрачности  $R_0$  имеет абсолютную погрешность и составляет 0.01 при  $R_0 \geq 0.6$  и 0.02 при  $R_0 < 0.6$ .

В ходе измерений получены значения  $y(R_0)$ , приведенные в таблице 1.

Таблица 1: Экспериментальные значения  $y(R_0)$ .

$R_0$	0.48	0.56	0.65	0.73	0.80	0.87	0.94
$y$	3.25	10.2	16.5	20.5	22.5	23.2	18.2

Таким образом, решается задача минимизации функционала (8) при

$$\begin{aligned} \boldsymbol{\omega} &= (\omega_1, \omega_2, \omega_3) = (\gamma, \alpha_0, g_0), \\ f(x, \boldsymbol{\omega}) &= y(R_0, \gamma, \alpha_0, g_0), \\ \sigma_{y_i} &= 0.02 y_i, \\ \sigma_{x_i} &= \begin{cases} 0.01 & | x_i \geq 0.6, \\ 0.02 & | x_i < 0.6. \end{cases} \end{aligned} \quad (11)$$

## 5.1 Оптимальные параметры модели

Кроме предложенного в настоящей работе функционала (8) рассмотрен также и классический функционал среднеквадратичной ошибки:

$$S = \sum_{i=1}^{\ell} (y_i - f(x_i, \omega))^2, \quad (12)$$

В таблице 2 приведены значения параметров  $\omega$  и  $\omega^0$  функции (10), минимизирующие (8) и (12) соответственно, а также относительные разности их компонент. Кроме того, приведены значения функционалов (8) и (12) для обоих векторов параметров.

Таблица 2: Оптимальные значения параметров модели.

	$g_0$	$\alpha_0$	$\gamma$	Значение (8)	Значение (12)
$\omega$	$2.920 \cdot 10^{-3}$	$2.156 \cdot 10^{-4}$	100.324	0.577	0.208
$\omega^0$	$2.917 \cdot 10^{-3}$	$2.129 \cdot 10^{-4}$	101.5	0.645	0.183
$ \omega_i - \omega_i^0 /\omega_i^0$	$1.03 \cdot 10^{-3}$	$1.27 \cdot 10^{-2}$	$1.16 \cdot 10^{-2}$	0.11	0.14

Отдельно отметим, что сравнивать непосредственные значения функционалов (8) и (12) не имеет смысла. Вместо этого необходимо сравнивать различные модели по каждому из этих функционалов в отдельности. Так, результаты, приведенные в таблице 2, показывают вполне естественный результат: каждый из двух векторов параметров ( $\omega$  и  $\omega^0$ ) является оптимальным лишь для того функционала, который он минимизирует.

Графики модели (10), соответствующие  $\omega$  и  $\omega^0$ , приведены на рис. 2.

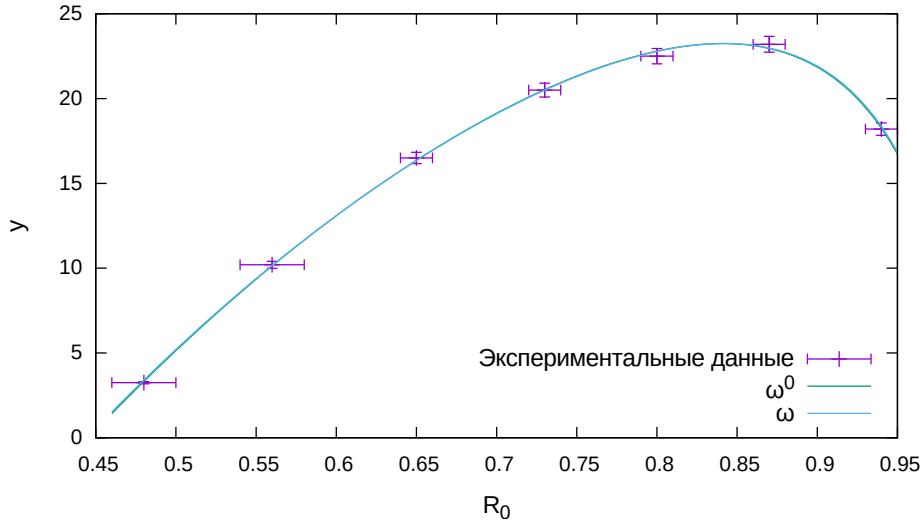


Рис. 2: Графики (10), соответствующие параметрам, минимизирующим (8) и (12).

## 5.2 Сходимость оптимальных параметров к классическим

Численно исследована зависимость сходимости параметров  $\omega$  к параметрам  $\omega^0$ , получаемым минимизацией функционалов (8) и (12) соответственно, от погрешности  $\sigma_y$  изме-



рения зависимой переменной  $y$ .

Разумно ожидать, что при увеличении погрешности измерения величины  $y$  при фиксированной погрешности измерения  $R_0$  оптимальный вектор  $\omega$  будет приближаться к  $\omega^0$ , так как тем более незначителен вклад ошибки измерения независимой переменной.

Рассматриваются два случая:

1. Погрешность  $i$ -го измерения  $y_i$  задается как  $\sigma_{y_i} = 0.02ky_i$ , т. е. погрешность зависит от значения самого  $y_i$ .
2. Погрешность  $i$ -го измерения  $y_i$  задается как  $\sigma_{y_i} = 0.02ky_{\max}$ , т. е. погрешность от значения конкретного  $y_i$  не зависит. Заметим, что выбор конкретного значения  $y$ , определяющего погрешность, является в данном случае достаточно произвольным и соответствует умножению всех погрешностей на некоторую константу.

В первом случае ошибки измерения  $y$  распределены неодинаково, следовательно, применение стандартного метода наименьших квадратов не обосновано. В то же время во втором случае ошибки принадлежат одному и тому же распределению, и, кроме того, независимы, поэтому в данном случае МНК-оценка применима (с точностью до ошибки измерения независимой переменной).

Для обоих случаев подробно рассматривалась область  $k \in [1; 100]$ , значение  $k$  изменялось с шагом 0.01. Отметим, что уже при  $k \approx 25$  характерная погрешность измерения величины  $y$  сопоставима с самой величиной  $y$ , а при  $k > 50$  превышает ее.

Результаты приведены на рис. 3. На графиках отображены компоненты вектора  $\omega$ , нормированные на соответствующие значения  $\omega^0$ , в зависимости от значения  $k$ .

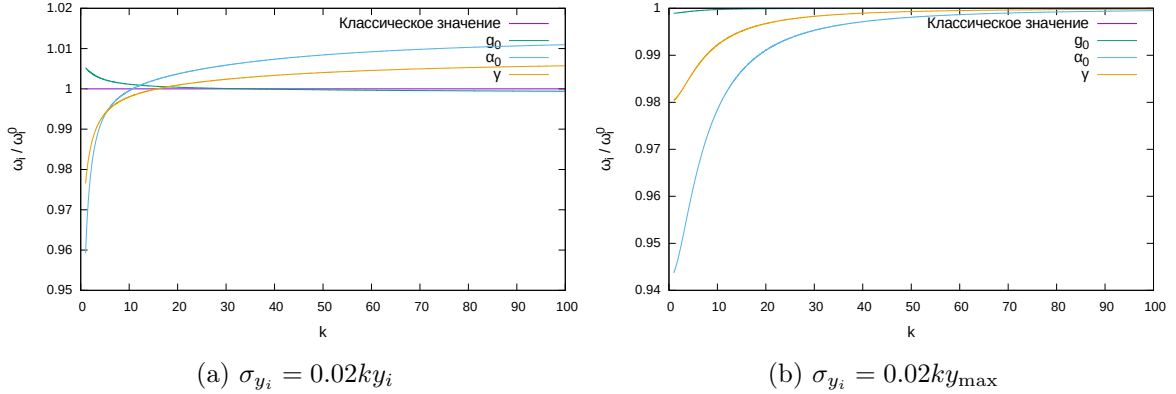


Рис. 3: Зависимость оптимальных параметров от  $k \in [1; 100]$ .

Видно, что в случае фиксированной погрешности  $\sigma_{y_i}$  значения  $\omega$  действительно стремятся к  $\omega^0$  для разумных значений  $k$ , а в случае гетероскедастичности ошибки такой зависимости не наблюдается, хотя значения  $\omega$  и оказываются достаточно близки к  $\omega^0$ .

### 5.3 Сходимость параметров к истинным

Численно исследована зависимость сходимости параметров  $\omega = \arg \min \check{S}$  и  $\omega^0 = \arg \min S$  к некоторому «истинному» значению вектора параметров  $\hat{\omega}$ , от количества точек  $\ell$  в обучающей выборке и от погрешности определения независимой переменной.

Для этого вектор параметров  $\omega$ , полученный минимизацией обучающей выборки из таблицы 1, принимается за некоторый «истинный» вектор параметров  $\hat{\omega}$ , и на каждой  $j$ -й итерации генерируется обучающая выборка  $D_j(\ell, k)$ :

$$D_j(\ell, k) = \{(x_i + \xi_i^x, y(x_i, \hat{\omega}) + \xi_i^y) \mid \xi_i^x \sim \mathcal{N}(0, k\sigma_{x_i}), \xi_i^y \sim \mathcal{N}(0, \sigma_{y_i}), i \in \{1, \dots, \ell\},$$

где  $y(x, \omega)$  задано соотношением (10), а  $\sigma_{x_i}$  и  $\sigma_{y_i}$  определяются соотношением (11).

Иными словами, генерируется обучающая выборка согласно искомой функции зависимости с известным и фиксированным вектором параметров, которая затем зашумляется нормально распределенным шумом. При этом стандартное отклонение шума для зависимой величины совпадает с экспертно предложенным для реального эксперимента, а стандартное отклонение независимой величины отличается от экспертно предложенного в  $k$  раз.

После генерации выборки  $D_j(\ell, k)$  по ней находятся значения  $\omega_j = \arg \min \check{S}$  и  $\omega_j^0 = \arg \min S$ , что повторяется  $N$  раз, и для  $\forall i$  рассматриваются значения

$$\overline{\delta\omega_i} = \frac{\sum_{j=1}^N (\omega_{ji} - \hat{\omega}_i)}{N},$$

$$\overline{\delta\omega_i^0} = \frac{\sum_{j=1}^N (\omega_{ji}^0 - \hat{\omega}_i)}{N}.$$

Обозначим, кроме того,  $\overline{\delta\omega} = \{\overline{\delta\omega_1}, \overline{\delta\omega_2}, \overline{\delta\omega_3}\}$ .

Таким образом, варьируя  $k$  и изучая поведение  $\overline{\delta\omega}$  и  $\overline{\delta\omega^0}$ , можно исследовать влияние погрешности определения независимой переменной на разность между  $\hat{\omega}$  и оптимальными параметрами  $\omega$  и  $\omega^0$  согласно (8) и (12) соответственно.

Так как при уменьшении  $k$  монотонно уменьшается погрешность определения независимой переменной, естественно ожидать, что различие между  $\omega$  и  $\omega^0$  будет уменьшаться. Однако, наш вычислительный эксперимент демонстрирует, что это не так.

В нашем эксперименте  $N = 1000$ ,  $\ell \in \{10, \dots, 5000\}$ ,  $k \in \{0.2, 0.25, \dots, 1\}$ .

Результаты приведены на рис. 4-6.

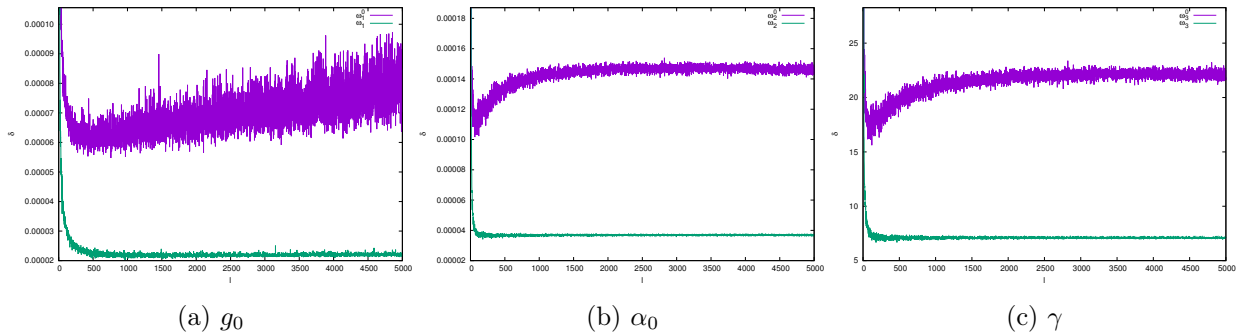


Рис. 4: Сходимость параметров к истинным при  $k = 0.2$ .

В частности, результаты на рис. 4 характерны для всех значений  $k \in [0.2, 0.6]$ . Однако, при  $k \geq 0.65$  поведение оптимальных параметров резко меняется. Так, на рис. 5 приведены графики сходимости для  $k = 0.65$ . Видно, что поведение параметров, оптимизированных согласно классическому функционалу качества  $S$ , является существенно

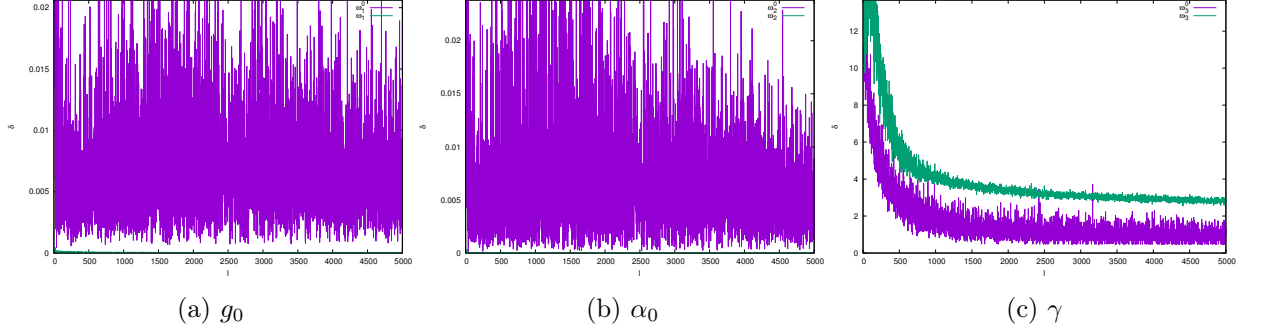


Рис. 5: Сходимость параметров к истинным при  $k = 0.65$ .

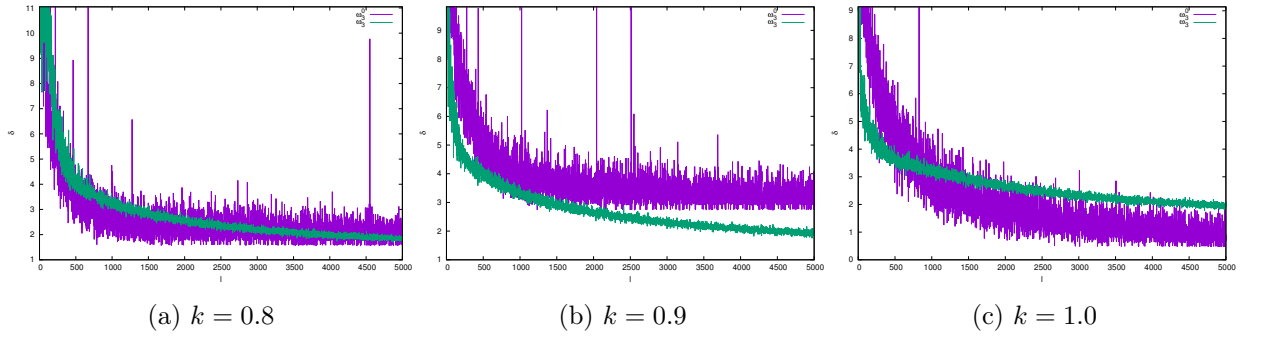


Рис. 6: Сходимость параметра  $\gamma$  при некоторых  $k$ .

более хаотическим, что может говорить о меньшей устойчивости [9] модели, оптимизированной согласно  $S$ .

Более того, для оценок параметров  $g_0$  и  $\alpha_0$  соответствующее приближение на несколько порядков хуже, чем полученное минимизацией  $\tilde{S}$ , вплоть до того, что кривые, соответствующие минимизирующим  $\tilde{S}$  параметрам, практически не видно на графиках (рис. 5а и 5b), поскольку в выбранном масштабе они практически совпадают с осью абсцисс. С другой стороны, важно отметить, что оценка параметра  $\gamma$ , полученная минимизацией  $S$ , является несколько лучшей для  $k = 0.65$  (для  $k = 0.7$  график выглядит аналогично), но минимизация  $\tilde{S}$  дает все лучшие и лучшие приближения с ростом  $k$  (рис. 6).

Стоит отметить следующее:

- Практически во всех случаях (кроме оценки  $\gamma$  для  $k = 0.8$ ) предложенный в настоящей работе функционал (8) дает лучшее приближение, в том числе, при разумно малом объеме обучающей выборки. Кроме того, в подавляющем большинстве случаев предпочтительность предложенного функционала сохраняется и для большего числа экспериментальных точек.
- Для малых  $k \leq 0.6$  ошибка оценки параметров при помощи классического функционала (12) имеет ярко выраженный минимум в окрестности 60-100 для  $\alpha_0$  и  $\gamma$  и 400 для  $g_0$  экспериментальных точек (рис. 4). Дальнейшее увеличение обучающей выборки ведет к ухудшению приближения, получаемого минимизацией (12).
- Для некоторых  $k$  ошибка приближения, получаемого минимизацией предложенного функционала (8), имеет явную горизонтальную асимптоту (рис. 4, 5а и 5b).

Вопрос о причинах подобного поведения оптимальных параметров является предметом дальнейшей работы.

## 6 Заключение

Предложен модифицированный функционал среднеквадратичной ошибки для существенно нелинейных моделей, применимый в случае наличия ошибок измерения независимых переменных и различия распределения, к которому принадлежат ошибки, в разных точках обучающей выборки.

Показана сходимость предложенного функционала к классическому функционалу среднеквадратичной ошибки для случая гомоскедастичности погрешностей зависимой переменной и пренебрежимо малой погрешности измерения независимых переменных.

Представляется разумным использовать предложенный в настоящей работе функционал качества при оптимизации параметров регрессионных моделей и анализе их устойчивости к погрешностям как в зависимых, так и независимых переменных [1, 9].

## Список литературы

- [1] Рудой, Г. И.: *О возможности применения методов Монте-Карло в анализе нелинейных регрессионных моделей*. Сибирский Журнал Вычислительной Математики, 4:425–434, 2015.
- [2] Jukić, Dragan: *On nonlinear weighted least squares estimation of Bass diffusion model*. Applied mathematics and computation, 219(14):7891–7900, 2013.
- [3] Jukić, Dragan и Marković, Darija: *On nonlinear weighted errors-in-variables parameter estimation problem in the three-parameter Weibull model*. Applied mathematics and computation, 215(10):3599–3609, 2010.
- [4] Kiryati, Nahum и Bruckstein, Alfred M: *Heteroscedastic Hough transform (HtHT): An efficient method for robust line fitting in the ‘errors in the variables’ problem*. Computer Vision and Image Understanding, 78(1):69–83, 2000.
- [5] Marquardt, D. W.: *An algorithm for least-squares estimation of non-linear parameters*. Journal of the Society of Industrial and Applied Mathematics, 11(2):431–441, 1963.
- [6] King, Davis E.: *Dlib-ml: A Machine Learning Toolkit*. Journal of Machine Learning Research, 10:1755–1758, 2009.
- [7] Александров, А. Ю., Долгих, В. А., и др.: *Кинетика возбуждаемого электронным пучком лазера высокого давления на "желтой" линии неона*. Квантовая электроника, 18(9):1029–1033, 1991.
- [8] Champagne, LF: *Transient optical absorption in the ultraviolet*. In McDaniel, E. W. and Nighan, William L. (editors): *Applied Atomic Collision Physics, Volume 3: Gas Lasers*, volume 3, pages 349–386, 1982.
- [9] Рудой, Г. И.: *Анализ устойчивости существенно нелинейных регрессионных моделей к погрешностям в измеряемых данных*. Pattern Recognition and Image Analysis, 2, 2016.