

# МОДИФИКАЦИЯ АЛГОРИТМА ЛЕВЕНБЕРГА-МАРКВАРДТА ДЛЯ ЗАДАЧ НЕЛИНЕЙНОЙ РЕГРЕССИИ С УЧЕТОМ ПОГРЕШНОСТЕЙ КАК ЗАВИСИМЫХ, ТАК И НЕЗАВИСИМЫХ ДАННЫХ

Г. И. Рудой

## 1 Введение

В ряде экспериментальных приложений возникает задача нахождения оптимальных коэффициентов  $\omega$  некоторой фиксированной регрессионной модели  $f$ , представленной в виде аналитической формулы, по набору экспериментальных данных. Для этого строится функционал, являющийся суммой квадратов отклонений экспериментальных точек  $y_i$  от регрессионной кривой:  $(y_i - f(x_i, \omega))^2$ . Для численного решения этой задачи широко применяется алгоритм Левенберга-Марквардта [1]. Однако, данный функционал построен и статистически обоснован в предположении о нормальности распределения регрессионных остатков и точно измеренных независимых переменных — иными словами, учитываются и рассматриваются только ошибки измерения зависимой переменной. Более того, предполагается, что ошибки для всех точек принадлежат одному и тому же распределению с одними и теми же параметрами.

В большинстве естественнонаучных приложений это предположение не выполняется. Например, в задаче нахождения дисперсионной зависимости прозрачного полимера (то есть, зависимости коэффициента преломления  $n$  от длины волны  $\lambda$ ) погрешности измерения различных физических параметров, вообще говоря, различны. Так, например, 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.

Таким образом, возникает задача поиска оптимальных коэффициентов регрессионной формулы с учетом отличающихся погрешностей различных экспериментальных точек. Для некоторых частных случаев эта задача уже была решена: например, в работе [?] вводится предположение, что зависимые переменные  $y_i$  измеряются неточно, и каждая переменная  $y_i$  имеет свою собственную погрешность измерения  $\sigma_{y_i}$ . Далее в [?] показывается, что обычный функционал суммы квадратов регрессионных остатков, где каждый остаток нормирован на соответствующую величину  $\sigma_{y_i}^2$ , корректен и статистически состоятелен.

В настоящей работе рассмотрена более общая ситуация, в которой независимые переменные также определяются неточно в процессе эксперимента, и каждая переменная имеет свою собственную погрешность измерения. Предлагается функционал качества и модифицированный алгоритм Левенберга-Марквардта, позволяющий найти оптимальные коэффициенты регрессионной формулы и опирающийся на классический алгоритм Левенберга-Марквардта (в дальнейшем будем называть их мАЛМ и АЛМ соответственно). Доказывается сходимость модифицированного алгоритма и приводятся результаты на экспериментальных данных по измерению параметров усиливающей среды лазерного излучателя.

## 2 Постановка задачи

Дана обучающая выборка  $D = \{\mathbf{x}_i, y_i\} | i \in \{1, \dots, \ell\}, x_i \in \mathbb{R}^m, y_i \in \mathbb{R}$ . Для каждой зависимой переменной 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, дана некоторая параметрическая регрессионная модель  $y = f(\mathbf{x}, \boldsymbol{\omega})$ .

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 компонент независимых переменных  $\sigma_{x_{ij}}$ :

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

Требуется построить функционал ошибки  $S(\boldsymbol{\omega})$  вектора параметров  $\boldsymbol{\omega}$  модели  $f$ , учитывающий ошибки измерений  $\sigma_{y_i}$  и  $\sigma_{x_{ij}}$ :

$$S(\boldsymbol{\omega}) = S(\boldsymbol{\omega}, \boldsymbol{\sigma}_y, \Sigma_x), \quad (1)$$

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

$$\hat{\boldsymbol{\omega}} = \arg \min_{\boldsymbol{\omega}} S(\boldsymbol{\omega}) \quad (2)$$

## 3 Модифицированный функционал качества

Воспользуемся следующим естественным качественным физическим соображением: чем больше погрешность определения зависимой и независимых переменных для некоторой экспериментальной точки, тем меньше соответствующий регрессионный остаток должен учитываться при оптимизации параметров модели.

Рассмотрим для простоты изложения случай одной независимой переменной:  $x \in \mathbb{R}$ . Введем следующее определение расстояния  $\rho(x, i)$  от точки  $(x_i, y_i)$  из обучающей выборки до некоторой точки  $(x, f(x, \boldsymbol{\omega}))$  на кривой, описываемой регрессионной моделью:

$$\rho(x, i) = \frac{(x_i - x)^2}{\sigma_{x_i}^2} + \frac{(y_i - y)^2}{\sigma_{y_i}^2}, \quad (3)$$

где  $y = f(x, \boldsymbol{\omega})$ .

Непосредственное точное определение расстояния представляется отдельной сложной вычислительной задачей, однако, в подавляющем большинстве практических приложений регрессионные зависимости достаточно гладкие, а погрешности измерения достаточно малы. Пользуясь этим, линеаризуем  $f(x, \boldsymbol{\omega})$  в окрестности точки  $x_i$ :

$$f(x, \boldsymbol{\omega}) = f(x_i, \boldsymbol{\omega}) + (x - x_i) \frac{\partial f}{\partial x}(x_i, \boldsymbol{\omega}). \quad (4)$$

Введем для удобства обозначение  $k = \frac{\partial f}{\partial x}(x_i, \boldsymbol{\omega})$ .

Тогда расстояние (3) можно выразить через линейризованную функцию (4) следующим образом:

$$\rho(x, i) = \frac{(x_i - x)^2}{\sigma_{x_i}^2} + \frac{(y_i - f(x_i, \omega) - k(x - x_i))^2}{\sigma_{y_i}^2}. \quad (5)$$

Минимизируя это выражение, находим выражение для расстояния от точки  $(x_i, y_i)$  из обучающей выборки до линейризованной в ее окрестности экспериментальной зависимости:

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

Аналогичным образом можно получить выражение для случая, когда  $\mathbf{x}$  — вектор  $m$ -мерном пространстве  $\mathbb{R}^m$ :

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

Таким образом, функционал, минимизирующий сумму введенных согласно (3) расстояний, выглядит следующим образом:

$$S(\omega) = \sum_{i=1}^{\ell} \frac{(y_i - f(\mathbf{x}_i, \omega))^2}{\sigma_{y_i}^2 + \sum_{j=1}^m \left( \frac{\partial f}{\partial x_j}(\mathbf{x}_i, \omega) \right)^2 \sigma_{x_{ij}}^2}. \quad (7)$$

Отметим следующие наблюдения:

- Функционал (7) 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 (6) и, соответственно, в функционале (7) 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 физической интуиции: чем меньше наклон регрессионной зависимости в окрестности данной точки, тем меньше влияние неточного измерения соответствующей независимой переменной на значение регрессионной зависимости в этой точке.
- Кроме того, если все независимые переменные измерены точно, то есть,  $\forall i, j : \sigma_{x_{ij}} = 0$ , то предложенный функционал переходит в предложенный в [?]. 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  $\sigma_y$ , влияющего на значение функционала, но не на точку его минимума.

## 4 Модифицированный алгоритм Левенберга-Марквардта

Для минимизации функционала (7) предлагается следующий итеративный алгоритм, предназначенный для использования с уже имеющимися реализациями АЛМ <sup>1</sup>:

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

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

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

Отметим следующее:

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

Сходимость предложенного алгоритма можно показать, сведя его к классическому АЛМ. Для этого вместо объектов  $(\mathbf{x}_i, y_i)$  в обучающей выборке будем формально рассматривать объекты  $(\tilde{\mathbf{x}}_i, \tilde{y}_i)$ , где  $\tilde{y}_i = 0$ , а  $\tilde{\mathbf{x}}_i$  — вектор  $\mathbf{x}_i$  с дополнительно приписанным к нему значением  $y_i$ . Примем

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

Тогда минимизация функционала (7) возможна средствами классического АЛМ, так как легко видеть, что (7) в этом случае эквивалентен

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

Кроме того, при возможности выполнить аналитическое дифференцирование функции  $f$ , этот метод показывает еще один способ практической минимизации функционала (7) без постоянной корректировки значений  $y_i$  и  $f$ , как в предложенном выше алгоритме.

---

<sup>1</sup>Предполагается, что реализация АЛМ «принимает на вход» массив значений  $y_i$ , функцию вычисления значения  $f$  в точках  $\mathbf{x}_i$  с вектором параметров  $\omega$ , и критерий останова в виде числа итераций. Примером такой реализации, использовавшейся авторами, может служить [2].

## 5 Вычислительный эксперимент

В вычислительном эксперименте рассматриваются данные, полученные в ходе анализа 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, что означает, что накачка однородна.

Пусть насыщающая переход интенсивность излучения —  $I_s$ , наблюдаемая интенсивность —  $I_l$ . В таком случае для безразмерной величины  $y = \frac{I_l}{I_s}$  можно получить нелинейное уравнение [?]:

$$\alpha_0 L - \frac{1}{2} \ln R_0 = g_0 L \frac{1 + \sqrt{R_0}}{1 - \sqrt{R_0}} \frac{1}{y} \ln \left( 1 + \frac{y^{\frac{1-\sqrt{R_0}}{1+\sqrt{R_0}}}}{1 + y^{\frac{2\sqrt{R_0}}{1-R_0}}} \right), \quad (8)$$

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

Для достаточно больших  $R_0$ , близких к единице, можно заменить  $2\sqrt{R_0} \approx 1 + R_0$  в (8), получив выражение

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

где  $\gamma$  — нормировочный коэффициент.

Длина активной среды  $L = 150$  см, точность определения мощности лазера зафиксирована как относительная погрешность в 2% ( $\sigma_y = 0.02y$ ), точность определения  $R_0$  зафиксирована как абсолютная погрешность и составляет 0.01 см ( $\sigma_{R_0} = 0.01$ ).

В ходе физических 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 (7) при  $f(R_0, \omega) = y(R_0, \gamma, \alpha_0, g_0)$ , то есть, вектор параметров  $\omega$  состоит из трех компонент  $\gamma, \alpha_0, g_0$ . Предполагается, что указанные выше погрешности определяют дисперсию соответствующего распределения ошибок измерений:  $\sigma_{y_i} = 0.02y_i$ ,  $\sigma_{x_i} = 0.01$ .

### 5.1 Оптимальные параметры модели

В таблице 2 приведены оптимальные значения параметров модели, а также соответствующие значения MSE, для модифицированного функционала (7). В таблице также приведены для сравнения соответствующие значения, получающиеся при минимизации классического функционала

$$S = \sum_{i=1}^{\ell} (y_i - f(x_i, \omega))^2, \quad (10)$$

а также относительная разность соответствующих величин.

Таблица 2: Оптимальные значения параметров модели.

	$g_0$	$\alpha_0$	$\gamma$	MSE
(7)	$2.932 \cdot 10^{-3}$	$2.219 \cdot 10^{-4}$	99.2	0.282
(10)	$2.917 \cdot 10^{-3}$	$2.129 \cdot 10^{-4}$	101.5	0.183
Разность	$5.2 \cdot 10^{-3}$	$4.2 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$	0.54

## 5.2 Сходимость оптимальных параметров

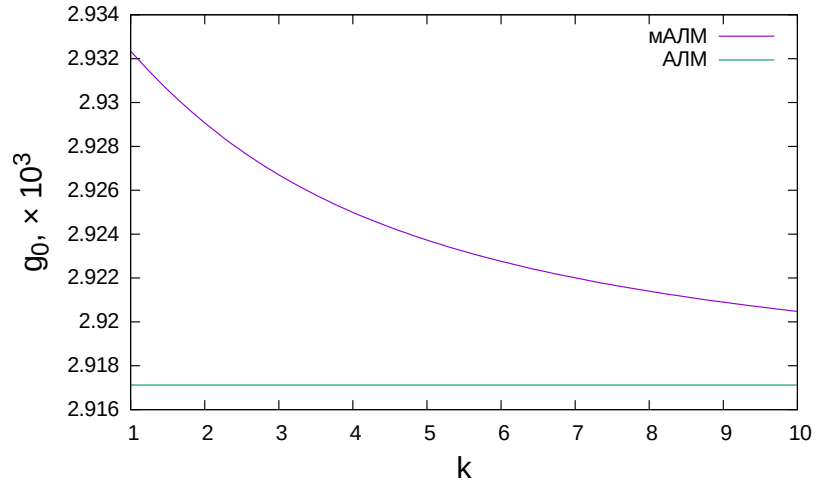
Численно исследована зависимость сходимости параметров  $\omega$  к параметрам  $\omega_0$ , получаемым минимизацией функционала (10).

Разумно предположить, 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 незначительным будет вклад ошибки измерения независимой переменной.

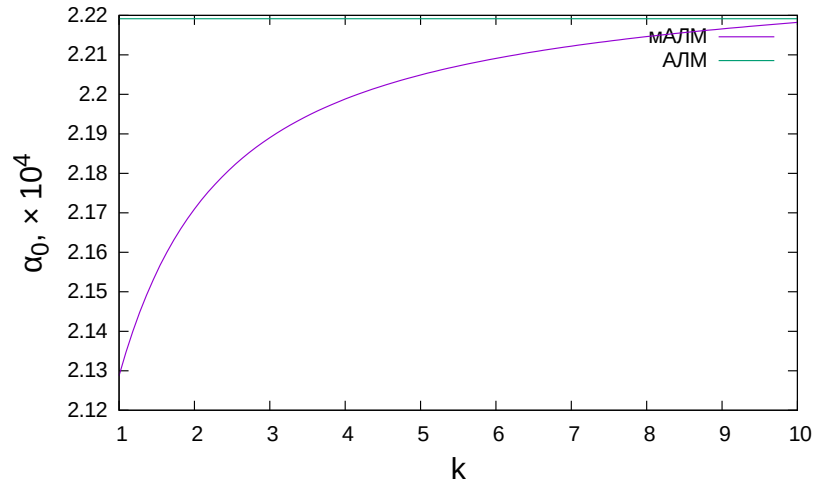
Рассматривается зависимость оптимальных коэффициентов  $g_0$ ,  $\alpha_0$  и  $\gamma$  от погрешности измерения  $y$ , задаваемой как  $k\sigma_y = 0.02ky$ , где  $k$  в одном эксперименте менялось от 1 до 10, а в другом — от 1 до  $10^4$ , в обоих случаях с шагом 0.1. Отметим, что при  $k = 50$  характерная погрешность измерения величины  $y$  сопоставима с самой величиной  $y$ , а при больших значениях  $k$  превышает ее.

Результаты приведены на рис. 1 и 2 соответственно.

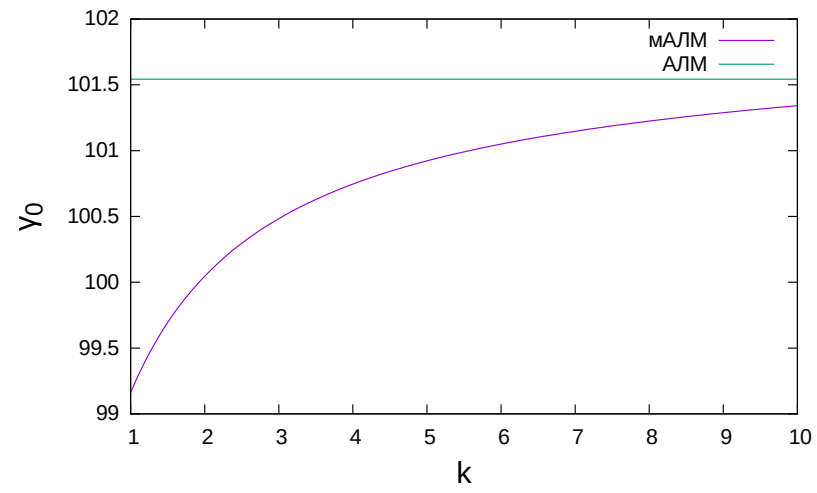
Видно, что для достаточно больших  $k$  оптимальные значения параметров  $\omega$  существенно отличаются от  $\omega_0$ . Соответствующая зависимость имеет сложный характер с несколькими (как минимум, двумя) пересечениями с искомыми оптимальными значениями. Подробное исследование подобной зависимости является целью будущих работ.



(a)  $g_0$

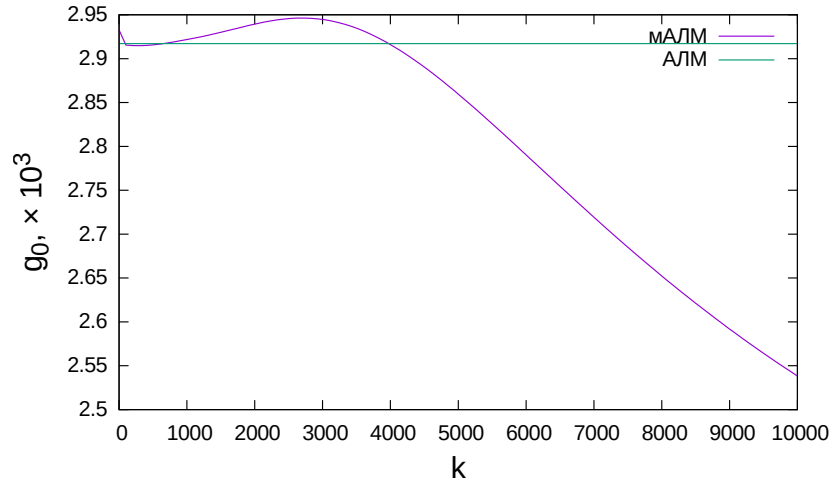


(b)  $\alpha_0$

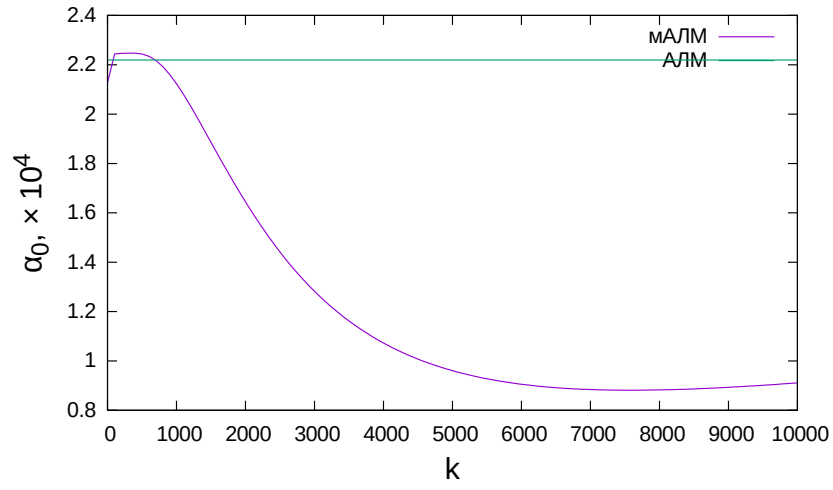


(c)  $\gamma$

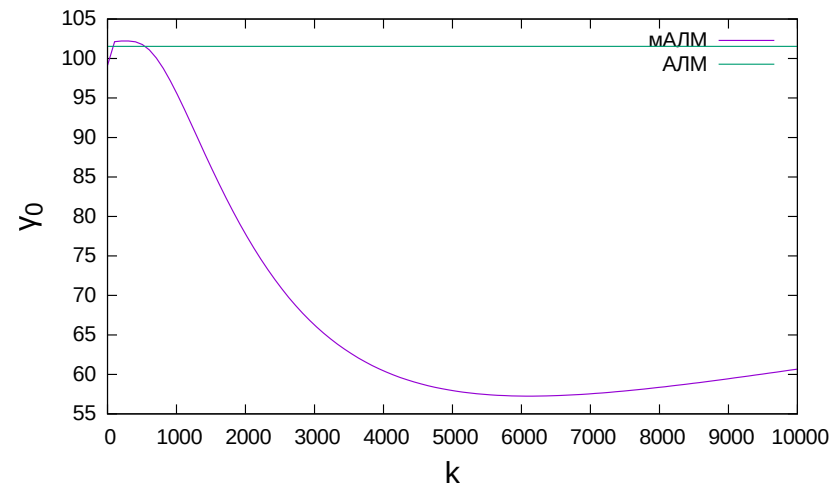
Рис. 1: Зависимость оптимальных параметров от  $\sigma_y = 0.02ky, k \in [1; 10]$ .



(a)  $g_0$



(b)  $\alpha_0$



(c)  $\gamma$

Рис. 2: Зависимость оптимальных параметров от  $\sigma_y = 0.02ky, k \in [1; 10000]$ .



## Список литературы

- [1] 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.
- [2] King, Davis E.: *Dlib-ml: A Machine Learning Toolkit*. Journal of Machine Learning Research, 10:1755–1758, 2009.