ISSN 2686-9667. Вестник российских университетов. Математика

Том 29, № 145 2024

НАУЧНАЯ СТАТЬЯ

https://doi.org/10.20310/2686-9667-2024-29-145-51-64
УДК 519

Методы с суженной матрицей Гессе как возмущенный метод Ньютона–Лагранжа

Андрей Андреевич ВОЛКОВ[1] **, Алексей Феридович ИЗМАИЛОВ**[1] **, Евгений Иванович УСКОВ**[2]

[1] ФГБОУ ВО «Московский государственный университет им. М. В. Ломоносова»
119992, 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. Для задачи оптимизации с ограничениями-равенствами обсуждается возможность интерпретации методов последовательного квадратичного программирования, использующих суженную на ядро матрицы Якоби ограничений матрицу Гессе функции Лагранжа, как возмущенного метода Ньютона–Лагранжа. Показано, что такая интерпретация с нужными оценками на возмущения возможна для определенных последовательностей, генерируемых вариантами метода с поправками второго порядка. Это позволяет с общих позиций установить сверхлинейную скорость сходимости таких последовательностей, вообще говоря отсутствующую для основных последовательностей рассматриваемых методов.

Ключевые слова: задача оптимизации с ограничениями-равенствами, последовательное квадратичное программирование, суженная матрица Гессе функции Лагранжа, схема возмущенного метода Ньютона–Лагранжа, поправки второго порядка, сверхлинейная сходимость

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

Для цитирования: *Волков А.А., Измаилов А.Ф., Усков Е.И.* Методы с суженной матрицей Гессе как возмущенный метод Ньютона–Лагранжа // Вестник российских университетов. Математика. 2024. Т. 29. № 145. С. 51–64. https://doi.org/10.20310/2686-9667-2024-29-145-51-64

SCIENTIFIC ARTICLE

https://doi.org/10.20310/2686-9667-2024-29-145-51-64

Reduced Hessian methods as a perturbed Newton–Lagrange method

Andrey A. VOLKOV[1] **, Alexey F. IZMAILOV**[1] **, Evgeniy I. USKOV**[2]

[1] Lomonosov Moscow State University
1 Leninskiye Gory, Moscow 119991, Russian Federation
[2] Derzhavin Tambov State University
33 Internatsionalnaya St., Tambov 392000, Russian Federation

Abstract. For an equality-constrained optimization problem, we consider the possibility to interpret sequential quadratic programming methods employing the Hessian of the Lagrangian reduced to the null space of the constraints' Jacobian, as a perturbed Newton–Lagrange method. We demonstrate that such interpretation with required estimates on perturbations is possible for certain sequences generated by variants of these methods making use of second-order corrections. This allows to establish, from a general perspective, superlinear convergence of such sequences, the property generally missing for the main sequences of the methods in question.

Keywords: equality-constrained optimization problem, sequential quadratic programming, reduced Hessian of the Lagrangian, perturbed Newton–Lagrange method framework, second-order corrections, superlinear convergence

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, 65K15.

For citation: Volkov A.A., Izmailov A.F., Uskov E.I. Reduced Hessian methods as a perturbed Newton–Lagrange method. *Vestnik rossiyskikh universitetov. Matematika = Russian Universities Reports. Mathematics*, **29**:145 (2024), 51–64. https://doi.org/10.20310/2686-9667-2024-29-145-51-64 (In Russian, Abstr. in Engl.)

Введение

Методы последовательного квадратичного программирования, использующие вместо настоящей (полной) матрицы Гессе функции Лагранжа ее суженную на ядро матрицы Якоби ограничений версию, направлены на снижения стоимости итерации таких методов, и являются важным инструментом практической условной оптимизации. В данной работе обсуждается возможность интерпретации таких методов как возмущенного метода Ньютона для системы уравнений Лагранжа. Будет показано, что такая интерпретация с нужными оценками на возмущения возможна для определенных последовательностей, генерируемых вариантами метода с поправками второго порядка. Это позволяет с общих позиций установить сверхлинейную скорость сходимости таких последовательностей, вообще говоря отсутствующую для основных последовательностей рассматриваемых методов.

Рассматривается задача оптимизации с ограничениями-равенствами

$$f(x) \to \min, \quad h(x) = 0, \tag{0.1}$$

где $f : \mathbb{R}^n \to \mathbb{R}$ и $h : \mathbb{R}^n \to \mathbb{R}^l$ дважды дифференцируемы вблизи $\bar{x} \in \mathbb{R}^n$. Метод последовательного квадратичного программирования для задачи (0.1) состоит в следующем [1, разд. 18.1], [2, разд. 4.3.1, 4.4.1], [3, разд. 4.1.1]: для текущего приближения $(x^k,\, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^l$ следующее приближение определяется как $(x^{k+1},\, \lambda^{k+1})$, где $x^{k+1} = x^k + \xi^k$, ξ^k является стационарной точкой квадратичной задачи

$$\langle f'(x^k),\, \xi\rangle + \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi,\, \xi\right\rangle \to \min, \quad h(x^k) + h'(x^k)\xi = 0, \tag{0.2}$$

а λ^{k+1} — отвечающим этой стационарной точке множителем Лагранжа. Здесь $L : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ — функция Лагранжа задачи (0.1):

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

Система уравнений Лагранжа задачи (0.2), характеризующая ее стационарные точки и отвечающие им множители, имеет вид

$$f'(x^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi + (h'(x^k))^\top \lambda = 0, \quad h(x^k) + h'(x^k)\xi = 0. \tag{0.3}$$

Отсюда следует, что итерация метода последовательного квадратичного программирования есть ни что иное, как итерация метода Ньютона–Лагранжа, т. е. метода Ньютона, применяемого к системе уравнений Лагранжа

$$\frac{\partial L}{\partial x}(x,\, \lambda) = 0, \quad h(x) = 0 \tag{0.4}$$

исходной задачи (0.1): уравнения в (0.3) получаются линеаризацией в текущем приближении $(x^k,\, \lambda^k)$ уравнений из (0.4).

Методы последовательного квадратичного программирования обладают характерной для ньютоновских методов сверхлинейной локальной сходимостью в естественных предположениях, и относятся к числу наиболее эффективных вычислительных технологий для задач условной оптимизации, и являются основой многих успешных оптимизационных солверов общего назначения. Однако, в своей базовой форме, итерация этих методов может быть весьма затратной, в частности, в связи с необходимостью вычисления полной матрицы Гессе функции Лагранжа, либо ее все более точных (по мере роста номера итерации) аппроксимаций.

Точнее, согласно [3, предложение 4.4], для сохранения сверхлинейной скорости сходимости генерируемой прямой последовательности $\{x^k\}$ к стационарной точке $\bar{x} \in \mathbb{R}^n$ задачи (0.1) (при наличии сходимости прямодвойственной последовательности $\{(x^k,\, \lambda^k)\}$ к $(\bar{x},\, \bar{\lambda})$, где $\bar{\lambda} \in \mathbb{R}^l$ — некоторый отвечающий $\bar{x}$ множитель Лагранжа) должны удовлетворять итерационной системе

$$f'(x^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(x - x^k) + (h'(x^k))^\top \lambda + \omega_1^k = 0, \quad h(x^k) + h'(x^k)(x - x^k) + \omega_2^k = 0, \tag{0.5}$$

возмущенного метода Ньютона–Лагранжа, получаемой добавлением к уравнениям в (0.3) параметров возмущения $\omega_1^k \in \mathbb{R}^n$ и $\omega_2^k \in \mathbb{R}^l$, которые должны удовлетворять оценкам

$$P\omega_1^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \tag{0.6}$$

$$\omega_2^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{0.7}$$

при $k \to \infty$; см. [3, (4.12), (4.13)]. Здесь P — оператор ортогонального проектирования на $\ker h'(\bar{x})$. Более того, при выполнении в стационарной точке $\bar{x}$ с множителем $\bar{\lambda}$ достаточного условия второго порядка оптимальности оценки (0.6), (0.7) являются и достаточными для сверхлинейной сходимости.

Методы с суженной матрицей Гессе, рассматриваемые в данной работе, позволяют избежать необходимости вычисления полной матрицы Гессе функции Лагранжа, поскольку на их итерациях используется только вводимая в разд. 1. так называемая суженная матрица Гессе, размерность которой определяется размерностью $\ker h'(x^k)$, что может быть (и обычно является) гораздо меньшим числом, чем n. Интерпретация этих методов как возмущенного метода Ньютона–Лагранжа не позволяет получить оценки (0.6), (0.7), и действительно, как демонстрирует пример, предложенный в [4] (см. пример 2.1 ниже), эти методы, вообще говоря, обладают лишь двухшаговой сверхлинейной скоростью сходимости, что является более слабым свойством, чем обычная (одношаговая) сверхлинейная скорость сходимости. (Отметим также аналогичный пример, построенный в [5]. Однако, корректность его конструкции и приведенные численные результаты вызывают определенные сомнения, главным образом в связи с очевидной ошибочностью выражений для производных по первой переменной в (2.14).)

Вместе с тем, как будет показано в разд. 2., если снабдить метод так называемыми поправками второго порядка, предназначенными для того, чтобы генерируемые прямые приближения были ближе к допустимому множеству задачи (0.1), то получаемый процесс можно интерпретировать как возмущенный метод Ньютона–Лагранжа с требуемыми оценками (0.6), (0.7), но не для основной последовательности метода, а для некоторой последовательности вспомогательных приближений, также генерируемых методом. Это позволяет получить новое общее понимание результата о сверхлинейной скорости сходимости последовательности таких вспомогательных приближений, доказанного в [6].

1. Методы с суженной матрицей Гессе

Обозначим через P_k оператор ортогонального проектирования на $\ker h'(x^k)$, и для всякого $\xi \in \mathbb{R}^n$ будем использовать разложение $\xi = \xi_1 + \xi_2$, $\xi_1 = (I - P_k)\xi$, $\xi_2 = P_k\xi$. Тогда ограничение в (0.2) принимает вид

$$h(x^k) + h'(x^k)\xi_1 = 0, \tag{1.1}$$

и если $\operatorname{rank} h'(x^k) = l$ (условие, которое всюду далее предполагается выполненным), то это уравнение однозначно определяет

$$\xi_1^k = -(h'(x^k))^\top (h'(x^k)(h'(x^k))^\top)^{-1} h(x^k) \tag{1.2}$$

как единственное нормальное решение уравнения в ограничении задачи (0.2). Фиксируя это ξ_1^k, из (0.2) получаем задачу для определения ξ_2^k:

$$\left\langle f'(x^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi_1^k,\, \xi_2 \right\rangle + \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi_2,\, \xi_2 \right\rangle \to \min, \quad \xi_2 \in \ker h'(x^k). \tag{1.3}$$

Ключевой момент в конструкции методов с суженной матрицей Гессе состоит в том, что слагаемое в (1.3), содержащее ξ_1^k и ξ_2, опускается. Используя равенство $\xi_2 = P_k\xi$, так модифицированная задача (1.3) вместе с уравнением (1.1) сводится к задаче

$$\left\langle f'(x^k),\, \xi \right\rangle + \frac{1}{2}\left\langle P\frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)P\xi,\, \xi \right\rangle \to \min, \quad h(x^k) + h'(x^k)\xi = 0, \tag{1.4}$$

записанной снова относительно исходной переменной ξ задачи (0.2).

Чтобы привести задачу (1.4) к более «практической» форме, потребуется конкретное представление проектора P_k. Пусть столбцы матрицы Z_k размеров $n \times (n-l)$ образуют некоторый базис линейного подпространства $\ker h'(x^k)$. (Напомним, что предполагается выполненным условие $\operatorname{rank} h'(x^k) = l$, и значит, $\dim\ker h'(x^k) = n-l$.) Тогда, как нетрудно видеть, $P_k = Z_k(Z_k^\top Z_k)^{-1}Z_k^\top$, и задача (1.4) принимает вид

$$\left\langle f'(x^k),\, \xi \right\rangle + \frac{1}{2}\left\langle Z_k(Z_k^\top Z_k)^{-1}H_k(Z_k^\top Z_k)^{-1}Z_k^\top\xi,\, \xi \right\rangle \to \min, \quad h(x^k) + h'(x^k)\xi = 0, \tag{1.5}$$

где симметричная $(n-l) \times (n-l)$ матрица

$$H_k = Z_k^\top \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k) Z_k \tag{1.6}$$

и есть *суженная матрица Гессе* функции Лагранжа.

Если столбцы матрицы Z_k образуют ортонормированную систему, то $Z_k^\top Z_k = I$, $P_k = Z_k Z_k^\top$, и задача (1.5) принимает вид

$$\langle f'(x^k),\, \xi\rangle + \frac{1}{2}\langle Z_k H_k Z_k^\top \xi,\, \xi\rangle \to \min, \quad h(x^k) + h'(x^k)\xi = 0. \tag{1.7}$$

Это в точности соответствует подзадаче, использованной в [7, (1.2)]. Заметим, что указанное свойство столбцов матрицы Z_k всегда выполняется, если матрица $(Y_k\ Z_k)$ с некоторой $n \times l$ матрицей Y_k получена QR-факторизацией $(h'(x^k))^\top$, что дает практический способ вычисления Z_k [1, разд. 15.3]. При этом столбцы Y_k образуют ортонормированный базис в $(\ker h'(x^k))^\perp$, и $I - P_k = Y_k Y_k^\top$.

Из сказанного выше ясно, что подзадача (1.4), а значит, и ее «практические» версии (1.5) и (1.7), отличаются от подзадачи (0.2) базового метода последовательного квадратичного программирования по существу только отсутствием в целевой функции слагаемого, содержащего оба ξ_1 и ξ_2. В частности, эти методы совпадают, если для каждого k

$$\left\langle \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi_1,\, \xi_2 \right\rangle = 0 \quad \forall\, \xi \in \mathbb{R}^n$$

(факт, отмеченный в [7, с. 286]). Если же

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi_1,\, \xi_2 \right\rangle = 0 \quad \forall\, \xi \in \mathbb{R}^n \tag{1.8}$$

для предельной пары $(\bar{x},\, \bar{\lambda})$, то методы, вообще говоря, не совпадают, но методы с суженной матрицей Гессе могут быть интерпретированы как возмущенный метод Ньютона–Лагранжа с требуемыми оценками (0.6), (0.7) (см. разд. 2.), а значит, имеют в соответствующих предположениях сверхлинейную скорость сходимости (что соответствует сказанному в [5, с. 231]).

Обращаясь к работам, предшествовавшим [7, (1.2)], в [4] обсуждался вариант метода последовательного квадратичного программирования с подзадачей

$$\langle f'(x^k),\, \xi\rangle + \frac{1}{2}\langle \mathcal{H}_k\xi,\, \xi\rangle \to \min, \quad h(x^k) + h'(x^k)\xi = 0, \tag{1.9}$$

в которой используется матрица $\mathcal{H}_k$, задаваемая равенством

$$(Z_k\ Y_k)^\top \mathcal{H}_k (Z_k\ Y_k) = \begin{pmatrix} H_k & 0 \\ 0 & I \end{pmatrix},$$

где H_k введено в (1.6). Если столбцы Z_k и Y_k ортонормированы, то

$$(Z_k\ Y_k)(Z_k\ Y_k)^\top = Z_k Z_k^\top + Y_k Y_k^\top = I,$$

и поэтому

$$\begin{gathered} \langle \mathcal{H}_k\xi,\, \xi\rangle = \langle \mathcal{H}_k(Z_k\ Y_k)(Z_k\ Y_k)^\top\xi,\, (Z_k\ Y_k)(Z_k\ Y_k)^\top\xi\rangle \\ = \langle (Z_k\ Y_k)^\top\mathcal{H}_k(Z_k\ Y_k)(Z_k\ Y_k)^\top\xi,\, (Z_k\ Y_k)^\top\xi\rangle = \langle (H_k Z_k^\top\xi,\, Y_k^\top\xi),\, (Z_k^\top\xi,\, Y_k^\top\xi)\rangle \\ = \langle Z_k H_k Z_k^\top\xi,\, \xi\rangle + \langle Y_k^\top\xi,\, Y_k^\top\xi\rangle, \end{gathered} \tag{1.10}$$

причем

$$\langle Y_k^\top\xi,\, Y_k^\top\xi\rangle = \langle Y_k^\top Y_k Y_k^\top\xi,\, Y_k^\top\xi\rangle = \langle Y_k Y_k^\top\xi,\, Y_k Y_k^\top\xi\rangle = \|(I - P_k)\xi\|^2 = \|\xi_1\|^2.$$

Вспоминая, что, как обсуждалось выше, ξ_1^k однозначно определяется ограничением в (1.7), а значит и идентичным ему в (1.9), отсюда получаем, что второе слагаемое в правой части (1.10) не играет в целевой функции подзадачи (1.9) никакой роли, и эта подзадача сводится к (1.7).

Далее, умножением обеих частей первого уравнения в системе Лагранжа

$$f'(x^k) + Z_k H_k Z_k^\top\xi + (h'(x^k))^\top\lambda = 0, \quad h(x^k) + h'(x^k)\xi = 0 \tag{1.11}$$

задачи (1.7) на $Z_k^\top$, необходимые условия первого порядка оптимальности для этой задачи могут быть записаны в следующей «прямой» форме, не использующей λ:

$$Z_k^\top f'(x^k) + H_k Z_k^\top \xi = 0, \quad h(x^k) + h'(x^k)\xi = 0, \tag{1.12}$$

откуда следует выполнение равенства (1.2), а в дополнительном предположении о невырожденности H_k и равенства

$$Z_k^\top \xi_2^k = -H_k^{-1} Z_k^\top f'(x^k).$$

Последнее также можно записать в виде явного выражения

$$\xi_2^k = P_k \xi_2^k = Z_k Z_k^\top \xi_2 = -Z_k H_k^{-1} Z_k^\top f'(x^k). \tag{1.13}$$

Вместе с тем, умножая обе части первого уравнения в (1.11) на $Y_k^\top$, получаем

$$Y_k^\top f'(x^k) + Y_k^\top (h'(x^k))^\top \lambda = 0, \tag{1.14}$$

Дальнейшее умножение на Y_k приводит это уравнение к «инвариантому» виду

$$(I - P_k) f'(x^k) + (h'(x^k))^\top \lambda = 0$$

(поскольку $(I - P_k)(h'(x^k))^\top = (h'(x^k)(I - P_k))^\top = (h'(x^k))^\top$), не зависящему от выбора Y_k. Если взять, например, $Y_k = (h'(x^k))^\top$, что обеспечивается требуемую здесь базисность столбцов этой матрицы в $\ker h'(x^k)$ (но, разумеется, не их ортонормированность), то из (1.14) получаем уравнение

$$h'(x^k)^\top f'(x^k) + h'(x^k)(h'(x^k))^\top \lambda = 0. \tag{1.15}$$

Отсюда следует явное выражение

$$\lambda^{k+1} = -(h'(x^k)(h'(x^k))^\top)^{-1} h'(x^k) f'(x^k). \tag{1.16}$$

Это соответствует [7, (2.13)], где, правда, x^k заменяется на x^{k+1}: λ^{k+1} (и x^{k+1}) используются для определения H_{k+1}, и это возможно, поскольку здесь, в отличие от обычного метода Ньютона–Лагранжа, итерация естественным образом разделяется на прямую и двойственную фазы. Использование (1.16) также соответствует выбору λ^{k+1} как решения задачи квадратичной задачи безусловной оптимизации

$$\|f'(x^k) + (h'(x^k))^\top \lambda\|^2 \to \min$$

относительно $\lambda \in \mathbb{R}^l$. Существуют и другие способы задания λ^{k+1}, например [7, (2.14)], где, впрочем, этот способ приводится без объяснения его происхождения.

Уравнения (1.12) и (1.15), порождающие явные выражения (1.2), (1.13) и (1.16), характеризуют класс методов последовательного квадратичного программирования, в форме методов Ньютона–Лагранжа, с суженной матрицей Гессе. В частности, (1.12) соответствует методу в [5, (1.5), (1.6)], который, в свою очередь, был взят из [8, алгоритм 4.1], где, собственно, и была установлена двухшаговая сверхлинейная сходимость методов такого типа. См. также [9, (14.41), теорема 14.7]. Следует также упомянуть так называемый «горизонтально-вертикальный» алгоритм из [10], который также рассматривался в [4] (см. также [9, (14.40)]): в нем используются выражения (1.2), (1.13), но только сначала вычисляется ξ_2^k согласно (1.13), и затем ξ_1^k согласно (1.2), где вместо $h(x^k)$ используется $h(x^k + \xi_2^k)$.

Разные варианты методов указанного класса могут отличаться способами выбора Z_k. Кроме того, важнейшее значение имеют методы, использующие не саму суженную матрицу Гессе из (1.6), а ее аппроксимации, и прежде всего квазиньютоновские. Эти вопросы обсуждаются [8] и в последующих цитированных публикациях, но здесь они не рассматриваются.

В некоторых изложениях (см., например, [1, разд. 18.3]) методов с суженной матрицей Гессе используют представления $\xi_1 = Y_k\eta$ и $\xi_2 = Z_k\zeta$, где $\eta \in \mathbb{R}^l$ и $\zeta \in \mathbb{R}^{n-l}$ однозначно определяются. Если столбцы Z_k ортонормированы, то, подставляя эти выражения в (1.12), получаем уравнения

$$Z_k^\top f'(x^k) + H_k\zeta = 0, \quad h(x^k) + h'(x^k)Y_k\eta = 0,$$

для определения η и ζ.

2. Методы с поправками второго порядка

Помимо метода из [10], обсуждавшегося в [4], в [6] рассматривается вариант метода последовательного квадратичного программирования с суженной матрицей Гессе, снабженный так называемыми поправками второго порядка [1, разд. 15.6], [2, разд. 5.4.2], [3, разд. 4.3.6, 6.2.2]. А именно, следующее приближение определяется равенством $x^{k+1} = x^k + \xi^k + \bar{\xi}^k$, где ξ^k — стационарная точка подзадачи (1.7), т. е. решение системы (1.12), а $\bar{\xi}^k$ — нормальное решение уравнения

$$h(x^k + \xi^k) + h'(x^k)\xi = 0, \tag{2.1}$$

т. е.

$$\bar{\xi}^k = -(h'(x^k))^\top (h'(x^k)(h'(x^k))^\top)^{-1} h(x^k + \xi^k). \tag{2.2}$$

Заметим, что при этом $\bar{\xi}^k \in \operatorname{im}(h'(x^k))^\top = (\ker h'(x^k))^\perp$, т. е. $\bar{\xi}_2^k = 0$. Новое двойственное приближение λ^{k+1} по-прежнему определяется уравнением (1.15).

Лемма 2.1. *Пусть функция $f : \mathbb{R}^n \to \mathbb{R}$ и отображение $h : \mathbb{R}^n \to \mathbb{R}^l$ дважды непрерывно дифференцируемы вблизи точки $\bar{x} \in \mathbb{R}^n$. Пусть $\bar{x}$ является стационарной точкой задачи (0.1), а $\bar{\lambda} \in \mathbb{R}^l$ — отвечающим ей множителем Лагранжа, причем выполнено условие регулярности ограничений $\operatorname{rank} h'(\bar{x}) = l$ и достаточное условие второго порядка оптимальности*

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\}. \tag{2.3}$$

Пусть последовательность $\{(x^k, \lambda^k)\}$, сгенерированная описанной выше процедурой, сходится к $(\bar{x}, \bar{\lambda})$.

Тогда

$$\xi^k = O(\|x^k - \bar{x}\|), \tag{2.4}$$

$$h(x^k + \xi^k) = O(\|x^k - \bar{x}\|^2), \tag{2.5}$$

$$\bar{\xi}^k = O(\|x^k - \bar{x}\|^2), \tag{2.6}$$

при $k \to \infty$.

Доказательство. Оценка (2.4) следует из (1.2) и (1.13), поскольку в сделанных предположениях матрицы $h'(x^k)(h'(x^k))^\top$ и H_k равномерно обратимы для $(x^k,\, \lambda^k)$ достаточно близких к $(\bar{x},\, \bar{\lambda})$.

Кроме того, в силу второго уравнения в (1.12),

$$h(x^k + \xi^k) = h(x^k) + h'(x^k)\xi^k + O(\|\xi^k\|^2) = O(\|\xi^k\|^2),$$

и поэтому, согласно (2.2),

$$\bar{\xi}^k = O(\|h(x^k + \xi^k)\|) = O(\|\xi^k\|^2).$$

Из последних двух оценок и из (2.4) следуют оценки (2.5), (2.6). □

Согласно первому равенству в (0.5), используя представление $P_k = Z_k Z_k^\top$ и оценку (2.6), выводим

$$\begin{gathered} P_k\omega_1^k = -P_k\Big(f'(x^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(\xi^k + \bar{\xi}^k) + (h'(x^k))^\top \lambda^{k+1}\Big) \\ = Z_k\Big(Z_k^\top f'(x^k) + Z_k^\top \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k) Z_k Z_k^\top \xi^k + Z_k^\top \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi_1^k\Big) + O(\|x^k - \bar{x}\|^2) \\ = P_k \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi_1^k + O(\|x^k - \bar{x}\|^2), \end{gathered} \tag{2.7}$$

где в последнем равенстве использовано определение H_k в (1.6) и первое уравнение в (1.12). Заметим, что в обеих частях этой оценки P_k можно заменить на P, поскольку

$$P_k - P = O(\|x^k - \bar{x}\|) \tag{2.8}$$

при $k \to \infty$. Это следует, например, из явного представления

$$P_k = (I - (h'(x^k))^\top (h'(x^k)(h'(x^k))^\top)^{-1} h'(x^k) \tag{2.9}$$

и равномерной обратимости матриц $(h'(x^k)(h'(x^k))^\top$ для x^k близких к $\bar{x}$.

Далее, согласно второму равенству в (0.5), используя второе уравнение в (1.12), (2.1), и оценку (2.5), выводим

$$\omega_2^k = -h(x^k) - h'(x^k)(\xi^k + \bar{\xi}^k) = -h'(x^k)\bar{\xi}^k = -h(x^k + \xi^k) = O(\|x^k - \bar{x}\|^2).$$

Таким образом, оценка (0.7) имеет место, а вот (0.6) можно гарантировать лишь при выполнении (1.8), причем эти рассуждения тем более проходят и при использовании $\bar{\xi}^k = 0$, т. е. сделанные выводы верны как для метода без поправок второго порядка, так и с такими поправками. В частности, гарантировать сверхлинейную скорость сходимости последовательности $\{x^k\}$ по-прежнему нельзя и для метода с поправками второго порядка.

Теперь обратимся к поведению последовательности $\{\widetilde{x}^k\}$, генерируемой по правилу

$$\widetilde{x}^{k+1} = x^k + \xi^k = x^{k-1} + \xi^{k-1} + \bar{\xi}^{k-1} + \xi^k = \widetilde{x}^k + \bar{\xi}^{k-1} + \xi^k.$$

Иными словами, будем рассматривать следующий двухшаговый процесс:

$$\begin{gathered} x^{k-1} \\ \downarrow \\ \widetilde{x}^k = x^{k-1} + \xi^{k-1} \\ \downarrow \\ x^k = \widetilde{x}^k + \bar{\xi}^{k-1} = x^{k-1} + \xi^{k-1} + \bar{\xi}^{k-1} \\ \downarrow \\ \widetilde{x}^{k+1} = x^k + \xi^k = \widetilde{x}^k + \bar{\xi}^{k-1} + \xi^k. \end{gathered}$$

Из оценок в лемме 2.1 вытекает, что последовательности $\{\widetilde{x}^k\}$ сходится к тому же пределу $\bar{x}$, что и последовательность $\{x^k\}$.

Помимо оценок из леммы 2.1, приведем некоторые дополнительные оценки на ингредиенты рассматриваемого итерационного процесса, следующие из [6, леммы 3.1, 3.3].

Лемма 2.2. *В предположениях леммы 2.1*

$$\bar{\xi}^{k-1} = O(\|\widetilde{x}^k - \bar{x}\|), \tag{2.10}$$

$$h'(x^k)\xi^k = o(\|\widetilde{x}^k - \bar{x}\|) \tag{2.11}$$

при $k \to \infty$.

Теорема 2.1. *В предположениях леммы 2.1 для*

$$\widetilde{\omega}_1^k = -f'(x^k) - \frac{\partial^2 L}{\partial x^2}(\widetilde{x}^k,\, \lambda^k)(\widetilde{x}^{k+1} - \widetilde{x}^k) - (h'(\widetilde{x}^k))^\top \lambda^{k+1} \tag{2.12}$$

и

$$\widetilde{\omega}_2^k = -h(\widetilde{x}^k) - h'(\widetilde{x}^k)(\widetilde{x}^{k+1} - \widetilde{x}^k) \tag{2.13}$$

выполняются оценки

$$P\widetilde{\omega}_1^k = o(\|\widetilde{x}^k - \bar{x}\|), \tag{2.14}$$

$$\widetilde{\omega}_2^k = o(\|\widetilde{x}^k - \bar{x}\|) \tag{2.15}$$

при $k \to \infty$, *и, в частности, последовательность* $\{\widetilde{x}^k\}$ *сходится к* $\bar{x}$ *со сверхлинейной скоростью.*

Последнее утверждение теоремы согласуется с [6, теорема 3.5].

Доказательство. Согласно (2.12), привлекая теорему о среднем, выводим

$$\begin{aligned} P_k\widetilde{\omega}_1^k &= -P_k\Big(f'(x^{k-1} + \xi^{k-1}) + \frac{\partial^2 L}{\partial x^2}(x^{k-1} + \xi^{k-1},\, \lambda^k)(\bar{\xi}^{k-1} + \xi^k) + (h'(x^{k-1} + \xi^{k-1}))^\top \lambda^{k+1}\Big) \\ &= -P_k\Big(\frac{\partial L}{\partial x}(x^{k-1} + \xi^{k-1},\, \lambda^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(\bar{\xi}^{k-1} + \xi^k) + (h'(x^k))^\top(\lambda^{k+1} - \lambda^k)\Big) \\ &\qquad + O(\|\xi^k\|\|\bar{\xi}^{k-1}\|) + O(\|\lambda^{k+1} - \lambda^k\|\|\bar{\xi}^{k-1}\|) + O(\|\bar{\xi}^{k-1}\|^2) \\ &= P_k\Big(\frac{\partial L}{\partial x}(x^{k-1} + \xi^{k-1} + \bar{\xi}^{k-1},\, \lambda^k) - \frac{\partial L}{\partial x}(x^{k-1} + \xi^{k-1},\, \lambda^k) - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\bar{\xi}^{k-1}\Big) \\ &\qquad - P_k\Big(\frac{\partial L}{\partial x}(x^{k-1} + \xi^{k-1} + \bar{\xi}^{k-1},\, \lambda^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi^k\Big) \end{aligned}$$

$$
\begin{aligned}
&+O(\|\xi^k\|\|\bar{\xi}^{k-1}\|) + O(\|\lambda^{k+1} - \lambda^k\|\|\bar{\xi}^{k-1}\|) + O(\|\bar{\xi}^{k-1}\|^2)\\
&= -P_k\Big(\frac{\partial L}{\partial x}(x^k,\, \lambda^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi^k\Big)\\
&+O(\|\xi^k\|\|\bar{\xi}^{k-1}\|) + O(\|\lambda^{k+1} - \lambda^k\|\|\bar{\xi}^{k-1}\|) + O(\|\bar{\xi}^{k-1}\|^2)\\
&= -P_k\Big(f'(x^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi^k + (h'(x^k))^\top\lambda^k\Big)\\
&+O(\|\xi^k\|\|\bar{\xi}^{k-1}\|) + O(\|\lambda^{k+1} - \lambda^k\|\|\bar{\xi}^{k-1}\|) + O(\|\bar{\xi}^{k-1}\|^2)\\
&= -P_k\frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi_1^k + O(\|\xi^k\|\|\bar{\xi}^{k-1}\|) + O(\|\lambda^{k+1} - \lambda^k\|\|\bar{\xi}^{k-1}\|) + O(\|\bar{\xi}^{k-1}\|^2),
\end{aligned}
$$

где последнее равенство выводится по аналогии с (2.7). Замечая, что согласно (2.9),

$$
\xi_1^k = (I - P_k)\xi^k = (h'(x^k))^\top(h'(x^k)(h'(x^k))^\top)^{-1}h'(x^k)\xi^k,
$$

отсюда и из оценок (2.10), (2.11) следует оценка

$$
P_k\widetilde{\omega}_1^k = o(\|\widetilde{x}^k - \bar{x}\|). \tag{2.16}
$$

Принимая во внимание (2.8) и вытекающую из (2.10) оценку,

$$
x^k - \bar{x} = \widetilde{x}^k - \bar{x} + \bar{\xi}^{k-1} = O(\|\widetilde{x}^k - \bar{x}\|),
$$

имеем

$$
P_k - P = O(\|\widetilde{x}^k - \bar{x}\|),
$$

откуда и из (2.16) следует оценка (2.14).

Далее, согласно (2.13), используя (2.1), выводим

$$
\begin{aligned}
\widetilde{\omega}_2^k &= -h(x^{k-1} + \xi^{k-1}) - h'(x^{k-1} + \xi^{k-1})(\bar{\xi}^{k-1} + \xi^k)\\
&= -h(x^{k-1} + \xi^{k-1}) - h'(x^{k-1})\bar{\xi}^{k-1} - (h'(x^{k-1} + \xi^{k-1}) - h'(x^{k-1}))\bar{\xi}^{k-1} - h'(x^{k-1} + \xi^{k-1})\xi^k\\
&= -(h'(x^{k-1} + \xi^{k-1}) - h'(x^k))\xi^k - h'(x^k)\xi^k + O(\|\xi^{k-1}\|\|\bar{\xi}^{k-1}\|)\\
&= -h'(x^k)\xi^k + O(\|\xi^{k-1}\|\|\bar{\xi}^{k-1}\|) + O(\|\xi^k\|\|\bar{\xi}^{k-1}\|),
\end{aligned}
$$

откуда и из оценок (2.10), (2.11) следует оценка (2.15).

Последнее утверждение теоремы вытекает из полученных оценок (2.14), (2.15) и из [3, предложение 4.4]. □

Следующий пример был предложен в [4, пример 2].

П р и м е р 2.1. Рассмотрим задачу (0.1), в которой $n = 2$, $l = 1$,

$$
f(x) = \frac{1}{2}x_1^2 - \alpha x_1x_2 + \frac{1}{2}x_2^2 + \frac{1}{3\alpha}(x_1 - \alpha)^3,
$$

$$
h(x) = \frac{1}{(x_2 - 2)^2} - 1,
$$

где α — вещественный параметр. Вблизи локального решения $\bar{x} = (\alpha,\, 1)$ такие f и h бесконечно дифференцируемы, причем в этом решении выполняется условие регулярности $\operatorname{rank} h'(\bar{x}) = l$ (поскольку $h'(\bar{x}) \neq 0$) и достаточное условие второго порядка оптимальности (2.3) с единственным отвечающим $\bar{x}$ множителем Лагранжа $\bar{\lambda} = \alpha^2 - 1$.

В таблицах 1, 2 для каждой итерации k и полученного на ней приближения $(x^k,\, \lambda^k)$ (и $\widetilde{x}^k$ для метода с поправками второго порядка) приводятся следующие сведения:

– “Pert” есть величина $\|P_k \omega_1^{k-1}\|/(\|x^k - x^{k-1}\| + \|x^{k-1} - \bar{x}\|)$ (см. (0.6); нормы везде евклидовы);

– “Res” есть невязка системы Лагранжа (0.4);

– “Dist” есть расстояние $\|x^k - \bar{x}\|$ до прямого решения;

– “PR” (от “primal ratio”) есть величина $\|x^k - \bar{x}\|/\|x^{k-1} - \bar{x}\|$;

– “PDR” (от “primal-dual ratio”) есть величина $\|(x^k - \bar{x},\, \lambda^k - \bar{\lambda})\|/\|(x^{k-1} - \bar{x},\, \lambda^{k-1} - \bar{\lambda})\|$;

– “PR-SOC” есть величина $\|\widetilde{x}^k - \bar{x}\|/\|\widetilde{x}^{k-1} - \bar{x}\|$ (для метода с поправками второго порядка).

Таблица 1

Численные результаты для примера 2.1: метод без поправок второго порядка

k	Pert	Res	Dist	PR	PDR
0	–	5.8e+0	1.0e-1	–	–
1	7.4e-1	6.7e+0	5.0e-1	5.0e+0	4.7e+1
2	4.9e-2	2.0e+0	1.0e-4	2.0e-4	4.2e-1
3	8.2e-1	7.4e-3	5.0e-4	5.0e+0	2.5e-3
4	5.0e-5	2.5e-3	9.2e-16	1.8e-12	5.1e-1
5	9.7e-1	1.1e-14	8.9e-16	9.7e-1	2.9e-12
6	0	3.6e-15	0.0e+0	0.0e+0	5.0e-1

Таблица 2

Численные результаты для примера 2.1: метод с поправками второго порядка

k	Pert	Res	Dist	PR	PDR	PR-SOC
0	–	5.8e+0	1.0e-1	–	–	–
1	7.4e-1	7.1e+0	5.0e-1	5.0e+0	4.7e+1	–
2	8.6e-3	2.7e+0	5.1e-2	1.0e-1	5.1e-1	5.1e-2
3	5.9e-7	2.5e-1	5.1e-4	1.0e-2	1.0e-1	8.5e-4
4	0.0e+0	2.5e-3	5.3e-8	1.0e-4	1.0e-2	1.0e-6
5	0.0e+0	2.6e-7	8.9e-16	1.7e-8	1.0e-4	1.7e-12
6	0.0e+0	3.6e-15	0.0e+0	0.0e+0	1.4e-8	0.0e+0

Результаты в таблицах 1, 2 получены с использованием $\alpha = 5$ и начального приближения $x^0 = \bar{x} + (0,\, 0.1)$, $\lambda^0 = \bar{\lambda}$.

Таблица 1 демонстрирует наличие у метода последовательного квадратичного программирования с суженной матрицей Гессе двухшаговой сверхлинейной сходимости, но отсутствие настоящей сверхлинейной сходимости: на каждой второй итерации расстояние до прямого решения (см. столбцы “Dist” и “PR”) возрастает, по крайней мере пока приближения не попадают в достаточно малую окрестность решения, где существенную роль начинают играть ошибки округления. Это согласуется с поведением величин в столбце “Pert”.

Согласно таблице 2, для варианта метода с поправками второго порядка очевидно имеет место сверхлинейная (квадратичная) сходимость последовательности $\{\widetilde{x}^k\}$ (столбец “PR-SOC”). Основная последовательность $\{x^k\}$ здесь тоже асимптотически сходится сверхлинейно (и даже квадратично; столбцы “Dist” и “PR”). Вместе с тем, шаг из точки x^0 в точку x^1 увеличивает расстояние до решения (примерно в 5 раз; столбцы “Dist” и “PR”), и эта ситуация сохраняется, каким бы близким x^0 не было к $\bar{x}$. Это является свидетельством принципиальных трудностей доказательства сверхлинейной сходимости основной последовательности метода последовательного квадратичного программирования с суженной матрицей Гессе, снабженного поправками второго порядка. Вместе с тем, примеры, демонстрирующие отсутствие сверхлинейной сходимости таких последовательностей, авторам не известны.

References

[1] J. Nocedal, S. J. Wright, *Numerical Optimization*, 2nd ed., Springer, New York, 2006.

[2] А. Ф. Измаилов, М. В. Солодов, *Численные методы оптимизации*, 2-е изд., перераб. и доп., Физматлит, М., 2008. [A. F. Izmailov, M. V. Solodov, *Chislenniye Metody Optimizatzii*, Fizmatlit Publ., Moscow, 2008 (In Russian)].

[3] 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.

[4] R. H. Byrd, “An example of irregular convergence in some constrained optimization methods that use the projected Hessian”, *Mathematical Programming*, **32** (1985), 232–237.

[5] Y. Yuan, “An only 2-step Q-superlinear convergence example for some algorithms that use reduced Hessian approximations”, *Mathematical Programming*, **32** (1985), 224–231.

[6] R. H. Byrd, “On the convergence of constrained optimization methods with accurate Hessian information on a subspace”, *SIAM J. on Numerical Analysis*, **27** (1990), 141–153.

[7] R. H. Byrd, J. Nocedal, “An analysis of reduced Hessian methods for constrained optimization”, *Mathematical Programming*, **49** (1990), 285–323.

[8] J. Nocedal, M. L. Overton, “Projected Hessian updating algorithms for nonlinearly constrained optimization”, *SIAM J. on Numerical Analysis*, **22** (1985), 821–850.

[9] J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal, C. Sagastizábal, *Numerical Optimization: Theoretical and Practical Aspects*, 2nd ed., Springer, Berlin, 2006.

[10] T. F. Coleman, A. R. Conn, “On the Local Convergence of a Quasi-Newton Method for a Nonlinear Programming Problem”, *SIAM Journal on Numerical Analysis*, **21**:4 (1984).

Информация об авторах

Волков Андрей Андреевич, магистр, кафедра исследования операций, Московский государственный университет им. М. В. Ломоносова, г. Москва, Российская Федерация. E-mail: and_rei99@mail.ru

Information about the authors

Andrey A. Volkov, Master, Operations Research Department, Lomonosov Moscow State University, Moscow, Russian Federation. E-mail: and_rei99@mail.ru

Измаилов Алексей Феридович, 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

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

Усков Евгений Иванович, научный сотрудник, 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

Evgeniy I. Uskov, Researcher, Derzhavin Tambov State University, Tambov, Russian Federation. E-mail: euskov@cs.msu.ru
ORCID: https://orcid.org/0000-0002-3639-0317

Конфликт интересов отсутствует.

There is no conflict of interests.

Для контактов:
Измаилов Алексей Феридович
E-mail: izmaf@cs.msu.ru

Corresponding author:
Alexey F. Izmailov
E-mail: izmaf@cs.msu.ru

Поступила в редакцию 21.01.2024 г.
Поступила после рецензирования 06.03.2024 г.
Принята к публикации 11.03.2024 г.

Received 21.01.2024
Reviewed 06.03.2024
Accepted for press 11.03.2024