

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

Г. И. Рудой<sup>1</sup>

**Аннотация:** Рассматривается задача восстановления нелинейной регрессионной зависимости по данным, имеющим погрешности определения как зависимых, так и независимых переменных, при этом распределения погрешностей различных измерений могут иметь различную дисперсию. Предлагается модифицированный функционал качества, учитывающий погрешности определения независимых переменных и различные распределения погрешностей в разных точках. Приводятся результаты численного моделирования на данных, полученных в ходе эксперимента по измерению зависимости мощности лазера от прозрачности резонатора. Рассматривается сходимость вектора параметров, минимизирующего предлагаемый функционал качества, к оптимальному для классического функционала среднеквадратичной ошибки. Сравнивается сходимость параметров, оптимальных для предлагаемого и классического функционалов, к некоторым «истинным» параметрам модели на данных, сгенерированных согласно этим «истинным» параметрам и зашумленным согласно предположениям о погрешностях измерений, в зависимости от параметров этих погрешностей.

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

## 1 Введение

В ряде приложений (например, [1, 2]) 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 значения регрессионной кривой  $f(x, \omega)$  в точке  $x_i$ , и находится вектор параметров  $\omega$ , его минимизирующий.

Однако данный функционал корректен только для точно измеренных независимых переменных и гомоскедастичных ошибок измерения зависимой переменной: в частности, для линейных моделей соответствующая оценка параметров является несмещенной, состоятельной и наиболее эффективной только при выполнении этих условий. В случае нелинейных моделей вывод функционала среднеквадратичной ошибки согласно методу наибольшего правдоподобия также опирается на эти предположения (с обобщением в виде взвешенного метода наименьших квадратов в случае различных стандартных отклонений зависимой переменной). Иными словами, предполагается существование лишь ошибок измерения зависимой переменной, распределение которых принимается одинаковым.

На практике, как правило, это предположение не выполняется, особенно при измерениях 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 [2]. 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.

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

Так, детальный обзор методов решения этой задачи в случае линейной регрессии приведен в [3]. В частности, для линейных моделей рассматривается даже более общая задача, когда распределение ошибок не является точно известным. Однако, по крайней мере, для ряда методов априорная информация все равно необходима, как то: значение отношения стандартных отклонений зависимой и независимой переменных в случае регрессии Деминга [4] либо наличие инструментальных

---

<sup>1</sup>Московский физико-технический институт, 0xd34df00d@gmail.com

переменных при использовании одноименного метода [5]. Отметим, что дополнительная априорная информация необходима для обеспечения возможности однозначного определения параметров модели, иначе модель становится неидентифицируемой. При этом условие идентифицируемости модели для случая многомерной линейной регрессии в общем виде до сих пор неизвестно [6].

Обзор методов решения аналогичной задачи для случая нелинейной регрессии приведен в [7]. Так, например, метод инструментальных переменных обобщается на случай нелинейных моделей, при этом опять же требуется наличие дополнительных наблюдаемых переменных, пропорциональных регрессору с точностью до некоторой аддитивной ошибки. Заметим, что условия идентифицируемости модели при этом неизвестны.

В ряде работ изучаются конкретные нелинейные регрессионные модели, и соответствующие ошибки измерений предполагаются экспертно заданными. Например, в [8] рассматривается модель Басса, описывающая динамику процесса распространения новых потребительских продуктов, 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 [9], где рассматривается задача оценки коэффициентов трехпараметрического распределения Вейбулла по неточно измеренным 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 [8].

В настоящей работе рассмотрена более общая ситуация, в которой не только зависимые, но и независимые переменные определяются неточно и каждая переменная имеет свою погрешность измерения, заданную экспертно. Исследуется случай нелинейной регрессионной зависимости, в отличие от, например, [10], где изучается линейная модель. Предлагается 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. В частности, оказывается, что весовые коэффициенты зависят не только от самой погрешности измерений в данной точке, но и от производных регрессионной модели в окрестности этой точки.

Предложенный функционал наиболее близок к описанному в [11]. Однако, кроме того, в настоящей работе предлагается вероятностная интерпретация предлагаемого функционала для случая нормально распределенных ошибок.

Во второй части настоящей работы формально поставлена задача нахождения оптимальных параметров регрессионной модели с учетом гетероскедастичных погрешностей определения как зависимых, так и независимых переменных. В третьей выводится предлагаемый функционал качества. Затем, в четвертой части, описывается метод использования имеющихся алгоритмов оптимизации, применяемых в подобных задачах (как, например, алгоритм Левенберга–Марквардта [12]), для минимизации предлагаемого функционала. В пятой части приводятся результаты вычислительного эксперимента, состоящие из трех частей: во-первых, приводятся результаты анализа экспериментальных данных по измерению параметров усиливающей среды газового лазера. Затем сравнивается сходимость оптимальных параметров для предложенного функционала качества к параметрам, минимизирующим классический функционал среднеквадратичной ошибки, в зависимости от параметров распределения ошибок. Кроме того, фиксируется некоторый вектор параметров модели, принимаемый «истинным», согласно которому генерируется набор зашумленных обучающих выборок, для которых анализируется сходимость параметров, оптимальных для классического и для предлагаемого функционалов качества, к «истинным» в зависимости от параметров шума. 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  $\sigma_{x_{ij}} | j \in \{1, \dots, m\}$ . При этом допускается, что близкие точки могут иметь сколь угодно различные ошибки. Кроме того, различные ошибки измерения независимы.

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

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

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

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

Отметим, что эта матрица не является ковариационной матрицей ошибок каждого конкретного объекта из обучающей выборки, поэтому нельзя утверждать, что она является диагональной (и, более того, квадратной).

Пусть выбрана некоторая регрессионная модель  $y = f(\mathbf{x}, \boldsymbol{\omega})$ , параметризованная вектором  $\boldsymbol{\omega}$ . Требуется построить функционал ошибки  $\check{S}(\boldsymbol{\omega})$  вектора параметров  $\boldsymbol{\omega}$  модели  $f$ , учитывающий ошибки измерений  $\boldsymbol{\sigma}_y$  и  $\Sigma_x$ :

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

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 качественным соображением: чем больше погрешность определения переменных (зависимых или независимых) 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 каждой из переменных.

Для упрощения изложения 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  $y = f(x, \boldsymbol{\omega})$ :

$$\tilde{\rho}^2(x, i) = \frac{(x_i - x)^2}{\sigma_{x_i}^2} + \frac{(y_i - f(x, \boldsymbol{\omega}))^2}{\sigma_{y_i}^2}, \quad (4)$$

Непосредственное точное определение расстояния от экспериментальной точки до регрессионной кривой представляется отдельной сложной вычислительной задачей оптимизации (решаемой, например, итерационно), поэтому предлагается рассматривать расстояние от точки не до самой кривой, а до линеаризованной кривой в окрестности этой точки. На рис. 1 показаны различные варианты определения расстояния, при этом в иллюстративных целях размерности и погрешности определения  $x$  и  $y$  приняты одинаковыми.

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

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

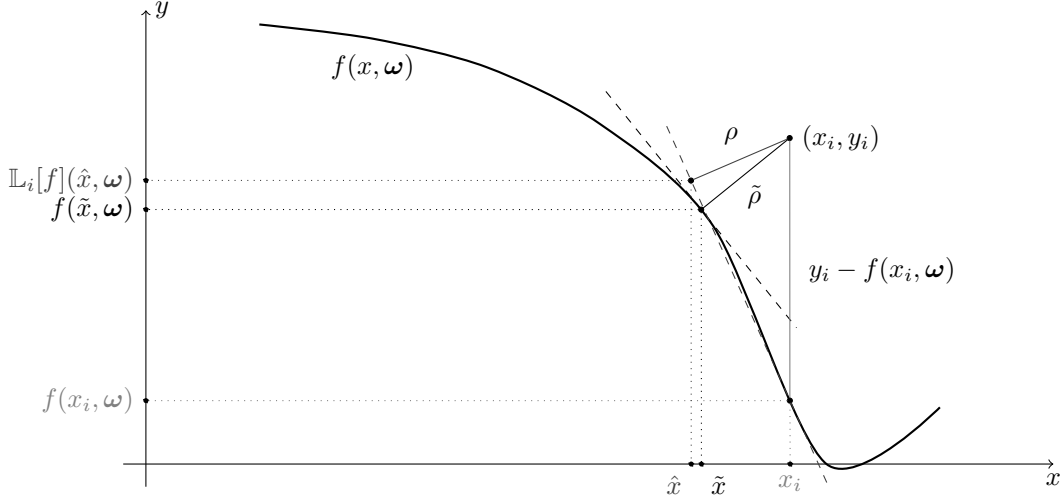


Рис. 1: Различные способы определения расстояния от точки до прямой:  $\tilde{\rho}$  — истинное расстояние как минимум расстояния от точки  $(x_i, y_i)$  до какой-либо точки на прямой,  $y_i - f(x_i, \omega)$  — расстояние в классическом функционале среднеквадратичной ошибки в предположении об отсутствии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

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

Отметим, что решение (7) корректируется при последовательном изменении линеаризации в связи с изменением вектора параметров  $\omega$  согласно выбранному итерационному методу решения этой задачи.

Аналогично можно получить выражение для расстояния в случае, когда объекты в обучающей выборке представлены  $m$  независимыми переменными ( $\mathbf{x} \in \mathbb{R}^m$ ):

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

Таким образом, предлагаемый функционал, минимизирующий сумму введенных согласно (4) расстояний с учетом их линеаризации, 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  $j$ -й 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 [8]. Если же, кроме того, все зависимые переменные имеют одну и ту же погрешность  $\sigma_y$ , то предложенный функционал переходит в известную сумму квадратов регрессионных остатков с точностью до некоторого множителя (а именно  $1/\sigma_y$ ), не влияющего на положения минимумов функционала среднеквадратичной ошибки.

Следует отметить возможность вероятностной интерпретации предложенного выражения для расстояния (4). Для случая одной независимой переменной предположим, что вероятность соответствия некоторой точки  $(\tilde{x}_i, f(\tilde{x}_i, \omega))$  на регрессионной кривой  $y = f(x, \omega)$  данной экспериментальной точке  $(x_i, y_i)$  описывается двумерным нормальным распределением с центром в этой экспериментальной точке  $(x_i, y_i)$  и диагональной ковариационной матрицей  $\Sigma_i = \begin{bmatrix} \sigma_{x_i}^2 & 0 \\ 0 & \sigma_{y_i}^2 \end{bmatrix}$  (то есть ошибки измерения каждой координаты независимы):

$$P(\tilde{x}_i, f(\tilde{x}_i, \omega)) \sim \mathcal{N}((x_i, y_i), \Sigma_i) = \frac{1}{2\pi\sqrt{\det \Sigma_i}} \exp \left\{ -\frac{1}{2} \left\| \begin{matrix} x_i - \tilde{x}_i \\ y_i - f(\tilde{x}_i, \omega) \end{matrix} \right\|_{\Sigma_i^{-1}}^2 \right\},$$

Максимизация логарифма правдоподобия с аналогичной (5) линеаризацией позволяет получить те же выражения (4), (8). Более подробное рассмотрение такого подхода и следствий из него станет предметом дальнейшей работы.

## 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

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

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

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

Тогда минимизация функционала (8) возможна известными методами оптимизации, так как прямой подстановкой можно убедиться, что (8) в этом случае эквивалентен

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

Для таким образом преобразованного функционала в качестве базового алгоритма оптимизации может быть использован любой метод решения задачи о наименьших квадратах, как, например, метод градиентного спуска или алгоритм Левенберга–Марквардта [13]. В этом случае 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 итеративный алгоритм, предназначенный для использования с уже имеющимися реализациями соответствующих методов оптимизации. Предполагается, что реализация «принимает на вход» массив значений  $y_i$ , функцию вычисления значения  $f$  в точках  $\mathbf{x}_i$  с вектором параметров  $\boldsymbol{\omega}$ .

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  $y_i$ , получая 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 [14].

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 [15]

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

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  $x$  согласно 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 [15]

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

где  $\gamma$  — нормировочный коэффициент.

В рассматриваемом физическом эксперименте 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 табл. 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

$$\begin{aligned} \omega &= (\omega_1, \omega_2, \omega_3) = (\gamma, \alpha_0, g_0), \\ f(x, \omega) &= y(R_0, \gamma, \alpha_0, g_0), \\ \sigma_{y_i} &= 0,02y_i, \\ \sigma_{x_i} &= \begin{cases} 0,01 & | x_i \geq 0,6, \\ 0,02 & | x_i < 0,6. \end{cases} \end{aligned} \quad (11)$$

## 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 табл. 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)
$\omega$	$2,93 \cdot 10^{-3}$	$2,07 \cdot 10^{-4}$	98,6	0,542	0,328
$\omega^0$	$2,92 \cdot 10^{-3}$	$2,22 \cdot 10^{-4}$	101,5	0,645	0,183
$ \omega_i - \omega_i^0 /\omega_i^0$	0,31%	6,59%	2,9%	16%	80%

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 табл. 2, показывают вполне естественный результат: каждый из двух векторов параметров ( $\omega$  и  $\omega^0$ ) является оптимальным лишь для того функционала, который он минимизирует.

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

## 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  $\sigma_y$  измерения зависимой переменной  $y$ .

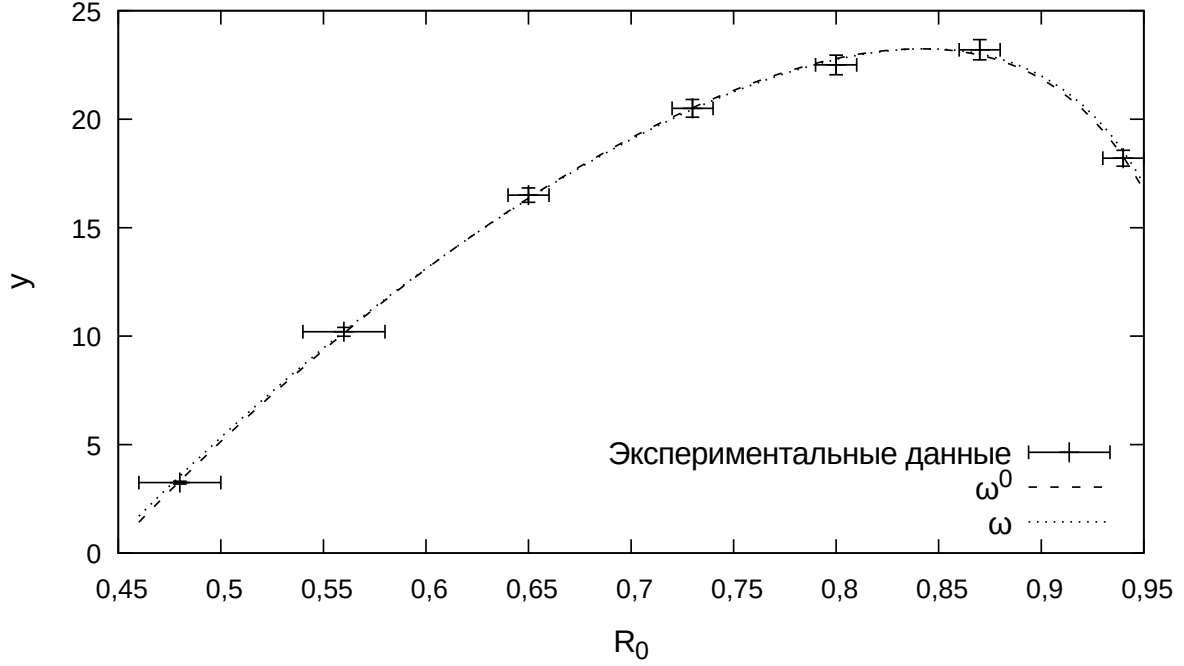


Рис. 2: Графики (10), соответствующие параметрам, минимизирующим (8) и (12)

Следует 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 (что нивелируется соответствующим изменением выбора диапазона  $k$ ).

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  $k$ .

В случае фиксированной погрешности  $\sigma_{y_i}$  значения  $\omega$  действительно стремятся к  $\omega^0$  для разумных значений  $k$ , а в случае гетероскедастичных ошибок такой зависимости не наблюдается, хотя значения  $\omega$  и оказываются достаточно близки к  $\omega^0$ . По мнению автора, такое поведение вектора оптимальных параметров является вполне ожидаемым и демонстрирует несостоятельность классического функционала качества в случае неодинаково распределенных ошибок.



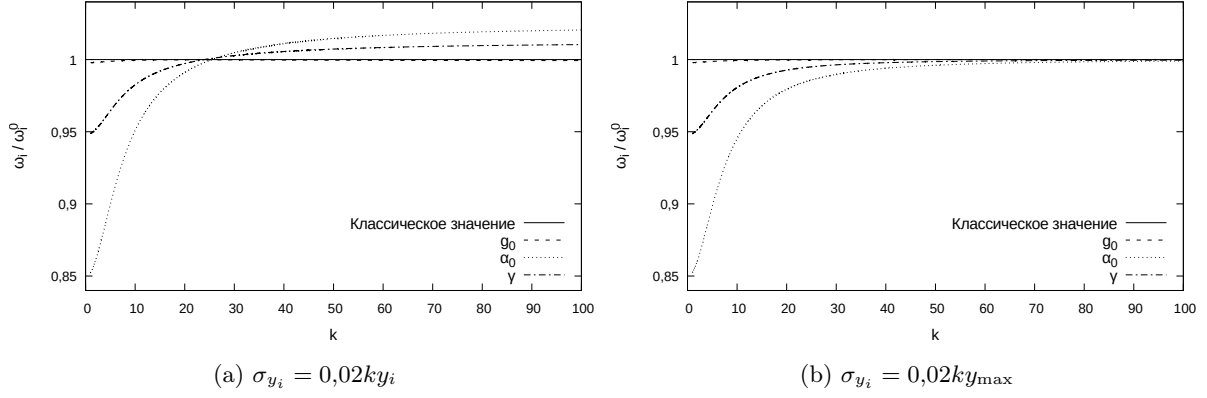


Рис. 3: Зависимость оптимальных параметров от  $k \in [1; 100]$

### 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 числа точек  $\ell$  в обучающей выборке и от погрешности определения независимой переменной.

Для этого вектор параметров  $\omega$ , полученный минимизацией обучающей выборки из табл. 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 модели с известным и фиксированным вектором параметров, которая затем зашумляется нормально распределенными случайными величинами. При этом стандартное отклонение шума для зависимой величины совпадает с экспертно предложенной погрешностью измерений для реального эксперимента, а стандартное отклонение независимой величины отличается от экспертно предложенной погрешности для этой величины в  $k$  раз.

После генерации выборки  $D_j(\ell, k)$  по ней находят значения  $\omega_j = \arg \min \check{S}(D_j)$  и  $\omega_j^0 = \arg \min S(D_j)$ , что повторяется  $N$  раз, и  $\forall i$  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, при варьировании  $k$  и изучении поведения  $\overline{\delta\omega}$  и  $\overline{\delta\omega^0}$  исследуется 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 вычислительный эксперимент демонстрирует, что это не так.

В проведенном эксперименте  $N = 1000$ ,  $\ell \in \{10; \dots; 5000\}$ ,  $k \in \{0,2; 0,25; \dots; 1\}$ .

Результаты представлены на рис. 4–6.

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 более хаотическим, что может говорить о меньшей устойчивости [16] модели, оптимизированной согласно  $S$ .

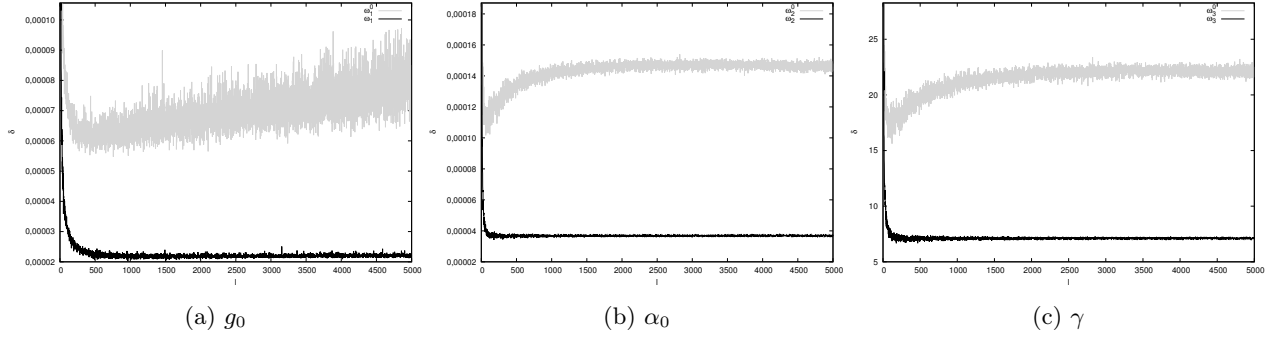


Рис. 4: Сходимость параметров к истинным при  $k = 0,2$

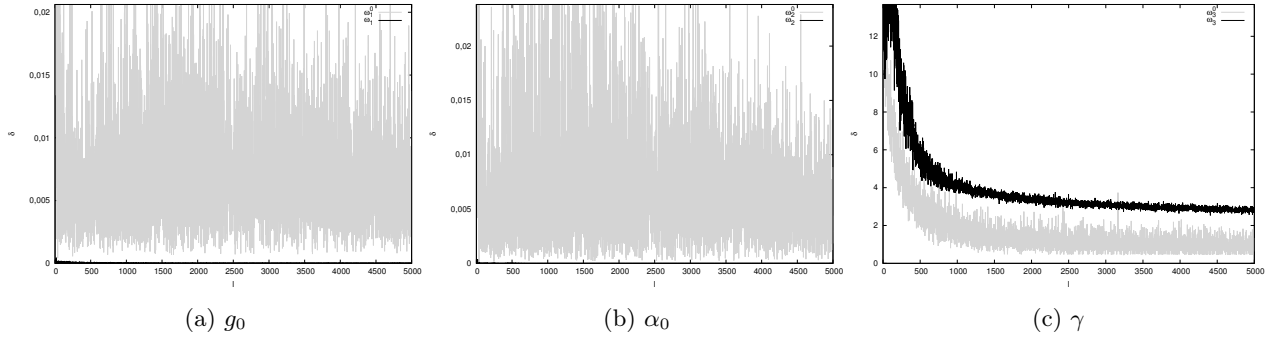


Рис. 5: Сходимость параметров к истинным при  $k = 0,65$

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 видны на графиках (рис. 5а и 5б), 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).

Отметим 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 5б).

Причины подобного поведения оптимальных параметров являются предметом дальнейших исследований.

## 6 Заключение

Предложен модифицированный функционал среднеквадратичной ошибки для задач регрессии, применимый в случае наличия ошибок измерения независимых переменных и различных распре-

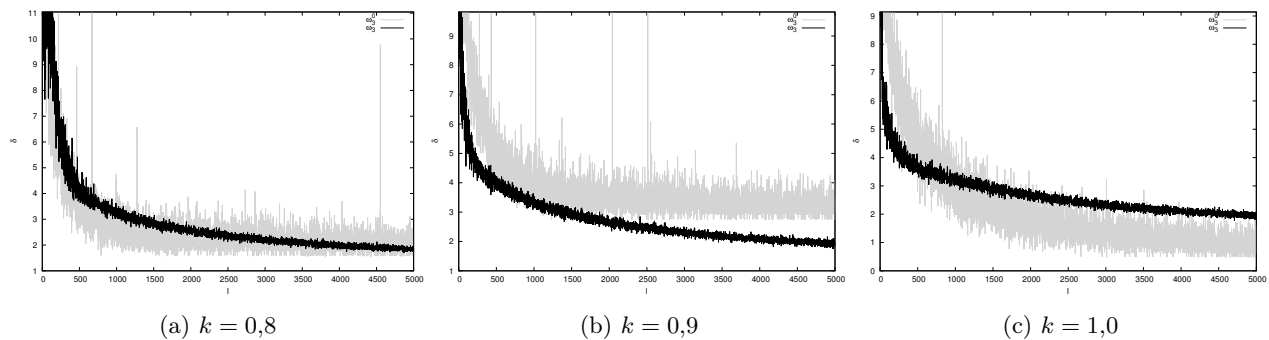


Рис. 6: Сходимость параметра  $\gamma$  при некоторых  $k$

делений, к которым принадлежат ошибки, в разных точках обучающей выборки. Предложена вероятностная интерпретация этого функционала для случая нормального распределения ошибок измерения.

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

Исследовано поведение оптимального вектора параметров для предлагаемого функционала в зависимости от параметров распределений ошибок независимых переменных, в том числе в сравнении с вектором параметров, минимизирующим классический функционал качества.

Представляется разумным использовать предложенный в настоящей работе функционал при оптимизации параметров регрессионных моделей и анализе их устойчивости к погрешностям как зависимых, так и независимых переменных [2, 16].

## Список литературы

- [1] *Гладун А. Д.* Лабораторный практикум по общей физике. — М.: МФТИ, 2004. 316 с.
- [2] *Рудой Г. И.* О возможности применения методов Монте–Карло в анализе нелинейных регрессионных моделей // *Сибирский журнал вычислительной математики*. 2015. Т. 4. С. 425–434.
- [3] *Gillard J. W.* An historical overview of linear regression with errors in both variables: technical report. — Cardiff University School of Mathematics, 2006.
- [4] *Deming W. E.* Statistical adjustment of data. — N. Y.: Wiley, 1943. 216 p.
- [5] *Bowden R. J., Turkington D. A.* Instrumental variables. — Cambridge: Cambridge University Press, 1990. 236 p.
- [6] *Bekker P. A.* Comment on identification in the linear errors in variables model // *Econometrica: Journal of the Econometric Society*. 1986. Vol. 54, no. 1. Pp. 215–217.
- [7] *Carroll R. J., Ruppert D., Stefanski L. A., Crainiceanu C. M.* Measurement error in nonlinear models: a modern perspective. — London: Chapman and Hall/CRC, 2006. 484 p.
- [8] *Jukić D.* On nonlinear weighted least squares estimation of Bass diffusion model // *Applied mathematics and computation*. 2013. Vol. 219, no. 14. Pp. 7891–7900.
- [9] *Jukić D., Marković D.* On nonlinear weighted errors-in-variables parameter estimation problem in the three-parameter Weibull model // *Applied mathematics and computation*. 2010. Vol. 215, no. 10. Pp. 3599–3609.
- [10] *Kiryati N., Bruckstein A. M.* Heteroscedastic Hough transform (HtHT): An efficient method for robust line fitting in the ‘errors in the variables’ problem // *Computer Vision and Image Understanding*. 2000. Vol. 78, no. 1. Pp. 69–83.

- [11] *Boggs P. T., Byrd R. H., Schnabel R. B.* A stable and efficient algorithm for nonlinear orthogonal distance regression // *SIAM Journal on Scientific and Statistical Computing*. 1987. Vol. 8, no. 6. Pp. 1052–1078.
- [12] *Marquardt D. W.* An Algorithm for Least-squares Estimation of Non-linear Parameters // *Journal of the Society of Industrial and Applied Mathematics*. 1963. Vol. 11, no. 2. Pp. 431–441.
- [13] *King D. E.* Dlib-ml: A Machine Learning Toolkit // *Journal of Machine Learning Research*. 2009. Vol. 10. Pp. 1755–1758.
- [14] *Александров А. Ю., Долгих В. А., Рудой И. Г., Сорока А. М.* Кинетика возбуждаемого электронным пучком лазера высокого давления на "желтой" линии неона // *Квантовая электроника*. 1991. Т. 18, № 9. С. 1029–1033.
- [15] *Champagne L. F.* Transient Optical Absorption in the Ultraviolet // *Applied Atomic Collision Physics, Volume 3: Gas Lasers* / Ed. by E. W. McDaniel, W. L. Nighan. Vol. 3. 1982. Pp. 349–386.
- [16] *Rudoy G. I.* Analysis of the stability of nonlinear regression models to errors in measured data // *Pattern Recognition and Image Analysis*. 2016. Vol. 26, no. 3. Pp. 608–616.

---



---

## ON MODIFICATION OF THE MSE LOSS FUNCTION FOR SOLVING NON-LINEAR HETEROSCEDASTIC ERRORS-IN-VARIABLES PROBLEMS

G. Rudoy<sup>1</sup>

<sup>1</sup>Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation

### Abstract

The paper considers the problem of finding the optimal parameters of a non-linear regression model accounting for errors in both dependent and independent variables. The errors of different measurements are assumed to belong to different probability distributions with different variances. A modified mean squared error-based loss function is derived and analyzed for this case.

In the computational experiment the measurements of the laser's radiation power as a non-linear function of the resonator's transparency are used to compare the parameters vectors minimizing the presented loss function and the classical mean squared error. The convergence of the parameters minimizing the presented loss function to the optimal parameters for the classical loss function is studied. In addition, some values of the parameters are considered to be «true» ones and are used to generate synthetic data using the physical model and Gaussian noise, which is then used to study the convergence of the parameters minimizing the presented and the classical loss function, respectively, as the function of the noise parameters.

**Keywords:** *errors-in-variables models, heteroscedastic errors, symbolic regression, non-linear regression.*

## References

- [1] Gladun, A. V. 2004. *Laboratornyy praktikum po obshchey fizike [Laboratory classes on general physics]*. Moscow: MFTI. 316 p.
- [2] Rudoy, G. I. 2015. Applying Monte Carlo methods to analysis of nonlinear regression models. *Numerical Analysis and Applications*, 4:344–350.
- [3] Gillard, J. W. 2006. An historical overview of linear regression with errors in both variables. *Cardiff University School of Mathematics Technical Report*.
- [4] Deming, W. E. 1943. *Statistical adjustment of data*. N.Y.: Wiley, 1943. 216 p.

- [5] Bowden, R. J., and D. A. Turkington 1990. *Instrumental variables*. Cambridge: Cambridge University Press. 236 p.
- [6] Bekker, P. A. 1986. Comment on identification in the linear errors in variables model. *Econometrica: Journal of the Econometric Society*, pages 215–217.
- [7] Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. 2006. *Measurement error in nonlinear models: a modern perspective*. London: Chapman and Hall/CRC. 484 p.
- [8] Jukić, D. 2013. On nonlinear weighted least squares estimation of bass diffusion model. *Applied mathematics and computation*, 219(14):7891–7900.
- [9] Jukić, D., and Marković, D. 2010. On nonlinear weighted errors-in-variables parameter estimation problem in the three-parameter weibull model. *Applied mathematics and computation*, 215(10):3599–3609.
- [10] Kiryati, N., and A. M. Bruckstein. 2000. 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.
- [11] Boggs, P. T., R. H. Byrd, and R. B. Schnabel. 1987. A stable and efficient algorithm for nonlinear orthogonal distance regression. *SIAM Journal on Scientific and Statistical Computing*, 8(6):1052–1078.
- [12] Marquardt, D. W. 1963. An algorithm for least-squares estimation of non-linear parameters. *Journal of the Society of Industrial and Applied Mathematics*, 11(2):431–441.
- [13] King, D. E. 2009. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758.
- [14] Aleksandrov, A. Yu., V. A. Dolgikh, I. G. Rudoy, and A. M. Soroka. 1991. Kinetics of a high-pressure electron-beam-excited laser emitting the “yellow” neon line. *Soviet Journal of Quantum Electronics*, 21(9):933–937.
- [15] Champagne, L. F. 1982. *Transient optical absorption in the ultraviolet*. In McDaniel, E. W., W. L. Nighan (editors): *Applied Atomic Collision Physics, Volume 3: Gas Lasers*, volume 3, pages 349–386.
- [16] Rudoy, G. I. 2016. Analysis of the stability of nonlinear regression models to errors in measured data. *Pattern Recognition and Image Analysis*, 26(3):608–616.

## Contributors

**Rudoy Georg I.** (b. 1991) – PhD student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; 0xd34df00d@gmail.com