

НАУЧНАЯ СТАТЬЯ

© Измаилов А.Ф., Усков Е.И., 2025

<https://doi.org/10.20310/2686-9667-2025-30-149-41-55>

УДК 519.6



Гибридная глобализация сходимости метода Левенберга–Марквардта для задач оптимизации с ограничениями-равенствами

Алексей Феридович ИЗМАЙЛОВ¹,
Евгений Иванович УСКОВ²

¹ ФГБОУ ВО «Московский государственный университет им. М.В. Ломоносова»
119991, ГСП-2, Российская Федерация, г. Москва, Ленинские горы, 1

² ФГБОУ ВО «Тамбовский государственный университет им. Г.Р. Державина»
392000, Российская Федерация, г. Тамбов, ул. Интернациональная, 33

Аннотация. Метод Левенберга–Марквардта обладает локальной сверхлинейной сходимостью для общих систем нелинейных уравнений при слабых предположениях, допускающих неизоллированность решений. Это обосновывает его применение к системам условий первого порядка оптимальности для задач условной оптимизации при возможном нарушении условий регулярности ограничений, что влечет неединственность множителей Лагранжа. Однако, существующие стратегии глобализации сходимости метода Левенберга–Марквардта не являются оптимизационными, т. е., в случае применения к задачам оптимизации, направлены на поиск не решений, а любых стационарных точек таких задач. В этой работе предлагаются оптимизационные стратегии глобализации сходимости метода Левенберга–Марквардта для задач оптимизации с ограничениями-равенствами. Стратегии носят гибридный характер, а именно, основаны на комбинировании глобально сходящегося оптимизационного метода внешней фазы с асимптотическим переключением на метод Левенберга–Марквардта. Установлены свойства глобальной сходимости и сверхлинейная скорость. Приводятся численные результаты, подтверждающие работоспособность предложенных гибридных алгоритмов.

Ключевые слова: задача оптимизации с ограничениями-равенствами, система Лагранжа, метод Левенберга–Марквардта, глобализация с возвратами, глобализация с рекордами

Благодарности: Исследование выполнено за счет гранта Российского научного фонда (проект № 24-21-00015, <https://rscf.ru/project/24-21-00015/>).

Для цитирования: Измаилов А.Ф., Усков Е.И. Гибридная глобализация сходимости метода Левенберга–Марквардта для задач оптимизации с ограничениями-равенствами // Вестник российских университетов. Математика. 2025. Т. 30. № 149. С. 41–55.

<https://doi.org/10.20310/2686-9667-2025-30-149-41-55>

SCIENTIFIC ARTICLE

© A. F. Izmailov, E. I. Uskov, 2025

<https://doi.org/10.20310/2686-9667-2025-30-149-41-55>

Hybrid globalization of convergence of the Levenberg–Marquardt method for equality-constrained optimization problems

Alexey F. IZMAILOV¹, Evgeniy I. USKOV²¹ Lomonosov Moscow State University

1 Leninskiye Gory, Moscow 119991, Russian Federation

² Derzhavin Tambov State University

33 Internatsionalnaya St., Tambov 392000, Russian Federation

Abstract. The Levenberg–Marquardt method possesses local superlinear convergence for general systems of nonlinear equations under weak assumptions allowing for nonisolated solutions. This justifies its application to first-order optimality systems of constrained optimization problems with possibly violated constraint qualifications, the latter giving rise to nonuniqueness of Lagrange multipliers. However, the existing strategies for globalization of convergence of the Levenberg–Marquardt method are not optimization-oriented by nature, i. e., when applied to optimization problems, they are intended not for finding solutions, but rather any stationary points of such problems. In this work, we propose optimization-oriented globalization strategies for the Levenberg–Marquardt method applied to optimization problems with equality constraints. The proposed strategies are hybrid by their character, i. e., they combine a globally convergent optimization outer phase method with asymptotic switching to the Levenberg–Marquardt method. Global convergence properties and superlinear rate of convergence are established. Numerical results are provided, demonstrating that the proposed hybrid algorithms are workable.

Keywords: equality-constrained optimization problem, Lagrange optimality system, Levenberg–Marquardt method, globalization with backups, globalization with records

Acknowledgements: The research was supported by the Russian Science Foundation (project no. 24-21-00015, <https://rscf.ru/en/project/24-21-00015/>).

Mathematics Subject Classification: 47J05, 49M05, 49M15, 65H10, 65K05.

For citation: Izmailov A.F., Uskov E.I. Hybrid globalization of convergence of the Levenberg–Marquardt method for equality-constrained optimization problems. *Vestnik Rossiyskikh universitetov. Matematika = Russian Universities Reports. Mathematics*, **30**:149 (2025), 41–55.

<https://doi.org/10.20310/2686-9667-2025-30-149-41-55>

Введение

Метод последовательного квадратичного программирования (SQP; от английского «Sequential Quadratic Programming») является одним из наиболее эффективных средств численного решения задач оптимизации с функциональными ограничениями; [1, разд. 4.4, 5.4], [2, разд. 4.2, 6.2]. В случае гладкой задачи с ограничениями-равенствами базовый вариант этого метода есть ни что иное, как метод Ньютона, применяемый к системе Лагранжа, характеризующей стационарные точки этой задачи и отвечающие им множители Лагранжа; см. [1, разд. 4.3], [2, разд. 4.1]. Результаты о локальной сверхлинейной сходимости таких методов предполагают выполнение в искомом решении условий регулярности ограничений, которые, в свою очередь, подразумевают единственность множителя Лагранжа. Если же множество множителей не является одноточечным, то оно может устойчивым образом содержать специальные множители, называемые критическими, которые притягивают двойственные последовательности, генерируемые методами SQP, и этот феномен приводит к потере сверхлинейной скорости сходимости; см. [2, разд. 7.1], [3] и цитированную там литературу.

Известны специальные модификации методов SQP, ориентированные на случай возможной неединственности множителей Лагранжа, и позволяющие ослабить указанный эффект притяжения к критическим множителям. Сюда относятся стабилизированный метод SQP [4–8], [2, разд. 7.7.2.], а также его версия со стабилизацией вдоль подпространства [9], в которых двойственные стабилизирующие механизмы препятствуют движению двойственных приближений вдоль множества множителей.

С другой стороны, численные результаты в [10] показывают, что метод Левенберга–Марквардта (LM), являющийся признанным средством поиска потенциально неизолированных решений нелинейных уравнений [11], демонстрирует вполне конкурентоспособное локальное поведение применительно к системе Лагранжа в случае неединственности множителей Лагранжа. В частности, в локальных экспериментах этот метод превосходит упомянутые выше метод SQP, стабилизированный метод SQP и метод SQP стабилизированный вдоль подпространства, как по робастности, так и по эффективности, несмотря на то, что в отличие от этих методов, метод LM предназначен для общих нелинейных уравнений, и никак не использует прямодвойственную структуру системы Лагранжа.

В связи с хорошим локальным поведением метода LM в данном контексте естественным образом возникает вопрос о глобализации сходимости этого метода, которая позволяла бы совместить высокую скорость локальной сходимости с гарантиями разумных свойств возможных предельных точек генерируемых последовательностей. В связи с этим, еще одно важное обстоятельство состоит в том, что подзадачи метода LM всегда разрешимы, в отличие от метода SQP, стабилизированного метода SQP и метода SQP стабилизированного вдоль подпространства, подзадачи которых гарантированно разрешимы только локально, вблизи квалифицированного (удовлетворяющего определенным предположениям) прямодвойственного решения. На самом деле, разрешимость подзадач — это первая трудность, возникающая при попытках глобализации сходимости этих методов. Для метода LM такой проблемы нет.

Способы глобализации сходимости метода LM для общих нелинейных уравнений известны [11] и могут применяться в данном контексте. Однако, эти способы основаны на последовательном уменьшении невязки уравнения, т. е., применительно к системе Лагранжа, направлены на поиск стационарных точек задачи оптимизации, а не ее решений, что,

мягко говоря, не идеально. Например, системе Лагранжа удовлетворяют как минимумы, так и максимумы, и такие глобализованные алгоритмы могут сходиться и к тем, и к другим, и вообще к любым стационарным точкам. С точки зрения теоретических гарантий, исключить возможность сходимости к любым таким точкам невозможно. Тем не менее, несомненный интерес представляют *оптимизационные* стратегии глобализации сходимости метода ЛМ, направленные на поиск именно решений задачи оптимизации (минимумов в задаче (1.1) ниже), и в частности, учитывающие значения целевой функции задачи, а не только невязку системы необходимых условий первого порядка, в которой значения целевой функции вообще не фигурируют, а только ее производная.

Один оптимизационный способ глобализации сходимости метода ЛМ был предложен в [10]; он основан на одномерном поиске для гладкой точной штрафной функции, включающей в себя два штрафных слагаемых (и, соответственно, два параметра штрафа): одно для невязки ограничений, а второе опять же для невязки системы Лагранжа, и это второе слагаемое может опять же играть доминирующую роль, если соответствующий параметр регуляризации бесконечно возрастает, что делает этот подход не вполне удовлетворительным. В этой работе предлагается альтернативный подход, основанный на гибридных стратегиях глобализации сходимости [2, разд. 5.3]. А именно, метод ЛМ определенным образом комбинируется с произвольным оптимизационным методом внешней фазы, обладающим разумными свойствами глобальной сходимости. Будет показано, что при определенных предположениях такие гибридные алгоритмы асимптотически превращаются в метод ЛМ, тем самым наследуя его высокую скорость сходимости. Для стабилизированного метода SQP родственные стратегии глобализации сходимости рассматривались в [12].

В разд. 1. приводится строгая постановка задачи и связанные с ней необходимые теоретические сведения. Разд. 2. содержит изложение метода ЛМ применительно к системе Лагранжа и наиболее точной известной на данный момент теории его локальной сверхлинейной сходимости. В разд. 3. предлагаются два способа гибридной глобализации сходимости этого метода, а именно, глобализация с возвратами и глобализация с рекордами, а также доказываются теоремы о глобальной сходимости и скорости сходимости соответствующих алгоритмов. Наконец, в разд. 4. приводятся результаты вычислительного эксперимента, позволяющие составить представление о практической перспективности гибридной глобализации в данном контексте.

Всюду далее через $\|\cdot\|$ обозначается евклидова норма в пространстве $\mathbb{R}^p$ соответствующей размерности p . Также будут использоваться нормы $\|\cdot\|_1$ и $\|\cdot\|_\infty$, т. е. сумма модулей компонент вектора и максимум из модулей его компонент, соответственно. Через $\text{dist}(v, U) = \inf_{u \in U} \|v - u\|$ обозначается расстояние от точки $v \in \mathbb{R}^p$ до множества $U \subset \mathbb{R}^p$, а через $B(v, \delta) = \{u \in \mathbb{R}^p \mid \|u - v\| \leq \delta\}$ — шар радиуса $\delta > 0$ с центром в точке v . Через I обозначается единичная матрица, размер которой всегда ясен из контекста, а через $\ker A$ и $\text{im } A$ — ядро (множество нулей) и образ (множество значений) линейного оператора A , соответственно. Для последовательности $\{u^k\} \subset \mathbb{R}^p$, сходящейся к некоторому $u \in \mathbb{R}^p$, будем говорить, что сходимость имеет Q -порядок $q > 1$, если существует $C > 0$ такое, что $\|u^{k+1} - u\| \leq C\|u^k - u\|^q$ для всех достаточно больших k .

1. Постановка задачи и предварительные сведения

Будем рассматривать задачу оптимизации с ограничениями-равенствами

$$f(x) \rightarrow \min, \quad h(x) = 0, \quad (1.1)$$

где функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ и отображение $h: \mathbb{R}^n \rightarrow \mathbb{R}^l$ по крайней мере дважды дифференцируемы.

Стационарные точки задачи (1.1) и отвечающие им множители Лагранжа характеризуются системой Лагранжа

$$\frac{\partial L}{\partial x}(x, \lambda) = 0, \quad h(x) = 0, \quad (1.2)$$

где $L: \mathbb{R}^n \times \mathbb{R}^l \rightarrow \mathbb{R}$ — функция Лагранжа задачи (1.1):

$$L(x, \lambda) = f(x) + \langle \lambda, h(x) \rangle.$$

Классический принцип Лагранжа состоит в том, что любое локальное решение $\bar{x}$ задачи (1.1), в котором выполнено условие регулярности ограничений

$$\text{rank } h'(\bar{x}) = l, \quad (1.3)$$

является стационарной точкой этой задачи, т. е. вместе с некоторым множителем Лагранжа $\bar{\lambda} \in \mathbb{R}^l$ удовлетворяет системе (1.2). Более того, (1.3) является необходимым и достаточным условием для единственности отвечающего $\bar{x}$ множителя Лагранжа $\bar{\lambda}$. В этой работе, однако, особый интерес представляет случай, когда множество множителей Лагранжа

$$\Lambda(\bar{x}) = \left\{ \lambda \in \mathbb{R}^l \mid \frac{\partial L}{\partial x}(\bar{x}, \lambda) = 0 \right\},$$

отвечающих стационарной точке $\bar{x}$ задачи (1.1), может не быть одноточечным, т. е. когда условие регулярности ограничений (1.3) может нарушаться, и, в частности, когда система (1.2) может иметь неизолированные решения.

В дальнейшем будет использоваться локальная липшицева оценка расстояния до множества $\{\bar{x}\} \times \Lambda(\bar{x})$ в стационарной точке $\bar{x}$ задачи (1.1) для множителя $\bar{\lambda} \in \Lambda(\bar{x})$, а именно, следующее свойство:

$$\|x - \bar{x}\| + \text{dist}(\lambda, \Lambda(\bar{x})) = O(\|\Phi(x, \lambda)\|) \quad (1.4)$$

при $(x, \lambda) \rightarrow (\bar{x}, \bar{\lambda})$, где отображение $\Phi: \mathbb{R}^n \times \mathbb{R}^l \rightarrow \mathbb{R}^n \times \mathbb{R}^l$ задается равенством

$$\Phi(x, \lambda) = \left(\frac{\partial L}{\partial x}(x, \lambda), h(x) \right). \quad (1.5)$$

Это свойство равносильно тому, что множитель $\bar{\lambda}$ является некритическим, т. е. условия

$$\xi \in \ker h'(\bar{x}), \quad \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi \in \text{im}(h'(\bar{x}))^\top$$

выполняются только с $\xi = 0$; см. [2, предложение 1.43]. Элементарно проверяется, что некритичность $\bar{\lambda}$ всегда имеет место при выполнении достаточного условия второго порядка оптимальности

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle > 0 \quad \forall \xi \in \ker h'(\bar{x}) \setminus \{0\}, \quad (1.6)$$

но обратная импликация не имеет места: $\bar{\lambda}$ может быть некритическим и при нарушении (1.6).

2. Метод Левенберга–Марквардта

Система Лагранжа (1.2) может быть записана в виде

$$\Phi(x, \lambda) = 0, \quad (2.1)$$

с отображением Φ , введенным в (1.5). Заметим, что Φ — это полный градиент функции Лагранжа L , и, в частности, матрица Якоби этого отображения в любой точке симметрична. С учетом этого обстоятельства, для текущего приближения $(x^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^l$, метод ЛМ для уравнения (2.1) (т. е. для системы Лагранжа (1.2)) определяет следующее приближение как $(x^{k+1}, \lambda^{k+1}) = (x^k + \xi^k, \lambda^k + \eta^k)$, где $v^k = (\xi^k, \eta^k)$ есть решение линейного уравнения

$$\Phi'(x^k, \lambda^k)\Phi(x^k, \lambda^k) + ((\Phi'(x^k, \lambda^k))^2 + \sigma(x^k, \lambda^k)I)v = 0 \quad (2.2)$$

относительно $v = (\xi, \eta)$. Здесь $\sigma(x^k, \lambda^k)$ — параметр регуляризации, который обычно определяется как некоторая степень невязки уравнения (2.1), т. е. с использованием функции $\sigma: \mathbb{R}^n \times \mathbb{R}^l \rightarrow \mathbb{R}_+$,

$$\sigma(x, \lambda) = \|\Phi(x, \lambda)\|^\theta, \quad (2.3)$$

где $\theta > 0$ — фиксированный параметр. При этом, если (x^k, λ^k) не является решением уравнения (2.1), то $\sigma(x^k, \lambda^k) > 0$, и уравнение (2.2) имеет единственное решение. Если же (x^k, λ^k) является решением уравнения (2.1), то будет удобно далее использовать соглашение, что в этом случае метод определяет $(x^{k+1}, \lambda^{k+1}) = (x^k, \lambda^k)$.

Из [11, теорема 3.2] и полученных при выводе этой теоремы промежуточных оценок (в частности, [11, (3.16)]) вытекает следующий результат о локальной сверхлинейной сходимости метода ЛМ.

Теорема 2.1. Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ и отображение $h: \mathbb{R}^n \rightarrow \mathbb{R}^l$ дважды дифференцируемы в окрестности стационарной точки $\bar{x}$ задачи (1.1), причем их вторые производные удовлетворяют условию Липшица вблизи $\bar{x}$. Пусть $\bar{\lambda} \in \Lambda(\bar{x})$ — некритический множитель. Пусть функция $\sigma: \mathbb{R}^n \times \mathbb{R}^l \rightarrow \mathbb{R}_+$ определена в (2.3) с некоторым фиксированным $\theta \in (0, 2]$.

Тогда для любой начальной точки $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^l$ метод ЛМ однозначно определяет последовательность $\{(x^k, \lambda^k)\}$. Более того, найдется функция $\varphi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ такая, что $\varphi(t) = O(t^{\min\{\theta+1, 2\}})$ при $t \rightarrow 0+$, и для любого $\delta > 0$ и любой начальной точки (x^0, λ^0) , достаточно близкой к $(\bar{x}, \bar{\lambda})$, последовательность $\{(x^k, \lambda^k)\}$ содержится в $B((\bar{x}, \bar{\lambda}), \delta)$ и сходится к $(\bar{x}, \lambda^*)$ с некоторым $\lambda^* \in \Lambda(\bar{x})$, причем скорость сходимости сверхлинейная с Q -порядком $\min\{\theta + 1, 2\}$, и

$$\|x^{k+1} - \bar{x}\| + \text{dist}(\lambda^{k+1}, \Lambda(\bar{x})) \leq \varphi(\|x^k - \bar{x}\| + \text{dist}(\lambda^k, \Lambda(\bar{x}))). \quad (2.4)$$

Заметим, что в сделанных предположениях для введенного в (1.5) отображения Φ функция $\|\Phi(\cdot)\|$ удовлетворяет условию Липшица в окрестности множества $\{\bar{x}\} \times \mathbb{R}^l$. Отсюда следует, что

$$\|\Phi(x, \lambda)\| = O(\|x - \bar{x}\| + \text{dist}(\lambda, \Lambda(\bar{x}))) \quad (2.5)$$

при $(x, \lambda) \rightarrow (\bar{x}, \bar{\lambda})$.

Комбинируя оценки (1.4) (эквивалентную некритичности множителя $\bar{\lambda}$) и (2.5) с оценкой (2.4), получаем такое следствие теоремы 2.1, которое потребуется в дальнейшем анализе.

Следствие 2.1. В предположениях теоремы 2.1 найдется функция $\varphi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ такая, что $\varphi(t) = o(t)$ при $t \rightarrow 0+$, и для любой точки $(x^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^l$, достаточно близкой к $(\bar{x}, \bar{\lambda})$, итерация метода LM определяет приближение (x^{k+1}, λ^{k+1}) , для которого выполняется

$$\|\Phi(x^{k+1}, \lambda^{k+1})\| \leq \varphi(\|\Phi(x^k, \lambda^k)\|). \quad (2.6)$$

3. Гибридная глобализация

Ниже предлагаются два способа гибридной глобализации метода LM: алгоритм с возвратами и алгоритм с рекордами. Идея алгоритма с возвратами заключается в том, что на каждой итерации из текущей точки (x^k, λ^k) делается шаг метода LM, который принимается, если невязка системы Лагранжа в полученной пробной точке не превышает заданной доли от ее невязки в точке (x^k, λ^k) . Если на какой-то из последующих итераций шаг метода LM не принимается, то происходит возврат к той точке, откуда был сделан первый в данной серии шагов метода LM, и из этой точки делается шаг метода внешней фазы.

Алгоритм 1. Фиксируем параметры $\theta > 0$ и $\rho \in (0, 1)$. Выбираем $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^l$ и полагаем $k = 0$.

1. Полагаем $\hat{k} = k$ и $(\hat{x}, \hat{\lambda}) = (x^k, \lambda^k)$.
2. Если $\Phi(x^k, \lambda^k) = 0$, стоп.
3. Вычисляем $v^k = (\xi^k, \eta^k)$ как решение линейной системы (2.2) с функцией $\sigma: \mathbb{R}^n \times \mathbb{R}^l \rightarrow \mathbb{R}_+$, определенной в (2.3).
4. Если

$$\|\Phi(x^k + \xi^k, \lambda^k + \eta^k)\| \leq \rho \|\Phi(x^k, \lambda^k)\|, \quad (3.1)$$

то полагаем $(x^{k+1}, \lambda^{k+1}) = (x^k + \xi^k, \lambda^k + \eta^k)$, увеличиваем номер шага k на 1, и переходим к п. 2. (Шаг метода LM.)

5. Если $k > \hat{k}$ (т. е. текущая точка (x^k, λ^k) сгенерирована с помощью шага метода LM), то восстанавливаем сохраненные значения, полагая $k = \hat{k}$ и $(x^k, \lambda^k) = (\hat{x}, \hat{\lambda})$.
6. Вычисляем (x^{k+1}, λ^{k+1}) с помощью шага метода внешней фазы, увеличиваем номер шага k на 1 и переходим к п. 1. (Шаг метода внешней фазы.)

Свойства глобальной сходимости алгоритма 1 вытекают из следующих наблюдений. Могут реализоваться только два возможных сценария: либо все итерации осуществляются с помощью метода внешней фазы, и при этом алгоритм 1 наследует свойства глобальной сходимости этого метода, либо все итерации, начиная с некоторого номера k , являются итерациями метода LM, и тогда из (3.1) вытекает следующая теорема.

Теорема 3.1. Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ и отображение $h: \mathbb{R}^n \rightarrow \mathbb{R}^l$ дважды дифференцируемы на $\mathbb{R}^n$. Пусть последовательность $\{(x^k, \lambda^k)\}$ сгенерирована алгоритмом 1, и пусть все члены этой последовательности, начиная с некоторого, получены шагами метода LM.

Тогда

$$\Phi(x^k, \lambda^k) \rightarrow 0$$

при $k \rightarrow \infty$, и, в частности, если $(\bar{x}, \bar{\lambda})$ является предельной точкой последовательности $\{(x^k, \lambda^k)\}$, то $\bar{x}$ — стационарная точка задачи (1.1), а $\bar{\lambda}$ — отвечающий ей множитель Лагранжа.

В следующей теореме устанавливается сверхлинейная скорость сходимости алгоритма 1.

Теорема 3.2. Пусть в предположениях теоремы 2.1 точка $(\bar{x}, \bar{\lambda})$ является предельной точкой последовательности $\{(x^k, \lambda^k)\}$, сгенерированной алгоритмом 1.

Тогда вся последовательность $\{(x^k, \lambda^k)\}$ сходится к $(\bar{x}, \bar{\lambda})$ со сверхлинейной скоростью.

Доказательство. Из следствия 2.1 вытекает, что найдутся окрестность U точки $(\bar{x}, \bar{\lambda})$ и функция $\varphi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ такие, что $\varphi(t) = o(t)$ при $t \rightarrow 0+$, и для любой точки $(x^k, \lambda^k) \in U$ итерация метода LM определяет приближение (x^{k+1}, λ^{k+1}) , для которого выполняется неравенство (2.6). Кроме того, из непрерывности Φ следует, что, сужая при необходимости окрестность U , можно гарантировать, что для любой точки $(x, \lambda) \in U$ справедливо неравенство

$$\varphi(\|\Phi(x, \lambda)\|) \leq \rho \|\Phi(x, \lambda)\|. \quad (3.2)$$

Далее, согласно теореме 2.1, найдется окрестность V точки $(\bar{x}, \bar{\lambda})$ такая, что для любой начальной точки из V метод LM определяет последовательность, содержащуюся в U . Поскольку $(\bar{x}, \bar{\lambda})$ является предельной точкой последовательности $\{(x^k, \lambda^k)\}$, то для некоторого k выполняется $(x^k, \lambda^k) \in V$, а значит, метод LM, запущенный из точки (x^k, λ^k) как из начальной точки, определяет последовательность, содержащуюся в U . Из неравенств (2.6) и (3.2) при этом вытекает выполнение вдоль этой последовательности условия (3.1). Поэтому все последующие итерации алгоритма 1 являются итерациями метода LM, и сходимость $\{(x^k, \lambda^k)\}$ к $(\bar{x}, \bar{\lambda})$ со сверхлинейной скоростью следует из теоремы 2.1. $\square$

Алгоритм с возвратами имеет по крайней мере один концептуальный недостаток, состоящий в том, что ранние попытки переключения на метод LM могут приводить к напрасным вычислительным затратам на итерации метода LM, результаты которых впоследствии отбрасываются. В связи с этим рассмотрим альтернативную стратегию гибридной глобализации, в которой вместо того, чтобы сравнивать невязку системы Лагранжа в пробной точке с невязкой, полученной на предыдущей итерации, будем сравнивать ее с рекордом, т. е. с наименьшим значением невязки на всех предыдущих итерациях, и принимать шаг метода LM только в том случае, если невязка в пробной точке не превышает заданной доли от рекорда.

Алгоритм 2. Фиксируем параметры $\theta > 0$ и $\rho \in (0, 1)$. Выбираем $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^l$ и полагаем $k = 0$. Вычисляем $\varphi_{\text{rec}} = \|\Phi(x^0, \lambda^0)\|$.

1. Если $\Phi(x^k, \lambda^k) = 0$, стоп.
2. Вычисляем $v^k = (\xi^k, \eta^k)$ как решение линейной системы (2.2) с функцией $\sigma: \mathbb{R}^n \times \mathbb{R}^l \rightarrow \mathbb{R}_+$, определенной в (2.3).
3. Если

$$\|\Phi(x^k + \xi^k, \lambda^k + \eta^k)\| \leq \rho \varphi_{\text{rec}}, \quad (3.3)$$

то полагаем $(x^{k+1}, \lambda^{k+1}) = (x^k + \xi^k, \lambda^k + \eta^k)$, $\varphi_{\text{rec}} = \|\Phi(x^{k+1}, \lambda^{k+1})\|$, увеличиваем номер шага k на 1 и переходим к п. 1. (Шаг метода LM.)

4. Вычисляем (x^{k+1}, λ^{k+1}) с помощью шага метода внешней фазы. Если

$$\|\Phi(x^{k+1}, \lambda^{k+1})\| < \varphi_{\text{rec}},$$

то полагаем $\varphi_{\text{rec}} = \|\Phi(x^{k+1}, \lambda^{k+1})\|$. Увеличиваем номер шага k на 1 и переходим к п. 1. (Шаг метода внешней фазы.)

Как и для алгоритма 1, возможны лишь два сценария: либо все итерации алгоритма 2, начиная с некоторой, осуществляются с помощью метода внешней фазы, либо бесконечное число итераций алгоритма 2 являются итерациями метода LM. В первом случае алгоритм 2 наследует свойства глобальной сходимости метода внешней фазы. Во втором случае из условия (3.3) легко выводится следующий результат о глобальной сходимости алгоритма 2.

Теорема 3.3. Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ и отображение $h: \mathbb{R}^n \rightarrow \mathbb{R}^l$ дважды дифференцируемы на $\mathbb{R}^n$. Пусть последовательность $\{(x^k, \lambda^k)\}$ сгенерирована алгоритмом 2, и пусть все члены (x^{k_j}, λ^{k_j}) этой последовательности получены шагами метода LM для бесконечного числа номеров j .

Тогда

$$\Phi(x^{k_j}, \lambda^{k_j}) \rightarrow 0$$

при $j \rightarrow \infty$, и, в частности, если $(\bar{x}, \bar{\lambda})$ является предельной точкой последовательности $\{(x^{k_j}, \lambda^{k_j})\}$, то $\bar{x}$ — стационарная точка задачи (1.1), а $\bar{\lambda}$ — отвечающий ей множитель Лагранжа.

Свойства глобальной сходимости алгоритма 2, установленные в теореме 3.3, несколько слабее соответствующих свойств алгоритма 1, установленных в теореме 3.1, но тем не менее, они вполне разумны. А именно, если вместо условия останова в п. 1 алгоритма 2 использовать практическое правило, состоящее в достаточной малости $\|\Phi(x^k, \lambda^k)\|$, то либо алгоритм успешно завершается после конечного числа итераций, либо все его итерации начиная с некоторой являются шагами метода внешней фазы.

Результат о сверхлинейной скорости сходимости алгоритма 2 использует несколько более сильные предположения, чем для алгоритма 1. А именно, в отличие от теоремы 3.2, в следующей теореме предполагается (а не утверждается) сходимость всей сгенерированной алгоритмом последовательности.

Теорема 3.4. Пусть в предположениях теоремы 2.1 последовательность $\{(x^k, \lambda^k)\}$, сгенерированная алгоритмом 2, сходится к точке $(\bar{x}, \bar{\lambda})$.

Тогда скорость сходимости сверхлинейная.

До к а з а т е л ь с т в о. Если рекордное значение φ_{rec} меняется бесконечное число раз, то в последовательности $\{(x^k, \lambda^k)\}$ найдется точка (x^k, λ^k) , сколь угодно близкая к $(\bar{x}, \bar{\lambda})$, для которой $\varphi_{\text{rec}} = \|\Phi(x^k, \lambda^k)\|$. Тогда аналогично доказательству теоремы 3.2 демонстрируется, что точка (x^{k+1}, λ^{k+1}) , полученная шагом метода LM, принимается алгоритмом 2, как и на всех последующих итерациях, и требуемый результат получается применением теоремы 2.1.

Если же предположить, что начиная с некоторой итерации рекордное значение $\varphi_{\text{rec}} > 0$ не меняется (что, в частности, подразумевает, что начиная с этой итерации шаги метода LM не принимаются алгоритмом 2), то приходим к противоречию. Действительно

$\Phi(x^k, \lambda^k) \rightarrow 0$ при $k \rightarrow \infty$, поскольку $\{(x^k, \lambda^k)\}$ сходится к точке $(\bar{x}, \bar{\lambda})$, отображение Φ непрерывно в этой точке, и $\Phi(\bar{x}, \bar{\lambda}) = 0$, и поэтому для достаточно большого k условие (3.3) должно выполняться, каким бы ни было фиксированное значение $\varphi_{\text{гес}} > 0$, а значит, рекордное значение $\varphi_{\text{гес}}$ должно измениться. $\square$

4. Численные результаты

Численный эксперимент проводился на всех задачах с ограничениями-равенствами из коллекции DEGEN¹⁾, в которой собраны задачи оптимизации с нерегулярными ограничениями в некотором известном решении, и, соответственно, с неединственным множителем Лагранжа, отвечающим этому решению. Всего использовалось 35 задач.

Сравнение проводилось для следующих алгоритмов.

- QNSQP — глобализованный посредством одномерного поиска квазиньютоновский метод SQP, приводимый ниже как алгоритм 3.
- LM backups — алгоритм 1, в котором в качестве метода внешней фазы используется QNSQP.
- LM records 1 — алгоритм 2, в котором в качестве метода внешней фазы используется QNSQP, причем если после принятого шага метода LM осуществлялся шаг метода внешней фазы, то он выполнялся с матрицей $H_k = I$ (см. описание метода внешней фазы ниже).
- LM records 2 — то же, что и LM records 1, но со следующим отличием: если после принятого шага метода LM осуществлялся шаг метода внешней фазы, то он выполнялся с матрицей H_k , получаемой по формуле BFGS с модификацией Пауэлла (см. ниже) из последней матрицы, использовавшейся на шаге метода внешней фазы.

Вычисления проводились в среде MATLAB, с использованием стандартных средств этой среды для решения систем линейных уравнений.

Конкретный вариант алгоритма QNSQP, использовавшийся в экспериментах, соответствует [1, алгоритм 5.4.1], с вычислением аппроксимаций H_k матрицы Гессе функции Лагранжа по правилу Бroyдена–Флетчера–Голдфарба–Шанно (BFGS) с модификацией Пауэлла (см. [2, р. 213]), с начальной матрицей $H_0 = I$. Определим семейство штрафных функций $\varphi_c : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\varphi_c(x) = f(x) + c\|h(x)\|_1.$$

Алгоритм 3. Фиксируем параметры $\bar{c} > 0$, $\bar{\delta} > 0$ и $\varepsilon, \varkappa, \nu \in (0, 1)$, и полагаем $H_0 = I$. Выбираем $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^l$ и полагаем $k = 0$.

1. Если $\Phi(x^k, \lambda^k) = 0$, стоп.
2. Вычисляем $v^k = (\xi^k, \eta^k)$ как решение линейной системы

$$\begin{pmatrix} \frac{\partial L}{\partial x}(x^k, \lambda^k) \\ h(x^k) \end{pmatrix} + \begin{pmatrix} H_k & (h'(x^k))^\top \\ h'(x^k) & 0 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = 0.$$

¹⁾ http://w3.impa.br/~optim/DEGEN_collection.zip

Иными словами, ξ^k и $\lambda^k + \eta^k$ есть стационарная точка и отвечающий ей множитель Лагранжа задача квадратичного программирования

$$\langle f'(x^k), \xi \rangle + \frac{1}{2} \langle H_k \xi, \xi \rangle \rightarrow \min, \quad h(x^k) + h'(x^k) \xi = 0.$$

Если v^k вычислить не удастся, или если $v^k = 0$, стоп.

3. Полагаем $\lambda^{k+1} = \lambda^k + \eta^k$,

$$c_k = \begin{cases} \|\lambda^{k+1}\|_\infty + \bar{c} + \bar{\delta}, & \text{если } k = 0 \text{ или } c_{k-1} < \|\lambda^{k+1}\|_\infty + \bar{c}, \\ c_{k-1}, & \text{иначе,} \end{cases}$$

$$\Delta_k = \langle f'(x^k), \xi^k \rangle - c_k \|h(x^k)\|_1.$$

4. Полагаем $\alpha = 1$. Если выполняется неравенство

$$\varphi_{c_k}(x^k + \alpha \xi^k) \leq \varphi_{c_k}(x^k) + \varepsilon \alpha \Delta_k, \quad (4.1)$$

полагаем $\alpha_k = \alpha$. В противном случае заменяем α на $\varkappa \alpha$ и проверяем снова неравенство (4.1) до тех пор, пока оно не выполнится, после чего полагаем $\alpha_k = \alpha$.

5. Полагаем $x^{k+1} = \alpha_k \xi^k$,

$$s^k = x^{k+1} - x^k, \quad r^k = \frac{\partial L}{\partial x}(x^{k+1}, \lambda^{k+1}) - \frac{\partial L}{\partial x}(x^k, \lambda^{k+1}),$$

$$\tau_k = \begin{cases} 1, & \text{если } \langle r^k, s^k \rangle \geq \nu \langle H_k s^k, s^k \rangle, \\ \frac{(1 - \nu) \langle H_k s^k, s^k \rangle}{\langle H_k s^k, s^k \rangle - \langle r^k, s^k \rangle}, & \text{иначе,} \end{cases}$$

$$\tilde{r}^k = \tau_k r^k + (1 - \tau_k) H_k s^k,$$

$$H_{k+1} = H_k + \frac{\tilde{r}^k (\tilde{r}^k)^\top}{\langle \tilde{r}^k, s^k \rangle} - \frac{H_k s^k (H_k s^k)^\top}{\langle H_k s^k, s^k \rangle},$$

увеличиваем номер шага k на 1 и переходим к п. 1.

Параметры в алгоритме 3 были выбраны следующим образом: $\bar{c} = \bar{\delta} = 1$, $\varepsilon = 0.01$, $\varkappa = 0.5$, $\nu = 0.2$. В алгоритмах 1 и 2 использовался параметр $\rho = 0.9$. Кроме того, в этих алгоритмах правило (2.3) для параметра регуляризации метода ЛМ было модифицировано следующим образом:

$$\sigma(x, \lambda) = \min\{0.1, \|\Phi(x, \lambda)\|^\theta\},$$

поскольку большие значения параметра регуляризации в удаленных от решений точках (с большими невязками) могут приводить к слишком коротким шагам метода ЛМ.

Запуск считался неудачным, если условие остановки

$$\|\Phi(x^k, \lambda^k)\| \leq 10^{-8}$$

ни разу не выполнялось за 500 итераций, или если очередная итерация метода внешней фазы оказывалась неудачной (а именно, если неудачей заканчивалось решение какой-либо из итерационных линейных систем метода для определения направления (ξ^k, η^k) , либо если в процессе одномерного поиска в алгоритме 3 получалось пробное значение параметра α длины шага, для которого $\alpha \|\xi^k\| \leq 10^{-12}$).

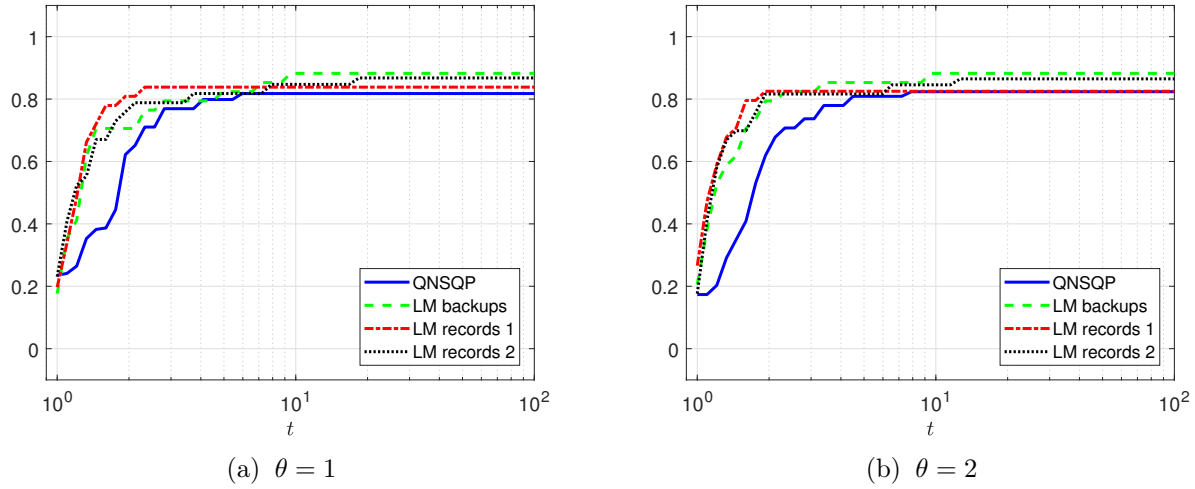


Рис. 1. Алгоритм 3 на внешней фазе; RSP = 1

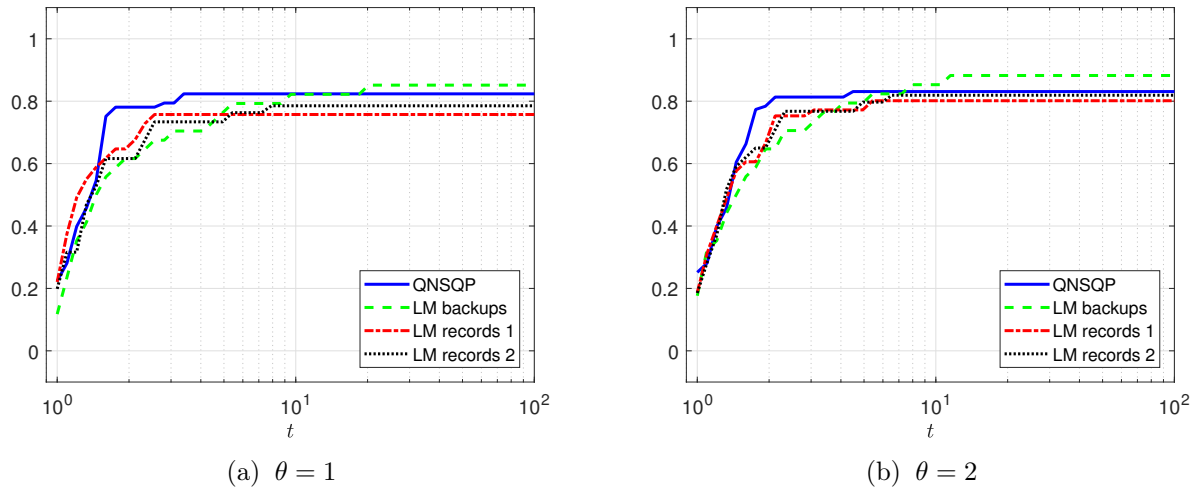


Рис. 2. Алгоритм 3 на внешней фазе; RSP = 10

Для каждой тестовой задачи осуществлялось 20 запусков из случайно сгенерированных начальных точек (x^0, λ^0) , равномерно распределенных в l_∞ -шаре (гиперкубе), радиус которого определялся параметром RSP (от «Range of Starting Points»), с центром в точке $(\bar{x}, 0)$, где $\bar{x}$ — известное решение, в котором нарушается условие регулярности (1.3).

Для представления результатов эксперимента использовались так называемые performance profiles, предложенные в [13], и адаптированные в [14] для случая многократных запусков для каждой тестовой задачи, с использованием средних характеристик эффективности и процентов успешных запусков. Полное описание этой конструкции, позволяющей одновременно отображать на рисунке сравнительную робастность (т. е. долю успешных запусков) и эффективность сравниваемых алгоритмов, можно найти в [15]. В качестве характеристики эффективности использовалось количество итераций (для алгоритма с возвратами это число учитывает отброшенные шаги метода LM), поскольку итерации всех сравниваемых алгоритмов имеют сопоставимую трудоемкость.

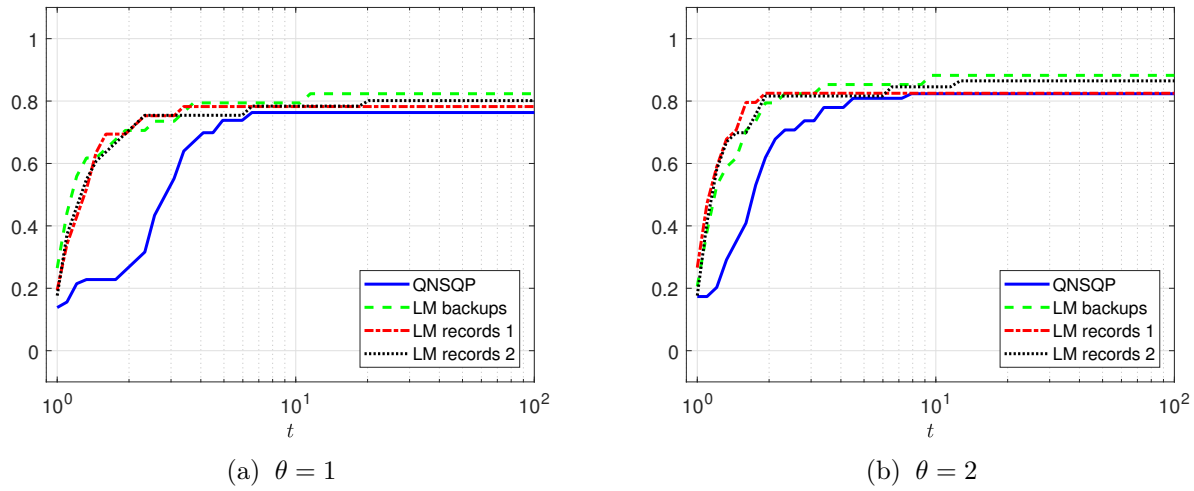


Рис. 3. Метод линейаризации на внешней фазе; RSP = 1

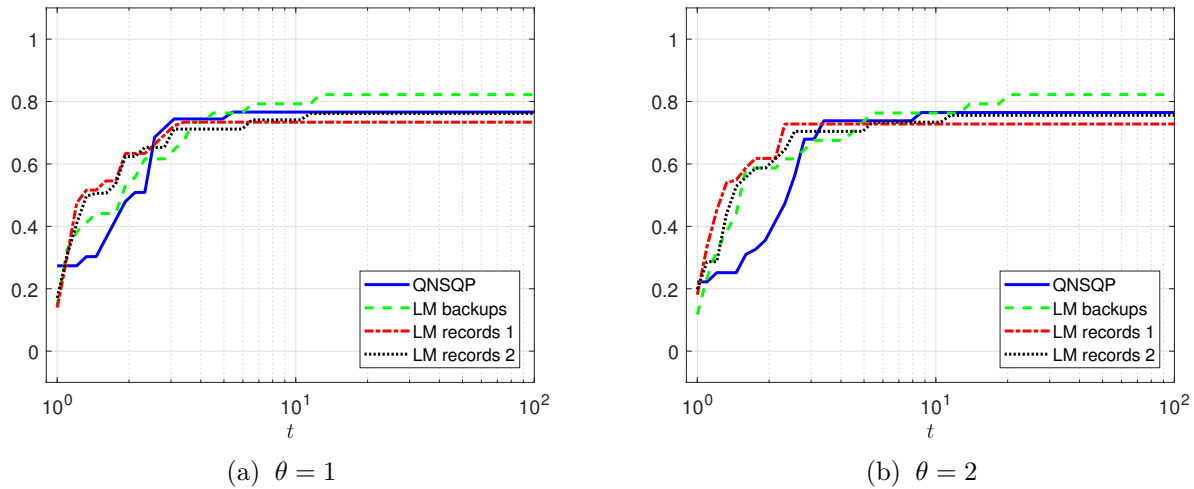


Рис. 4. Метод линейаризации на внешней фазе; RSP = 10

Из рис. 1 и 2 видно, что сравниваемые алгоритмы имеют на данном тестовом наборе схожую робастность, хотя для LM backups эта характеристика всегда является наилучшей. Кроме того, при RSP = 1 варианты глобализованного метода LM вполне конкурентоспособны и по эффективности, особенно с при использовании показателя степени $\theta = 2$ в параметре регуляризации. Однако, при RSP = 10 варианты метода LM уже несколько уступают по эффективности алгоритму QNSQP.

Дело, однако, в том, что какие-либо гарантии глобальной сходимости алгоритма 3 неизвестны, и проблема здесь в отсутствии гарантий равномерной положительной определенности и ограниченности матриц H_k , генерируемых этим алгоритмом по правилу BFGS. Поэтому использование этого алгоритма в качестве метода внешней фазы не вполне оправдано, по крайней мере формально. В связи с этим были также проведены эксперименты с вариантом алгоритма QNSQP, в котором в алгоритме 3 в качестве H_k всегда бралась единичная матрица. Метод SQP с $H_k = I$ называют методом линейаризации. Глобальная сходимость такого алгоритма следует, например, из [1, теорема 5.4.1].

Результаты для алгоритма QNSQP, использующего метод линейаризации, представлены на рис. 3 и 4, и они демонстрируют, что соответствующие гибридные глобализации метода LM по-прежнему сравнимы с этим алгоритмом по робастности, причем метод LM backups превосходит остальные алгоритмы по этой характеристике. При этом гибридные глобализации метода LM и существенно более эффективными при $RSP = 1$, и эта картина сохраняется и при $RSP = 10$, особенно для показателя степени $\theta = 2$ в параметре регуляризации, хотя и не столь выраженным образом.

References

- [1] А. Ф. Измаилов, М. В. Солодов, *Численные методы оптимизации*, 2-е изд., перераб. и доп., Физматлит, М., 2008. [A. F. Izmailov, M. V. Solodov, *Chislennyye Metody Optimizatsii*, 2nd edition, revised and enlarged, Fizmatlit, Moscow, 2008 (In Russian)].
- [2] A. F. Izmailov, M. V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, Springer, Cham, 2014.
- [3] A. F. Izmailov, M. V. Solodov, “Critical Lagrange multipliers: what we currently know about them, how they spoil our lives, and what we can do about it”, *TOP*, **23**:1 (2015), 1–26.
- [4] S. J. Wright, “Superlinear convergence of a stabilized SQP method to a degenerate solution”, *Computational Optimization and Applications*, **11**:3 (1998), 253–275.
- [5] W. W. Hager, “Stabilized sequential quadratic programming”, *Computational Optimization and Applications*, **12**:1–3 (1999), 253–273.
- [6] S. J. Wright, “Modifying SQP for degenerate problems”, *SIAM J. on Optimization*, **13**:2 (2002), 470–497.
- [7] D. Fernández, M. V. Solodov, “Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems”, *Mathematical Programming*, **125**:1 (2010), 47–73.
- [8] A. F. Izmailov, M. V. Solodov, “Stabilized SQP revisited”, *Mathematical Programming*, **133**:1–2 (2012), 93–120.
- [9] A. F. Izmailov, E. I. Uskov, “Subspace-stabilized sequential quadratic programming”, *Computational Optimization and Applications*, **67**:1 (2017), 129–154.
- [10] A. F. Izmailov, M. V. Solodov, E. I. Uskov, “A globally convergent Levenberg–Marquardt method for equality-constrained optimization”, *Computational Optimization and Applications*, **72**:1 (2019), 215–239.
- [11] A. Fischer, A. F. Izmailov, M. V. Solodov, “The Levenberg–Marquardt method: an overview of modern convergence theories and more”, *Computational Optimization and Applications*, **89**:1 (2024), 33–67.
- [12] А. Ф. Измаилов, А. М. Крылова, Е. И. Усков, “Гибридная глобализация стабилизированного метода последовательного квадратичного программирования”, *Теоретические и прикладные задачи нелинейного анализа*, ВЦ РАН, М., 2011, 47–66. [A. F. Izmailov, A. M. Krylova, E. I. Uskov, “Hybrid globalization of the stabilized method of sequential quadratic programming”, *Theoretical and Applied Problems of Nonlinear Analysis*, CC RAS, Moscow, 2011, 47–66 (In Russian)].
- [13] E. D. Dolan, J. J. Moré, “Benchmarking optimization software with performance profiles”, *Mathematical Programming*, **91**:2 (2002), 201–213.
- [14] A. F. Izmailov, M. V. Solodov, E. I. Uskov, “Combining stabilized SQP with the augmented Lagrangian algorithm”, *Computational Optimization and Applications*, **62**:2 (2015), 405–429.
- [15] A. F. Izmailov, E. I. Uskov, Yan Zhibai, “The piecewise Levenberg–Marquardt method for generalized Nash equilibrium problems.”, *Advances in System Sciences and Applications*, **24**:2 (2024), 19–31 (to appear).

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@cs.msu.ru

ORCID: <https://orcid.org/0000-0001-9851-0524>

Усков Евгений Иванович, кандидат физико-математических наук, научный сотрудник НОЦ «Фундаментальные математические исследования», Тамбовский государственный университет им. Г.Р. Державина, г. Тамбов, Российская Федерация. E-mail: euskov@cs.msu.ru

ORCID: <https://orcid.org/0000-0002-3639-0317>

Конфликт интересов отсутствует.

Для контактов:

Измаилов Алексей Ферилович

E-mail: izmaf@cs.msu.ru

Поступила в редакцию 26.01.2025 г.

Поступила после рецензирования 04.03.2025 г.

Принята к публикации 13.03.2025 г.

Information about the authors

Alexey F. Izmailov, Doctor of Physical and Mathematical Sciences, Professor of the Operations Research Department, Lomonosov Moscow State University, Moscow, Russian Federation. E-mail: izmaf@cs.msu.ru

ORCID: <https://orcid.org/0000-0001-9851-0524>

Evgeniy I. Uskov, Candidate of Physical and Mathematical Sciences, Researcher at the Scientific and Educational Center “Fundamental Mathematical Research”, Derzhavin Tambov State University, Tambov, Russian Federation. E-mail: euskov@cs.msu.ru

ORCID: <https://orcid.org/0000-0002-3639-0317>

There is no conflict of interests.

Corresponding author:

Alexey F. Izmailov

E-mail: izmaf@cs.msu.ru

Received 26.01.2025

Reviewed 04.03.2025

Accepted for press 13.03.2025