

© Измаилов А.Ф., Куренной А.С., Стецюк П.И., 2019

DOI 10.20310/1810-0198-2019-24-125-60-74

УДК 519

## Метод Левенберга–Марквардта для задач безусловной оптимизации

Алексей Феридович ИЗМАИЛОВ<sup>1</sup>, Алексей Святославович КУРЕННОЙ<sup>2</sup>  
Петр Иванович СТЕЦЮК<sup>3</sup>

<sup>1</sup> ФГБОУ ВО «Московский государственный университет им. М. В. Ломоносова»  
119992, ГСП-2, Российская Федерация, г. Москва, Ленинские горы, факультет ВМК  
ORCID: <https://orcid.org/0000-0001-9851-0524>, e-mail: izmaf@ccas.ru

<sup>2</sup> ФГБОУ ВО «Тамбовский государственный университет им. Г.Р. Державина»  
392000, Российская Федерация, г. Тамбов, ул. Интернациональная, 33  
ORCID: <https://orcid.org/0000-0001-8216-9659>, e-mail: akurennoy@cs.msu.ru

<sup>3</sup> Институт кибернетики им. В. М. Глушкова НАН Украины  
03187, Украина, г. Киев, проспект Академика Глушкова, 40  
ORCID: <https://orcid.org/0000-0003-4036-2543>, e-mail: stetsyukp@gmail.com

## Levenberg–Marquardt method for unconstrained optimization

Alexey F. IZMAILOV<sup>1</sup>, Alexey S. KURENNOY<sup>2</sup>, Petr I. STETSYUK<sup>3</sup>

<sup>1</sup> Lomonosov Moscow State University

VMK Faculty, Leninskiye Gory, Moscow 119991, Russian Federation  
ORCID: <https://orcid.org/0000-0001-9851-0524>, e-mail: izmaf@ccas.ru

<sup>2</sup> Tambov State University named after G.R. Derzhavin  
33 Internatsionalnaya St., Tambov 392000, Russian Federation  
ORCID: <https://orcid.org/0000-0001-8216-9659>, e-mail: akurennoy@cs.msu.ru

<sup>3</sup> V. M. Glushkov Institute of Cybernetics of NAS of Ukraine  
40 Akademika Glushkova Ave., Kiev 03187, Ukraine  
ORCID: <https://orcid.org/0000-0003-4036-2543>, e-mail: stetsyukp@gmail.com

**Аннотация.** В работе предлагается и исследуется глобализованный одномерным поиском метод Левенберга–Марквардта для задач безусловной оптимизации с возможно неизолированными решениями. Хорошо известно, что этот метод является эффективным средством решения систем нелинейных уравнений, особенно в случаях наличия вырожденных и даже неизолированных решений. Традиционные способы глобализации сходимости метода Левенберга–Марквардта основаны на одномерном поиске для квадрата евклидовой невязки решаемого уравнения, в роли которого в случае задачи безусловной оптимизации выступает вытекающее из принципа Ферма условие равенства нулю градиента целевой функции. В контексте задач оптимизации такие способы глобализации не вполне адекватны, так как соответствующие алгоритмы не имеют «предпочтений» в плане сходимости к минимумам, максимумам, и вообще любым ста-

ционарным точкам. В связи со этим, в данной работе рассматривается другой способ глобализации сходимости метода Левенберга–Марквардта, использующий одномерный поиск для самой целевой функции исходной задачи. В работе показано, что предложенный алгоритм обладает разумными свойствами глобальной сходимости, а также сохраняет высокую скорость локальной сходимости метода Левенберга–Марквардта в слабых предположениях.

**Ключевые слова:** задач безусловной оптимизации; неизолированные решения; метод Левенберга–Марквардта; глобализация сходимости

**Благодарности:** Работа выполнена при поддержке РФФИ (проекты № 17-01-00125\_a и № 19-51-12003 ННИО\_a) и фонда Volkswagen (грант 90306).

**Для цитирования:** *Измайлов А. Ф., Куренной А. С., Стецюк П. И.* Метод Левенберга–Марквардта для задач безусловной оптимизации // Вестник Тамбовского университета. Серия: естественные и технические науки. Тамбов, 2019. Т. 24. № 125. С. 60–74. DOI 10.20310/1810-0198-2019-24-125-60-74

**Abstract.** We propose and study the Levenberg–Marquardt method globalized by means of linesearch for unconstrained optimization problems with possibly nonisolated solutions. It is well-recognized that this method is an efficient tool for solving systems of nonlinear equations, especially in the presence of singular and even nonisolated solutions. Customary globalization strategies for the Levenberg–Marquardt method rely on linesearch for the squared Euclidean residual of the equation being solved. In case of unconstrained optimization problem, this equation is formed by putting the gradient of the objective function equal to zero, according to the Fermat principle. However, these globalization strategies are not very adequate in the context of optimization problems, as the corresponding algorithms do not have “preferences” for convergence to minimizers, maximizers, or any other stationary points. To that end, in this work we consider a different technique for globalizing convergence of the Levenberg–Marquardt method, employing linesearch for the objective function of the original problem. We demonstrate that the proposed algorithm possesses reasonable global convergence properties, and preserves high convergence rate of the Levenberg–Marquardt method under weak assumptions.

**Keywords:** unconstrained optimization problem; nonisolated solutions; Levenberg–Marquardt method; globalization of convergence

**Acknowledgements:** The work is partially supported by the Russian Foundation for Basic Research (projects no. 17-01-00125\_a and 19-51-12003 ННИО\_a) and by the Volkswagen Foundation (grant 90306).

**For citation:** Izmailov A. F., Kurennoy A. S., Stetsyuk P. I. Metod Levenberga–Markvardta dlya zadach bezuslovnoy optimizatsii [Levenberg–Marquardt Method for Unconstrained Optimization]. *Vestnik Tambovskogo universiteta. Seriya: estestvennye i tekhnicheskie nauki – Tambov University Reports. Series: Natural and Technical Sciences*, 2019, vol. 24, no. 125, pp. 60–74. DOI 10.20310/1810-0198-2019-24-125-60-74 (In Russian, Abstr. in Engl.)

## Введение

В этой работе исследуются численные методы решения задачи безусловной оптимизации

$$f(x) \rightarrow \min, \quad x \in \mathbb{R}^n, \quad (1)$$

с по крайней мере дважды дифференцируемой целевой функцией  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , и при условиях допускающих, что задача (1) может иметь неизолированные решения.

Один успешный подход к решению нелинейных задач с возможно неизолрованными решениями основан на методе Левенберга–Марквардта [1], [2], который, однако, в принципе предназначен для решения систем нелинейных уравнений, а не задач оптимизации. Для уравнения

$$f'(x) = 0, \quad (2)$$

описывающего стационарные точки задачи (1), и для текущего приближения  $x^k \in \mathbb{R}^n$ , этот метод определяет следующее приближение как  $x^{k+1} = x^k + p^k$ , где  $p^k$  является единственным решением линейного уравнения

$$(H_k^2 + \sigma_k I)p = -H_k f'(x^k), \quad (3)$$

в котором базовым выбором симметричной  $n \times n$  матрицы  $H_k$  является  $f''(x^k)$ , а  $\sigma_k > 0$  играет роль параметра регуляризации. Из результатов в [3], [4] следует, что при правильном управлении этим параметром метод Левенберга–Марквардта обладает локальной квадратичной сходимостью к стационарной точке задачи (1) при очень слабых предположениях, допускающих неизолрованность стационарных точек. Более того, сходимость метода может быть глобализована одномерным поиском для квадрата невязки уравнения (2) в роли функции качества.

Однако, в данном (оптимизационном) контексте указанный способ глобализации сходимости метода Левенберга–Марквардта не вполне удовлетворителен, поскольку получаемый глобализованный алгоритм направлен на поиск стационарных точек задачи (1), а не ее решений. В частности, такой алгоритм «не различает» максимумы и минимумы функции  $f$ , как и любые стационарные точки задачи (1). В связи с этим обстоятельством в настоящей работе разрабатывается «более оптимизационный» алгоритм, использующий направления  $p^k$  метода Левенберга–Марквардта, но с одномерным поиском для целевой функции  $f$  задачи (1) вместо невязки уравнения (2).

### 1. Глобализованный алгоритм

Ключевое наблюдение заключается в следующем. Направление  $p^k$  метода Левенберга–Марквардта можно явно выразить из (3) как

$$p^k = -Q_k f'(x^k), \quad (4)$$

где

$$Q_k = (H_k^2 + \sigma_k I)^{-1} H_k.$$

Предположим, что матрица  $H_k$  положительно определена; тогда матрица  $Q_k$  невырождена (как произведение двух невырожденных матриц), и

$$Q_k^{-1} = H_k^{-1}(H_k^2 + \sigma_k I) = H_k + \sigma_k H_k^{-1}. \quad (5)$$

Но полученная таким образом матрица является симметричной и положительно определенной, откуда следует, что таковой является и матрица  $Q_k$ , а значит,  $p^k$  является направлением убывания функции  $f$  в точке  $x^k$ , если  $f'(x^k) \neq 0$ .

Разумеется, матрица  $H_k = f''(x^k)$  может не быть положительно определенной, даже при  $x^k$  сколь угодно близком к множеству решений. Поэтому глобализованный алгоритм должен включать в себя механизм выбора подходящего  $H_k$ , причем допускающий базовый выбор вблизи решений.

**А л г о р и т м.** Выбираем параметры  $\rho_1, \rho_2 > 0$ ,  $\tau_1 > 0$ ,  $\tau_2 > 1$ ,  $\bar{\sigma} > 0$ ,  $q > 0$ ,  $\varepsilon, \theta \in (0, 1)$ . Выбираем  $x^0 \in \mathbb{R}^n$  и полагаем  $k = 0$ .

1. Если  $f'(x^k) = 0$ , стоп.

2. Полагаем  $H_k = f''(x^k)$ . Если

$$\|H_k f'(x^k)\| \geq \rho_1 \|f'(x^k)\|^{\tau_1}, \quad (6)$$

вычисляем  $p^k$  как решение уравнения (3) при  $\sigma_k = \min\{\bar{\sigma}, \|f'(x^k)\|^q\}$ . Иначе переходим к шагу 4.

3. Если

$$\langle f'(x^k), p^k \rangle \leq -\rho_2 \|p^k\|^{\tau_2}, \quad (7)$$

переходим к шагу 5.

4. Вычисляем симметричную  $n \times n$  матрицу  $H_k$  и соответствующее решение  $p^k$  задачи (3), удовлетворяющие (6) и (7).

5. Вычисляем  $\alpha_k = \theta^j$ , где  $j$  — наименьшее неотрицательное целое число, при котором выполняется неравенство Армихо

$$f(x^k + \theta^j p^k) \leq f(x^k) + \varepsilon \theta^j \langle f'(x^k), p^k \rangle. \quad (8)$$

6. Полагаем  $x^{k+1} = x^k + \alpha_k p^k$ , увеличиваем  $k$  на 1, и переходим к шагу 1.

Ключевой вопрос состоит в том как реализовать шаг 4. Очевидно, что (6) может выполняться для любого  $\rho_1 > 0$ , если  $H_k$  достаточно положительно определена, а именно, ее минимальное собственное значение не меньше  $\rho_1$ . Более того, из (4) и (5) следует, что (7) также достигается достаточной положительной определенностью  $H_k$ . Последнего же можно добиться посредством модифицированного разложения Холецкого [5, разд. 4.4.2.2], [6], или посредством модифицированного симметричного знаконеопределенного разложения [7] матрицы Гессе  $f''(x^k)$ ; см. также [8, разд. 3.4]. При этом на каждой итерации алгоритма потребуется решить не более двух систем линейных уравнений, как и в случае, когда для второй системы берется просто  $H_k = \rho I$  с достаточно большим  $\rho > 0$ .

Альтернативным образом можно выбирать  $H_k$  последовательно, заменяя  $H_k$  на  $H_k + \omega I$  с некоторым  $\omega > 0$ , до тех пор, пока (6) и (7) не окажутся выполнены. Разумеется, при таком подходе итерация алгоритма может потребовать решения более двух систем линейных уравнений, но получаемое в результате направление может быть лучше, так как выбранная в итоге матрица  $H_k$  может быть ближе к истинной матрице Гессе функции  $f$ .

## 2. Глобальная сходимость

Следующая теорема описывает свойства глобальной сходимости предложенного алгоритма.

**Теорема 1.** Пусть  $f$  дважды дифференцируема на  $\mathbb{R}^n$ .

Тогда алгоритм корректно определен, и либо останавливается после конечного числа итераций на шаге 1 в некоторой стационарной точке  $x^k$  задачи (1), либо генерирует такую бесконечную последовательность  $\{x^k\}$ , что, в случае ограниченности  $\{H_k\}$ , любая предельная точка этой последовательности является стационарной точкой задачи (1).

**Доказательство.** Корректная определенность алгоритма вытекает из его конструкции.

Допустим, что алгоритм не останавливается на шаге 1, а значит, генерирует бесконечную последовательность  $\{x^k\}$ . Предположим далее, что последовательность  $\{H_k\}$  ограничена. Покажем, что последовательность  $\{p^k\}$  направлений поиска является равномерно градиентной в терминологии [9, с. 24], т. е. если некоторая подпоследовательность  $\{x^{k_j}\}$  сходится к  $\bar{x} \in \mathbb{R}^n$ , причем  $f'(\bar{x}) \neq 0$ , то последовательность  $\{p^{k_j}\}$  ограничена и

$$\limsup_{j \rightarrow \infty} \langle f'(x^{k_j}), p^{k_j} \rangle < 0. \quad (9)$$

Из (7) получаем, что для всех  $k$

$$\|f'(x^k)\| \|p^k\| \geq \rho_2 \|p^k\|^2.$$

Отсюда и из непрерывности  $f'$ , которая автоматически следует из двукратной дифференцируемости  $f$ , вытекает ограниченность последовательности  $\{p^{k_j}\}$ . Кроме того, в силу (3) и (6), для всех  $k$  справедливо соотношение

$$\|(H_k^2 + \sigma_k I)p^k\| = \|H_k f'(x^k)\| \geq \rho_1 \|f'(x^k)\|.$$

Из ограниченности  $\{H_k\}$  и  $\{\sigma_{k_j}\}$  (где последнее следует из определения  $\sigma_k$  и, опять же, непрерывности  $f'$ ) вытекает, что если  $\{p^{k_j}\}$  имеет 0 в качестве предельной точки, то  $f'(\bar{x}) = 0$ , что противоречит сделанному предположению. Следовательно, норма  $p^{k_j}$  отделена от нуля, и тогда из (7) вытекает (9).

Таким образом,  $\{p^k\}$  является равномерно градиентной последовательностью направлений поиска. А значит, согласно [9, теорема 1.8], каждая предельная точка  $\{x^k\}$  является стационарной точкой задачи (1).  $\square$

## 3. Квадратичная скорость сходимости

Обозначим через  $S$  множество стационарных точек задачи (1):

$$S = \{x \in \mathbb{R}^n \mid f'(x) = 0\}.$$

**Лемма 1.** Пусть  $f$  дважды непрерывно дифференцируема в окрестности стационарной точки  $\bar{x}$  задачи (1), и пусть выполняется следующая локальная липшицева оценка расстояния до множества стационарных точек:

$$\text{dist}(x, S) = O(\|f'(x)\|) \quad (10)$$

при  $x \rightarrow \bar{x}$ .

Тогда существует такое  $\rho > 0$ , что

$$\|f''(x)f'(x)\| \geq \rho\|f'(x)\|$$

выполняется для любого  $x \in \mathbb{R}^n$ , достаточно близкого к  $\bar{x}$ .

**Доказательство.** Согласно [10, следствие 2], в сделанных предположениях выполняется

$$\text{dist}(x, S) = O(\|f''(x)f'(x)\|)$$

при  $x \rightarrow \bar{x}$ . Нужное утверждение получается комбинированием данной оценки с оценкой, следующей из непрерывности  $f''$ :

$$\|f'(x)\| = O(\text{dist}(x, S))$$

при  $x \rightarrow \bar{x}$ . □

Таким образом, в предположениях этой леммы, если в алгоритме либо  $\tau_1 > 1$ , либо  $\tau_1 = 1$  и  $\rho_1 > 0$  достаточно мало, то при  $H_k = f''(x^k)$  (6) выполняется для всех  $x^k \in \mathbb{R}^n$  достаточно близких к  $\bar{x}$ .

Во избежание ненужных сложностей, в утверждениях ниже будет удобно формально предполагать, что если  $x^k \in S$  (и, соответственно,  $\sigma_k = 0$ ), то  $p^k = 0$ . Заметим, что если  $x^k \in S$ , то алгоритм останавливается на шаге 1, и вычисления  $p^k$  на самом деле больше не выполняются. С другой стороны,  $p^k = 0$  всегда является решением (3) при  $x^k \in S$ , и для целей анализа удобно предполагать, что в таких случаях выбирается именно это решение.

**Лемма 2.** Пусть в дополнение к предположениям леммы 1 вторая производная  $f$  удовлетворяет условию Липшица в окрестности  $\bar{x}$ .

Тогда для любого  $q \in [1, 2]$  существует такое  $\rho > 0$ , что для решения  $p^k$  уравнения (3) при  $H_k = f''(x^k)$  выполнено

$$\|f''(\bar{x})p^k\| \geq \rho\|p^k\|$$

для всех  $x^k \in \mathbb{R}^n$ , достаточно близких к  $\bar{x}$ .

**Доказательство.** Согласно [4, леммы 2.1, 2.2], в сделанных предположениях выполняются оценки

$$\|p^k\| = O(\text{dist}(x^k, S)), \quad (11)$$

$$\text{dist}(x^k + p^k, S) = O((\text{dist}(x^k, S))^{(2+q)/2}) \quad (12)$$

при  $x^k \rightarrow \bar{x}$ . Точнее, в [4, лемма 2.2] дополнительно предполагается, что  $x^k + p^k$  остается достаточно близким к  $\bar{x}$ . Однако, в доказательстве в [4, теорема 2.1] установлено, что любая необходимая близость  $x^k + p^k$  к  $\bar{x}$  гарантируется достаточной близостью  $x^k$  к  $\bar{x}$ , а значит, это предположение можно опустить.

От противного: предположим, что существует последовательность  $\{x^k\} \subset \mathbb{R}^n$ , сходящаяся к  $\bar{x}$  и такая, что

$$f''(\bar{x})p^k = o(\|p^k\|)$$

при  $k \rightarrow \infty$ . Тогда согласно (11)

$$f''(\bar{x})p^k = o(\text{dist}(x^k, S)), \quad (13)$$

в то время как согласно (12)

$$f'(x^k + p^k) = O(\text{dist}(x^k + p^k, S)) = o(\text{dist}(x^k, S)). \quad (14)$$

С помощью теоремы о среднем, вновь привлекая (11), получаем

$$f'(x^k + p^k) - f'(x^k) - f''(\bar{x})p^k = o(\|p^k\|) = o(\text{dist}(x^k, S)),$$

а значит, принимая во внимание (13) и (14),

$$f'(x^k) = o(\text{dist}(x^k, S))$$

при  $k \rightarrow \infty$ , что противоречит (10). □

**Лемма 3.** Пусть  $f$  дважды дифференцируема в окрестности локального решения  $\bar{x}$  задачи (1), а ее вторая производная удовлетворяет условию Липшица в этой окрестности. Допустим, что выполняется локальная липшицева оценка расстояния (10) при  $x \rightarrow \bar{x}$ .

Тогда для любого  $q \in [1, 2]$  существует такое  $\rho > 0$ , что для решения  $p^k$  уравнения (3) при  $H_k = f''(x^k)$  выполнено

$$\langle f''(\bar{x})p^k, p^k \rangle \geq \rho \|p^k\|^2$$

для всех  $x^k \in \mathbb{R}^n$ , достаточно близких к  $\bar{x}$ .

**Доказательство.** В силу необходимого условия оптимальности второго порядка  $f''(\bar{x})$  является неотрицательно определенной матрицей. Но тогда  $\langle f''(\bar{x})\xi, \xi \rangle > 0$  для всех  $\xi \in \mathbb{R}^n$ , при которых  $f''(\bar{x})\xi \neq 0$ . Требуемый результат теперь получается из леммы 2 с привлечением свойств однородности и компактности. □

Следующий результат вытекает из (12) и леммы 3 посредством рассуждения в [11, следствие 5.1].

**Лемма 4.** В предположениях леммы 3, для каждого  $q \in [1, 2]$ , если либо  $\tau_2 > 2$ , либо  $\tau_2 = 2$  и  $\rho_2 > 0$  достаточно мало, то (7) выполняется для решения  $p^k$  уравнения (3) с  $H_k = f''(x^k)$  при любом  $x^k \in \mathbb{R}^n$ , достаточно близком к  $\bar{x}$ . Более того, если  $\varepsilon \in (0, 1/2)$ , то (8) выполняется при  $j = 0$ , т. е. на шаге 5 алгоритма принимается  $\alpha_k = 1$ .

**Теорема 2.** Пусть в предположениях леммы 3 последовательность  $\{x^k\}$  получена алгоритмом при  $q \in [1, 2]$ ,  $\varepsilon \in (0, 1/2)$  и либо при  $\tau_1 > 1$ , либо при  $\tau_1 = 1$  и достаточно малом  $\rho_1 > 0$ , и либо при  $\tau_2 > 2$ , либо при  $\tau_2 = 2$  и достаточно малом  $\rho_2 > 0$ . Предположим, что при некотором  $k$  приближение  $x^k$  оказывается достаточно близким к  $\bar{x}$ .

Тогда последовательность  $\{x^k\}$  сходится к некоторой стационарной точке задачи (1), и для любого достаточно большого  $k$  выполняется  $H_k = f''(x^k)$ ,  $\alpha_k = 1$ , причем скорость сходимости квадратичная.

**Доказательство.** Это следует из лемм 1 и 4, а также из [4, теорема 2.2], где нужно принять во внимание, что, согласно рассуждению в [4, теорема 2.1], достаточная близость  $x^k$  к  $\bar{x}$  обеспечивает нужную близость всех последующих приближений к  $\bar{x}$ .  $\square$

Покажем теперь, что оценка расстояния (10) обеспечивается выполнением условия квадратичного роста в точке  $\bar{x}$ , что означает существование такого  $\gamma > 0$ , что

$$f(x) - f(\hat{x}) \geq \gamma(\text{dist}(x, S))^2 \quad (15)$$

для любого  $x \in \mathbb{R}^n$ , достаточно близкого к  $\bar{x}$ , и для некоторой проекции  $\hat{x}$  точки  $x$  на  $S$ .

**Предложение 1.** Пусть  $f$  дважды дифференцируема в окрестности стационарной точки  $\bar{x}$  задачи (1), а ее вторая производная непрерывна в  $\bar{x}$ , и предположим, что в точке  $\bar{x}$  выполнено условие квадратичного роста.

Тогда имеет место локальная липшицева оценка расстояния (10) при  $x \rightarrow \bar{x}$ .

**Доказательство.** По теореме о среднем, и принимая во внимание, что  $\hat{x} \rightarrow \bar{x}$  при  $x \rightarrow \bar{x}$ , получаем оценки

$$\begin{aligned} f'(x) - f''(\hat{x})(x - \hat{x}) &= f'(x) - f'(\hat{x}) - f''(\hat{x})(x - \hat{x}) = o(\text{dist}(x, S)), \\ f(x) - f(\hat{x}) - \frac{1}{2}\langle f''(\hat{x})(x - \hat{x}), x - \hat{x} \rangle &= f(x) - f(\hat{x}) - \langle f'(\hat{x}), x - \hat{x} \rangle \\ &\quad - \frac{1}{2}\langle f''(\hat{x})(x - \hat{x}), x - \hat{x} \rangle \\ &= o((\text{dist}(x, S))^2) \end{aligned}$$

при  $x \rightarrow \bar{x}$ . Следовательно, в силу (15),

$$\begin{aligned} \gamma(\text{dist}(x, S))^2 &\leq f(x) - f(\hat{x}) \\ &= \frac{1}{2} \langle f'(x), x - \hat{x} \rangle + o((\text{dist}(x, S))^2) \\ &\leq \frac{1}{2} \|f'(x)\| \|x - \hat{x}\| + o((\text{dist}(x, S))^2) \\ &= \frac{1}{2} \|f'(x)\| \text{dist}(x, S) + o((\text{dist}(x, S))^2), \end{aligned}$$

откуда очевидным образом следует (10).  $\square$

#### 4. Численные примеры

Помимо глобализации метода Левенберга–Марквардта посредством одномерного поиска для квадрата евклидовой невязки уравнения (2), естественным конкурентом предложенного в этой работе алгоритма является регуляризованный метод Ньютона, направление  $p^k$  которого в текущем приближении  $x^k$  вычисляется из линейного уравнений

$$(H_k + \sigma_k I)p = -f'(x^k), \quad (16)$$

а глобализация сходимости осуществляется посредством одномерного поиска для функции  $f$ ; см. [12], [13]. Этот метод также может требовать модификации базового выбора  $H_k = f''(x^k)$  для обеспечения того, чтобы  $p^k$  было направлением убывания функции  $f$  в точке  $x^k$ . Более того, с учетом неотрицательной определенности матрицы  $H_k^2$ , есть основания ожидать, что итерация метода Левенберга–Марквардта будет требовать меньшего количества последовательных модификаций  $H_k$ , а значит, решения меньшего количества систем линейных уравнений. Помимо этого, важным теоретическим преимуществом метода Левенберга–Марквардта является то, что его итерационные системы (3) всегда разрешимы, поскольку их матрицы положительно определены (а значит, невырождены) при  $\sigma_k > 0$ .

Таким образом, в этом разделе рассматриваются следующие алгоритмы.

1. Алгоритм из разд. 1 с последовательной модификацией матрицы  $H_k$  по формуле  $H_k = H_k + \omega I$  на шаге 4, в котором (3) заменено на итерационное уравнение (16) регуляризованного метода Ньютона, а тест (6) опущен, но процедура модификации  $H_k$  иницируется и в тех случаях, когда уравнение (16) решить не удастся. Такой алгоритм соответствует глобализованным версиям регуляризованного метода Ньютона в [12], [13]. Варианты этого алгоритма с выбором  $q = 1$  и  $q = 2$  будем обозначать RNM1 и RNM2, соответственно.

2. Алгоритм из разд. 1 с последовательной модификацией матрицы  $H_k$  по той же формуле. Варианты этого алгоритма с выбором  $q = 1$  и  $q = 2$  будем обозначать L-MM1-obj и L-MM2-obj, соответственно.

3. Алгоритм из разд. 1, но с опущенными тестами (6) и (7), и опущенным шагом 4, а также с функцией  $f$  на шаге 5, замененной на квадрат невязки уравнения (2), т. е.

на функцию  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\varphi(x) = \frac{1}{2} \|f'(x)\|^2.$$

Такой алгоритм соответствует глобализованным версиям метода Левенберга–Марквардта в [3], [4]. Варианты этого алгоритма с выбором  $q = 1$  и  $q = 2$  будем обозначать L-MM1-res и L-MM2-res, соответственно.

В алгоритмах использовались следующие значения параметров:  $\rho_1 = \rho_2 = 10^{-9}$ ,  $\tau_1 = 1.1$ ,  $\tau_2 = 2.1$ ,  $\bar{\sigma} = 1$ ,  $\varepsilon = 0.01$ ,  $\theta = 0.5$ ,  $\omega = 10$ .

Запуск считался успешным, если для некоторого  $k \leq 500$  реализовалось неравенство

$$\|f'(x^k)\| < 10^{-8}.$$

Если на некоторой итерации на шаге 5 алгоритма  $\alpha$  становилось меньше  $10^{-12}$ , то запуск считался неудачным и прекращался.

Для каждой из рассматриваемых ниже трех задач выполнялось 1000 запусков каждого алгоритма из случайных начальных точек, равномерно распределенных в области  $\|x\|_\infty \leq 100$ . Значения столбцов в приводимых ниже таблицах таковы

- «S» (от «Successes»): процент успешных запусков;
- «I» (от «Iterations»): среднее количество итераций на один успешный запуск;
- «LS» (от «Linear Systems»): среднее количество решенных линейных систем на один успешный запуск;
- «OV» (от «Objective function Values»): средняя величина натуральных логарифмов от значений целевой функции в точках завершения алгоритма (вне зависимости от успешности запусков). Эта величина характеризует качество получаемых приближений: чем она меньше, тем в среднем лучше получаемые приближения. При этом она более устойчива по отношению к единичным статистическим выбросам, чем средняя величина самих значений целевой функции.

Таблица 1: Результаты для примера 1

Method	S (%)	I	LS	OV
RNM1	100	32	32	-61.81
RNM2	95	32	32	-61.57
L-MM1-obj	100	32	32	-61.47
L-MM2-obj	100	32	32	-61.64
L-MM1-res	100	32	32	-61.78
L-MM2-res	96	32	32	-59.92

**Пример 1.** Пусть  $n = 2$ ,  $f(x) = ((x_1^2 + x_2^2)^2 - 2(x_1^2 - x_2^2))^2$ . Тогда множество решений задачи (1) — это лемниската Бернулли.

Лемниската Бернулли содержится в круге с центром в нуле радиуса  $\sqrt{2}$ . В определенном смысле это объясняет то, что при запусках из удаленных от нуля начальных точек все рассматриваемые алгоритмы ведут себя в этом примере схожим образом; см. таблицу 1.

Таблица 2: Результаты для примера 2

Method	S (%)	I	LS	OV
RNM1	100	20	21	-44.60
RNM2	100	26	27	-27.36
L-MM1-obj	100	18	18	-53.29
L-MM2-obj	100	18	18	-51.81
L-MM1-res	100	18	18	-53.61
L-MM2-res	100	18	18	-53.16

**Пример 2.** Пусть  $n = 2$ ,  $f(x) = x_1^2 x_2^2$ . Тогда множество решений задачи (1) задается уравнением  $x_1 x_2 = 0$ .

Как видно из таблицы 2, в этом примере алгоритмы, основанные на методе Левенберга–Марквардта, ведут себя схожим образом и превосходят RNM как по эффективности, так и по качеству получаемых приближений. При этом все алгоритмы демонстрируют абсолютную робастность.

Таблица 3: Результаты для примера 3

Method	S (%)	I	LS	OV
RNM1	100	28	28	-27.44
RNM2	100	27	27	-27.54
L-MM1-obj	100	17	17	-57.65
L-MM2-obj	100	19	19	-52.57
L-MM1-res	100	17	17	-57.82
L-MM2-res	99	19	19	-53.45

**Пример 3.** Пусть  $n = 3$ ,  $f(x) = (x_1^2 + x_2^2 - x_3^2)^2$ . Тогда множество решений задачи (1) задается уравнением  $x_1^2 + x_2^2 = x_3^2$ .

Как видно из таблицы 3, в этом примере ситуация та же, что и в примере 2.

Чтобы продемонстрировать различие в поведении двух рассмотренных глобализаций методов Левенберга–Марквардта, рассмотрим еще один простой (одномерный) пример.

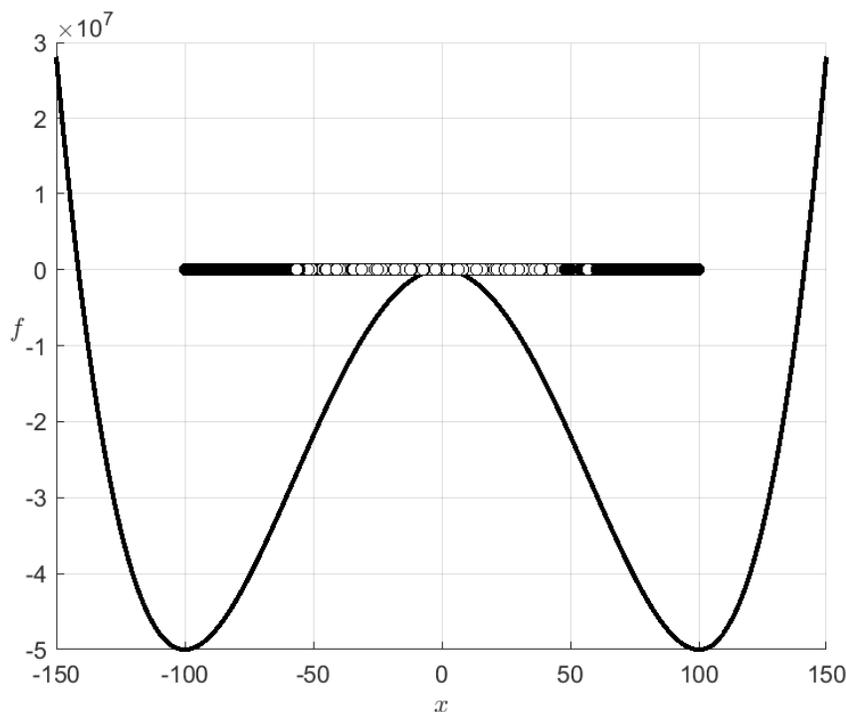


Рис. 1: Пример 4

Таблица 4: Результаты для примера 4

Method	S (%)	I	LS	OV	CS (%)
RNM1	77	6	6	-113.97	100
RNM2	77	5	6	-113.39	100
L-MM1-obj	80	5	6	-112.14	100
L-MM2-obj	80	5	6	-112.24	100
L-MM1-res	100	4	5	-45.72	49
L-MM2-res	100	4	5	-44.80	48

Пример 4. Пусть  $n = 1$ ,  $f(x) = x^4/2 - 10^4 x^2$ . Тогда множество решений задачи (1) состоит из двух точек:  $-100$  и  $100$ . Кроме того, функция  $f$  имеет локальный максимум в точке  $0$  (см. рис. 1).

Помимо данных, сообщавшихся в таблицах для предыдущих примеров, таблица 4 имеет еще столбец «CS» (от «Convergences to Solution»), в котором указан процент

успешных запусков со сходимостью к одному из решений, а именно, закончившихся в точке, значение функции  $f$  в которой отличалось от оптимального значения  $-10^8/2$  не более чем на  $10^{-5}$ .

Из таблицы 4 видно, что алгоритмы с одномерным поиском для целевой функции  $f$  существенно превосходят алгоритмы с одномерным поиском для квадрата невязки уравнения (2) в плане качества получаемых приближений.

На рис. 1 также показано распределение начальных точек, из которых имела место сходимость к одному из решений (черные точки), и тех, из которых имела место сходимость к локальному максимуму (белые точки). Распределение становится весьма сложным вблизи  $\pm 100/\sqrt{3}$ . Данные приведены для L-ММ1-res; для L-ММ2-res картина аналогичная.

Разумеется, приведенные численные результаты предназначены лишь для иллюстрации теории, разработанной в разд. 1–3, и, в частности, не подразумевают каких-либо далеко идущих выводов. Систематическое сравнительное численное тестирование описанных алгоритмов составит одно из направлений дальнейшей работы авторов.

### Список литературы

- [1] K. Levenberg, “A method for the solution of certain non-linear problems in least squares”, *Quarterly of Appl. Math.*, **2** (1944), 164–168.
- [2] D. W. Marquardt, “An algorithm for least-squares estimation of nonlinear parameters”, *J. SIAM*, **11** (1963), 431–441.
- [3] N. Yamashita, M. Fukushima, “On the rate of convergence of the Levenberg–Marquardt method”, *Computing*, 2001, 15, 237–249.
- [4] J.-Y. Fan, Y.-X. Yuan, “On the quadratic convergence of the Levenberg–Marquardt method”, *Computing*, **74** (2005), 23–39.
- [5] P. E. Gill, W. Murray, M. H. Wright, *Practical Optimization*, Academic Press, San Diego, 1981.
- [6] R. B. Schnabel, E. Eskow, “A new modified Cholesky factorization”, *SIAM J. Sci. Statist. Comput.*, **11** (1990), 1136–1158.
- [7] S. H. Cheng, N. J. Higham, “A modified Cholesky algorithm based on a symmetric indefinite factorization”, *SIAM J. Matrix. Anal. Appl.*, **9** (1998), 1097–1110.
- [8] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2, Heidelberg: Springer-Verlag, New York, Berlin, 2006.
- [9] Д. Бертсекас, *Условная оптимизация и методы множителей Лагранжа*, Радио и связь, М., 1987.
- [10] A. Fischer, “Local behavior of an iterative framework for generalized equations with nonisolated solutions”, *Math. Program.*, **94** (2002), 91–124.
- [11] A. F. Izmailov, M. V. Solodov, E. I. Uskov, “Globalizing stabilized SQP by smooth primal-dual exact penalty function”, *J. Optim. Theory Appl.*, **169** (2016), 148–178.
- [12] K. Ueda, N. Yamashita, “Convergence properties of the regularized Newton method for the unconstrained nonconvex optimization”, *Appl. Math. Optim.*, **62** (2010), 27–46.
- [13] C. Shen, X. Chen, Y. Liang, “A regularized Newton method for degenerate unconstrained optimization problems”, *Optim. Lett.*, **6** (2012), 1913–1933.

## References

- [1] K. Levenberg, “A method for the solution of certain non-linear problems in least squares”, *Quarterly of Appl. Math.*, **2** (1944), 164–168.
- [2] D. W. Marquardt, “An algorithm for least-squares estimation of nonlinear parameters”, *J. SIAM*, **11** (1963), 431–441.
- [3] N. Yamashita, M. Fukushima, “On the rate of convergence of the Levenberg–Marquardt method”, *Computing*, 2001, 15, 237–249.
- [4] J.-Y. Fan, Y.-X. Yuan, “On the quadratic convergence of the Levenberg–Marquardt method”, *Computing*, **74** (2005), 23–39.
- [5] P. E. Gill, W. Murray, M. H. Wright, *Practical Optimization*, Academic Press, San Diego, 1981.
- [6] R. B. Schnabel, E. Eskow, “A new modified Cholesky factorization”, *SIAM J. Sci. Statist. Comput.*, **11** (1990), 1136–1158.
- [7] S. H. Cheng, N. J. Higham, “A modified Cholesky algorithm based on a symmetric indefinite factorization”, *SIAM J. Matrix. Anal. Appl.*, **9** (1998), 1097–1110.
- [8] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2, Heidelberg: Springer-Verlag, New York, Berlin, 2006.
- [9] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.
- [10] A. Fischer, “Local behavior of an iterative framework for generalized equations with nonisolated solutions”, *Math. Program.*, **94** (2002), 91–124.
- [11] A. F. Izmailov, M. V. Solodov, E. I. Uskov, “Globalizing stabilized SQP by smooth primal-dual exact penalty function”, *J. Optim. Theory Appl.*, **169** (2016), 148–178.
- [12] K. Ueda, N. Yamashita, “Convergence properties of the regularized Newton method for the unconstrained nonconvex optimization”, *Appl. Math. Optim.*, **62** (2010), 27–46.
- [13] C. Shen, X. Chen, Y. Liang, “A regularized Newton method for degenerate unconstrained optimization problems”, *Optim. Lett.*, **6** (2012), 1913–1933.

## Информация об авторах

**Измаилов Алексей Ферилович**, доктор физико-математических наук, профессор кафедры исследования операций. Московский государственный университет им. М. В. Ломоносова, г. Москва, Российская Федерация. E-mail: izmaf@ccas.ru  
**ORCID:** <https://orcid.org/0000-0001-9851-0524>

**Куренной Алексей Святославович**, кандидат физико-математических наук, научный сотрудник. Тамбовский государственный университет им. Г. Р. Державина, г. Тамбов, Российская Федерация. E-mail: akurennoy@cs.msu.ru  
**ORCID:** <https://orcid.org/0000-0001-8216-9659>

## Information about the authors

**Alexey F. Izmailov**, Doctor of Physics and Mathematics, Professor of the Operations Research Department. Lomonosov Moscow State University, Moscow, the Russian Federation. E-mail: izmaf@ccas.ru  
**ORCID:** <https://orcid.org/0000-0001-9851-0524>

**Alexey S. Kurennoy**, Candidate of Physics and Mathematics, Researcher. Tambov State University named after G. R. Derzhavin, Tambov, the Russian Federation. E-mail: akurennoy@cs.msu.ru  
**ORCID:** <https://orcid.org/0000-0001-8216-9659>

**Стецюк Петр Иванович**, доктор физико-математических наук, заведующий отделом методов негладкой оптимизации. Институт кибернетики им. В. М. Глушкова НАН Украины, г. Киев, Украина. E-mail: stetsyukp@gmail.com  
**ORCID:** <https://orcid.org/0000-0003-4036-2543>

Конфликт интересов отсутствует.

**Для контактов:**

Измаилов Алексей Феридович  
E-mail: izmaf@ccas.ru

Поступила в редакцию 10.01.2019 г.  
Поступила после рецензирования 11.02.2019 г.  
Принята к публикации 14.03.2019 г.

**Petr I. Stetsyuk**, Doctor of Physics and Mathematics, Head of the Nonsmooth Optimization Methods Department. V.M. Glushkov Institute of Cybernetics of NAS of Ukraine, Kiev, Ukraine. E-mail: stetsyukp@gmail.com  
**ORCID:** <https://orcid.org/0000-0003-4036-2543>

There is no conflict of interests.

**Corresponding author:**

Alexey F. Izmailov  
E-mail: izmaf@ccas.ru

Received 10 January 2019  
Reviewed 11 February 2019  
Accepted for press 14 March 2019