Московский государственный университет им. М. В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра исследования операций

На правах рукописи

Куренной Алексей Святославович

НЬЮТОНОВСКИЕ МЕТОДЫ РЕШЕНИЯ ЗАДАЧ ОПТИМИЗАЦИИ С ЛИПШИЦЕВЫМИ ПРОИЗВОДНЫМИ

Специальность 01.01.09 — дискретная математика и математическая кибернетика

Диссертация на соискание учёной степени
кандидата физико-математических наук

Научный руководитель:
профессор, д.ф.-м.н.
Измаилов Алексей Феридович

Москва

2014

Оглавление

Введение

Имеющийся в литературе локальный анализ наиболее эффективных численных методов оптимизации традиционно предполагает двукратную непрерывную дифференцируемость целевой функции и ограничений задачи. Настоящая работа посвящена распространению результатов о локальной сходимости этих методов на задачи с более слабыми свойствами гладкости и одновременно построению единой теории их локальной сходимости. Кроме того, целью работы является изучение свойств локальной сходимости этих алгоритмов в различных предположениях о регулярности задачи, и, в частности, ослабление требований регулярности, на которые опираются существующие результаты об их локальной сходимости.

В работе рассматриваются *задачи с липшицевыми производными*, т. е. задачи оптимизации, производные целевой функции и ограничений которых являются локально липшицевыми. Напомним, что отображение $\Phi\colon \mathbb{R}^s \mapsto \mathbb{R}^r$ называется локально липшицевым в точке $z \in \mathbb{R}^s$, если оно удовлетворяет условию Липшица в некоторой окрестности этой точки, т. е. если существует число $\ell > 0$ и окрестность $U \subset \mathbb{R}^s$ точки z такие, что

$$\|\Phi(z_1) - \Phi(z_2)\| \leqslant \ell \|z_1 - z_2\| \quad \forall\, z_1,\, z_2 \in U.$$

При этом отображение называется локально липшицевым, если оно локально липшицево в каждой точке своей области определения.

Приведем типичный пример возникновения задачи оптимизации с липшицевыми производными. Предположим, что фирма может производить n типов товаров, и ее выпуск задается вектором $x \in \mathbb{R}^n$, j-я компонента которого равна производимому количеству j-го товара. Пусть при производстве товаров может выделяться N различных вредных веществ. Объем i-го вредного вещества, выделяемый при производстве единицы j-го товара равен a_{ij}. Если выделенное количество i-го вредного вещества превышает установленный государством допустимый уровень b_i, то фирма должна понести затраты по утилизации излишка, равные его квадрату. Тогда, если $p \in \mathbb{R}^n$ — вектор цен, а $c\colon \mathbb{R}^n \mapsto \mathbb{R}$ — функция издержек, то функция прибыли фирмы $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ имеет вид

$$f(x) = \langle p,\, x\rangle - c(x) - \sum_{i=1}^{N}\left(\max\left\{0,\, \sum_{j=1}^{n} a_{ij}x_j - b_i\right\}\right)^2.$$

Как легко проверить, функция f является дифференцируемой, и ее производная локально липшицева, но при этом функция f, вообще говоря, не является дважды дифференцируемой. Фирма может поставить перед собой задачу максимизации функции f при тех или иных ограничениях на выпуск. Если эти ограничения обладают соответствующими свойствами гладкости, то в результате возникает задача с липшицевыми производными.

Другим важным примером задач оптимизации с липшицевыми производными являются так называемые поднятые переформулировки задач оптимизации с комплементарными ограничениями. Задачей оптимизации с комплементарными ограничениями (см. [64, 71]) называется задача вида

$$f(x) \to \min, \quad G(x) \geqslant 0,\ H(x) \geqslant 0,\ \langle G(x),\, H(x)\rangle = 0,$$

где $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ — заданная функция, а $G,\, H\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ — заданные отображения. Один из подходов к решению таких задач оптимизации состоит в их переформулировке в виде задач с ограничениями равенствами (см. [49, 88]):

$$f(x) \to \min, \quad (\min\{0,\, y\})^2 - G(x) = 0,\ (\max\{0,\, y\})^2 - H(x) = 0,$$

где $y \in \mathbb{R}^m$ — дополнительная переменная, и операции взятия максимума/минимума и возведения в степень осуществляются покомпонентно. Такая переформулировка называется поднятой. Какими бы гладкими ни были отображения G и H, ограничения поднятой переформулировки не являются дважды дифференцируемыми. В то же время, если функция f и отображения G и H дифференцируемы, и их производные локально липшицевы, то переформулированная задача представляет собой задачу с липшицевыми производными.

Еще одним источником возникновения задач оптимизации с липшицевыми производными являются методы поиска обобщенного равновесия Нэша. Пусть имеется N агентов (игроков), и пусть стратегия i-го игрока представляет собой вектор размерности n_i. Требуется найти набор стратегий $\bar{x} = (\bar{x}_1,\, \bar{x}_2,\, \ldots,\, \bar{x}_N) \in \mathbb{R}^n$, $n = n_1 + \ldots + n_N$, $\bar{x}_i \in \mathbb{R}^{n_i}$, такой, что для каждого $i = 1,\, \ldots,\, N$ точка $\bar{x}_i$ является решением задачи оптимизации

$$f_i(\bar{x}_1,\, \ldots,\, \bar{x}_{i-1},\, x_i,\, \bar{x}_{i+1},\, \ldots,\, \bar{x}_N) \to \min, \quad (\bar{x}_1,\, \ldots,\, \bar{x}_{i-1},\, x_i,\, \bar{x}_{i+1},\, \ldots,\, \bar{x}_N) \in X,$$

где $f_i\colon \mathbb{R}^n \mapsto \mathbb{R}$ — функция потерь i-го игрока, непрерывная по совокупности переменных и выпуклая по переменной x_i, а $X \subset \mathbb{R}^n$ — непустое замкнутое и выпуклое множество.

Оказывается, что поиск обобщенного равновесия Нэша (см. [90, теорема 3.3]) может быть основан на решении задачи оптимизации

$$V_\alpha(x) \to \min, \quad x \in X,$$

целевая функция $V_\alpha\colon \mathbb{R}^n \mapsto \mathbb{R}$ которой задается формулой $V_\alpha(x) = \max_{y \in X} \Psi_\alpha(x,\, y)$, где $\Psi_\alpha\colon \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$ — регуляризованная функция Никайдо–Изода,

$$\Psi_\alpha(x,\, y) = \sum_{i=1}^{N} \Big(f_i(x_1,\, \ldots, x_{i-1},\, x_i,\, x_{i+1},\, \ldots,\, x_N) - \\ - f_i(x_1,\, \ldots, x_{i-1},\, y_i,\, x_{i+1},\, \ldots,\, x_N) - \frac{\alpha}{2}\|x_i - y_i\|^2 \Big),$$

($\alpha > 0$ — фиксированный параметр). Функция V_α корректно определена, поскольку при любом $\alpha > 0$ и любом $x \in \mathbb{R}^n$ функция $\Psi_\alpha(x,\, \cdot)$ сильно вогнута, а множество X непусто и замкнуто. Более того, поскольку X выпукло, существует единственный вектор $y_\alpha(x)$ такой, что $\Psi_\alpha(x,\, y_\alpha(x)) = V_\alpha(x)$. Предположим дополнительно, что функции f_i дважды непрерывно дифференцируемы, а $X = \{x \in \mathbb{R}^n \mid g(x) \leqslant 0\}$, где $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ — дважды непрерывно дифференцируемое отображение с выпуклыми компонентами. Тогда, как показано в [89], если $\bar{x} \in \mathbb{R}^n$ — обобщенное равновесие Нэша, и градиенты $g_j'(y_\alpha(\bar{x}))$, $j \in \{k = 1,\, \ldots,\, m \mid g_k(y_\alpha(\bar{x})) = 0\}$, линейно независимы, то функция V_α дифференцируема, и ее производная локально липшицева в точке $\bar{x}$, а значит, задача оптимизации

$$V_\alpha(x) \to \min, \quad g(x) \leqslant 0$$

локально (вблизи $\bar{x}$) представляет собой задачу оптимизации с липшицевыми производными.

Наконец, помимо указанных приложений, задачи оптимизации с липшицевыми производными возникают в оптимальном управлении (речь идет о так называемых обобщенных линейно-квадратичных задачах) [84, 85], в машинном обучении [65, 91] и др.

Подчеркнем, что задачи оптимизации, целевая функция и ограничения которых дважды непрерывно дифференцируемы, образуют подкласс задач оптимизации с липшицевыми производными. Несмотря на то, что этот подкласс хорошо изучен, многие результаты, полученные в настоящей работе, являются новыми и для него.

Итак, **объектом исследования** в диссертационной работе являются задачи оптимизации с липшицевыми производными, а также численные методы их решения.

Основной целью диссертационного исследования является построение единой теории сходимости ряда эффективных численных методов оптимизации, пригодных для решения задач с липшицевыми производными, и анализ сходимости этих методов применительно к задачам этого класса в как можно более слабых предположениях регулярности.

Актуальность темы диссертационной работы обусловлена тем фактом, что задачи с липшицевыми производными, с одной стороны, имеют широкий спектр приложений, а с другой стороны, являются малоизученными, особенно в плане обоснования соответствующих численных методов оптимизации.

Методику исследования составляют средства современного негладкого анализа, нелинейного анализа, теории оптимизации, а также подходы современной численной оптимизации. Исследование локальной сходимости численных методов оптимизации осуществляется в диссертации путем представления их в виде частного случая предварительно разработанных и проанализированных абстрактных итерационных алгоритмов.

Опишем **структуру диссертации**. Диссертационная работа состоит из введения, трех глав, заключения и списка литературы из 91 источника.

Первая глава посвящена ряду вспомогательных вопросов вариационного и негладкого анализа.

В разделе 1.1 получена полная картина соотношений между различными условиями регулярности для смешанных комплементарных задач. В разделе 1.2 изучаются соотношения между несколькими обобщенными дифференциальными объектами. В разделе 1.3 доказывается оценка расстояния до решения системы Каруша–Куна–Таккера задачи с липшицевыми производными.

Во второй главе разрабатываются методы решения обобщенных уравнений.

Методы, разрабатываемые в разделе 2.1, имеют абстрактный характер и представляют собой инструмент теоретического анализа численных методов для задач оптимизации и вариационного анализа. В разделе 2.2 рассматривается более специальная схема, являющаяся обобщением метода Джозефи–Ньютона на негладкий случай.

В третьей главе осуществляется анализ ряда эффективных численных методов оптимизации применительно к задачам с липшицевыми производными.

Раздел 3.1 посвящен изучению локальной сходимости метода модифицированных функций Лагранжа в различных предположениях. В частности, в нем на задачи с липшицевыми производными обобщается наиболее тонкий на текущий момент результат о локальной сходимости этого метода. Кроме того, для задач с ограничениями равенствами доказываются результаты о сходимости в еще более слабых предположениях. Эти результаты являются новыми и для задач, целевая функция и ограничения которых являются дважды непрерывно дифференцируемыми.

В разделе 3.2 рассматривается метод множителей с линеаризованными ограничениями. Для него, как и для метода модифицированных функций, на задачи с липшицевыми производными обобщается наиболее сильный на текущий момент результат о локальной сходимости. При этом улучшается оценка скорости сходимости, что делает соответствующий результат новым и в дважды дифференцируемом случае.

В разделе 3.3 изучается метод последовательного квадратичного программирования. Для него устанавливаются необходимые и достаточные условия прямой сверхлинейной сходимости.

В заключении перечислены основные результаты, полученные в диссертации.

Научная новизна исследования состоит в следующем:

- в диссертации разработаны новые итерационные схемы решения обобщенных уравнений;

- результаты о локальной сходимости метода модифицированных функций Лагранжа, полученные для задач со смешанными ограничениями, являются новыми в контексте задач с липшицевыми производными;

- результаты о локальной сходимости метода модифицированных функций, полученные для задач с ограничениями-равенствами, являются первыми результатами о его локальной сходимости без требования регулярности ограничений и в предположениях, более слабых, чем достаточное условие второго порядка (они новы даже в дважды дифференцируемом случае);

- результаты о локальной сходимости метода множителей с линеаризованными ограничениями являются новыми в контексте задач с липшицевыми производными;

- необходимые и достаточные условия прямой локальной сверхлинейной сходимости полугладкого метода последовательного квадратичного программирования представляют собой новые теоретические результаты;

- в работе установлены неизвестные ранее соотношения между важнейшими условиями регулярности для смешанных комплементарных задач;

- доказанная в работе оценка расстояния до решения системы Каруша–Куна–Таккера является новой в контексте задач с липшицевыми производными;

- установленные в разделе 1.2 соотношения между различными обобщенными дифференциальными объектами не были известны ранее и могут быть полезными при их использовании.

Перечислим **основные результаты, выносимые на защиту**.

1. Разработаны абстрактные схемы решения обобщенных уравнений, которые могут быть использованы для теоретического анализа различных численных методов решения задач оптимизации и вариационного анализа.

2. Развита теория локальной сходимости метода модифицированных функций Лагранжа и метода множителей с линеаризованными ограничениями применительно к задачам с липшицевыми производными.

3. Теория локальной сходимости полугладкого метода последовательного квадратичного программирования дополнена необходимыми и достаточными условиями его прямой сверхлинейной сходимости.

Достоверность научных положений обусловлена строгостью математических доказательств и использованных научных методов.

Список публикаций. Полученные в работе результаты опубликованы в [2]–[7], [43]–[48], в том числе 6 статей опубликовано в журналах из списка ВАК [5, 43, 45, 46, 47, 48].

Апробация результатов. Результаты, полученные в диссертации, были представлены на XXI международном симпозиуме по математическому программированию «ISMP2012» (Берлин, Германия), на международной конференции по непрерывной оптимизации «ICC OPT2013» (Лиссабон, Португалия), на X всемирном конгрессе по структурной и междисциплинарной оптимизации «WCSMO-10» (Орландо, США, 2013), на ежегодных международных научных конференциях студентов и молодых ученых «Ломоносов-2011», «Ломоносов-2012» (Москва), на ежегодной научной конференции «Тихоновские чтения» (Москва, 2012), на ежегодной научной конференции «Ломоносовские чтения» (Москва, 2011), а также на VII Московской международной конференции по исследованию операций «ORM2013».

Общепринятые обозначения специально не оговариваются, их пояснение вынесено в список обозначений. В работе используется следующая система нумерации ее частей. Номер раздела состоит из двух цифр, первая из которых обозначает номер главы, в которой расположен раздел. Аналогично, номер пункта состоит из трех цифр, первые две из которых составляют номер раздела, в котором находится этот пункт. Нумерация объектов (формул, теорем и т. д.) в каждой главе независимая. При ссылке на объект извне главы, в которой он находится, используется номер, состоящий из двух цифр, первая из которых является номером главы, а вторая номером объекта в главе. Под «условиями утверждения» (теоремы, предложения, леммы) всегда понимается все то, что сказано в этом утверждении до слова

«Тогда».

Автор выражает огромную благодарность своему научному руководителю Алексею Феридовичу Измаилову, а также своим родителям.

Список основных обозначений

$\mathbb{R}$ — множество вещественных чисел;

$\mathbb{R}_+$ — множество неотрицательных вещественных чисел;

$\mathbb{R}^n$ — n-мерное арифметическое пространство, снабженное евклидовым скалярным произведением и соответствующей нормой;

$\langle \cdot, \cdot \rangle$ — евклидово скалярное произведение;

$\| \cdot \|$ — евклидова норма вектора (если не оговорено иначе);

$\operatorname{conv} S$ — выпуклая оболочка множества S (минимальное выпуклое множество, содержащее S);

$\ker A$ — ядро (множество нулей) матрицы (линейного оператора) A;

A^{T} — матрица, транспонированная к матрице A;

$\operatorname{rank} A$ — ранг матрицы (линейного оператора) A;

$\pi_S(x)$ — проекция точки x на замкнутое выпуклое множество S;

$\operatorname{dist}(x, S) = \inf_{\xi \in S} \|x - \xi\|$ — расстояние от точки x до множества S;

$\operatorname{diag}(u)$ — диагональная $s \times s$-матрица с диагональю u, где $u \in \mathbb{R}^s$;

y_I — вектор с компонентами y_i, $i \in I$, где $y \in \mathbb{R}^s$, $I \subset \{1, \ldots, s\}$;

$M_{K_1 K_2}$ — подматрица матрицы M, отвечающая номерам строк $i \in K_1$ и номерам столбцов $j \in K_2$;

$B(u, \delta)$ — замкнутый шар радиуса δ с центром в точкеu;

$\blacksquare$ — знак окончания доказательства.

Глава 1

Элементы вариационного и негладкого анализа

Данная глава посвящена ряду вспомогательных вопросов и состоит из трех разделов. Первый раздел существенно отличается от двух других. В то время как материал второго и третьего раздела непосредственно относится к задачам оптимизации с липшицевыми производными, в первом разделе рассматриваются гладкие комплементарные задачи. Появление этого раздела объясняется следующими обстоятельствами. Один из подходов к решению задач оптимизации состоит в численном решении соответствующей системы Каруша–Куна–Таккера, которая представляет собой частный случай смешанной комплементарной задачи. При этом задачам оптимизации с пониженной гладкостью соответствуют негладкие системы Каруша–Куна–Таккера. Теория таких систем является естественным продолжением теории гладких комплементарных задач. Однако даже в последней, как оказалось, имелся ряд «белых пятен», устранение которых и стало предметом первого раздела данной главы.

1.1. Условия регулярности для смешанных комплементарных задач

В настоящем разделе будет рассмотрен ряд широко используемых условий регулярности для (гладких) смешанных комплементарных задач и получена полная картина взаимоотношений между этими условиями.

Напомним, что *смешанной комплементарной задачей* (СКЗ) называется вариационное неравенство на прямоугольнике:

$$z \in [\ell,\, u], \quad \langle \Phi(z),\, y - z \rangle \geqslant 0 \quad \forall\, y \in [\ell,\, u]. \tag{1}$$

Здесь $\Phi\colon \mathbb{R}^s \mapsto \mathbb{R}^s$ — заданное отображение, а $[\ell,\, u]$ есть (обобщенный) прямоугольник,

$$[\ell,\, u] = \{z \in \mathbb{R}^s \mid \ell_i \leqslant z_i \leqslant u_i,\ i = 1,\, \ldots,\, s\},$$

задаваемый числами $\ell_i \in \mathbb{R} \cup \{-\infty\}$ и $u_i \in \mathbb{R} \cup \{+\infty\}$, $\ell_i < u_i$, $i = 1,\, \ldots,\, s$. Эквивалентным образом СКЗ может быть сформулирована так:

$$z \in [\ell,\, u], \quad \Phi_i(z) \begin{cases} \geqslant 0 & \text{if } z_i = \ell_i, \\ = 0 & \text{if } z_i \in (\ell_i,\, u_i), \\ \leqslant 0 & \text{if } z_i = u_i, \end{cases} \quad i = 1,\, \ldots,\, s. \tag{2}$$

Важными частными случаями смешанной комплементарной задачи являются *нелинейная комплементарная задача* (НКЗ)

$$z \geqslant 0, \quad \Phi(z) \geqslant 0, \quad \langle z,\, \Phi(z)\rangle = 0, \tag{3}$$

соответствующая случаю, когда $\ell_i = 0$, $u_i = +\infty$, $i = 1,\, \ldots,\, s$, и система Каруша–Куна–Таккера (ККТ)

$$\begin{gathered} F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu = 0, \quad h(x) = 0, \\ \mu \geqslant 0, \quad g(x) \leqslant 0, \quad \langle \mu,\, g(x)\rangle = 0, \end{gathered} \tag{4}$$

относительно неизвестных $(x,\, \lambda,\, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$. Здесь $F\colon \mathbb{R}^n \to \mathbb{R}^n$, $h\colon \mathbb{R}^n \to \mathbb{R}^l$ и $g\colon \mathbb{R}^n \to \mathbb{R}^m$ — заданные отображения, причем последние два предполагаются дифференцируемыми. Для того чтобы представить систему ККТ (4) в форме смешанной комплементарной задачи (2), достаточно положить $s = n + l + m$,

$$\begin{gathered} \ell_i = -\infty,\ i = 1,\, \ldots,\, n + l, \quad \ell_i = 0,\ i = n + l + 1,\, \ldots,\, n + l + m, \\ u_i = +\infty,\ i = 1,\, \ldots,\, n + l + m, \end{gathered}$$

и определить отображение Φ по правилу

$$\Phi(z) = (G(x,\, \lambda,\, \mu),\, h(x),\, -g(x)), \quad z = (x,\, \lambda,\, \mu),$$

где $G\colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^n$ — отображение вида

$$G(x,\, \lambda,\, \mu) = F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu.$$

Отметим, что СКЗ (1) с $\Phi(z) = \varphi'(z)$, $z \in \mathbb{R}^s$, представляет собой набор условий первого порядка оптимальности в задаче

$$\varphi(z) \to \min, \quad z \in [\ell,\, u],$$

где $\varphi : \mathbb{R}^s \to \mathbb{R}$ — некоторая гладкая функция.

Другой хорошо известный факт (см., например, [15, 34]) состоит в том, что СКЗ (2) может быть переформулирована в виде системы нелинейных уравнений с использованием так называемой *функции дополнительности* $\psi \colon \mathbb{R}^2 \to \mathbb{R}$, удовлетворяющей условию

$$\psi(a,\, b) = 0 \quad \Longleftrightarrow \quad a \geqslant 0,\ b \geqslant 0,\ ab = 0.$$

Предполагая наличие у этой функции дополнительных свойств

$$\psi(a,\, b) < 0 \quad \forall\, a > 0,\, b < 0, \quad \psi(a,\, b) > 0 \quad \forall\, a > 0,\, b > 0, \tag{5}$$

с учетом эквивалентной переформулировки (2) смешанной комплементарной задачи (1) легко видеть, что множество ее решений совпадает с множеством решений уравнения

$$\Psi(z) = 0 \tag{6}$$

с оператором $\Psi \colon \mathbb{R}^s \to \mathbb{R}^s$ вида

$$\Psi_i(z) = \begin{cases} \Phi_i(z) & \text{if } i \in I_\Phi, \\ \psi(z_i - l_i, \Phi_i(z)) & \text{if } i \in I_\ell, \\ -\psi(u_i - z_i, -\Phi_i(z)) & \text{if } i \in I_u, \\ \psi(z_i - l_i, -\psi(u_i - z_i, -\Phi_i(z))) & \text{if } i \in I_{\ell u}, \end{cases} \tag{7}$$

где

$$\begin{aligned} I_\Phi &= \{i = 1,\, \dots,\, s \mid \ell_i = -\infty,\, u_i = +\infty\}, \\ I_\ell &= \{i = 1,\, \dots,\, s \mid \ell_i > -\infty,\, u_i = +\infty\}, \\ I_u &= \{i = 1,\, \dots,\, s \mid \ell_i = -\infty,\, u_i < +\infty\}, \\ I_{\ell u} &= \{i = 1,\, \dots,\, s \mid \ell_i > -\infty,\, u_i < +\infty\}. \end{aligned}$$

Двумя наиболее часто используемыми функциями дополнительности (обе они обладают свойствами (5)) является функция естественной невязки (или функция минимума)

$$\psi(a,\, b) = \min\{a,\, b\}$$

и функция Фишера–Бурмейстера

$$\psi(a,\, b) = a + b - \sqrt{a^2 + b^2}.$$

Отображение Ψ, заданное согласно (7) с использованием функции естественной невязки, будет обозначаться через Ψ_{NR}, а с использованием функции Фишера–Бурмейстера — через Ψ_{FB}. Отметим, что решение уравнения (6) с $\Psi = \Psi_{NR}$ или $\Psi = \Psi_{FB}$ методами ньютоновского типа является одним из наиболее эффективных численных подходов к решению комплементарных задач.

Для заданного решения $\bar{z} \in \mathbb{R}^s$ СКЗ (1) (или, что то же самое, (2)) определим множества индексов

$$I_+ = I_+(\bar{z}) = \{i = 1, \ldots, s \mid \Phi_i(\bar{z}) = 0, \, \bar{z}_i \in (\ell_i, \, u_i)\},$$
$$I_0 = I_0(\bar{z}) = \{i = 1, \ldots, s \mid \Phi_i(\bar{z}) = 0, \; \bar{z}_i \in \{\ell_i, \, u_i\}\},$$
$$I_N = I_N(\bar{z}) = \{i = 1, \ldots, s \mid \Phi_i(\bar{z}) \neq 0\}.$$

Заметим, что, вне зависимости от свойств гладкости отображения Φ, при нарушении условия строгой дополнительности $I_0 = \emptyset$ как отображение $\Psi = \Psi_{NR}$, так и $\Psi = \Psi_{FB}$ может быть недифференцируемым в точке $\bar{z}$. Однако, эти отображения являются локально липшицевыми в точке $\bar{z}$, если отображение Φ локально липшицево в этой точке. В связи с этим, дифференциальными объектами, подходящими для изучения указанных отображений, являются B-дифференциал и дифференциал Кларка. Напомним, что B-дифференциал отображения $F\colon \mathbb{R}^s \mapsto \mathbb{R}^r$ в точке $\bar{z} \in \mathbb{R}^s$ определяется как множество

$$\partial_B F(\bar{z}) = \{M \in \mathbb{R}^{r \times s} \mid \exists \, \{z^k\} \subset \mathcal{D}_F : \{z^k\} \to \bar{z}, \, \{F'(z^k)\} \to M\},$$

где $\mathcal{D}_F \subset \mathbb{R}^s$ — множество точек дифференцируемости отображения F. Дифференциалом Кларка называется выпуклая оболочка B-дифференциала:

$$\partial F(\bar{z}) = \operatorname{conv} \partial_B F(\bar{z}).$$

(см. [21, раздел 2.6.1], [32, раздел 7.1]).

Пусть отображение Φ дифференцируемо вблизи $\bar{z}$, причем его производная непрерывна в точке $\bar{z}$ (что влечет за собой локальную липшицевость отображения Φ в точке $\bar{z}$). Из определения отображения Ψ_{NR} непосредственно выводится следующая верхняя оценка для множества $\partial_B \Psi_{NR}(\bar{z})$ (см., например, [55]): строки любой матрицы из $J \in \partial_B \Psi_{NR}(\bar{z})$ удовлетворяют соотношениям

$$J_i \begin{cases} = \Phi_i'(\bar{z}) & \text{if } i \in I_+, \\ \in \{\Phi_i'(\bar{z}), \, e^i\} & \text{if } i \in I_0, \\ = e^i & \text{if } i \in I_N, \end{cases} \tag{8}$$

где через e^i обозначена i-я строка единичной матрицы размера $s \times s$, $i = 1, \ldots, s$. Обозначим множество матриц из $\mathbb{R}^{s \times s}$, строки которых удовлетворяют соотношениям (8), символом $\Delta_{NR}(\bar{z})$. Таким образом, $\partial_B \Psi_{NR}(\bar{z}) \subset \Delta_{NR}(\bar{z})$.

Напрямую из определений может быть выведена верхняя оценка и для множества $\partial_B \Psi_{FB}(\bar{z})$. А именно, строки любой матрицы из $J \in \partial_B \Psi_{FB}(\bar{z})$ удовлетворяют соотношениям

$$J_i = \begin{cases} \Phi_i'(\bar{z}) & \text{if } i \in I_+, \\ \alpha_i \Phi_i'(\bar{z}) + \beta_i e^i & \text{if } i \in I_0, \\ e^i & \text{if } i \in I_N, \end{cases} \tag{9}$$

где пара чисел $(\alpha_i,\, \beta_i) \in \mathbb{R} \times \mathbb{R}$ принадлежит окружности

$$S = \{(a,\, b) \in \mathbb{R}^2 \mid (a-1)^2 + (b-1)^2 = 1\} \tag{10}$$

для каждого $i \in I_0$. Отметим, что аналогичная оценка для множества $\partial \Psi_{FB}(\bar{z})$ была получена в [15]. Обозначая через $\Delta_{FB}(\bar{z})$ множество матриц, строки которых удовлетворяют соотношениям (9) при некоторых $(\alpha_i,\, \beta_i) \in S$, $i \in I_0$, можем записать: $\partial_B \Psi_{FB}(\bar{z}) \subset \Delta_{FB}(\bar{z})$.

Ниже будет приведен список наиболее широко используемых условий регулярности для смешанных комплементарных задач, включая условия, основанные на понятиях упомянутых выше обобщенных дифференциальных объектов. В пункте 1.1.1 будет получена полная картина взаимоотношений между условиями регулярности для общего случая СКЗ, а также для специального случая НКЗ. Пункт 1.1.2 посвящен системам ККТ.

Следует отметить, что все рассматриваемые условия регулярности имеют большую важность, поскольку каждое из них является ключевым предположением в анализе локальной сходимости определенных численных методов решения комплементарных задач и задач оптимизации. Полная картина соотношений между условиями регулярности позволяет сравнивать между собой теории этих методов, и этим обусловлено ее большое теоретическое и практическое значение. Понимание того, как соотносятся различные теории численных методов между собой, во-первых, важно с теоретической точки зрения, а во-вторых, может быть полезно при выборе алгоритмов для решения комплементарных задач и задач оптимизации на практике.

Будем говорить, что множество (квадратных) матриц невырождено, если все принадлежащие ему матрицы невырождены. Следующие условия регулярности играют ключевую роль в обосновании локальной сверхлинейной сходимости полугладкого метода Ньютона, применяемого к уравнению (6) с локально липшицевым оператором (см. [60, 61, 75, 78], а также [32, Раздел 7.5]).

Определение 1. Отображение $\Psi\colon \mathbb{R}^s \to \mathbb{R}^s$ называется *BD-регулярным* (*CD-регулярным*) в точке $\bar{z} \in \mathbb{R}^s$, если множество $\partial_B \Psi(\bar{z})$ $(\partial \Psi(\bar{z}))$ не вырождено.

Если $\bar{z} \in \mathbb{R}^s$ — решение СКЗ (1), то в силу верхних оценок для множеств $\partial_B \Psi_{NR}(\bar{z})$ и $\partial_B \Psi_{FB}(\bar{z})$, приведенных выше, BD-регулярность отображения Ψ_{NR} (Ψ_{FB}) в точке $\bar{z}$ следует из невырожденности множества $\Delta_{NR}(\bar{z})$ (соответственно, $\Delta_{FB}(\bar{z})$), в то время как CD-регулярность отображения Ψ_{NR} (Ψ_{FB}) в точке $\bar{z}$ следует из невырожденности $\operatorname{conv} \Delta_{NR}(\bar{z})$ (соответственно, $\operatorname{conv} \Delta_{FB}(\bar{z})$).

Определим множества

$$\Sigma = \{(a,\, b) \in \mathbb{R}^2 \mid a + b = 1,\ a \geqslant 0,\ b \geqslant 0\},$$

и

$$B = \operatorname{conv} S = \{(a,\, b) \in \mathbb{R}^2 \mid (a-1)^2 + (b-1)^2 \leqslant 1\}.$$

Из определения множества $\Delta_{NR}(\bar{z})$ следует, что $\operatorname{conv} \Delta_{NR}(\bar{z})$ есть совокупность всех матриц $J \in \mathbb{R}^{s \times s}$, строки которых удовлетворяют условию (9) с некоторыми числами $(\alpha_i,\, \beta_i) \in \Sigma$, $i \in I_0$. Аналогичным образом, легко установить, что множество $\operatorname{conv} \Delta_{FB}(\bar{z})$ совпадает с множеством всех матриц $J \in \mathbb{R}^{s \times s}$, строки которых удовлетворяют соотношению (9) с некоторыми числами $(\alpha_i,\, \beta_i) \in B$, $i \in I_0$.

Отметим, что в случае нелинейной комплементарной задачи (3) невырожденность множества $\Delta_{NR}(\bar{z})$ известна в литературе под названием *b-регулярности* решения $\bar{z}$ [72]. Это условие эквивалентно тому, что матрица

$$\begin{pmatrix} (\Phi'(\bar{z}))_{I_+ I_+} & (\Phi'(\bar{z}))_{I_+ K} \\ (\Phi'(\bar{z}))_{K I_+} & (\Phi'(\bar{z}))_{KK} \end{pmatrix}$$

является невырожденной для любого подмножества $K \subset I_0$.

В случае систем Каруша–Куна–Таккера (4), невырожденность множества $\Delta_{NR}(\bar{z})$ есть то же самое, что и *квази-регулярность* решения $\bar{z} = (\bar{x},\, \bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ [31]. Предположим, что отображение F дифференцируемо вблизи $\bar{x}$, причем его производная является непрерывной в точке $\bar{x}$, а отображения h и g являются дважды дифференцируемыми вблизи $\bar{x}$, и их вторые производные непрерывны в точке $\bar{x}$. Определим множества индексов

$$\begin{gathered} A = A(\bar{x}) = \{i = 1,\, \ldots,\, m \mid g_i(\bar{x}) = 0\}, \\ A_+ = A_+(\bar{x},\, \bar{\mu}) = \{i \in A \mid \bar{\mu}_i > 0\}, \quad A_0 = A_0(\bar{x},\, \bar{\mu}) = \{i \in A \mid \bar{\mu}_i = 0\}. \end{gathered} \tag{11}$$

Квази-регулярность эквивалентна тому, что матрица

$$\begin{pmatrix} \dfrac{\partial G}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) & (h'(\bar{x}))^{\mathrm{T}} & (g'_{A_+ \cup K}(\bar{x}))^{\mathrm{T}} \\ h'(x) & 0 & 0 \\ g'_{A_+ \cup K}(\bar{x}) & 0 & 0 \end{pmatrix}$$

является невырожденной для любого множества индексов $K \subset A_0$.

Возвращаясь к общему случаю смешанных комплементарных задач, в дополнение к перечисленным выше условиям регулярности рассмотрим следующие условия.

Определение 2. Решение $\bar{z}$ СКЗ (1) называется *сильно регулярным*, если для каждого $r \in \mathbb{R}^s$, достаточно близкого к 0, возмущенная линеаризованная СКЗ

$$z \in [\ell,\, u], \quad \langle \Phi(\bar{z}) + \Phi'(\bar{z})(z - \bar{z}) - r,\, y - z\rangle \geqslant 0 \quad \forall\, y \in [\ell,\, u],$$

имеет вблизи $\bar{z}$ единственное решение $z(r)$, и отображение $z(\cdot)$ является локально липшицевым в точке 0.

Понятие сильной регулярности было введено в [81] и продолжает играть важную роль в вариационном анализе (см., например, [27, Глава 2], [32, Глава 5], а также библиографические ссылки в этих работах). В частности, оно является ключевым предположением в анализе локальной сходимости различных итерационных схем решения вариационных задач (см. [27, Глава 6]).

В [81] была получена простая алгебраическая характеризация сильной регулярности для НКЗ. Эта характеризация была позднее обобщена на случай СКЗ в [30]. Для того, чтобы сформулировать ее, напомним, что квадратная матрица M называется P-матрицей, если все ее главные миноры положительны. Решение СКЗ $\bar{z}$ является сильно регулярным тогда и только тогда, когда матрица $(\Phi'(\bar{z}))_{I_+I_+}$ невырождена, и ее дополнение Шура

$$(\Phi'(\bar{z}))_{I_0I_0} - (\Phi'(\bar{z}))_{I_0I_+}(\Phi'(\bar{z}))^{-1}_{I_+I_+}(\Phi'(\bar{z}))_{I_+I_0}$$

в матрице

$$\begin{pmatrix} (\Phi'(\bar{z}))_{I_+I_+} & (\Phi'(\bar{z}))_{I_+I_0} \\ (\Phi'(\bar{z}))_{I_0I_+} & (\Phi'(\bar{z}))_{I_0I_0} \end{pmatrix}$$

является P-матрицей.

Следующее более слабое условие регулярности было введено в [17] и является основным ингредиентом ряда тонких результатов о локальной сходимости методов ньютоновского типа для вариационных задач. Другие приложения этого свойства связаны с теорией чувствительности, см. [32, Разделы 5.3, 6.2].

Определение 3. Решение $\bar{z}$ СКЗ (1) называется *полуустойчивым*, если существует константа $C > 0$ такая, что для любого $r \in \mathbb{R}^s$ всякое решение $z(r)$ возмущенной комплементарной задачи

$$z \in [\ell,\, u], \quad \langle \Phi(z) - r,\, y - z\rangle \geqslant 0 \quad \forall\, y \in [\ell,\, u],$$

достаточно близкое к $\bar{z}$, удовлетворяет оценке

$$\|z(r) - \bar{z}\| \leqslant C\|r\|.$$

1.1.1. Смешанные комплементарные задачи

Известные соотношения

Начнем с некоторых соотношений между BD-регулярностью и CD-регулярностью. Прежде всего, очевидно, что $\Delta_{NR}(\bar{z}) \subset \Delta_{FB}(\bar{z})$, а значит, справедливо и соответствующее включение для выпуклых оболочек этих множеств: $\operatorname{conv} \Delta_{NR}(\bar{z}) \subset \operatorname{conv} \Delta_{FB}(\bar{z})$. В частности, невырожденность множества $\Delta_{FB}(\bar{z})$ влечет за собой невырожденность множества $\Delta_{NR}(\bar{z})$.

Обратное включение, вообще говоря, не имеет места. Более того, из невырожденности множества $\Delta_{NR}(\bar{z})$ не следует ни CD-регулярность отображения Ψ_{NR} в точке $\bar{z}$, ни BD-регулярность отображения Ψ_{FB} в точке $\bar{z}$, что демонстрирует следующий пример, взятый из [24, Пример 2.1].

Пример 1. Пусть $s = 2$. Рассмотрим НКЗ (3) с $\Phi(z) = (-z_1 + z_2, -z_2)$. Точка $\bar{z} = 0$ является единственным решением этой НКЗ.

Непосредственно проверяется, что

$$\partial_B \Psi_{NR}(\bar{z}) = \Delta_{NR}(\bar{z}) = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} \right\},$$

а значит, множество $\Delta_{NR}(\bar{z})$ не вырождено. С другой стороны, матрица

$$\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 \\ 0 & 1 \end{pmatrix}$$

вырождена, а значит, отображение Ψ_{NR} не является CD-регулярным в точке $\bar{z}$.

Для каждого $k = 1, 2, \ldots$ положим $z^k = (1/k, 2/k)$. Тогда отображение Ψ_{FB} дифференцируемо в точках z^k,

$$\Psi'_{FB}(z^k) = \begin{pmatrix} 0 & 1 - 1/\sqrt{2} \\ 0 & -\sqrt{2} \end{pmatrix},$$

и последовательность $\{z^k\}$ сходится к $\bar{z}$. Следовательно, вырожденная матрица в правой части последнего выражения лежит в $\partial_B \Psi_{FB}(\bar{z})$, а значит, отображение Ψ_{FB} не является BD-регулярным в точке $\bar{z}$.

Следующий пример, позаимствованный из [23, Пример 2], показывает, что множество $\partial_B \Psi_{FB}(\bar{z})$ может быть собственным подмножеством $\Delta_{FB}(\bar{z})$. Более того, может случиться, что каждое из отображений Ψ_{NR} и Ψ_{FB} является CD-регулярным в точке $\bar{z}$, но при этом множества $\Delta_{NR}(\bar{x})$ и $\Delta_{FB}(\bar{x})$ содержат вырожденные матрицы.

Пример 2. Пусть $s = 2$, и $\Phi(z) = ((z_1 + z_2)/2,\, (z_1 + z_2)/2)$. Тогда точка $\bar{z} = 0$ является единственным решением НКЗ (3).

Непосредственной проверкой убеждаемся, что

$$\partial_B \Psi_{NR}(\bar{z}) = \left\{ \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix},\, \begin{pmatrix} 1/2 & 1/2 \\ 0 & 1 \end{pmatrix} \right\},$$

а значит,

$$\partial \Psi_{NR}(\bar{z}) = \left\{ \begin{pmatrix} t + (1-t)/2 & (1-t)/2 \\ t/2 & (1-t) + t/2 \end{pmatrix} \middle|\, t \in [0,\, 1] \right\}.$$

Путем непосредственного вычисления получаем, что $\det J = 1/2$ для любой матрицы $J \in \partial \Psi_{NR}(\bar{z})$, откуда следует CD-регулярность (а значит и BD-регулярность) отображения Ψ_{NR} в точке $\bar{z}$. В то же время,

$$\Delta_{NR}(\bar{z}) = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\, \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix},\, \begin{pmatrix} 1/2 & 1/2 \\ 0 & 1 \end{pmatrix},\, \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \right\},$$

где матрица

$$J_0 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \tag{12}$$

является вырожденной.

Далее, множество $\Delta_{FB}(\bar{z})$ состоит из матриц вида

$$J = J(\alpha,\, \beta) = \begin{pmatrix} \alpha_1/2 + \beta_1 & \alpha_1/2 \\ \alpha_2/2 & \alpha_2/2 + \beta_2 \end{pmatrix} \tag{13}$$

для всех $(\alpha_i,\, \beta_i) \in S$, $i = 1,\, 2$. Путем непосредственного вычисления получаем:

$$\det J(\alpha,\, \beta) = \frac{1}{2}(\alpha_1 \beta_2 + \alpha_2 \beta_1 + 2\beta_1 \beta_2).$$

Поскольку из включения $(\alpha_i,\, \beta_i) \in S$ следует, что $\alpha_i \geqslant 0$, $\beta_i \geqslant 0$ для $i = 1,\, 2$, указанный определитель равен нулю тогда и только тогда, когда

$$\alpha_1 \beta_2 = 0, \quad \alpha_2 \beta_1 = 0, \quad \beta_1 \beta_2 = 0,$$

т. е. $\alpha_1 = \alpha_2 = 1$, $\beta_1 = \beta_2 = 0$, откуда следует, что матрица J_0, определенная в (12), является единственной вырожденной матрицей во множестве $\Delta_{FB}(\bar{z})$. Более того, в свете приведенного выше представления для множества $\operatorname{conv} \Delta_{FB}(\bar{z})$ становится ясно, что J_0 — единственная вырожденная матрица и в этом множестве. При этом, как легко проверить, эта матрица не принадлежит множеству $\partial_B \Psi_{FB}(\bar{z})$. Но тогда эта матрица не принадлежит и множеству $\partial \Psi_{FB}(\bar{z})$, ведь, как видно из (13), J_0 не может не быть крайней точкой множества $\partial \Psi_{FB}(\bar{z})$. Таким образом, отображение Ψ_{FB} является CD-регулярным (а значит, и BD-регулярным) в точке $\bar{z}$.

Можно предположить, что BD-регулярность (или, по крайней мере, CD-регулярность) отображения Ψ_{FB} в точке $\bar{z}$ влечет за собой CD-регулярность (или хотя бы BD-регулярность) отображения Ψ_{NR} в точке $\bar{z}$, но, как показывает следующий пример, взятый из [1], это не так.

Пример 3. Пусть $s = 2$, $\Phi(z) = (z_2, \, -z_1 + z_2)$. Тогда $\bar{z} = 0$ является единственным решением НКЗ (3), причем, как легко видеть, множество $\partial_B \Psi_{NR}(\bar{z})$ содержит вырожденную матрицу

$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix},$$

в то время как отображение Ψ_{FB} является CD-регулярным в точке $\bar{z}$. Действительно, несложно убедиться в том, что указанная матрица является единственной вырожденной матрицей в $\mathrm{conv}\, \Delta_{FB}(\bar{z})$ и при этом не принадлежит множеству $\partial_B \Psi_{FB}(\bar{z})$. Кроме того, она не может быть некрайней точкой множества $\partial \Psi_{FB}(\bar{z})$. Следовательно, эта матрица не лежит в $\partial \Psi_{FB}(\bar{z})$, и отображение Ψ_{FB} является CD-регулярным в точке $\bar{z}$.

Помимо указанных фактов, известно, что BD-регулярность отображения Ψ_{FB} в точке $\bar{z}$ не влечет за собой CD-регулярность отображения Ψ_{FB} в $\bar{z}$. Это демонстрируется следующим простым примером.

Пример 4. Пусть $s = 1$, $\Phi(z) = -z$. Тогда точка $\bar{z} = 0$ является единственным решением НКЗ (3).

Очевидно, что $\Psi_{FB}(z) = -\sqrt{2}|z|$, $\partial_B \Psi_{FB}(\bar{z}) = \{-\sqrt{2}, \, \sqrt{2}\}$, откуда следует BD-регулярность отображения Ψ_{FB} в точке $\bar{z}$. В то же время, множество $\partial \Psi_{FB}(\bar{z}) = \Delta_{FB}(\bar{z}) = [-\sqrt{2}, \, \sqrt{2}]$ содержит 0.

Наконец, приведем известные соотношения, касающиеся сильной регулярности и полуустойчивости.

В [23] было показано, что BD-регулярность любого из отображений Ψ_{NR} и Ψ_{FB} в решении СКЗ $\bar{z}$ влечет полуустойчивость этого решения. Обратная импликация не имеет места, как видно из примеров 1 и 2.

Что касается сильной регулярности, как следует из доказательства [34, Теорема 1], если $\bar{z}$ является сильно регулярным решением СКЗ (1), то множество $\operatorname{conv}\Delta_{FB}(\bar{z})$ (а значит, и $\operatorname{conv}\Delta_{NR}(\bar{z})$) не вырождено. В частности, сильная регулярность не может следовать из условий, которые не влекут за собой невырожденность множества $\operatorname{conv}\Delta_{FB}(\bar{z})$.

Соотношения между условиями регулярности, о которых шла речь в настоящем пункте, представлены в таблице 1.1. Знак «+» («−») в ячейке означает, что свойство, соответствующее строке данной ячейки, влечет (не влечет) за собой свойство, указанное в заголовке ее столбца. Знак вопроса говорит о том, что до настоящего момента вопрос о наличии или отсутствии соответствующей импликации оставался открытым.

Таблица 1.1. Условия регулярности для СКЗ: известные соотношения.

Свойство	Полуустойчивость	BD-регулярность Ψ_{NR}	BD-регулярность Ψ_{FB}	CD-регулярность Ψ_{NR}	CD-регулярность Ψ_{FB}	Невырожденность Δ_{NR}	Невырожденность Δ_{FB}	Невырожденность $\operatorname{conv}\Delta_{NR}$	Невырожденность $\operatorname{conv}\Delta_{FB}$	Сильная регулярность
Semistability	+	−	−	−	−	−	−	−	−	−
BD-регулярность Ψ_{NR}	+	+	−	−	−	−	−	−	−	−
BD-регулярность Ψ_{FB}	+	−	+	−	−	−	−	−	−	−
CD-регулярность Ψ_{NR}	+	+	?	+	?	−	−	−	−	−
CD-регулярность Ψ_{FB}	+	−	+	−	+	−	−	−	−	−
Невырожденность Δ_{NR}	+	+	−	−	−	+	−	−	−	−
Невырожденность Δ_{FB}	+	+	+	?	?	+	+	?	?	?
Невырожденность $\operatorname{conv}\Delta_{NR}$	+	+	?	+	?	+	?	+	?	?
Невырожденность $\operatorname{conv}\Delta_{FB}$	+	+	+	+	+	+	+	+	+	?
Сильная регулярность	+	+	+	+	+	+	+	+	+	+

Недостающие соотношения

Займемся устранением знаков вопроса в таблице 1.1. Начнем с доказательства следующего факта.

Предложение 1. *Для решения $\bar{z}$ СКЗ* (1) *следующие три свойства эквивалентны:*

а) *множество* $\operatorname{conv}\Delta_{NR}(\bar{z})$ *невырождено;*

б) *множество* $\Delta_{FB}(\bar{z})$ *невырождено;*

в) *множество* $\operatorname{conv}\Delta_{FB}(\bar{z})$ *невырождено.*

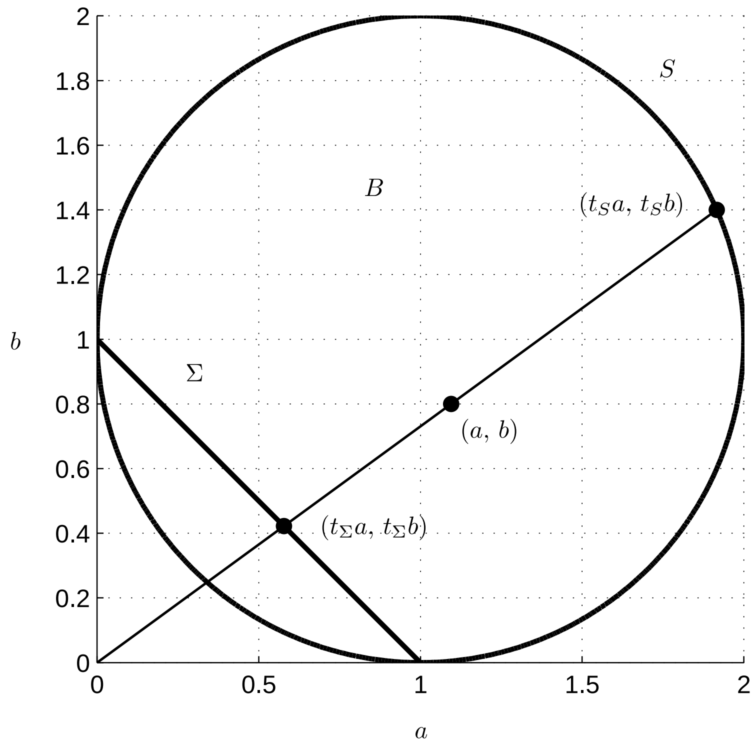

Рис. 1.1. Множества Σ, S и B.

Доказательство. Ключевое наблюдение, делающее сформулированное утверждение очевидным, состоит в том, что для всякой пары $(a,\, b) \in B$ можно указать число $t_\Sigma > 0$ такое, что $(t_\Sigma a,\, t_\Sigma b) \in \Sigma$, и число $t_S > 0$ такое, что $(t_S a,\, t_S b) \in S$ (см. рисунок 1.1). Следовательно, любая матрица J вида (9) с $(\alpha_i,\, \beta_i) \in B$ для всех $i \in I_0$ путем домножения некоторых ее строк на положительные числа может быть преобразована в матрицу такого же вида, но с $(\alpha_i,\, \beta_i) \in \Sigma$ или $(\alpha_i,\, \beta_i) \in S$ для всех $i \in I_0$. Отсюда сразу получаем требуемое. ∎

Далее, любое из эквивалентных свойств а)–в) в утверждении 1 влечет за собой сильную регулярность решения $\bar{z}$. Доказательству этого факта предпошлем следующую лемму.

Лемма 1. *Если матрица $M \in \mathbb{R}^{p\times p}$ не является P-матрицей, то найдутся наборы чисел $\alpha,\, \beta \in \mathbb{R}^p$ такие, что $(\alpha_i,\, \beta_i) \in S$ для всех $i = 1,\, \dots,\, p$, где множество S определено в* (10)*, и при этом матрица*

$$M(\alpha,\, \beta) = \mathrm{diag}(\alpha) M + \mathrm{diag}(\beta) \tag{14}$$

вырождена.

Доказательство. Проведем индукцию по p. Если $p = 1$, предположение о том, что матрица M не является P-матрицей, означает, что M есть неположительное число. Приравняв правую часть формулы (14) к нулю, получим уравнение прямой в плоскости $(\alpha,\, \beta)$, которая имеет непустое пересечение с окружностью S. Любая из точек в этом пересечении является искомой парой $(\alpha,\, \beta)$.

Предположим теперь, что утверждение леммы выполнено для любой матрицы размера $(p-1)\times(p-1)$, и что матрица $M \in \mathbb{R}^{p\times p}$ с элементами $m_{i,\,j}$, $i,\, j = 1,\, \dots,\, p$, имеет неположительный главный минор. Если единственный такой минор есть $\det M$, положим $\alpha_i = 1$, $\beta_i = 0$, $i = 2,\, \dots,\, p$, и разложим определитель $\det M(\alpha,\, \beta)$ по первой строке:

$$\begin{aligned}
\det M(\alpha,\, \beta) &= \det \begin{pmatrix} \alpha_1 m_{11} + \beta_1 & \alpha_1 m_{12} & \dots & \alpha_1 m_{1p} \\ m_{21} & m_{22} & \dots & m_{2p} \\ \dots & \dots & \dots & \dots \\ m_{p1} & m_{p2} & \dots & m_{pp} \end{pmatrix} \\
&= \alpha_1 \det M + \beta_1 \det M_{\{2,\,\dots,\,p\}\{2,\,\dots,\,p\}}.
\end{aligned}$$

Поскольку $\det M \leqslant 0$, а $\det M_{\{2,\,\dots,\,p\}\{2,\,\dots,\,p\}} > 0$, прямая, задаваемая в плоскости $(\alpha_1,\, \beta_1)$ уравнением $\det M(\alpha,\, \beta) = 0$, имеет непустое пересечение с окружностью S, и, взяв в качестве α_1 и β_1 компоненты произвольной точки из этого пересечения, получим искомые векторы α и β.

Остается рассмотреть случай существования подмножества $K \subset \{1,\, \dots,\, p\}$ такого, что $\det M_{KK} \leqslant 0$, и найдется $k \in \{1,\, \dots,\, p\} \setminus K$. Удалив из матрицы M строку и столбец с номером k, получим матрицу $\tilde{M} \in \mathbb{R}^{(p-1)\times(p-1)}$ с неположительным главным минором. По предположению индукции, существуют векторы $\tilde{\alpha} = (\alpha_1,\, \dots,\, \alpha_{k-1},\, \alpha_{k+1},\, \dots,\, \alpha_p) \in \mathbb{R}^{p-1}$, $\tilde{\beta} = (\beta_1,\, \dots,\, \beta_{k-1},\, \beta_{k+1},\, \dots,\, \beta_p) \in \mathbb{R}^{p-1}$ такие, что $(\alpha_i,\, \beta_i) \in S$ для всех $i = 1,\, \dots,\, p$, $i \neq k$, и матрица $\mathrm{diag}(\tilde{\alpha})\tilde{M} + \mathrm{diag}(\tilde{\beta})$ вырождена. Полагая $\alpha_k = 0$, $\beta_k = 1$, получаем векторы α и β с требуемыми свойствами. ■

Предложение 2. *Если точка $\bar{z}$ является решением СКЗ* (1)*, и при этом множество $\Delta_{FB}(\bar{z})$ невырождено, то решение $\bar{z}$ сильно регулярно.*

Доказательство. Невырожденность всех матриц в множестве $\Delta_{FB}(\bar{z})$ эквивалентна тому, что матрица

$$D(\alpha,\,\beta) = \begin{pmatrix} (\Phi'(\bar{z}))_{I_+I_+} & (\Phi'(\bar{z}))_{I_+I_0} \\ \mathrm{diag}(\alpha)(\Phi'(\bar{z}))_{I_0I_+} & \mathrm{diag}(\alpha)(\Phi'(\bar{z}))_{I_0I_0} + \mathrm{diag}(\beta) \end{pmatrix} \tag{15}$$

невырождена при любых $\alpha = (\alpha_i,\ i \in I_0)$ и $\beta = (\beta_i,\ i \in I_0)$ таких, что $(\alpha_i,\,\beta_i) \in S$, $i \in I_0$. Полагая $\alpha_i = 0$, $\beta_i = 1$, $i \in I_0$, отсюда сразу получаем, что матрица $(\Phi'(\bar{z}))_{I_+I_+}$ невырождена.

Предположим теперь, что решение СКЗ $\bar{z}$ не является сильно регулярным. Тогда из невырожденности матрицы $(\Phi'(\bar{z}))_{I_+I_+}$ следует, что матрица

$$(\Phi'(\bar{z}))_{I_0I_0} - (\Phi'(\bar{z}))_{I_0I_+}((\Phi'(\bar{z}))_{I_+I_+})^{-1}(\Phi'(\bar{z}))_{I_+I_0}$$

не является P-матрицей. Тогда, по лемме 1, существуют наборы чисел $\alpha = (\alpha_i,\ i \in I_0)$ и $\beta = (\beta_i,\ i \in I_0)$ такие, что $(\alpha_i,\,\beta_i) \in S$, $i \in I_0$, и при этом матрица

$$\mathrm{diag}(\alpha)((\Phi'(\bar{z}))_{I_0I_0} - (\Phi'(\bar{z}))_{I_0I_+}((\Phi'(\bar{z}))_{I_+I_+})^{-1}(\Phi'(\bar{z}))_{I_+I_0}) + \mathrm{diag}(\beta)$$

вырождена. Но эта матрица представляет собой дополнение Шура невырожденной матрицы $(\Phi'(\bar{z}))_{I_+I_+}$ в $D(\alpha,\,\beta)$, и следовательно, матрица $D(\alpha,\,\beta)$ также вырождена (см., к примеру, [8, Теорема 1.3.1.1]), что не возможно. ■

Объединяя предложения 1 и 2 и тот факт, что сильная регулярность решения $\bar{z}$ влечет за собой невырожденность множества $\mathrm{conv}\,\Delta_{FB}(\bar{z})$, заключаем, что свойства а)–в) из предложения 1 эквиваленты сильной регулярности.

Теперь остается лишь выяснить, следует ли из CD-регулярности отображения Ψ_{NR} в точке $\bar{z}$, что отображение Ψ_{FB} является CD-регулярным (или хотя бы BD-регулярным) в точке $\bar{z}$. Следующий пример показывает, что это не так. Более того, даже комбинация CD-регулярности отображения Ψ_{NR} в точке $\bar{z}$ и невырожденности множества $\Delta_{NR}(\bar{z})$ не дает BD-регулярность (а уж тем более CD-регулярность) отображения Ψ_{FB} в точке $\bar{z}$.

Пример 5. Пусть $n = 2$, $\Phi(z) = (-z_1 + 3z_2/(2\sqrt{2}),\ 2z_1 + (1 - 3/(2\sqrt{2}))z_2)$. Тогда точка $\bar{z} = 0$ является единственным решением соответствующей НКЗ.

Рассмотрим произвольную последовательность $\{z^k\} \subset \mathbb{R}^2$ такую, что $z_1^k < 0$, $z_2^k = 0$ для всех k, и z_1^k стремится к 0 при $k \to \infty$. Тогда для всех k отображение Ψ_{FB} дифференцируемо в точке z^k, и

$$\Psi'_{FB}(z^k) = \begin{pmatrix} -\dfrac{2z_1^k}{\sqrt{2(z_1^k)^2}} & \dfrac{3}{2\sqrt{2}}\left(1 + \dfrac{z_1^k}{\sqrt{2(z_1^k)^2}}\right) \\ 2\left(1 - \dfrac{2z_1^k}{\sqrt{(2z_1^k)^2}}\right) & 1 + \left(1 - \dfrac{3}{2\sqrt{2}}\right)\left(1 - \dfrac{2z_1^k}{\sqrt{(2z_1^k)^2}}\right) \end{pmatrix} =$$

$$= \begin{pmatrix} \sqrt{2} & \dfrac{3}{2\sqrt{2}}\left(1 - \dfrac{1}{\sqrt{2}}\right) \\ 4 & 1 + 2\left(1 - \dfrac{3}{2\sqrt{2}}\right) \end{pmatrix},$$

а значит, вырожденная матрица в правой части последнего равенства лежит в $\partial_B \Psi_{FB}(\bar{z})$.

В то же время, непосредственной проверкой можно установить, что

$$\partial_B \Psi_{NR}(\bar{z}) = \left\{ \begin{pmatrix} 1 & 0 \\ 2 & 1 - \dfrac{3}{2\sqrt{2}} \end{pmatrix},\ \begin{pmatrix} -1 & \dfrac{3}{2\sqrt{2}} \\ 0 & 1 \end{pmatrix} \right\},$$

и следовательно,

$$\partial \Psi_{NR}(\bar{z}) = \left\{ \begin{pmatrix} t - (1-t) & (1-t)\dfrac{3}{2\sqrt{2}} \\ 2t & t\left(1 - \frac{3}{2\sqrt{2}}\right) + (1-t) \end{pmatrix} \,\middle|\, t \in [0,\,1] \right\}.$$

Для любой матрицы $J(t)$ в правой части последнего соотношения имеем

$$\det J(t) = \left(2 - \frac{3}{2\sqrt{2}}\right)t - 1 \leqslant 1 - \frac{3}{2\sqrt{2}} < 0 \quad \forall\, t \in [0,\,1],$$

откуда следует CD-регулярность отображения Ψ_{NR} в точке $\bar{z}$.

Наконец, заметим, что

$$\Delta_{NR}(\bar{z})$$
$$= \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ \begin{pmatrix} 1 & 0 \\ 2 & 1 - \dfrac{3}{2\sqrt{2}} \end{pmatrix},\ \begin{pmatrix} -1 & \dfrac{3}{2\sqrt{2}} \\ 0 & 1 \end{pmatrix},\ \begin{pmatrix} -1 & \dfrac{3}{2\sqrt{2}} \\ 2 & 1 - \dfrac{3}{2\sqrt{2}} \end{pmatrix} \right\},$$

и значит, множество $\Delta_{NR}(\bar{z})$ невырождено, т. е. точка $\bar{z}$ является b-регулярным решением рассматриваемой НКЗ.

Пример 5 показывает, что предполагая невырожденность всех матриц из выпуклой оболочки множества $\partial_B \Psi_{NR}(\bar{z})$ и невырожденность всех матриц из другого расширения этого множества — $\Delta_{NR}(\bar{z})$, нельзя гарантировать даже BD-регулярность отображения Ψ_{FB} в точке $\bar{z}$. В то же время, как показывают утверждения 1 и 2, используя оба вида расширения множества $\partial_B \Psi_{NR}(\bar{z})$ *вместе* (т. е. предполагая невырожденность матриц из множества $\operatorname{conv} \Delta_{NR}(\bar{z})$), мы приходим к сильной регулярности решения $\bar{z}$.

1.1.2. Системы Каруша–Куна–Таккера

Некоторые импликации, неверные в общем случае СКЗ, оказываются справедливыми в специальном случае систем ККТ.

Дело в том, что в отличие от случая НКЗ, для систем ККТ формула (8) дает не просто верхнюю оценку B-дифференциала отображения Ψ_{NR}, а точную характеризацию этого множества: для всякого решения $\bar{z} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ системы ККТ (4) справедливо равенство $\partial_B \Psi_{NR}(\bar{z}) = \Delta_{NR}(\bar{z})$, которое влечет за собой равенство $\partial \Psi_{NR}(\bar{z}) = \operatorname{conv} \Delta_{NR}(\bar{z})$. Данный факт объясняется тем, что прямая переменная x и двойственная переменная μ «разделяются» по разным аргументам функции естественной невязки, и для любой матрицы $J \in \Delta_{NR}(\bar{z})$ легко построить последовательность $\{z^k\} \subset \mathcal{S}_{\Psi_{NR}}$ такую, что $\{z^k\} \to \bar{z}$ и $\{\Psi'_{NR}(z^k)\} \to J$.

Как следствие, для систем ККТ BD-регулярность отображения Ψ_{NR} в решении $\bar{z}$ эквивалентна невырожденности множества $\Delta_{NR}(\bar{z})$ (т. е. квази-регулярности решения), а CD-регулярность отображения Ψ_{NR} в решении $\bar{z}$ эквивалентна невырожденности множества $\operatorname{conv} \Delta_{NR}(\bar{z})$.

Обратимся теперь к отображению Ψ_{FB}. Легко видеть, что формула (9) дает точную характеризацию B-дифференциала этого отображения в решении системы ККТ $\bar{z}$, если градиенты $g'_i(\bar{x})$, $i \in A_0$, линейно независимы. Покажем, что последнее условие автоматически выполняется в решении $\bar{z}$, если отображение Ψ_{FB} является BD-регулярным в этом решении.

Предложение 3. *Пусть отображение $F\colon \mathbb{R}^n \to \mathbb{R}^n$ дифференцируемо в окрестности точки $\bar{x} \in \mathbb{R}^n$, причем его производная непрерывна в этой точке, и пусть отображения $h\colon \mathbb{R}^n \to \mathbb{R}^l$ и $g\colon \mathbb{R}^n \to \mathbb{R}^m$ дважды дифференцируемы в окрестности точки $\bar{x}$, причем их вторые производные непрерывны в этой точке. Пусть вектор $\bar{z} = (\bar{x}, \bar{\lambda}, \bar{\mu})$, где $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$, является решением системы ККТ* (4).

Тогда, если отображение Ψ_{FB} является BD-регулярным в точке $\bar{z}$, то в точке $\bar{x}$ выполнено условие линейной независимости: градиенты $h'_j(\bar{x})$, $j = 1, \ldots, l$, $g'_i(\bar{x})$, $i \in A$, линейно независимы.

Доказательство. Зафиксируем последовательность $\{\mu^k\} \subset \mathbb{R}^m$ такую, что

$$\mu_A^k > 0,\ \mu_{\{1,\,\ldots,\,m\}\setminus A}^k = 0 \quad \forall\, k,$$

и $\{\mu_A^k\} \to \bar{\mu}_A$. Тогда последовательность $\{(\bar{x},\, \bar{\lambda},\, \mu^k)\}$ сходится к $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, и, как легко видеть, для каждого номера k отображение Ψ_{FB} дифференцируемо в точке $(\bar{x},\, \bar{\lambda},\, \mu^k)$, причем соответствующая последовательность матриц Якоби этого отображения сходится к матрице

$$\begin{pmatrix} \dfrac{\partial G}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) & (h'(\bar{x}))^{\mathrm{T}} & (g_A'(\bar{x}))^{\mathrm{T}} & (g_{\{1,\,\ldots,\,m\}\setminus A}'(\bar{x}))^{\mathrm{T}} \\ h'(x) & 0 & 0 & 0 \\ -g_A'(\bar{x}) & 0 & 0 & 0 \\ 0 & 0 & 0 & I \end{pmatrix}$$

(с точностью до перестановки строк и столбцов), где символом I обозначена единичная матрица соответствующего размера. Отсюда получаем требуемое. ∎

Из этого предложения и предшествующих обсуждений следует, то для систем ККТ BD-регулярность отображения Ψ_{FB} в решении $\bar{z}$ влечет за собой невырожденность множества $\Delta_{FB}(\bar{z})$, а значит, в силу предложения 1, невырожденность множества $\operatorname{conv}\Delta_{FB}(\bar{z})$ (из которой в свою очередь следует CD-регулярность отображения Ψ_{FB} в точке $\bar{z}$). Кроме того, ясно, что CD-регулярность отображения Ψ_{FB} в решении $\bar{z}$ системы ККТ влечет невырожденность множества $\Delta_{FB}(\bar{z})$.

Таким образом, для решения $\bar{z}$ системы ККТ можно выделить три группы эквивалентных свойств. Первая группа состоит из BD-регулярности отображения Ψ_{FB} в точке $\bar{z}$, CD-регулярности отображения Ψ_{NR} в точке $\bar{z}$, CD-регулярности отображения Ψ_{FB} в точке $\bar{z}$, невырожденности множества $\Delta_{FB}(\bar{z})$, невырожденности множества $\operatorname{conv}\Delta_{NR}(\bar{z})$, невырожденности множества $\operatorname{conv}\Delta_{FB}(\bar{z})$ и сильной регулярности решения $\bar{z}$. Вторая группа объединяет в себе BD-регулярность отображения Ψ_{NR} в точке $\bar{z}$ и невырожденность множества $\Delta_{NR}(\bar{z})$. Наконец, последняя группа образована полуустойчивостью. Условия первой группы влекут за собой условия второй, а из условий второй группы следует полуустойчивость.

Обратные импликации не имеют места. Действительно, НКЗ из примера 4 соответствует прямым условиям первого порядка оптимальности в задаче

$$-\tfrac{1}{2}x^2 \to \min,$$
$$x \geqslant 0.$$

Единственным решением соответствующей системы Каруша–Куна–Таккера является точка $\bar{z} = (\bar{x},\, \bar{\mu}) = (0,\, 0)$, и, как легко видеть, это решение является квази-регулярным, но при этом множество $\Delta_{FB}(\bar{z})$ содержит вырожденную матрицу.

Тот факт, что из полуустойчивости решения системы ККТ не следует BD-регулярность отображения Ψ_{NR}, может быть продемонстрирован примером 1 из [50].

Завершая раздел, представим соотношения между условиями регулярности в виде диаграммы (см. рисунок 1.2). Сплошные линии на этой диаграмме соответствуют импликациям, которые верны для общего случая СКЗ, а штриховыми линиями обозначены дополнительные импликации, справедливые в специальном случае систем ККТ. Приведенная диаграмма является полной: существование пути из одного ее блока в другой *эквивалентно* тому, что из свойства, соответствующего первому блоку, следует свойство, соответствующее второму блоку. Последнее справедливо как для общего случая СКЗ, так и для НКЗ (в силу того, что во всех контрпримерах, представленных в пункте 1.1.1, фигурируют только НКЗ). С учетом штриховых стрелок, то же самое верно и для систем ККТ.

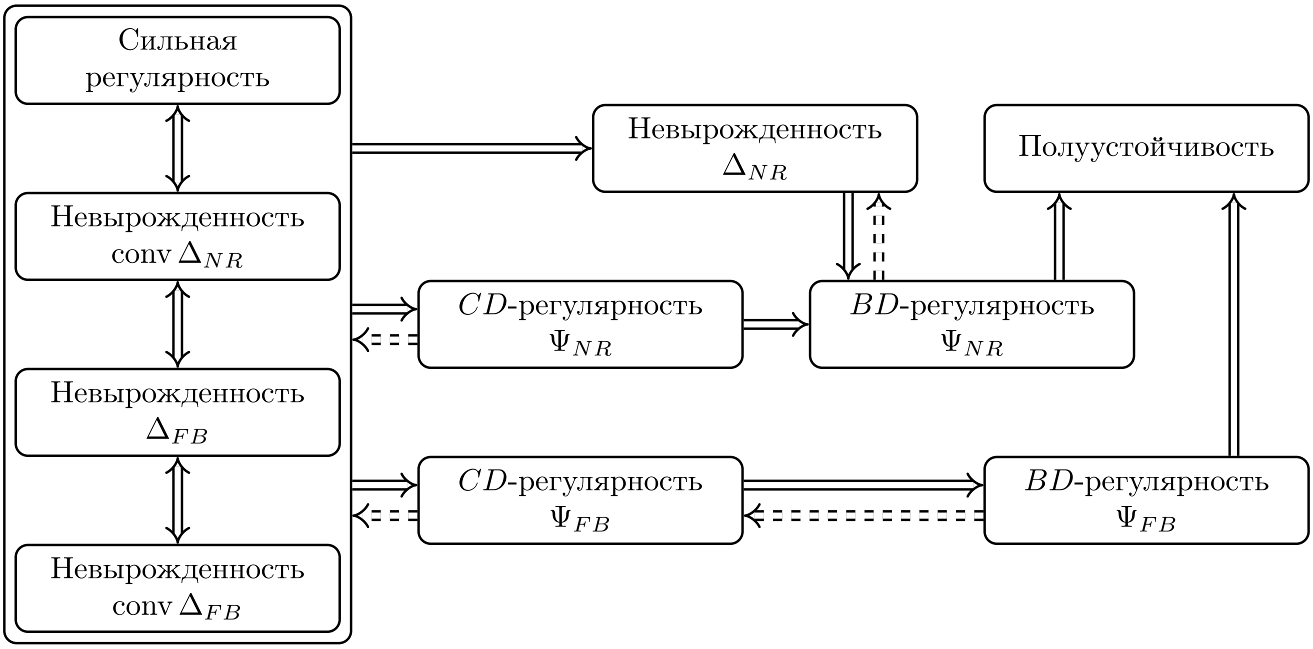

Рис. 1.2. Условия регулярности: полная диаграмма взаимосвязей.

1.2. Обобщенные дифференциалы

В контексте задач оптимизации, целевая функция и ограничения которых являются дважды непрерывно дифференцируемыми, наибольшую эффективность имеют численные методы, использующие информацию второго порядка, т. е. задействующие матрицу Гессе функции Лагранжа задачи либо ее аппроксимации. В случае задач с липшицевыми производными указанная матрица Гессе может не существовать, и ее роль выполняют обобщенные дифференциальные объекты, определения которых, как правило, основаны на понятиях B-

дифференциала и дифференциала Кларка (см. с. 14).

Пусть $\Phi\colon \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^r$. Частным B-дифференциалом отображения Φ по переменной x в точке $\bar{z} = (\bar{x},\, \bar{y}) \in \mathbb{R}^p \times \mathbb{R}^q$ называется множество

$$(\partial_B)_x \Phi(\bar{x},\, \bar{y}) = \partial_B \Phi(\cdot,\, \bar{y})\big|_{\bar{x}},$$

а частным дифференциалом Кларка — множество

$$\partial_x \Phi(\bar{x},\, \bar{y}) = \partial \Phi(\cdot,\, \bar{y})\big|_{\bar{x}} \ (= \operatorname{conv}(\partial_B)_x \Phi(\bar{x},\, \bar{y})).$$

Также введем в рассмотрение оператор Π_x проектирования пространства $\mathbb{R}^{r\times p} \times \mathbb{R}^{r\times q}$ на подпространство $\mathbb{R}^{r\times p}$:

$$\Pi_x[M_1 \ M_2] = M_1, \quad M_1 \in \mathbb{R}^{r\times p},\ M_2 \in \mathbb{R}^{r\times q}.$$

Основной целью настоящего раздела является изучение взаимосвязи между частным дифференциалом Кларка $\partial_x \Phi(\bar{x},\, \bar{y})$ и проекцией полного дифференциала Кларка $\Pi_x \partial \Phi(\bar{x},\, \bar{y})$. Множества как первого, так и второго типа используются в ряде численных методов решения задач с липшицевыми производными (см., например, [39, 76]). В связи с этим важно понимать, как они соотносятся друг с другом, и какие последствия имеет замена одного на другое в практических алгоритмах.

Помимо $\partial_x \Phi(\bar{x},\, \bar{y})$ и $\Pi_x \partial \Phi(\bar{x},\, \bar{y})$ будем использовать следующие обобщенные дифференциалы

$$(\tilde{\partial}_B)_x \Phi(\bar{x},\, \bar{y}) = \left\{ M \in \mathbb{R}^{r\times p} \,\middle|\, \exists \{(x^k,\, y^k)\} \subset \mathbb{R}^p \times \mathbb{R}^q : \right.$$
$$\Phi \ \text{ дифференцируемо по } x \text{ в } (x^k,\, y^k)\ \forall\, k,\ \{(x^k,\, y^k)\} \to (\bar{x},\, \bar{y}),$$
$$\left. \left\{ \frac{\partial \Phi}{\partial x}(x^k,\, y^k) \right\} \to M \right\},$$

$$\tilde{\partial}_x \Phi(\bar{x},\, \bar{y}) = \operatorname{conv}(\tilde{\partial}_B)_x \Phi(\bar{x},\, \bar{y}),$$

$$(\hat{\partial}_B)_x \Phi(\bar{x},\, \bar{y}) = \left\{ M \in \mathbb{R}^{r\times p} \,\middle|\, \exists \{(x^k,\, y^k)\} \subset \mathcal{D}_\Phi : \right.$$
$$\left. \{(x^k,\, y^k)\} \to (\bar{x},\, \bar{y}),\ \left\{ \frac{\partial \Phi}{\partial x}(x^k,\, y^k) \right\} \to M \right\},$$

$$\hat{\partial}_x \Phi(\bar{x},\, \bar{y}) = \operatorname{conv}(\hat{\partial}_B)_x \Phi(\bar{x},\, \bar{y}).$$

1.2.1. Общий случай

Начнем со следующего вспомогательного факта.

Лемма 2. *Пусть отображение* $\Phi\colon \mathbb{R}^p \times \mathbb{R}^q \mapsto \mathbb{R}^r$ *является липшицевым в некоторой окрестности точки* $(\bar{x},\, \bar{y}) \in \mathbb{R}^p \times \mathbb{R}^q$.

Тогда

$$\Pi_x \partial_B \Phi(\bar{x},\, \bar{y}) = (\hat{\partial}_B)_x \Phi(\bar{x},\, \bar{y}) \subset (\tilde{\partial}_B)_x \Phi(\bar{x},\, \bar{y}), \tag{16}$$

$$(\partial_B)_x \Phi(\bar{x},\, \bar{y}) \subset (\tilde{\partial}_B)_x \Phi(\bar{x},\, \bar{y}). \tag{17}$$

Доказательство. Начнем с доказательства равенства в (16).

Включение $\Pi_x \partial_B \Phi(\bar{x},\, \bar{y}) \subset (\hat{\partial}_B)_x \Phi(\bar{x},\, \bar{y})$ очевидно. Покажем обратное включение.

Принадлежность матрицы M множеству $(\hat{\partial}_B)_x \Phi(\bar{x},\, \bar{y})$ означает, что найдется последовательность

$$\{(x^k,\, y^k)\} \subset \mathcal{D}_\Phi : \{(x^k,\, y^k)\} \to (\bar{x},\, \bar{y}),\ \left\{ \frac{\partial \Phi}{\partial x}(x^k,\, y^k) \right\} \to M. \tag{18}$$

Из липшицевости отображения Φ в окрестности точки $(\bar{x},\, \bar{y})$ следует, что его производная в любой точке $z = (x,\, y) \in \mathcal{D}_\Phi$ из этой окрестности не превосходит по норме константу Липшица $\ell > 0$ отображения Φ. За неимением адекватной ссылки, приведем доказательство этого (безусловно хорошо известного) факта. Действительно, зафиксируем произвольный элемент $\zeta \in \mathbb{R}^p \times \mathbb{R}^q$, $\|\zeta\| = 1$. При всех достаточно малых $t > 0$ выполняется

$$t\|\Phi'(z)\zeta\| \leqslant \|\Phi(z + t\zeta) - \Phi(z)\| + o(t) \leqslant t\ell\|\zeta\| + o(t).$$

Разделив левую и правую части на t и перейдя к пределу при $t \to 0+$, получаем неравенство $\|\Phi'(z)\zeta\| \leqslant \ell$, что с учетом произвольности ζ дает требуемую оценку $\|\Phi'(z)\| \leqslant \ell$.

Таким образом, последовательность $\{\Phi'(x^k,\, y^k)\}$ ограничена, а значит, имеет предельную точку $\tilde{M}$. Тогда $\tilde{M} \in \partial_B \Phi(\bar{x},\, \bar{y})$, причем из (18) следует, что $\tilde{M} = [M\ \ \tilde{M}_2]$, где $\tilde{M}_2 \in \mathbb{R}^{r \times q}$ — некоторая матрица. Поэтому $M \in \Pi_x \partial_B \Phi(\bar{x},\, \bar{y})$. Тем самым доказано включение $(\hat{\partial}_B)_x \Phi(\bar{x},\, \bar{y}) \subset \Pi_x \partial_B \Phi(\bar{x},\, \bar{y})$, что и завершает доказательство равенства в (16).

Включение в (16) и включение (17) очевидны. ■

Таким образом, обобщенные дифференциалы

$$\Pi_x \partial_B \Phi(\bar{x},\, \bar{y}),\ (\tilde{\partial}_B)_x \Phi(\bar{x},\, \bar{y}),\ (\partial_B)_x \Phi(\bar{x},\, \bar{y})$$

отличаются тем, в каком множестве должны лежать последовательности, фигурирующие в определении этих обобщенных дифференциалов. Для $\Pi_x \partial_B \Phi(\bar{x},\, \bar{y})$ таким «допустимым» множеством является $\mathcal{D}_\Phi$, для $(\tilde{\partial}_B)_x \Phi(\bar{x},\, \bar{y})$ — множество точек, в которых Φ дифференцируемо по x. В случае $(\partial_B)_x \Phi(\bar{x},\, \bar{y})$ это множество точек вида $(x,\, \bar{y})$, в которых Φ дифференцируемо по x.

Замечание 1. Из (16) и (17) легко вытекают аналогичные соотношения для множеств $\Pi_x \partial \Phi(\bar{x},\, \bar{y})$, $\tilde{\partial}_x \Phi(\bar{x},\, \bar{y})$ и $\partial_x \Phi(\bar{x},\, \bar{y})$,

$$\Pi_x \partial \Phi(\bar{x},\, \bar{y}) = \hat{\partial}_x \Phi(\bar{x},\, \bar{y}) \subset \tilde{\partial}_x \Phi(\bar{x},\, \bar{y}), \tag{19}$$

$$\partial_x \Phi(\bar{x},\, \bar{y}) \subset \tilde{\partial}_x \Phi(\bar{x},\, \bar{y}). \tag{20}$$

Для доказательства достаточно учесть два факта. Первый состоит в том, что для любых множеств $S_1 \subset \mathbb{R}^{r\times p}$ и $S_2 \subset \mathbb{R}^{r\times p}$ из $S_1 \subset S_2$ следует $\operatorname{conv} S_1 \subset \operatorname{conv} S_2$. Второй факт заключается в том, что для любого линейного оператора $\Lambda\colon \mathbb{R}^{r\times(p+q)} \mapsto \mathbb{R}^{r\times p}$ и любого множества $S \subset \mathbb{R}^{r\times(p+q)}$ выполняется равенство $\Lambda(\operatorname{conv} S) = \operatorname{conv} \Lambda(S)$.

Для кусочно-линейной функции $\Phi\colon \mathbb{R}^2 \mapsto \mathbb{R}$, построенной в примере 2.5.2 из [21], выполнено

$$\partial_x \Phi(0,0) \subset \Pi_x \partial \Phi(0,0),\ \ \partial_x \Phi(0,0) \neq \Pi_x \partial \Phi(0,0),$$

из чего следует, что для нее включения (17) и (20) выполняются строгим образом, и, значит, в приведенных утверждениях они не могут быть заменены на равенства.

1.2.2. Отображения специального вида

Перейдем к рассмотрению отображений $\Phi\colon \mathbb{R}^p \times \mathbb{R}^q \mapsto \mathbb{R}^r$ вида

$$\Phi(x,\, y) = K(x)y + b(x), \tag{21}$$

где $K\colon \mathbb{R}^p \mapsto \mathbb{R}^{r\times q}$ и $b\colon \mathbb{R}^p \mapsto \mathbb{R}^r$ — заданные отображения.

Оказывается, что для отображений такого вида дифференцируемость по переменной x и дифференцируемость по совокупности переменных эквивалентны.

Лемма 3. *Пусть отображение $\Phi\colon \mathbb{R}^p \times \mathbb{R}^q \mapsto \mathbb{R}^r$ имеет вид* (21)*, причем отображение $K\colon \mathbb{R}^p \mapsto \mathbb{R}^{r\times q}$ непрерывно в точке $\bar{x}$. Тогда если $\Phi(x, y)$ дифференцируемо по x в точке $(\bar{x},\, \bar{y})$, то оно дифференцируемо по совокупности переменных в этой точке. При этом*

$$\Phi'(\bar{x},\, \bar{y}) = [\Phi'_x(\bar{x},\, \bar{y})\ \ K(\bar{x})].$$

Доказательство. Дифференцируемость Φ по x в $(\bar{x},\, \bar{y})$ означает, что

$$K(\bar{x}+\xi)\bar{y} + b(\bar{x}+\xi) - K(\bar{x})\bar{y} - b(\bar{x}) - \Phi'_x(\bar{x},\, \bar{y})\xi = \bar{o}(\|\xi\|).$$

Тогда

$$\begin{aligned} &K(\bar{x}+\xi)(\bar{y}+\eta) + b(\bar{x}+\xi) - K(\bar{x})\bar{y} - b(\bar{x}) = \\ &\qquad = K(\bar{x}+\xi)\bar{y} + b(\bar{x}+\xi) - K(\bar{x})\bar{y} - b(\bar{x}) + K(\bar{x}+\xi)\delta y = \\ &\qquad\qquad = \Phi'_x(\bar{x},\, \bar{y})\xi + K(\bar{x}+\xi)\eta + \bar{o}(\|\xi\|) = \\ &\qquad\qquad\qquad = \Phi'_x(\bar{x},\, \bar{y})\xi + K(\bar{x})\eta + (K(\bar{x}+\xi) - K(\bar{x}))\eta + \bar{o}(\|\xi\|). \end{aligned}$$

Но в силу непрерывности K в точке $\bar{x}$ имеем $(K(\bar{x}+\xi)-K(\bar{x}))\eta = \bar{o}(\|\eta\|)$, что и дает требуемое. ∎

В следующей лемме устанавливается еще одно дифференциальное свойство отображений вида (21).

Лемма 4. *Пусть отображение* $\Phi\colon \mathbb{R}^p \times \mathbb{R}^q \mapsto \mathbb{R}^r$ *имеет вид* (21)*, причем отображение* $K\colon \mathbb{R}^p \mapsto \mathbb{R}^{r\times q}$ *является липшицевым в некоторой окрестности точки* $\bar{x} \in \mathbb{R}^p$ *с константой* $\ell_K > 0$.

Тогда если Φ *дифференцируемо по* x *в точках* $(\bar{x},\, y^1)$ *и* $(\bar{x},\, y^2)$ *при некоторых* $y^1,\, y^2 \in \mathbb{R}^q$*, то*

$$\left\| \frac{\partial \Phi}{\partial x}(\bar{x},\, y^1) - \frac{\partial \Phi}{\partial x}(\bar{x},\, y^2) \right\| \leqslant \ell_K \|y^1 - y^2\|. \tag{22}$$

Доказательство. Дифференцируемость Φ по переменной x в точках $(\bar{x}, y^1)$ и $(\bar{x}, y^2)$ означает, что для $\xi \in \mathbb{R}^p$

$$\begin{aligned} &K(\bar{x}+\xi)y^j + b(\bar{x}+\xi) - K(\bar{x})y^j - b(\bar{x}) - \frac{\partial \Phi}{\partial x}(\bar{x},\, y^j)\xi = \\ &\qquad = \Phi(\bar{x}+\xi,\, y^j) - \Phi(\bar{x},\, y^j) - \frac{\partial \Phi}{\partial x}(\bar{x},\, y^j)\xi = o(\|\xi\|), \quad j = 1,\, 2. \end{aligned}$$

Отсюда следует соотношение

$$(K(\bar{x}+\xi) - K(\bar{x}))(y^1 - y^2) - \left(\frac{\partial \Phi}{\partial x}(\bar{x},\, y^1) - \frac{\partial \Phi}{\partial x}(\bar{x},\, y^2) \right) \xi = o(\|\xi\|). \tag{23}$$

Зафиксируем произвольный элемент $\xi \in \mathbb{R}^p$, $\|\xi\| = 1$. В силу (23), используя липшицевость отображения K в окрестности точки $\bar{x}$, при $t > 0$ имеем

$$t \left\| \left(\frac{\partial \Phi}{\partial x}(\bar{x},\, y^1) - \frac{\partial \Phi}{\partial x}(\bar{x},\, y^2) \right) \xi \right\| \leqslant$$
$$\leqslant \|K(\bar{x} + t\xi) - K(\bar{x})\| \|y^1 - y^2\| + o(t) \leqslant \ell_K t \|y^1 - y^2\| + o(t).$$

Разделив левую и правую части на t и перейдя к пределу при $t \to 0+$, получаем неравенство

$$\left\| \left(\frac{\partial \Phi}{\partial x}(\bar{x}, y^1) - \frac{\partial \Phi}{\partial x}(\bar{x}, y^2) \right) \xi \right\| \leqslant \ell_K \|y^1 - y^2\|,$$

что с учетом произвольности ξ и дает требуемую оценку (22). ∎

Основной результат данного раздела содержится в следующей теореме.

Теорема 1. *Пусть отображение* $\Phi\colon \mathbb{R}^p \times \mathbb{R}^q \mapsto \mathbb{R}^r$ *имеет вид* (21)*, причем отображения* $K\colon \mathbb{R}^p \mapsto \mathbb{R}^{r\times q}$ *и* $b\colon \mathbb{R}^p \mapsto \mathbb{R}^r$ *являются локально липшицевыми в точке* $\bar{x} \in \mathbb{R}^p$*.*

Тогда для любого $\bar{y} \in \mathbb{R}^q$ *справедливы равенства*

$$\partial_x \Phi(\bar{x},\, \bar{y}) = \Pi_x \partial \Phi(\bar{x},\, \bar{y}) = \hat{\partial}_x \Phi(\bar{x},\, \bar{y}) = \tilde{\partial}_x \Phi(\bar{x},\, \bar{y}). \tag{24}$$

Доказательство. Из леммы 3 следует, что для любого $\bar{y} \in \mathbb{R}^q$

$$(\partial_B)_x \Phi(\bar{x},\, \bar{y}) \subset (\hat{\partial}_B)_x \Phi(\bar{x},\, \bar{y}).$$

Это влечет за собой включение

$$\partial_x \Phi(\bar{x},\, \bar{y}) \subset \hat{\partial}_x \Phi(\bar{x},\, \bar{y}). \tag{25}$$

Поскольку $K(x)$ и $b(x)$ локально липшицевы в точке $\bar{x}$, $\Phi(x,\, y)$ является локально липшицевым в точке $(\bar{x},\, \bar{y})$ для любого $\bar{y} \in \mathbb{R}^q$. Поэтому, в силу замечания 1 выполняются соотношения

$$\Pi_x \partial \Phi(\bar{x},\, \bar{y}) = \hat{\partial} \Phi(\bar{x},\, \bar{y}) \subset \tilde{\partial}_x \Phi(\bar{x},\, \bar{y}).$$

С учетом этих соотношений и (25), для завершения доказательства теоремы достаточно показать, что при любом $\bar{y} \in \mathbb{R}^q$

$$\tilde{\partial}_x \Phi(\bar{x},\, \bar{y}) \subset \partial_x \Phi(\bar{x},\, \bar{y}). \tag{26}$$

Пусть $J \in (\tilde{\partial}_B)_x \Phi(\bar{x},\, \bar{y})$. Тогда найдется последовательность $\{(x^k,\, y^k)\} \to (\bar{x},\, \bar{y})$, в каждой точке которой Φ дифференцируемо по x, причем

$$\left\{ \frac{\partial \Phi}{\partial x}(x^k,\, y^k) \right\} \to J, \tag{27}$$

и, кроме того, для любого k точка x^k принадлежит окрестности $\mathcal{U}$ точки $\bar{x}$, в которой отображения K и b являются липшицевыми. Заметим, что при всех k отображения $\Phi(\cdot,\, y^k)$ и $\Phi(\cdot,\, \bar{y})$ являются липшицевыми на $\mathcal{U}$.

Определим Γ как множество точек, в которых отображение $\Phi(\cdot,\, \bar{y})$ не дифференцируемо: $\Gamma = \mathbb{R}^p \setminus \mathcal{D}_{\Phi(\cdot,\, \bar{y})}$. По теореме Радемахера [9, теорема 3.1.6] мера Лебега множества $\Gamma \cap \mathcal{U}$ равна нулю. Теперь заметим, что

$$\frac{\partial \Phi}{\partial x}(x^k,\, y^k) \in \partial_x \Phi(x^k,\, y^k) \quad \forall\, k. \tag{28}$$

Как показано в [29], дифференциал Кларка не зависит от множеств нулевой меры Лебега. Отсюда и из (28) вытекает, что

$$\frac{\partial \Phi}{\partial x}(x^k,\, y^k) \in \partial_x^{\Gamma} \Phi(x^k,\, y^k)\ \forall\, k,$$

где

$$\partial_x^{\Gamma} \Phi(x^k,\, y^k) = \operatorname{conv}(\partial_B^{\Gamma})_x \Phi(x^k,\, y^k),$$

$$(\partial_B^{\Gamma})_x \Phi(x^k,\, y^k) = \Big\{ M \in \mathbb{R}^{r \times p} \mid \exists\ \{x_j^k\} \subset \mathcal{D}_{\Phi(\cdot,\, y^k)} \setminus \Gamma : \\ \{x_j^k\} \to x^k,\ \left\{ \frac{\partial \Phi}{\partial x}(x_j^k,\, y^k) \right\} \to M\ (j \to \infty) \Big\}. \tag{29}$$

Отсюда и из теоремы Каратеодори следует, что для всякого k найдутся

$$\alpha_i^k \geqslant 0, \quad J_i^k \in (\partial_B^{\Gamma})_x \Phi(x^k,\, y^k), \quad i = 1,\, \ldots,\, rp+1, \tag{30}$$

такие, что

$$\sum_{i=1}^{rp+1} \alpha_i^k = 1, \quad \frac{\partial \Phi}{\partial x}(x^k, y^k) = \sum_{i=1}^{rp+1} \alpha_i^k J_i^k. \tag{31}$$

Из (29), (30) вытекает, что для любого $i = 1,\, \ldots,\, rp+1$ и любого k существует последовательность $\{x_j^{k,\, i}\} \subset (\mathcal{D}_{\Phi(\cdot,\, y^k)} \cap \mathcal{U}) \setminus \Gamma$ такая что

$$\{x_j^{k,\, i}\} \to x^k, \quad \left\{ \frac{\partial \Phi}{\partial x}(x_j^{k,\, i},\, y^k) \right\} \to J_i^k \quad (j \to \infty),$$

и значит, можно выбрать номер $j(k)$ так, чтобы выполнялось

$$\|x_{j(k)}^{k,\, i} - x^k\| < \frac{1}{k}, \quad \left\| \frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,\, i},\, y^k) - J_i^k \right\| < \frac{1}{k}. \tag{32}$$

Далее, для любого k имеет место цепочка равенств

$$\sum_{i=1}^{rp+1} \alpha_i^k \frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, \bar{y}) = \sum_{i=1}^{rp+1} \alpha_i^k \frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, y^k) + \\ + \sum_{i=1}^{rp+1} \alpha_i^k \left(\frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, \bar{y}) - \frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, y^k) \right) = \\ = \sum_{i=1}^{rp+1} \alpha_i^k J_i^k + \sum_{i=1}^{rp+1} \alpha_i^k \left(\frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, y^k) - J_i^k \right) + \\ + \sum_{i=1}^{rp+1} \alpha_i^k \left(\frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, \bar{y}) - \frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, y^k) \right) = \\ = \frac{\partial \Phi}{\partial x}(x^k, y^k) + \sum_{i=1}^{rp+1} \alpha_i^k \left(\frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, y^k) - J_i^k \right) + \\ + \sum_{i=1}^{rp+1} \alpha_i^k \left(\frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, \bar{y}) - \frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, y^k) \right), \quad (33)$$

где последнее равенство следует из второго равенства в (31). С учетом первого равенства в (31) и второго неравенства в (32) имеем

$$\left\| \sum_{i=1}^{rp+1} \alpha_i^k \left(\frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, y^k) - J_i^k \right) \right\| \leqslant \sum_{i=1}^{rp+1} \alpha_i^k \left\| \frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, y^k) - J_i^k \right\| < \frac{1}{k} \sum_{i=1}^{rp+1} \alpha_i^k = \frac{1}{k},$$

и поэтому

$$\sum_{i=1}^{rp+1} \alpha_i^k \left(\frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, y^k) - J_i^k \right) \to 0. \quad (34)$$

Кроме того, с учетом леммы 4 выводим оценку

$$\left\| \sum_{i=1}^{rp+1} \alpha_i^k \left(\frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, \bar{y}) - \frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, y^k) \right) \right\| \leqslant \\ \leqslant \sum_{i=1}^{rp+1} \alpha_i^k \left\| \frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, \bar{y}) - \frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, y^k) \right\| < \\ < \ell_K \|y^k - \bar{y}\| \sum_{i=1}^{rp+1} \alpha_i^k = \ell_K \|y^k - \bar{y}\|,$$

где $\ell_K > 0$ — константа Липшица для K на $\mathcal{U}$. Отсюда следует, что

$$\sum_{i=1}^{rp+1} \alpha_i^k \left(\frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, \bar{y}) - \frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, y^k) \right) \to 0. \quad (35)$$

Из (27) и (33)–(35) вытекает предельное соотношение

$$\sum_{i=1}^{rp+1} \alpha_i^k \frac{\partial \Phi}{\partial x}(x_{j(k)}^{k,i}, \bar{y}) \to J. \quad (36)$$

Учитывая липшицевость $\Phi(\cdot,\,\bar{y})$ на $\mathcal{U}$, для каждого индекса $i=1,\,\ldots,\,rp+1$ последовательность $\{\frac{\partial\Phi}{\partial x}(x^{k,\,i}_{j(k)},\,\bar{y})\}$ ограничена. Значит, в силу очевидной ограниченности последовательностей $\{\alpha^k_i\}$, переходя при необходимости к подпоследовательностям, можем считать, что

$$\alpha^k_i\to\alpha_i,\quad \left\{\frac{\partial\Phi}{\partial x}(x^{k,\,i}_{j(k)},\,\bar{y})\right\}\to J_i\quad(k\to\infty)\tag{37}$$

при некоторых $\alpha_i\geqslant 0$ и $J_i\in\mathbb{R}^{r\times p}$, $i=1,\,\ldots,\,rp+1$, причем из первого соотношения в (31), из (36) и из (37) вытекают равенства

$$\sum_{i=1}^{rp+1}\alpha_i=1,\quad \sum_{i=1}^{rp+1}\alpha_iJ_i=J.\tag{38}$$

При этом, в силу второго предельного соотношения в (37), для всякого $i=1,\,\ldots,\,rp+1$ матрица J_i лежит в $(\partial_B)_x\Phi(\bar{x},\,\bar{y})$, так как по построению

$$x^{k,\,i}_{j(k)}\in\mathbb{R}^p\setminus\Gamma=\mathcal{D}_{\Phi(\cdot,\,\bar{y})}\quad\forall\,k,$$

причем в силу первого неравенства в (32) справедливо $\{x^{k,\,i}_{j(k)}\}\to\bar{x}$ $(k\to\infty)$. Но тогда из (38) получаем, что

$$J\in\operatorname{conv}(\partial_B)_x\Phi(\bar{x},\,\bar{y})=\partial_x\Phi(\bar{x},\,\bar{y}).$$

Это доказывает включение

$$(\tilde{\partial}_B)_x\Phi(\bar{x},\,\bar{y})\subset\partial_x\Phi(\bar{x},\,\bar{y}),$$

из которого, в силу выпуклости множества $\partial_x\Phi(\bar{x},\,\bar{y})$, вытекает требуемое включение (26). ■

В завершение пункта приведем еще одно утверждение, справедливое для отображений вида (21), которое понадобится нам в дальнейшем.

Лемма 5. *Пусть отображение* $\Phi\colon\mathbb{R}^p\times\mathbb{R}^q\mapsto\mathbb{R}^r$ *имеет вид* (21)*, причем отображения* $K\colon\mathbb{R}^p\to\mathbb{R}^{r\times q}$ *и* $b\colon\mathbb{R}^p\to\mathbb{R}^r$ *локально липшицевы в точке* $\bar{x}\in\mathbb{R}^p$*. Пусть последовательности* $\{(x^k,\,y^k_1)\}\subset\mathbb{R}^p\times\mathbb{R}^q$ *и* $\{(x^k,\,y^k_2)\}\subset\mathbb{R}^p\times\mathbb{R}^q$ *сходятся к* $(\bar{x},\,\bar{y})$*, где* $\bar{y}\in\mathbb{R}^q$*.*

Тогда для любой последовательности матриц $\{W^1_k\}\subset\mathbb{R}^{r\times p}$ *такой, что*

$$W^1_k\in\partial_x\Phi(x^k,\,y^k_1)$$

для всех k*, существует последовательность матриц* $\{W^2_k\}\subset\mathbb{R}^{r\times p}$*, такая что* $W^2_k\in\partial_x\Phi(x^k,\,y^k_2)$ *для всех достаточно больших* k*, и*

$$\|W^1_k-W^2_k\|=O(\|y^k_1-y^k_2\|).$$

Доказательство. Рассмотрим произвольную окрестность U точки $\bar{x}$, в которой отображения K и b удовлетворяют условию Липшица. Тогда для всех k отображение $\Phi(\cdot,\, y_2^k)$ является липшицевым на U, поэтому, по теореме Радемахера, оно дифференцируемо всюду в $U \setminus \Gamma$, где $\Gamma \subset U$ — некоторое множество, мера Лебега которого равна нулю. Пусть $\mathcal{D}_k \subset U$ обозначает множество точек дифференцируемости отображения $\Phi(\cdot,\, y_1^k)$. Поскольку дифференциал Кларка не зависит от множества нулевой меры Лебега [29], для всех достаточно больших k (таких, чтобы выполнялось $x^k \in U$) и любых матриц $W_k^1 \in \partial_x \Phi(x^k,\, y_1^k)$ существуют $s_k \in \mathbb{N}$, матрицы $W_{k,\,i}^1 \in \mathbb{R}^{r\times q}$ и числа $\alpha_{k,\,i} \geqslant 0$, $i = 1,\, \ldots,\, s_k$, такие что $\sum_{i=1}^{s_k} \alpha_{k,\,i} = 1$, $W_k^1 = \sum_{i=1}^{s_k} \alpha_{k,\,i} W_{k,\,i}^1$, и для всех $i = 1,\, \ldots,\, s_k$ найдется последовательность $\{x_j^{k,\,i}\} \subset \mathcal{D}_k \setminus \Gamma$, сходящаяся к x^k и такая, что $\{\frac{\partial \Phi}{\partial x}(x_j^{k,\,i},\, y_1^k)\} \to W_{k,\,i}^1$ при $j \to \infty$.

Далее, по лемме 4, при всех достаточно больших k, для всех j, начиная с некоторого номера, справедливо неравенство

$$\left\| \frac{\partial \Phi}{\partial x}(x_j^{k,\,i},\, y_1^k) - \frac{\partial \Phi}{\partial x}(x_j^{k,\,i},\, y_2^k) \right\| \leqslant \ell_K \|y_1^k - y_2^k\| \quad \forall\, i = 1,\, \ldots,\, s_k, \tag{39}$$

где ℓ_K — произвольная константа Липшица отображения K на U. Для всех больших k, поскольку $\Phi(\cdot,\, y_2^k)$ локально липшицево в точке x^k, для всех $i = 1,\, \ldots,\, s_k$ последовательность $\{\frac{\partial \Phi}{\partial x}(x_j^{k,\,i},\, y_2^k)\}$ ограничена, и, следовательно, при необходимости переходя к подпоследовательности, можно считать, что при $j \to \infty$, она сходится к некоторой матрице $W_{k,\,i}^2$. Тогда, переходя к пределу в (39), получаем оценку

$$\|W_{k,\,i}^1 - W_{k,\,i}^2\| \leqslant \ell_K \|y_1^k - y_2^k\| \quad \forall\, i = 1,\, \ldots,\, s_k. \tag{40}$$

По определению B-дифференциала, имеем $W_{k,\,i}^2 \in (\partial_x)_B \Phi(x^k,\, y_2^k)$. Но тогда, по определению дифференциала Кларка, выпуклая комбинация $W_k^2 = \sum_{i=1}^{s_k} \alpha_k^i W_{k,\,i}^2$ принадлежит $\partial_x \Phi(x^k,\, y_2^k)$, и при этом, в силу (40),

$$\|W_k^1 - W_k^2\| = \left\| \sum_{i=1}^{s_k} \alpha_{k,\,i} W_{k,\,i}^1 - \sum_{i=1}^{s_k} \alpha_{k,\,i} W_{k,\,i}^2 \right\| \leqslant \sum_{i=1}^{s_k} \alpha_{k,\,i} \|W_{k,\,i}^1 - W_{k,\,i}^2\| \leqslant \ell_K \|y_1^k - y_2^k\|.$$

■

1.3. Оценки расстояния до решений

При анализе численных методов оптимизации важную роль играют так называемые оценки расстояния до решения системы Каруша–Куна–Таккера (ККТ) задачи оптимизации. Однако, имеющиеся в литературе оценки такого рода (см. [26, 31, 56, 57, 63, 82], а также обзор [50]) устанавливаются либо в предположении регулярности ограничений, либо для систем ККТ, соответствующих задачам оптимизации, целевая функция и ограничения которых являются дважды непрерывно дифференцируемыми. В настоящем разделе будет доказана оценка расстояния до множества решений системы ККТ задачи с липшицевыми производными без требования каких бы то ни было условий регулярности ограничений. Эта оценка станет одним из инструментов получения тонких результатов о сходимости численных методов решения задач с липшицевыми производными в главе 3.

Рассмотрим задачу оптимизации

$$\begin{gathered} f(x) \to \min, \\ h(x) = 0,\ g(x) \leqslant 0, \end{gathered} \tag{41}$$

в которой целевая функция $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ и отображения $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ и $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ дифференцируемы. Система Каруша–Куна–Таккера задачи (41) имеет вид

$$\begin{gathered} \frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) = 0, \\ h(x) = 0, \quad \mu \geqslant 0, \quad g(x) \leqslant 0, \quad \langle \mu,\, g(x)\rangle = 0, \end{gathered} \tag{42}$$

где $L\colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ — функция Лагранжа задачи (41),

$$L(x,\, \lambda,\, \mu) = f(x) + \langle \lambda,\, h(x)\rangle + \langle \mu,\, g(x)\rangle. \tag{43}$$

Пусть $\mathcal{M}(\bar{x}) \subset \mathbb{R}^l \times \mathbb{R}^m$ — совокупность множителей Лагранжа, отвечающих точке $\bar{x} \in \mathbb{R}^n$ (т. е. множество пар $(\lambda,\, \mu) \in \mathbb{R}^l \times \mathbb{R}^m$ таких, что тройка $(\bar{x},\, \lambda,\, \mu)$ удовлетворяет системе (42)). Точка $\bar{x} \in \mathbb{R}^n$ называется стационарной точкой задачи (41), если $\mathcal{M}(\bar{x}) \neq \varnothing$.

В случае, когда f, h и g дважды непрерывно дифференцируемы в точке $\bar{x} \in \mathbb{R}^n$, из результатов [36, 38, 51, 52] следует эквивалентность следующих трех свойств.

Свойство 1 (верхняя липшицева устойчивость решений системы ККТ при канонических возмущениях). Существует окрестность $\mathcal{U}$ точки $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ и число $\ell > 0$ такие, что для любого $\sigma = (a,\, b,\, c) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, достаточно близкого к точке $(0,\, 0,\, 0)$, всякое решение $(x(\sigma),\, \lambda(\sigma),\, \mu(\sigma)) \in \mathcal{U}$ возмущенной системы ККТ

$$\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) = a, \quad h(x) = b, \quad \mu \geqslant 0, \quad g(x) \leqslant c, \quad \langle \mu,\, g(x) - c\rangle = 0 \tag{44}$$

удовлетворяет оценке

$$\|x(\sigma) - \bar{x}\| + \operatorname{dist}((\lambda(\sigma),\, \mu(\sigma)),\, \mathcal{M}(\bar{x})) \leqslant \ell \|\sigma\|. \tag{45}$$

Свойство 2 (оценка расстояния для системы ККТ). Существует окрестность $\mathcal{U}$ точки $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ и число $\ell > 0$ такие, что для любого вектора $(x,\, \lambda,\, \mu) \in \mathcal{U}$ выполняется неравенство

$$\|x - \bar{x}\| + \mathrm{dist}((\lambda,\, \mu),\, \mathcal{M}(\bar{x})) \leqslant \ell \left\| \begin{pmatrix} \dfrac{\partial L}{\partial x}(x,\, \lambda,\, \mu) \\ h(x) \\ \min\{\mu,\, -g(x)\} \end{pmatrix} \right\|. \tag{46}$$

Свойство 3 (некритичность множителя $(\bar{\lambda},\, \bar{\mu})$**).** Не существует тройки $(\xi,\, \eta,\, \zeta) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, с $\xi \neq 0$, которая бы удовлетворяла системе

$$\frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi + (h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \quad h'(\bar{x})\xi = 0, \quad g'_{A_+}(\bar{x})\xi = 0, \tag{47}$$

$$\zeta_{A_0} \geqslant 0, \quad g'_{A_0}(\bar{x})\xi \leqslant 0, \quad \zeta_i\langle g'_i(\bar{x}),\, \xi\rangle = 0,\ i \in A_0, \quad \zeta_{\{1,\, \ldots,\, m\}\setminus A} = 0, \tag{48}$$

где множества индексов A, A_+ и A_0 введены в (11).

Замечание 2. Эквивалентность свойств 1 и 2 не требует двукратной дифференцируемости функции f и отображений h и g. Тот факт, что из свойства 2 следует свойство 1, доказывается тривиальным образом. Обратная импликация была получена в [36, Теорема 2].

Однако свойство 3, являющееся наиболее конструктивным, для задач с липшицевыми производными вообще говоря не имеет смысла, поскольку в этом случае матрица Гессе $\frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ может не существовать. В связи с этим возникает вопрос о том, какие конструктивные условия, гарантирующие (без каких-либо дополнительных предположений) выполнение оценки расстояния (т. е. свойства 2), можно предъявить для задач с липшицевыми производными. Нахождение таких условий и составляет цель настоящего раздела.

Напомним, что, согласно [58, (6.6)], контингентной производной отображения $\Phi\colon \mathbb{R}^p \mapsto \mathbb{R}^r$ в точке u (в которой это отображение является локально липшицевым) называется точечно-множественное отображение $C\Phi(u)$ из $\mathbb{R}^p$ во множество подмножеств $\mathbb{R}^r$, заданное по правилу

$$\begin{aligned} C\Phi(u)(v) = \{w \in \mathbb{R}^r \quad | \quad & \exists\, \{t_k\} \subset \mathbb{R}_+,\, \{t_k\} \to 0+ : \\ & \{(\Phi(u + t_k v) - \Phi(u))/t_k\} \to w\}. \end{aligned} \tag{49}$$

В частности, если Φ дифференцируемо в точке u по направлению v, то $C\Phi(u)(v)$ состоит из единственного элемента — производной отображения Φ в точке u по направлению v. Заметим, что

$$\forall\, w \in C\Phi(u)(v) \quad \exists\, J \in \partial\Phi(u)\colon w = Jv, \tag{50}$$

(см. [58, (6.5), (6.6), (6.16)]).

Кроме того, частной контингентной производной отображения $\Phi : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^r$ в точке $(\bar{u},\, \bar{v}) \in \mathbb{R}^p \times \mathbb{R}^q$ по переменной u называется контингентная производная отображения $\Phi(\cdot,\, \bar{v})$ в точке $\bar{u}$. Будем обозначать эту производную через $C_u\Phi(\bar{u},\, \bar{v})$.

Наконец, будем говорить, что отображение $\Phi : \mathbb{R}^p \to \mathbb{R}^r$ удовлетворяет условию Липшица относительно точки $u \in \mathbb{R}^p$ в окрестности $\mathcal{U}$ этой точки с константой $\ell > 0$, если

$$\|\Phi(v) - \Phi(u)\| \leqslant \ell \|v - u\| \quad \forall\, v \in \mathcal{U}.$$

Следующее условие является обобщением свойства 3 на случай задач с липшицевыми производными

Определение 4. Множитель Лагранжа $(\bar{\lambda},\, \bar{\mu})$, отвечающий стационарной точке $\bar{x} \in \mathbb{R}^n$ задачи (41), называется *некритическим*, если не существует тройки $(\xi,\, \eta,\, \zeta) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ с $\xi \neq 0$, удовлетворяющей системе

$$d + (h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \quad h'(\bar{x})\xi = 0, \quad g'_{A_+}(\bar{x})\xi = 0, \tag{51}$$

$$\zeta_{A_0} \geqslant 0, \quad g'_{A_0}(\bar{x})\xi \leqslant 0, \quad \zeta_i \langle g'_i(\bar{x}),\, \xi\rangle = 0,\ i \in A_0, \quad \zeta_{\{1,\, \ldots,\, m\}\setminus A} = 0 \tag{52}$$

с некоторым $d \in C_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})(\xi)$. В противном случае множитель $(\bar{\lambda},\, \bar{\mu})$ называется *критическим*.

Замечание 3. Из (50) следует, что множитель $(\bar{\lambda},\, \bar{\mu})$ является некритическим, если не существует тройки $(\xi,\, \eta,\, \zeta) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ с $\xi \neq 0$, удовлетворяющей системе (51)–(52) при $d = H\xi$, где H — некоторая матрица из множества $\partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})$. Множитель, удовлетворяющий последнему условию, будет называться *строго некритическим*.

С учетом замечания 3 становится очевидным, что некритичность множителя $(\bar{\lambda},\, \bar{\mu})$ следует из так называемого достаточного условия второго порядка:

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \quad \langle H\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{53}$$

где $C(\bar{x})$ — критический конус задачи (41) в точке $\bar{x}$,

$$\begin{aligned} C(\bar{x}) = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_A(\bar{x})\xi \leqslant 0,\ \langle f'(\bar{x}),\, \xi\rangle \leqslant 0\} = \\ = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+}(\bar{x})\xi = 0,\ g'_{A_0}(\bar{x})\xi \leqslant 0\}. \end{aligned} \tag{54}$$

Отметим, что, как показано в [59], условие (53) в действительности является достаточным для локальной оптимальности точки $\bar{x}$ в задаче (41).

Рассмотрим следующую параметрическую версию системы ККТ (42):

$$\frac{\partial f}{\partial x}(\sigma,\, x) + \left(\frac{\partial h}{\partial x}(\sigma,\, x)\right)^{\mathrm{T}} \lambda + \left(\frac{\partial g}{\partial x}(\sigma,\, x)\right)^{\mathrm{T}} \mu = 0,$$
$$h(\sigma,\, x) = 0, \quad \mu \geqslant 0, \quad g(\sigma,\, x) \leqslant 0, \quad \langle \mu,\, g(\sigma,\, x)\rangle = 0. \tag{55}$$

В этой системе $\sigma \in \mathbb{R}^s$ является параметром, а функция $f\colon \mathbb{R}^s \times \mathbb{R}^n \mapsto \mathbb{R}$ и отображения $h\colon \mathbb{R}^s \times \mathbb{R}^n \mapsto \mathbb{R}^l$ и $g\colon \mathbb{R}^s \times \mathbb{R}^n \mapsto \mathbb{R}^m$ дифференцируемы по переменной x. Отметим, что система (55) является системой ККТ параметрической задачи

$$f(\sigma,\, x) \to \min$$
$$h(\sigma,\, x) = 0,\ g(\sigma,\, x) \leqslant 0 \tag{56}$$

относительно неизвестной x. Пусть $L\colon \mathbb{R}^s \times \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \mapsto \mathbb{R}$ — функция Лагранжа этой параметрической задачи:

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

В контексте оптимизационных задач, следующий результат о чувствительности обобщает [51, Теорема 2.3] (см. также связанный с этим результат в [38]).

Теорема 2. *Пусть функция $f\colon \mathbb{R}^s \times \mathbb{R}^n \mapsto \mathbb{R}$ и отображения $h\colon \mathbb{R}^s \times \mathbb{R}^n \mapsto \mathbb{R}^l$ и $g\colon \mathbb{R}^s \times \mathbb{R}^n \mapsto \mathbb{R}^m$ дифференцируемы по переменной x вблизи точки $(\bar\sigma,\, \bar x)$, и, кроме того, выполнены следующие предположения:*

а) *отображения $\frac{\partial f}{\partial x}(\bar\sigma,\, \cdot)$, $\frac{\partial h}{\partial x}(\bar\sigma,\, \cdot)$ и $\frac{\partial g}{\partial x}(\bar\sigma,\, \cdot)$ локально липшицевы в точке $\bar x$;*

б) *существует окрестность $\mathcal{U}$ точки $\bar\sigma$ и число $\ell > 0$ такие, что для любого фиксированного x, достаточно близкого к точке $\bar x$, отображения $\frac{\partial f}{\partial x}(\cdot,\, x)$, $g(\cdot,\, x)$, $h(\cdot,\, x)$, $\frac{\partial h}{\partial x}(\cdot,\, x)$ и $\frac{\partial g}{\partial x}(\cdot,\, x)$ удовлетворяют условию Липшица относительно точки $\bar\sigma$ на $\mathcal{U}$ с константой ℓ.*

Пусть $\bar x \in \mathbb{R}^n$ — стационарная точка задачи (56) при $\sigma = \bar\sigma$, а $(\bar\lambda,\, \bar\mu) \in \mathbb{R}^l \times \mathbb{R}^m$ — отвечающий ей множитель Лагранжа.

Тогда, если множитель $(\bar\lambda,\, \bar\mu)$ некритический, то для любого σ, достаточно близкого к $\bar\sigma$, и любого решения $(x(\sigma),\, \lambda(\sigma),\, \mu(\sigma))$ системы (55), достаточно близкого к $(\bar x,\, \bar\lambda,\, \bar\mu)$, выполняется оценка

$$\|x(\sigma) - \bar x\| + \mathrm{dist}((\lambda(\sigma),\, \mu(\sigma)),\, \mathcal{M}(\bar\sigma,\, \bar x)) = O(\|\sigma - \bar\sigma\|),$$

где $\mathcal{M}(\bar\sigma,\, \bar x)$ есть множество пар $(\lambda,\, \mu) \in \mathbb{R}^l \times \mathbb{R}^m$ таких, что тройка $(\bar x,\, \lambda,\, \mu)$ является решением системы (55) при $\sigma = \bar\sigma$.

Доказательство. Сначала докажем, что

$$\|x(\sigma)-\bar{x}\| = O(\|\sigma-\bar{\sigma}\|). \tag{57}$$

Предположим, что (57) не выполняется, т. е. найдутся последовательности $\{\sigma^k\} \subset \mathbb{R}^s$ и $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ такие, что $\{\sigma^k\} \to \bar{\sigma}$, $\{(x^k,\, \lambda^k,\, \mu^k)\} \to (\bar{x},\, \bar{\lambda},\, \bar{\mu})$, и для каждого k точка $(x^k,\, \lambda^k,\, \mu^k)$ является решением системы (55) при $\sigma = \sigma^k$, и при этом $\|x^k - \bar{x}\| > \gamma_k\|\sigma^k - \bar{\sigma}\|$, где $\{\gamma_k\}$ — некоторая последовательность положительных чисел, стремящаяся к бесконечности. Тогда выполняется оценка

$$\|\sigma^k-\bar{\sigma}\| = o(\|x^k-\bar{x}\|). \tag{58}$$

Определим множества индексов

$$A = A(\bar{\sigma},\, \bar{x}) = \{i = 1,\, \ldots,\, m \mid g_i(\bar{\sigma},\, \bar{x}) = 0\},$$
$$A_+ = A_+(\bar{\sigma},\, \bar{x},\, \bar{\mu}) = \{i \in A \mid \bar{\mu}_i > 0\},$$
$$A_0 = A_0(\bar{\sigma},\, \bar{x},\, \bar{\mu}) = A \setminus A_+.$$

Поскольку существует лишь конечное число способов разбить множество индексов A_0 на два непересекающихся подмножества, без ограничения общности можно считать, что для каждого k справедливы соотношения

$$\mu_i^k > 0 \quad \forall\, i \in I_1, \quad \mu_i^k = 0 \quad \forall\, i \in I_2, \tag{59}$$

где I_1 и I_2 — фиксированные множества индексов такие, что $I_1 \cup I_2 = A_0$, $I_1 \cap I_2 = \emptyset$. Далее, в предположениях теоремы отображение g является непрерывным в точке $(\bar{\sigma},\, \bar{x})$. Следовательно, в силу условия дополняющей нежесткости в (55) и того факта, что последовательность $\{(x^k,\, \lambda^k,\, \mu^k)\}$ сходится к $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, без потери общности можно полагать, что для любого k верно, что

$$\mu_i^k > 0 \quad \forall\, i \in A_+, \quad \mu_i^k = 0 \quad \forall\, i \in N. \tag{60}$$

Кроме того, при необходимости переходя к подпоследовательности, можно считать, что $\{(x^k - \bar{x})/\|x^k - \bar{x}\|\}$ сходится к некоторому вектору $\xi \in \mathbb{R}^n$, $\|\xi\| = 1$. Тогда, полагая $t_k = \|x^k - \bar{x}\|$ и вновь переходя к подпоследовательности, если требуется, в силу (49) и локальной липшицевости отображения $\frac{\partial L}{\partial x}(\bar{\sigma},\, \cdot,\, \bar{\lambda},\, \bar{\mu})$ в точке $\bar{x}$ (которая следует из предположения a)), получаем существование вектора $d \in C_x \frac{\partial L}{\partial x}(\bar\sigma,\, \bar x,\, \bar\lambda,\, \bar\mu)(\xi)$ такого, что

$$\begin{aligned}\frac{\partial L}{\partial x}(\bar\sigma,\, x^k,\, \bar\lambda,\, \bar\mu) = \frac{\partial L}{\partial x}(\bar\sigma,\, \bar x + t_k\xi,\, \bar\lambda,\, \bar\mu) - \frac{\partial L}{\partial x}(\bar\sigma,\, \bar x,\, \bar\lambda,\, \bar\mu) +\\ + \frac{\partial L}{\partial x}(\bar\sigma,\, x^k,\, \bar\lambda,\, \bar\mu) - \frac{\partial L}{\partial x}(\bar\sigma,\, \bar x + t_k\xi,\, \bar\lambda,\, \bar\mu) =\\ = t_k d + o(t_k) + O(\|x^k - \bar x - t_k\xi\|) =\\ = t_k d + o(t_k) + O\left(t_k \left\| \frac{x^k - \bar x}{\|x^k - \bar x\|} - \xi \right\|\right) =\\ = \|x^k - \bar x\| d + o(\|x^k - \bar x\|). \quad (61)\end{aligned}$$

Первое уравнение в (55) может быть записано в виде

$$\frac{\partial L}{\partial x}(\sigma,\, x,\, \lambda,\, \mu) = 0.$$

Тогда, используя (59)–(61) и предположения а) и б), а также (58), получаем, что

$$\begin{aligned}0 = \frac{\partial L}{\partial x}(\sigma^k,\, x^k,\, \lambda^k,\, \mu^k) =\\ = \frac{\partial L}{\partial x}(\sigma^k,\, x^k,\, \bar\lambda,\, \bar\mu) + \left(\frac{\partial h}{\partial x}(\sigma^k,\, x^k)\right)^{\mathrm T}(\lambda^k - \bar\lambda) + \left(\frac{\partial g}{\partial x}(\sigma^k,\, x^k)\right)^{\mathrm T}(\mu^k - \bar\mu) =\\ = \frac{\partial L}{\partial x}(\bar\sigma,\, x^k,\, \bar\lambda,\, \bar\mu) + \left(\frac{\partial h}{\partial x}(\bar\sigma,\, x^k)\right)^{\mathrm T}(\lambda^k - \bar\lambda) + \left(\frac{\partial g}{\partial x}(\bar\sigma,\, x^k)\right)^{\mathrm T}(\mu^k - \bar\mu) +\\ + O(\|\sigma^k - \bar\sigma\|) = \|x^k - \bar x\| d + \left(\frac{\partial h}{\partial x}(\bar\sigma,\, \bar x)\right)^{\mathrm T}(\lambda^k - \bar\lambda) +\\ + \left(\frac{\partial g_{A_+\cup I_1}}{\partial x}(\bar\sigma,\, \bar x)\right)^{\mathrm T}(\mu^k_{A_+\cup I_1} - \bar\mu_{A_+\cup I_1}) + o(\|x^k - \bar x\|) =\\ = \|x^k - \bar x\| d + \left(\frac{\partial h}{\partial x}(\bar\sigma,\, \bar x)\right)^{\mathrm T}(\lambda^k - \bar\lambda) +\\ + \left(\frac{\partial g_{A_+}}{\partial x}(\bar\sigma,\, \bar x)\right)^{\mathrm T}(\mu^k_{A_+} - \bar\mu_{A_+}) + \left(\frac{\partial g_{I_1}}{\partial x}(\bar\sigma,\, \bar x)\right)^{\mathrm T}\mu^k_{I_1} + o(\|x^k - \bar x\|).\end{aligned}$$

Следовательно,

$$\begin{aligned}-\operatorname{im}\left(\frac{\partial h}{\partial x}(\bar\sigma,\, \bar x)\right)^{\mathrm T} - \operatorname{im}\left(\frac{\partial g_{A_+}}{\partial x}(\bar\sigma,\, \bar x)\right)^{\mathrm T} - \left(\frac{\partial g_{I_1}}{\partial x}(\bar\sigma,\, \bar x)\right)^{\mathrm T}\mathbb{R}^{|I_1|}_+ \ni\\ \ni \|x^k - \bar x\| d + o(\|x^k - \bar x\|), \quad (62)\end{aligned}$$

где множество в левой части (будучи суммой линейного попространства и полиэдрального конуса) является замкнутым конусом.

Воспользуясь второй строкой в (55), и предположением б), а также (58), заключаем, что

$$\begin{aligned}0 = h(\sigma^k,\, x^k) = \frac{\partial h}{\partial x}(\bar\sigma,\, \bar x)(x^k - \bar x) + O(\|\sigma^k - \bar\sigma\|) + o(\|x^k - \bar x\|) =\\ = \frac{\partial h}{\partial x}(\bar\sigma,\, \bar x)(x^k - \bar x) + o(\|x^k - \bar x\|). \quad (63)\end{aligned}$$

Аналогично, применяя (59) и (60), выводим оценки

$$0 = g_{A_+\cup I_1}(\sigma^k,\, x^k) = \frac{\partial g_{A_+\cup I_1}}{\partial x}(\bar{\sigma},\, \bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \tag{64}$$

$$0 \geqslant g_{I_2}(\sigma^k,\, x^k) = \frac{\partial g_{I_2}}{\partial x}(\bar{\sigma},\, \bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|). \tag{65}$$

Разделив (62)–(65) на $\|x^k - \bar{x}\|$ и перейдя к пределу при $k \to \infty$, получаем, что

$$d \in -\operatorname{im}\left(\frac{\partial h}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} - \operatorname{im}\left(\frac{\partial g_{A_+}}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} - \left(\frac{\partial g_{I_1}}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} \mathbb{R}_+^{|I_1|}, \tag{66}$$

$$\frac{\partial h}{\partial x}(\bar{\sigma},\, \bar{x})\xi = 0, \quad \frac{\partial g_{A_+}}{\partial x}(\bar{\sigma},\, \bar{x})\xi = 0, \tag{67}$$

$$\frac{\partial g_{I_1}}{\partial x}(\bar{\sigma},\, \bar{x})\xi = 0, \quad \frac{\partial g_{I_2}}{\partial x}(\bar{\sigma},\, \bar{x})\xi \leqslant 0. \tag{68}$$

Включение (66) означает, что существует вектор $(\eta,\, \zeta) \in \mathbb{R}^l \times \mathbb{R}^m$, удовлетворяющий

$$d + \left(\frac{\partial h}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} \eta + \left(\frac{\partial g}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} \zeta = 0 \tag{69}$$

и такой, что

$$\zeta_{I_1} \geqslant 0, \quad \zeta_{I_2\cup(\{1,\,\ldots,\,m\}\setminus A)} = 0.$$

Тогда, с учетом (68), тройка $(\xi,\, \eta,\, \zeta)$ удовлетворяет системе

$$\begin{gathered} \zeta_{A_0} \geqslant 0, \quad \frac{\partial g_{A_0}}{\partial x}(\bar{\sigma},\, \bar{x})\xi \leqslant 0, \quad \zeta_i\left\langle \frac{\partial g_i}{\partial x}(\bar{\sigma},\, \bar{x}),\, \xi\right\rangle = 0,\ i \in A_0, \\ \zeta_{\{1,\,\ldots,\,m\}\setminus A} = 0. \end{gathered} \tag{70}$$

Поскольку $\xi \neq 0$, соотношения (67), (69), (70) противоречат некритичности множителя $(\bar{\lambda},\, \bar{\mu})$. Это доказывает (57).

Для завершения доказательства, заметим, что для всякого σ, достаточно близкого к $\bar{\sigma}$, и всякого решения $(x(\sigma),\, \lambda(\sigma),\, \mu(\sigma))$ системы (55), достаточно близкого к $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, имеем

$$\frac{\partial L}{\partial x}(\sigma,\, x(\sigma),\, \lambda(\sigma),\, \mu(\sigma)) = 0, \quad \mu(\sigma) \geqslant 0, \quad \mu_{\{1,\,\ldots,\,m\}\setminus A}(\sigma) = 0,$$

где последнее равенство справедливо в силу непрерывности отображения g в точке $(\bar{\sigma},\, \bar{x})$. Так как $\mathcal{M}(\bar{\sigma},\, \bar{x})$ есть множество решения линейной системы

$$\frac{\partial L}{\partial x}(\bar{\sigma},\, \bar{x},\, \lambda,\, \mu) = 0, \quad \mu_A \geqslant 0, \quad \mu_{\{1,\,\ldots,\,m\}\setminus A} = 0,$$

по лемме Хоффмана (см., например, [32, лемма 3.2.3]) получаем, что

$$\begin{aligned} \operatorname{dist}((\lambda(\sigma),\, \mu(\sigma)),\, \mathcal{M}(\bar{\sigma},\, \bar{x})) = O\Bigg(\left\|\frac{\partial L}{\partial x}(\bar{\sigma},\, \bar{x},\, \lambda(\sigma),\, \mu(\sigma))\right\| + \\ + \sum_{i\in A}\min\{0,\, \mu_i(\sigma)\} + \|\mu_N(\sigma)\|\Bigg) = O\Bigg(\bigg\| \frac{\partial L}{\partial x}(\sigma,\, x(\sigma),\, \lambda(\sigma),\, \mu(\sigma)) - \\ - \frac{\partial L}{\partial x}(\bar{\sigma},\, \bar{x},\, \lambda(\sigma),\, \mu(\sigma))\bigg\| \Bigg) = O(\|\sigma - \bar{\sigma}\|) + O(\|x(\sigma) - \bar{x}\|), \end{aligned}$$

где были также использованы предположения а) и б). Вместе с (57), это дает утверждение теоремы. ■

Замечание 4. Применительно к каноническим возмущениям, теорема 2 демонстрирует, что некритичность множителя достаточна для выполнения свойства 1. На самом деле, оказывается, что она является также и необходимой для этого. Чтобы убедиться в этом, возьмем произвольную тройку $(\xi,\, \eta,\, \zeta) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, удовлетворяющую системе (51)–(52) с некоторым $d \in C_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})(\xi)$, и зафиксируем последовательность $\{t_k\}$, соответствующую этому d в определении контингентной производной (49). Тогда, как легко проверить, для всех достаточно больших k тройка $(\bar{x} + t_k\xi,\, \bar{\mu} + t_k\eta,\, \bar{\lambda} + t_k\zeta)$ удовлетворяет системе (44) с некоторыми $a = o(t_k)$, $b = o(t_k)$ и $c = o(t_k)$. Следовательно, если $\xi \neq 0$, мы приходим к противоречию с (45).

Замечание 5. В отличие от случая, когда целевая функция и ограничения задачи оптимизации являются дважды непрерывно дифференцируемыми, строгая некритичность множителя, определенная в замечании 3, является строго более сильной, чем некритичность, введенная в определении 4, а значит, строго более сильной, чем свойство 1.

К примеру, положим $n = m = 1$, $l = 0$, $f(x) = \frac{1}{2}(\max\{0,\, x\})^2$, $g(x) = -x$. Тогда пара $(\bar{x},\, \bar{\mu}) = (0,\, 0)$ является единственным решением системы (42). Непосредственно проверяется, что множитель $\bar{\mu}$ является некритическим, и что свойство 1 для указанной пары $(\bar{x},\, \bar{\mu})$ выполняется. В самом деле, система (44) в данном примере имеет вид

$$\max\{0,\, x\} - \mu = a, \quad \mu \geqslant 0, \quad -x \leqslant c, \quad \mu(x + c) = 0.$$

Отсюда следует, что либо $\mu = 0$, и в этом случае x удовлетворяет по крайней мере одному из соотношений $x = a$ и $-c \leqslant x < 0$, либо $\mu > 0$, и тогда $x = -c$ и μ равно либо $-a$, либо $-a - c$. Поэтому $|x| + |\mu| \leqslant |a| + 2|c|$, что дает (45) с $\mathcal{U} = \mathbb{R} \times \mathbb{R}$ и некоторой константой $\ell > 0$ (зависящей от выбора нормы в правой части (45)). Однако, множество $(\partial_B)_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\mu}) = \{0,\, 1\}$ содержит ноль, и любой вектор $\xi \geqslant 0$ удовлетворяет системе (51)–(52) с $d = 0$ и $\eta = 0$.

С учетом замечаний 4 и 2 получаем следующий результат.

Следствие 1. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображения* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *и* $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ *дифференцируемы вблизи точки* $\bar{x} \in \mathbb{R}^n$*, причем их производные являются локально липшицевыми в этой точке. Пусть при этом* $\bar{x}$ *является стационарной точкой задачи* (41)*, а пара* $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ *— отвечающим ей множителем Лагранжа.*

Тогда следующие три свойства эквивалентны:

1. *Свойство 1 (верхняя липшицева устойчивость решений системы ККТ при канонических возмущениях).*

2. *Свойство 2 (оценка расстояния для системы ККТ).*

3. *Некритичность множителя* $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ *(в смысле определения 4).*

Глава 2

Итерационные схемы для решения обобщенных уравнений

Настоящая глава посвящена разработке и анализу итерационных схем решения *обобщенных уравнений*, т. е. включений вида

$$\Phi(u) + N(u) \ni 0, \tag{1}$$

где $\Phi\colon \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$ — однозначное отображение (именуемое *базовым* или *базой*), а $N\colon \mathbb{R}^\nu \mapsto 2^{\mathbb{R}^\nu}$ — многозначное (иногда называемое *полем*). Получаемые здесь теоремы о сходимости этих итерационных схем станут основным средством анализа численных методов решения задач оптимизации с липшицевыми производными в главе 3.

Глава состоит из двух разделов. Итерационные схемы, рассматриваемые в первом разделе, предназначены для решения обобщенных уравнений с произвольной непрерывной базой и носят абстрактный характер. Во втором разделе речь пойдет о более специальной схеме для обобщенных уравнений, база которых удовлетворяет некоторым дополнительным условиям.

2.1. Абстрактные ньютоновские схемы

Пусть Π есть множество произвольной природы. Рассмотрим следующий итерационный процесс. По текущей точке $u^k \in \mathbb{R}^\nu$ и текущему значению параметра $\pi^k \in \Pi$ будем генерировать очередную точку u^{k+1} как решение подзадачи

$$\mathcal{A}(\pi^k,\, u^k,\, u) + N(u) \ni 0, \tag{2}$$

где для любых $\pi \in \Pi$ и $\tilde{u} \in \mathbb{R}^\nu$ точечно-множественное отображение $\mathcal{A}(\pi,\, \tilde{u},\, \cdot)$ из $\mathbb{R}^\nu$ в $2^{\mathbb{R}^\nu}$ представляет собой аппроксимацию отображения Φ вблизи $\tilde{u}$. Свойства, которым должна удовлетворять эта аппроксимация, будут сформулированы ниже.

В следующих трех пунктах для представленного процесса будет установлена локальная сходимость к решению обобщенного уравнения (1) при различных наборах предположений.

2.1.1. Сходимость к сильно метрически регулярным решениям

В настоящем разделе рассматривается вариант итерационной схемы из [27, раздел 6C], которая берет свое начало в [83].

Для каждого значения параметра $\pi \in \Pi$ и каждой точки $\tilde{u} \in \mathbb{R}^\nu$ определим множество

$$U(\pi,\, \tilde{u}) = \{u \in \mathbb{R}^\nu \mid \mathcal{A}(\pi,\, \tilde{u},\, u) + N(u) \ni 0\}, \tag{3}$$

то есть множество решения подзадачи (2) при $\pi^k = \pi$ и $u^k = \tilde{u}$. Как бы близко точка $\tilde{u}$ ни лежала к искомому решению обобщенного уравнения (1), множество (3) может содержать далекие от этого решения точки. В связи с этим, для того, чтобы процесс (2) обладал локальной сходимостью, его итерацию приходится снабжать условием отбора решений подзадачи. Условие такого рода называется *условием локализации*, и, как правило, состоит в ограничении расстояния между двумя последовательными итерациями. Простейший вариант условия локализации имеет вид

$$\|u^{k+1} - u^k\| \leqslant \delta, \tag{4}$$

где $\delta > 0$ — фиксированное достаточно малое число.

Дополняя итерацию процесса (2) условием локализации (4), приходим к следующей схеме

$$u^{k+1} \in U(\pi^k,\, u^k) \cap B(u^k,\, \delta), \quad k = 0,\, 1,\, \ldots, \tag{5}$$

где $\{\pi^k\} \subset \Pi$ — некоторая фиксированная последовательность значений параметра.

В сформулированном ниже результате о сходимости предполагается, что отображение $\mathcal{A}$ является однозначным и аппроксимирует отображение Φ в достаточно сильном смысле: разность $\Phi(\cdot) - \mathcal{A}(\pi,\, \tilde{u},\, \cdot)$ должна быть локально липшицевой в искомом решении с малой константой липшица равномерно по параметру π и точкам $\tilde{u} \in \mathbb{R}^\nu$, близким к этому решению.

Теорема 1. *Пусть отображение* $\Phi\colon \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$ *непрерывно в окрестности точки* $\bar{u} \in \mathbb{R}^\nu$, *и пусть* $N(\cdot)$ *— точечно-множественное отображение из* $\mathbb{R}^\nu$ *в* $2^{\mathbb{R}^\nu}$. *Пусть точка* $\bar{u} \in \mathbb{R}^\nu$ *является решением обобщенного уравнения* (1). *Пусть также заданы множество* Π *и отображение* $\mathcal{A}\colon \Pi \times \mathbb{R}^\nu \times \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$.

Предположим выполнение следующих свойств:

1) (*сильная метрическая регулярность решения*) *Существует константа* $\ell > 0$ *такая, что для всякого* $r \in \mathbb{R}^\nu$, *достаточно близкого к* 0, *возмущенное обобщенное уравнение*

$$\Phi(u) + N(u) \ni r \tag{6}$$

имеет вблизи точки $\bar{u}$ *единственное решение* $u(r)$, *причем отображение* $u(\cdot)$ *является локально липшицевым в точке* 0 *с константой* ℓ.

2) (*качество аппроксимации*) *Существует число* $\bar{\varepsilon} > 0$ *такое, что*

   а) $\mathcal{A}(\pi, \tilde{u}, \tilde{u}) = \Phi(\tilde{u})$ *для всех* $\pi \in \Pi$ *и* $\tilde{u} \in B(\bar{u}, \bar{\varepsilon})$;

   б) *существует функция* $\omega \colon \Pi \times \mathbb{R}^\nu_+ \times \mathbb{R}^\nu \times \mathbb{R}^\nu \mapsto \mathbb{R}$, *удовлетворяющая условию*

$$q := \ell \sup \left\{ \omega(\pi, \tilde{u}, u^1, u^2) \mid \pi \in \Pi,\ \tilde{u}, u^1, u^2 \in B(\bar{u}, \bar{\varepsilon}) \right\} < \frac{1}{2},$$

   и такая, что оценка

$$\|(\Phi(u^1) - \mathcal{A}(\pi, \tilde{u}, u^1)) - (\Phi(u^2) - \mathcal{A}(\pi, \tilde{u}, u^2))\| \leqslant \omega(\pi, \tilde{u}, u^1, u^2)\|u^1 - u^2\| \tag{7}$$

   выполняется для всех $\pi \in \Pi$ *и* $\tilde{u}, u^1, u^2 \in B(\bar{u}, \bar{\varepsilon})$.

Тогда найдутся числа $\delta > 0$ *и* $\varepsilon_0 > 0$ *такие, что для любого начального приближения* $u^0 \in B(\bar{u}, \varepsilon_0)$ *и любой последовательности* $\{\pi^k\} \subset \Pi$ *существует единственная последовательность* $\{u^k\} \subset \mathbb{R}^\nu$, *генерируемая схемой* (5); *эта последовательность сходится к решению* $\bar{u}$, *и для всех* k *выполняются оценки*

$$\|u^{k+1} - \bar{u}\| \leqslant \frac{\ell\omega(\pi^k, u^k, u^k, u^{k+1})}{1 - \ell\omega(\pi^k, u^k, u^k, u^{k+1})}\|u^k - \bar{u}\| \leqslant \frac{q}{1-q}\|u^k - \bar{u}\|.$$

В частности, скорость сходимости последовательности $\{u^k\}$ *линейная. Более того, скорость сходимости является сверхлинейной, если* $\omega(\pi^k, u^k, u^k, u^{k+1}) \to 0$ *при* $k \to \infty$, *и квадратичной, если*

$$\omega(\pi^k, u^k, u^k, u^{k+1}) = O(\|u^k - \bar{u}\|).$$

Приведенное утверждение следует из доказываемой ниже теоремы 2 и из теоремы 1.4 в [28], которая по сути говорит об устойчивости свойства сильной метрической регулярности при однозначных возмущениях с малой константой Липшица.

Теорема 1 показывает, что сверхлинейная скорость сходимости достигается, если точность аппроксимации отображения Φ возрастает с ростом k. Общность ее постановки позволяет увидеть, что существует два источника увеличения точности аппроксимации: $\omega(\pi^k, u^k, u^k, u^{k+1})$ может уменьшаться либо естественным образом по мере приближения u^k и u^{k+1} к

$\bar{u}$, либо искусственным, за счет управления параметром π. К примеру, если отображение Φ дифференцируемо вблизи $\bar{u}$, а его производная непрерывна в точке $\bar{u}$, то Φ можно аппроксимировать его линеаризацией $\mathcal{A}(\tilde{u},\, u) = \Phi(\tilde{u}) + \Phi'(\tilde{u})(u - \tilde{u})$, без каких-либо параметров. В этом случае, по теореме о среднем, предположение 2) теоремы 1 выполняется с функцией $\omega(\tilde{u},\, u^1,\, u^2) = \sup_{t\in[0,\,1]} \|\Phi'(tu^1 + (1-t)u^2) - \Phi'(\tilde{u})\|$, которая стремится к нулю при $\tilde{u}$, u^1 и u^2, стремящихся к $\bar{u}$. При таком выборе аппроксимации $\mathcal{A}$ итерационная схема (5) соответствует так называемому методу Джозефи–Ньютона, а теорема 1 включает в себя результат о сходимости этого метода, полученный в [53, 54].

Как будет показано ниже, метод множителей с линеаризованными ограничениями набирает требуемое качество аппроксимации естественным путем, в то время как в методе модифицированных функций Лагранжа сверхлинейная скорость сходимости достигается за счет стремления параметра штрафа к бесконечности.

Как показано в [42], если Φ является локально липшицевым в решении $\bar{u}$ обобщенного уравнения (1), сильная метрическая регулярность этого решения следует из так называемой CD-регулярности этого решения (см. с. 58).

В следующем пункте предположения 1) и 2) теоремы 1 будут ослаблены, но «не бесплатно»: разрешимость подзадач итерационной схемы придется требовать явно. Кроме того, нельзя будет гарантировать единственность траектории, которую генерирует схема. Условие регулярности решения будет далее ослаблено в пункте 2.1.3. В частности, искомое решение там не будет предполагаться изолированным. Однако это потребует использования более сильного условия локализации, чем (4).

2.1.2. Сходимость к полуустойчивым решениям

Следующий результат представляет собой развитие анализа из [17].

Теорема 2. *Пусть отображение* $\Phi\colon \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$ *является непрерывным в окрестности точки* $\bar{u} \in \mathbb{R}^\nu$*, и пусть* $N(\cdot)$ *— точечно-множественное отображение из* $\mathbb{R}^\nu$ *в* $2^{\mathbb{R}^\nu}$*. Пусть* $\bar{u} \in \mathbb{R}^\nu$ *является решением обобщенного уравнения* (1)*, и пусть задано точечно-множественное отображение* $\mathcal{A}$ *из* $\Pi \times \mathbb{R}^\nu \times \mathbb{R}^\nu$ *в* $2^{\mathbb{R}^\nu}$*, где* Π *— заданное множество.*

Предположим выполнение следующих свойств:

1) (*полуустойчивость решения*) *Существует число* $\ell > 0$ *такое, что для любого* $r \in \mathbb{R}^\nu$ *всякое решение* $u(r)$ *возмущенного обобщенного уравнения* (6)*, достаточно близкое к* $\bar{u}$*, удовлетворяет оценке*
$$\|u(r) - \bar{u}\| \leqslant \ell \|r\|.$$

2) (*качество аппроксимации*) *Существуют число* $\bar{\varepsilon} > 0$ *и функция* $\omega\colon \Pi \times \mathbb{R}^\nu \times \mathbb{R}^\nu \mapsto \mathbb{R}_+$ *такие, что*

$$q := \ell \sup \{\omega(\pi,\, \tilde{u},\, u) \mid \pi \in \Pi,\ \tilde{u},\, u \in B(\bar{u},\, \bar{\varepsilon})\} < \frac{1}{3}, \tag{8}$$

и оценка

$$\sup\{\|w\| \mid w \in \Phi(u) - \mathcal{A}(\pi,\, \tilde{u},\, u)\} \leqslant \omega(\pi,\, \tilde{u},\, u)(\|u - \tilde{u}\| + \|\tilde{u} - \bar{u}\|) \tag{9}$$

справедлива для всех $\pi \in \Pi$ *и* $\tilde{u},\, u \in B(\bar{u},\, \bar{\varepsilon})$.

3) (*разрешимость подзадач*) *Для любого* $\varepsilon > 0$ *существует* $\tilde{\varepsilon} > 0$ *такое, что*

$$U(\pi,\, \tilde{u}) \cap B(\bar{u},\, \varepsilon) \neq \emptyset \quad \forall\, \pi \in \Pi,\ \forall\, \tilde{u} \in B(\bar{u},\, \tilde{\varepsilon}). \tag{10}$$

Тогда найдутся числа $\delta > 0$ *и* $\varepsilon_0 > 0$ *такие, что для любого начального приближения* $u^0 \in B(\bar{u},\, \varepsilon_0)$ *и любой последовательности значений параметра* $\{\pi^k\} \subset \Pi$ *итерационная схема* (5) (*при произвольном выборе точки* u^k*, удовлетворяющей* (5)*, на каждой итерации*) *генерирует траекторию* $\{u^k\} \subset \mathbb{R}^\nu$; *любая такая траектория сходится к* $\bar{u}$*, и для всех* k *выполнено*:

$$\|u^{k+1} - \bar{u}\| \leqslant \frac{2\ell\omega(\pi^k,\, u^k,\, u^{k+1})}{1 - \ell\omega(\pi^k,\, u^k,\, u^{k+1})}\|u^k - \bar{u}\| \leqslant \frac{2q}{1-q}\|u^k - \bar{u}\|. \tag{11}$$

В частности, скорость сходимости последовательности $\{u^k\}$ *линейная. Кроме того, скорость сходимости является сверхлинейной, если* $\omega(\pi^k,\, u^k,\, u^{k+1}) \to 0$ *при* $k \to \infty$*, и квадратичной, если*

$$\omega(\pi^k,\, u^k,\, u^{k+1}) = O(\|u^k - \bar{u}\|).$$

Если же в дополнение к условиям теоремы для всякого достаточно малого $\varepsilon > 0$ *существует* $\tilde{\varepsilon} > 0$ *такое, что множество в* (10) *содержит ровно один элемент, то числа* δ *и* ε_0 *можно выбрать так, чтобы для любых* $u^0 \in B(\bar{u},\, \varepsilon_0)$ *и* $\{\pi^k\} \subset \Pi$ *последовательность* $\{u^k\}$*, удовлетворяющая* (5)*, была единственной.*

Доказательство. Согласно (3), для всякого $\pi \in \Pi$ и всякого $\tilde{u} \in \mathbb{R}^\nu$ любая точка $u \in U(\pi,\, \tilde{u})$ удовлетворяет обобщенному уравнению (6) с некоторым

$$r \in \Phi(u) - \mathcal{A}(\pi,\, \tilde{u},\, u), \tag{12}$$

и, кроме того, в силу предположения 2), существует число $\bar{\varepsilon} > 0$ такое, что

$$\|r\| \leqslant \omega(\pi,\, \tilde{u},\, u)(\|u - \tilde{u}\| + \|\tilde{u} - \bar{u}\|),$$

если $\tilde{u},\ u \in B(\bar{u},\ \bar{\varepsilon})$. Кроме того, по предположению 1), уменьшая $\bar{\varepsilon} > 0$ при необходимости, можно гарантировать, что для всех $\pi \in \Pi$, $\tilde{u} \in B(\bar{u},\ \bar{\varepsilon})$ и $u \in B(\bar{u},\ \bar{\varepsilon}) \cap U(\pi,\ \tilde{u})$ выполняются неравенства

$$\|u - \bar{u}\| \leqslant \ell\|r\| \leqslant \ell\omega(\pi,\ \tilde{u},\ u)(\|u - \bar{u}\| + 2\|\tilde{u} - \bar{u}\|).$$

С учетом (8), из последней формулы следует оценка

$$\|u - \bar{u}\| \leqslant \frac{2\ell\omega(\pi,\ \tilde{u},\ u)}{1 - \ell\omega(\pi,\ \tilde{u},\ u)}\|\tilde{u} - \bar{u}\|. \tag{13}$$

Согласно предположению 3), существуют чила $\tilde{\varepsilon},\ \varepsilon \in (0,\ \bar{\varepsilon}/3]$ такие, что выполнено (10). Положим $\delta = \varepsilon + \tilde{\varepsilon}$. Тогда для всех $\tilde{u} \in B(\bar{u},\ \tilde{\varepsilon})$ и $u \in B(\bar{u},\ \varepsilon)$ имеем

$$\|u - \tilde{u}\| \leqslant \|u - \bar{u}\| + \|\tilde{u} - \bar{u}\| \leqslant \varepsilon + \tilde{\varepsilon} = \delta,$$

а значит, (10) влечет за собой соотношение

$$U(\pi,\ \tilde{u}) \cap B(\tilde{u},\ \delta) \neq \emptyset \quad \forall\, \pi \in \Pi,\ \forall\, \tilde{u} \in B(\bar{u},\ \tilde{\varepsilon}). \tag{14}$$

Более того, для любых $\tilde{u} \in B(\bar{u},\ \tilde{\varepsilon})$ и $u \in B(\tilde{u},\ \delta)$ выполнено

$$\|u - \bar{u}\| \leqslant \|u - \tilde{u}\| + \|\tilde{u} - \bar{u}\| \leqslant \delta + \tilde{\varepsilon} = \varepsilon + 2\tilde{\varepsilon} \leqslant \bar{\varepsilon},$$

а значит, в силу (13), оценки

$$\|u - \bar{u}\| \leqslant \frac{2\ell\omega(\pi,\ \tilde{u},\ u)}{1 - \ell\omega(\pi,\ \tilde{u},\ u)}\|\tilde{u} - \bar{u}\| \leqslant \frac{2q}{1 - q}\|\tilde{u} - \bar{u}\| \tag{15}$$

выполняются для любого $\pi \in \Pi$, любого $\tilde{u} \in B(\bar{u},\ \tilde{\varepsilon})$ и любого $u \in U(\pi,\ \tilde{u}) \cap B(\tilde{u},\ \delta)$, где, согласно (8), $2q/(1 - q) < 1$.

Соотношения (14) и (15) очевидно влекут за собой требуемое с $\varepsilon_0 = \tilde{\varepsilon}$. В частности, если ε и $\tilde{\varepsilon}$ выбрать таким образом, чтобы $\tilde{\varepsilon} \leqslant \varepsilon$ и множество $U(\pi,\ \tilde{u}) \cap B(\bar{u},\ \varepsilon)$ содержало ровно один элемент для любого $\pi \in \Pi$ и любого $\tilde{u} \in B(\bar{u},\ \tilde{\varepsilon})$, можно дополнительно гарантировать, что множество в (14) содержит ровно один элемент для всех таких π и $\tilde{u}$. ∎

Если убрать слагаемое $\|\tilde{u} - \bar{u}\|$ из правой части неравенства (9), в условиях доказанной теоремы можно заменить ограничение $q < 1/3$ на $q < 1/2$ и убрать множитель 2 из правой части формулы (11). С учетом этого, теорема 1 может быть получена как следствие теоремы 2.

Из доказательства теоремы 2 видно, что предположение 2) может быть несколько ослаблено: (8) можно заменить на

$$q := \ell \sup\{\omega(\pi,\ \tilde{u},\ u) \mid \pi \in \Pi,\ u \in U(\pi,\ \tilde{u}),\ \tilde{u},\ u \in B(\bar{u},\ \bar{\varepsilon})\} < \frac{1}{3},$$

а вместо (9) достаточно предположить, что оценка

$$\|w\| \leqslant \omega(\pi,\, \tilde{u},\, u)(\|u-\tilde{u}\| + \|\tilde{u}-\bar{u}\|)$$

верна для всех $\pi \in \Pi$, любого $w \in \Phi(u) + (-\mathcal{A}(\pi,\, \tilde{u},\, u)) \cap N(u)$ и любых $\tilde{u},\, u \in B(\bar{u},\, \bar{\varepsilon})$. С учетом этого, теорема 2 покрывает результат из [51, теорема 2.1] о сходимости возмущенного метода Джозефи–Ньютона, который соответствует выбору $\mathcal{A}(\tilde{u},\, u) = \Phi(\tilde{u}) + \Phi'(\tilde{u})(u-\tilde{u}) + \Omega(\tilde{u},\, u-\tilde{u})$, где точечно-множественное отображение Ω из $\mathbb{R}^\nu \times \mathbb{R}^\nu$ во множество подмножеств $\mathbb{R}^\nu$ описывает неточность подзадачи метода.

Отметим, что в отличие от теоремы 1, более слабые предположения теоремы 2, вообще говоря, не обеспечивают единственности траектории итерационной схемы (5).

2.1.3. Случай возможно неизолированных решений

В этом пункте будет получено обобщение итерационной схемы из [36].

В случае, когда искомое решение может быть неизолированным, «параметр локализации» δ в (4) приходится стремить к нулю по мере увеличения номера итерации, причем со специальной скоростью. А именно, для произвольного, но фиксированного числа $c > 0$ зададим множество

$$U^c(\pi,\, \tilde{u}) = \{u \in U(\pi,\, \tilde{u}) \mid \|u-\tilde{u}\| \leqslant c \operatorname{dist}(\tilde{u},\, \bar{U})\}, \tag{16}$$

где $\bar{U}$ есть множество решений обобщенного уравнения (1), и рассмотрим итерационную схему

$$u^{k+1} \in U^c(\pi^k,\, u^k), \quad k = 0,\, 1,\, \ldots. \tag{17}$$

Теорема 3. *Пусть отображение $\Phi\colon \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$ непрерывно в окрестности точки $\bar{u} \in \mathbb{R}^\nu$, и пусть $N(\cdot)$ — точечно-множественное отображение из $\mathbb{R}^\nu$ в $2^{\mathbb{R}^\nu}$. Пусть $\bar{U}$ — множество решений обобщенного уравнения (1), и пускай $\bar{u} \in \bar{U}$. Предположим также, что множество $\bar{U} \cap B(\bar{u},\, \varepsilon)$ является замкнутым для любого достаточно малого $\varepsilon > 0$. Пусть $\mathcal{A}$ — точечно множественное отображение из $\Pi \times \mathbb{R}^\nu \times \mathbb{R}^\nu$ в $2^{\mathbb{R}^\nu}$.*

Предположим, что для некоторого числа $c > 0$ выполняются следующие свойства:

1) *(верхне-липшицево поведение решения при канонических возмущениях) Существует число $\ell > 0$ такое, что для любого $r \in \mathbb{R}^\nu$ всякое решение $u(r)$ возмущенного обобщенного уравнения (6), достаточно близкое к $\bar{u}$, удовлетворяет оценке*

$$\operatorname{dist}(u(r),\, \bar{U}) \leqslant \ell\|r\|.$$

2) (*качество аппроксимации*) *Существуют число* $\bar{\varepsilon} > 0$ *и функция* $\omega\colon \Pi \times \mathbb{R}^\nu \times \mathbb{R}^\nu \mapsto \mathbb{R}_+$ *такие, что*

$$q := \ell \sup\{\omega(\pi,\, \tilde{u},\, u) \mid \pi \in \Pi,\ \tilde{u} \in B(\bar{u},\, \bar{\varepsilon}),\ \|u - \tilde{u}\| \leqslant c\,\mathrm{dist}(\tilde{u},\, \bar{U})\} < 1, \tag{18}$$

и оценка

$$\sup\{\|w\| \mid w \in \Phi(u) - \mathcal{A}(\pi,\, \tilde{u},\, u)\} \leqslant \omega(\pi,\, \tilde{u},\, u)\,\mathrm{dist}(\tilde{u},\, \bar{U}) \tag{19}$$

справедлива для любых $\pi \in \Pi$ *и* $(\tilde{u},\, u) \in \mathbb{R}^\nu \times \mathbb{R}^\nu$*, удовлетворяющих условиям* $\tilde{u} \in B(\bar{u},\, \bar{\varepsilon})$*,* $\|u - \tilde{u}\| \leqslant c\,\mathrm{dist}(\tilde{u},\, \bar{U})$*.*

3) (*разрешимость подзадач*) *Для любого* $\pi \in \Pi$ *и любого* $\tilde{u} \in \mathbb{R}^\nu$*, достаточно близкого к* $\bar{u}$*, множество* $U^c(\pi,\, \tilde{u})$*, опредленное в* (3), (16)*, непусто.*

Тогда для любого $\varepsilon > 0$ *найдется* $\varepsilon_0 > 0$ *такое, что для любого начального приближения* $u^0 \in B(\bar{u},\, \varepsilon_0)$ *и любой последовательности* $\{\pi^k\} \subset \Pi$ *итерационная схема* (17) (*при произвольном выборе точки* u^k*, удовлетворяющей* (17)*, на итерациях*) *генерирует траекторию* $\{u^k\} \subset B(\bar{u},\, \varepsilon)$*; любая такая траектория сходится к некоторой точке* $u^* \in \bar{U}$*, и при всех* k *выполняются следующие оценки*:

$$\|u^{k+1} - u^*\| \leqslant \frac{c\ell\omega(\pi^k,\, u^k,\, u^{k+1})}{1-q}\,\mathrm{dist}(u^k,\, \bar{U}) \leqslant \frac{cq}{1-q}\,\mathrm{dist}(u^k,\, \bar{U}), \tag{20}$$

$$\mathrm{dist}(u^{k+1},\, \bar{U}) \leqslant \ell\omega(\pi^k,\, u^k,\, u^{k+1})\,\mathrm{dist}(u^k,\, \bar{U}) \leqslant q\,\mathrm{dist}(u^k,\, \bar{U}). \tag{21}$$

В частности, скорости сходимости последовательности $\{u^k\}$ *к* u^* *и последовательности* $\{\mathrm{dist}(u^k,\, \bar{U})\}$ *к нулю являются сверхлинейными, если* $\omega(\pi^k,\, u^k,\, u^{k+1}) \to 0$ *при* $k \to \infty$*. Более того, обе эти скорости являются квадратичными, если* $\omega(\pi^k,\, u^k,\, u^{k+1}) = O(\mathrm{dist}(u^k,\, \bar{U}))$*.*

Доказательство. Согласно (3) и (16), для любых $\pi \in \Pi$ и $\tilde{u} \in \mathbb{R}^\nu$ всякая точка $u \in U^c(\pi,\, \tilde{u})$ является решением обобщенного уравнения (6) с некоторым $r \in \mathbb{R}^\nu$, удовлетворяющим (12), и, кроме того, если $\tilde{u} \in B(\bar{u},\, \bar{\varepsilon})$, то в силу предположения 2) справедлива оценка

$$\|r\| \leqslant \omega(\pi,\, \tilde{u},\, u)\,\mathrm{dist}(\tilde{u},\, \bar{U}).$$

Поскольку для любого $\varepsilon > 0$ за счет близости $\tilde{u}$ к $\bar{u}$ можно обеспечить выполнение неравенства $\|u - \bar{u}\| < \varepsilon$, из предположений 1) и 3) следует, что для любого $\varepsilon > 0$ существуют числа $\tilde{\varepsilon} \in (0,\, \min\{\varepsilon,\, \bar{\varepsilon}\}]$ и $\ell > 0$ такие, что для всех $\pi \in \Pi$ верны соотношения

$$\begin{gathered} U^c(\pi,\, \tilde{u}) \neq \emptyset, \quad \mathrm{dist}(u,\, \bar{U}) \leqslant \ell\omega(\pi,\, \tilde{u},\, u)\,\mathrm{dist}(\tilde{u},\, \bar{U}) \quad \forall\, u \in U^c(\pi,\, \tilde{u}) \\ \forall\, \tilde{u} \in B(\bar{u},\, \tilde{\varepsilon}), \end{gathered} \tag{22}$$

и при этом множество $\bar{U} \cap B(\bar{u},\, 2\tilde{\varepsilon})$ замкнуто. Положим

$$\varepsilon_0 = \frac{\tilde{\varepsilon}(1-q)}{c+1-q}. \tag{23}$$

В силу (18) $\varepsilon_0 > 0$. Теперь докажем по индукции, что если $u^0 \in B(\bar{u},\, \varepsilon_0)$, то для любой последовательности $\{\pi^k\} \subset \Pi$ итерационная схема (17) генерирует последовательность $\{u^k\} \subset \mathbb{R}^\nu$ вне зависимости от выбора точки $u^{k+1} \in U^c(\pi^k,\, u^k)$ на каждой итерации, и что для любой такой последовательности верно, что $u^k \in B(\bar{u},\, \tilde{\varepsilon}) \subset B(\bar{u},\, \varepsilon)$ для всех k.

В самом деле, поскольку равенство (23) очевидно влечет за собой неравенство $\varepsilon_0 \leqslant \tilde{\varepsilon}$, база индукции выполняется: $u^0 \in B(\bar{u},\, \tilde{\varepsilon})$. Предположим теперь, что точки $u^j \in \mathbb{R}^n$, $j = 1,\, \ldots,\, k$, таковы, что

$$u^j \in U^c(\pi^{j-1},\, u^{j-1}) \cap B(\bar{u},\, \tilde{\varepsilon}) \quad \forall\, j = 1,\, \ldots,\, k.$$

Тогда $\operatorname{dist}(u^j,\, \bar{U}) \leqslant \|u^j - \bar{u}\| \leqslant \tilde{\varepsilon}$ для всех $j = 0,\, 1,\, \ldots,\, k$, и, согласно (18) и (22), существует $u^{k+1} \in U^c(\pi^k,\, u^k)$, причем для любой такой точки u^{k+1} справедливы соотношения

$$\operatorname{dist}(u^{j+1},\, \bar{U}) \leqslant \ell\omega(\pi^j,\, u^j,\, u^{j+1}) \operatorname{dist}(u^j,\, \bar{U}) \leqslant q \operatorname{dist}(u^j,\, \bar{U}) \quad \forall\, j = 0,\, 1,\, \ldots,\, k.$$

Тогда, принимая во внимание неравенство в (16), получаем

$$\begin{aligned} \|u^{k+1} - u^0\| &\leqslant \sum_{j=0}^{k} \|u^{j+1} - u^j\| \leqslant \sum_{j=0}^{k} c \operatorname{dist}(u^j,\, \bar{U}) \leqslant c \sum_{j=0}^{k} q^j \operatorname{dist}(u^0,\, \bar{U}) = \\ &= c\frac{1-q^{k+1}}{1-q} \operatorname{dist}(u^0,\, \bar{U}) \leqslant \frac{c}{1-q} \operatorname{dist}(u^0,\, \bar{U}) \leqslant \frac{c}{1-q}\|u^0 - \bar{u}\| \leqslant \frac{c\varepsilon_0}{1-q}. \end{aligned} \tag{24}$$

Следовательно,

$$\|u^{k+1} - \bar{u}\| \leqslant \|u^{k+1} - u^0\| + \|u^0 - \bar{u}\| \leqslant \left(\frac{c}{1-q} + 1\right) \varepsilon_0 = \tilde{\varepsilon},$$

где последнее равенство следует из (23). Таким образом, $u^{k+1} \in B(\bar{u},\, \tilde{\varepsilon})$, и шаг индукции доказан.

Далее, в силу (18) и (22), а также только что доказанного утверждения, для любого $u^0 \in B(\bar{u},\, \varepsilon_0)$, любой последовательности $\{\pi^k\} \subset \Pi$ и любой последовательности $\{u^k\} \subset \mathbb{R}^\nu$, удовлетворяющей (17), справедливы соотношения

$$\operatorname{dist}(u^{k+1},\, \bar{U}) \leqslant \ell\omega(\pi^k,\, u^k,\, u^{k+1}) \operatorname{dist}(u^k,\, \bar{U}) \leqslant q \operatorname{dist}(u^k,\, \bar{U}) \quad \forall\, k. \tag{25}$$

В частности, в силу (18) последовательность $\{\operatorname{dist}(u^k,\, \bar{U})\}$ сходится к нулю, и выполняется оценка (21).

Аналогично (24), легко установить оценку

$$\|u^{k+j} - u^k\| \leqslant \frac{c}{1-q}\,\mathrm{dist}(u^k,\, \bar{U}) \quad \forall\, k, \forall\, j, \tag{26}$$

где правая часть стремится к нулю при $k \to \infty$. Следовательно, последовательность $\{u^k\}$ является фундаментальной, а значит, она сходится к некоторой точке $u^* \in \mathbb{R}^\nu$. Более того, пользуясь сходимостью последовательности $\{\mathrm{dist}(u^k,\, \bar{U})\}$ к нулю, включением $\{u^k\} \subset B(\bar{u},\, \tilde{\varepsilon})$, а также замкнутостью множества $\bar{U} \cap B(\bar{u},\, 2\tilde{\varepsilon})$, заключаем, что $u^* \in \bar{U}$.

Для того, чтобы установить скорость сходимости последовательности $\{u^k\}$, задействуя (25) и (26), получаем

$$\|u^{k+j} - u^{k+1}\| \leqslant \frac{c}{1-q}\,\mathrm{dist}(u^{k+1},\, \bar{U}) \leqslant \frac{c\ell\omega(\pi^k,\, u^k,\, u^{k+1})}{1-q}\,\mathrm{dist}(u^k,\, \bar{U}) \quad \forall\, k, \forall\, j \geqslant 1.$$

Переходя к пределу при $j \to \infty$, имеем

$$\|u^* - u^{k+1}\| \leqslant \frac{c\ell\omega(\pi^k,\, u^k,\, u^{k+1})}{1-q}\,\mathrm{dist}(u^k,\, \bar{U}) \leqslant \frac{cq}{1-q}\,\mathrm{dist}(u^k,\, \bar{U}) \quad \forall\, k,$$

то есть выполнена оценка (20). ■

Из доказательства теоремы 3 видно, что предположение 2) может быть ослаблено: условие (18) можно заменить на

$$q := \ell \sup\{\omega(\pi,\, \tilde{u},\, u) \mid \pi \in \Pi,\ u \in U^c(\pi,\, \tilde{u}),\ \tilde{u} \in B(\bar{u},\, \bar{\varepsilon})\} < 1,$$

а выполнение условия (19) достаточно предполагать лишь для $w \in \Phi(\tilde{u}) + (-\mathcal{A}(\pi,\, \tilde{u},\, u)) \cap N(u)$.

2.2. Полугладкий метод Джозефи–Ньютона

В этом разделе будет рассмотрен еще один метод решения обобщенных уравнений. В отличие от итерационных схем, изученных в предыдущем разделе, этот метод является менее абстрактным. При этом он налагает на базу обобщенного уравнения более сильные ограничения гладкости, чем непрерывность вблизи искомого решения. А именно, предполагается, что отображение Φ является полугладким в этом решении.

Определение 1. Отображение $\Phi\colon \mathbb{R}^\nu \mapsto \mathbb{R}^q$ называется *полугладким* [32, раздел 7.4] в точке $u \in \mathbb{R}^\nu$, если оно является локально липшицевым в u, дифференцируемым в u по любому направлению и удовлетворяет условию

$$\sup_{J \in \partial\Phi(u+v)} \|\Phi(u+v) - \Phi(u) - Jv\| = o(\|v\|).$$

Если выполняется более сильное условие

$$\sup_{J \in \partial\Phi(u+v)} \|\Phi(u+v) - \Phi(u) - Jv\| = O(\|v\|^2),$$

отображение Φ называют *сильно полугладким* в точке u.

Отображение называется полугладким, если оно полугладко в каждой точке своей области определения.

Напомним, что итерация метода Джозефи–Ньютона для решения уравнения (1) [17, 51, 53, 54] состоит в решении (частично) линеаризованного обобщенного уравнения

$$\Phi(u^k) + J_k(u - u^k) + N(u) \ni 0, \tag{27}$$

где $u^k \in \mathbb{R}^\nu$ — текущее приближение к искомому решению уравнения (1), а $J_k \in \mathbb{R}^{\nu\times\nu}$ — некоторая матрица. В классическом варианте метода матрица J_k полагается равной $\Phi'(u^k)$, что, конечно же, предполагает дифференцируемость отображения Φ вблизи искомого решения. Заметим, что применительно к обычным нелинейным уравнениям, т. е. если $N(\cdot) = \{0\}$, такой способ выбора матрицы J_k приводит к классическому методу Ньютона.

Поскольку полугладкость отображения не влечет за собой его дифференцируемость, классический метод Джозефи–Ньютона к обобщенным уравнениям с полугладкой базой не применим. Однако он приводит к мысли выбирать матрицу J_k в (27) из обобщенных дифференциальных объектов, которые служат заменой производной для полугладких отображений. Один из таких вариантов метода Джозефи–Ньютона и будет изучен в настоящем разделе. Его итерация будет определена позже. Пока отметим лишь, что она обобщает, с одной стороны, итерацию классического метода Джозефи–Ньютона [17, 51, 53, 54] на случай негладкой базы, а с другой — итерацию так называемого полугладкого метода Ньютона для нелинейных уравнений [60, 61, 78, 75] на случай обобщенных уравнений.

В теории классического метода Джозефи–Ньютона важную роль играет понятие сильной регулярности решения обобщенного уравнения, введенное в [81] (хотя, как следует отметить, в настоящий момент сходимость классического метода Джозефи–Ньютона установлена в более слабых предположениях [17]). А именно, решение $\bar{u}$ обобщенного уравнения (1) называется *сильно регулярным*, если для любого $r \in \mathbb{R}^\nu$, достаточно близкого к 0, возмущенное (частично) линеаризованное обобщенное уравнение

$$\Phi(\bar{u}) + \Phi'(\bar{u})(u - \bar{u}) + N(u) \ni r$$

имеет вблизи $\bar{u}$ единственное решение $u(r)$, и при этом отображение $u(\cdot)$ является локально липшицевым в точке 0. Ясно, что точка $\bar{u}$ является сильно регулярным решением обобщенного уравнения (1) тогда и только тогда, когда она является сильно регулярным решением линеаризации этого обобщенного уравнения

$$\Phi(\bar{u}) + \Phi'(\bar{u})(u - \bar{u}) + N(u) \ni 0.$$

Характеризации свойства сильной регулярности для обобщенных уравнений в терминах обобщенных дифференциальных объектов были получены в [66] (см. также [67]).

Введем в рассмотрение обобщение свойства сильной регулярности на случай, когда отображение Φ может быть недифференцируемым.

Определение 2. Решение $\bar{u} \in \mathbb{R}^\nu$ обобщенного уравнения (1) называется *сильно регулярным относительно множества матриц* $\Delta \subset \mathbb{R}^{\nu\times\nu}$, если для любой матрицы $J \in \Delta$ точка $\bar{u}$ является сильно регулярным решением обобщенного уравнения

$$\Phi(\bar{u}) + J(u - \bar{u}) + N(u) \ni 0, \tag{28}$$

т. е. для любой матрицы $J \in \Delta$ и любого $r \in \mathbb{R}^\nu$, достаточно близкого 0, возмущенное частично линеаризованное уравнение

$$\Phi(\bar{u}) + J(u - \bar{u}) + N(u) \ni r$$

имеет вблизи $\bar{u}$ единственное решение $u_J(r)$, и отображение $u_J(\cdot)$ является локально липшицевым в точке 0.

Когда $\Delta = \partial_B\Phi(\bar{u})$ ($\Delta = \partial\Phi(\bar{u})$), будем говорить, что решение $\bar{u}$ является *BD-регулярным* (*CD-регулярным*) решением обобщенного уравнения (1).

Очевидно, что введенное понятие сильной регулярности обобщает следующие широко известные свойства: сильную регулярность в смысле [81], соответствующую случаю, когда отображение Φ дифференцируемо, а $\Delta = \{\Phi'(\bar{u})\}$, и BD-регулярность [73] и CD-регулярность [77] решения обычного уравнения, которые соответствуют случаю, когда $N(\cdot) = \{0\}$, а $\Delta = \partial_B\Phi(\bar{u})$ и, соответственно, $\Delta = \partial\Phi(\bar{u})$.

Напомним, что, как показано в [42], если отображение Φ является локально липшицевым в решении $\bar{u}$ обобщенного уравнения (1), CD-регулярность этого решения влечет за собой его сильную метрическую регулярность (см. с. 50).

Следующий результат об устойчивости свойства сильной регулярности при малых липшицевых возмущениях следует, к примеру, из [28, теорема 1.4].

Предложение 1. *Пусть для заданного отображения* $\Phi\colon \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$, *матрицы* $J \in \mathbb{R}^{\nu\times\nu}$ *и точечно-множественного отображения* N *из* $\mathbb{R}^\nu$ *в* $2^{\mathbb{R}^\nu}$ *точка* $\bar{u}$ *является сильно регулярным решением обобщенного уравнения* (28).

Тогда для любой фиксированной окрестности W точки $\bar{u}$ и любой достаточно малой константы $\ell \geqslant 0$ существуют число $\bar{\ell} > 0$ и окрестности U точки $\bar{u}$ и V точки 0 такие, что для всякого отображения $R\colon \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$, липшицевого на W с константой Липшица ℓ, и любого $r \in R(\bar{u}) + V$ обобщенное уравнение

$$R(u) + \Phi(\bar{u}) + J(u - \bar{u}) + N(u) \ni r$$

имеет в U единственное решение $u(r)$, и отображение $u(\cdot)$ удовлетворяет условию Липшица на $R(\bar{u}) + V$ с константой $\bar{\ell}$.

Пользуясь предложением 1, докажем разрешимость возмущенного линеаризованного уравнения для всех точек, достаточно близких к сильно регулярному решению исходного уравнения (1), и всех матриц J, достаточно близких ко множеству Δ, по отношению к которому это решение сильно регулярно.

Предложение 2. *Пусть отображение $\Phi\colon \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$ является непрерывным в точке $\bar{u} \in \mathbb{R}^\nu$. Пусть для заданного точечно-множественного отображения N из $\mathbb{R}^\nu$ в $2^{\mathbb{R}^\nu}$ точка $\bar{u}$ является сильно регулярным решением обобщенного уравнения* (1) *относительно компактного множества $\Delta \subset \mathbb{R}^{\nu\times\nu}$.*

Тогда существуют число $\varepsilon > 0$, число $\bar{\ell} > 0$ и окрестности $\tilde{U}$ и U точки $\bar{u}$, а также окрестность V точки 0 такие, что для любого $\tilde{u} \in \tilde{U}$, любой матрицы $J \in \mathbb{R}^{\nu\times\nu}$, удовлетворяющей условию

$$\operatorname{dist}(J,\, \Delta) < \varepsilon, \tag{29}$$

и любого $r \in V$ обобщенное уравнение

$$\Phi(\tilde{u}) + J(u - \tilde{u}) + N(u) \ni r \tag{30}$$

имеет в U единственное решение $u(r)$, причем отображение $u(\cdot)$ удовлетворяет условию Липшица на V с константой $\bar{\ell}$.

Доказательство. Зафиксируем произвольную матрицу $\bar{J} \in \Delta$. Для любых $\tilde{u} \in \mathbb{R}^\nu$ и $J \in \mathbb{R}^{\nu\times\nu}$ определим отображение $R\colon \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$,

$$R(u) = \Phi(\tilde{u}) - \Phi(\bar{u}) - J\tilde{u} + \bar{J}\bar{u} + (J - \bar{J})u. \tag{31}$$

Заметим, что для всякого фиксированного $\ell > 0$ отображение R удовлетворяет условию Липшица на $\mathbb{R}^\nu$ с константой ℓ, если матрица J лежит достаточно близко к $\bar{J}$. Кроме того,

$R(\bar{u}) = \Phi(\tilde{u}) - \Phi(\bar{u}) - J(\tilde{u} - \bar{u})$ стремится к 0 при $\tilde{u} \to \bar{u}$. Тогда, применяя предложение 1 с $W = \mathbb{R}^\nu$, получаем существование чисел $\varepsilon > 0$, $\bar{\ell} > 0$ и окрестностей $\tilde{U}$ и U точки $\bar{u}$, а также окрестности V точки 0 таких, что для любого $\tilde{u} \in \tilde{U}$ и любой матрицы $J \in \mathbb{R}^{\nu\times\nu}$, удовлетворяющей условию $\|J - \bar{J}\| < \varepsilon$, при каждом $r \in V$ обобщенное уравнение

$$R(u) + \Phi(\bar{u}) + \bar{J}(u - \bar{u}) + N(u) \ni r \tag{32}$$

имеет в U единственное решение $u(r)$, и при этом отображение $u(\cdot)$ удовлетворяет условию Липшица на V с константой $\bar{\ell}$. Подставляя (31) в (32), видим, что последнее уравнение совпадает с (30).

Рассмотрим теперь для каждой матрицы $\bar{J} \in \Delta$ открытый шар в пространстве $\mathbb{R}^{\nu\times\nu}$ с центром в $\bar{J}$ радиуса ε, где число ε для каждой матрицы $\bar{J}$ определяется в соответствии с проведенным рассуждением. Совокупность указанных шаров образует открытое покрытие компакта Δ, из которого можно выделить конечное подпокрытие. Теперь переопределим $\varepsilon > 0$ таким образом, чтобы любая матрица $J \in \mathbb{R}^{\nu\times\nu}$, удовлетворяющяя (29), принадлежала этому конечному подпокрытию. Далее, пусть $\bar{\ell} > 0$ — максимум из констант Липшица, соответствующих шарам подпокрытия, а окрестности $\tilde{U}$, U и V являются пересечением окрестностей, соответствующих этим шарам. При необходимости сужая окрестности $\tilde{U}$ и V для того, чтобы гарантировать, что для любого $\tilde{u} \in \tilde{U}$, любой матрицы $J \in \mathbb{R}^{\nu\times\nu}$, удовлетворяющей (29), и всякого $r \in V$ решение $u(r)$ уравнения (30) принадлежит U, получаем требуемое. ∎

Наряду с *полугладким* методом Джозефи–Ньютона, итерация которого задается формулой (27) с $J_k \in \partial\Phi(u^k)$, будем рассматривать его возмущенную версию. Ее итерация заключается в следующем: по текущему приближению к решению уравнения $u^k \in \mathbb{R}^\nu$ следующее приближение u^{k+1} ищется как решение обобщенного уравнения

$$\omega^k + \Phi(u^k) + J_k(u - u^k) + N(u) \ni 0 \tag{33}$$

с некоторой $J_k \in \partial\Phi(u^k)$, где $\omega^k \in \mathbb{R}^\nu$ есть возмущение итерации исходного метода. Возникновение возмущения, к примеру, может быть вызвано неточным решением подзадачи $\Phi(u^k) + J_k(u - u^k) + N(u) \ni 0$. Еще один пример возмущений связан с квазиньютоновскими вариантами полугладкого метода Джозефи–Ньютона, итерация которых может быть записана в общем виде $\Phi(u^k) + J(u - u^k) + N(u) \ni 0$, где матрица $J \notin \partial\Phi(u^k)$. В этом случае $\omega^k = (J - J_k)(u^{k+1} - u^k)$.

Начнем со следующего апостериорного результата о сверхлинейной сходимости алгоритма (33), т. е. результата в предположении о наличии сходимости как таковой.

Предложение 3. *Пусть отображение* $\Phi\colon \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$ *полугладко в точке* $\bar{u} \in \mathbb{R}^\nu$. *Пусть* $\bar{u}$ *является решением обобщенного уравнения* (1), *сильно регулярным относительно замкнутого множества* $\bar{\Delta} \subset \partial\Phi(\bar{u})$. *Пусть последовательность* $\{u^k\} \subset \mathbb{R}^\nu$ *сходится к* $\bar{u}$, *и при этом точка* u^{k+1} *удовлетворяет включению* (33) *при всех* $k = 0, 1, \ldots$ *с некоторой матрицей* $J_k \in \partial\Phi(u^k)$ *и возмущением* $\omega^k \in \mathbb{R}^\nu$ *такими, что*

$$\mathrm{dist}(J_k,\, \bar{\Delta}) \to 0 \text{ при } k \to \infty, \tag{34}$$

и

$$\omega^k = o(\|u^{k+1} - u^k\| + \|u^k - \bar{u}\|). \tag{35}$$

Тогда скорость сходимости последовательности $\{u^k\}$ *является сверхлинейной. Более того, скорость сходимости является квадратичной, если отображение* Φ *сильно полугладко в точке* $\bar{u}$, *и*

$$\omega^k = O(\|u^{k+1} - u^k\|^2 + \|u^k - \bar{u}\|^2). \tag{36}$$

Доказательство. Определим $\varepsilon > 0$, $\bar{\ell} > 0$, $\tilde{U}$, U и V в соответствии с предложением 2 при $\Delta = \bar{\Delta}$. Тогда для любого $u^k \in \tilde{U}$, любой матрицы $J_k \in \partial\Phi(u^k)$, удовлетворяющей условию $\mathrm{dist}(J_k,\, \bar{\Delta}) < \varepsilon$, и любого $r \in V$ обобщенное уравнение

$$\Phi(u^k) + J_k(u - u^k) + N(u) \ni r \tag{37}$$

имеет в U единственное решение $u(r)$, которое зависит от r на V липшицевым образом с константой $\bar{\ell}$. Для каждого k положим

$$r^k = \Phi(u^k) - \Phi(\bar{u}) - J_k(u^k - \bar{u}). \tag{38}$$

Заметим, что в силу полугладкости отображения Φ в точке $\bar{u}$

$$r^k = o(\|u^k - \bar{u}\|). \tag{39}$$

Кроме того, с учетом (38) имеем:

$$0 \in \Phi(\bar{u}) + N(\bar{u}) = \Phi(u^k) + J_k(\bar{u} - u^k) + N(\bar{u}) - r^k. \tag{40}$$

Так как последовательность $\{u^k\}$ сходится к $\bar{u}$, и выполнены соотношения (34), (35) и (39), для всех достаточно больших k верно, что u^k, $u^{k+1} \in \tilde{U} \cap U$, $\mathrm{dist}(J_k,\, \bar{\Delta}) < \varepsilon$, $-\omega^k \in V$, и $r^k \in V$. Следовательно, согласно предложению 2, u^{k+1} является единственным в U решением обобщенного уравнения (37) с $r = -\omega^k$, т.е. $u^{k+1} = u(-\omega^k)$, в то время как в силу (40)

$\bar{u}$ является единственным в U решением обобщенного уравнения (37) с $r = r^k$, а значит, $\bar{u} = u(r^k)$. Тогда

$$\|u^{k+1} - \bar{u}\| = \|u(-\omega^k) - u(r^k)\| \leqslant \bar{\ell}\|\omega^k + r^k\| = o(\|u^{k+1} - u^k\| + \|u^k - \bar{u}\|), \tag{41}$$

где последняя оценка следует из (35) и (39).

Тот факт, что из (41) следует сверхлинейная скорость сходимости, доказывается стандартным образом, см., например, [51, предложение 2.1]. Остается заметить, что если отображение Φ является сильно полугладким в точке $\bar{u}$, то для вектора r^k, определенного согласно (38), выполняется оценка

$$r^k = O(\|u^k - \bar{u}\|^2).$$

Эта оценка вместе с (36) и с неравенством в (41) влечет за собой квадратичную скорость сходимости. ∎

Из предложения 3 непосредственно следует результат типа Дениса–Морэ для метода Джозефи–Ньютона применительно к обобщенным уравнениям с полугладкой базой. А именно, пусть имеется последовательность матриц $\{J_k\} \subset \mathbb{R}^{\nu\times\nu}$. Пусть по текущему приближению $u^k \in \mathbb{R}^\nu$ к решению обобщенного уравнения (1) следующее приближение u^{k+1} вычисляется как решение обобщенного уравнения (27). Кроме того, предположим, что последовательность $\{J_k\}$ удовлетворяет следующему условию типа Дениса–Морэ:

$$\min_{J\in\partial\Phi(u^k)} \|(J_k - J)(u^{k+1} - u^k)\| = o(\|u^{k+1} - u^k\|). \tag{42}$$

Имеет место следующий апостериорный результат.

Теорема 4. *Пусть отображение* $\Phi\colon \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$ *является полугладким в точке* $\bar{u} \in \mathbb{R}^\nu$. *Пусть* $\bar{u}$ *— решение обобщенного уравнения* (1), *сильно регулярное относительно некоторого замкнутого множества* $\bar{\Delta} \subset \partial\Phi(\bar{u})$. *Пусть имеется последовательность матриц* $\{J_k\} \subset \mathbb{R}^{\nu\times\nu}$, *и пусть последовательность* $\{u^k\} \subset \mathbb{R}^\nu$ *сходится к* $\bar{u}$, *и при этом для всех достаточно больших* k *точка* u^{k+1} *удовлетворяет включению* (27), *и существует матрица* $\tilde{J}_k \in \partial\Phi(u^k)$ *такая, что*

$$\mathrm{dist}(\tilde{J}_k, \bar{\Delta}) \to 0 \text{ при } k \to \infty \tag{43}$$

и

$$(J_k - \tilde{J}_k)(u^{k+1} - u^k) = o(\|u^{k+1} - u^k\|). \tag{44}$$

Тогда скорость сходимости последовательности $\{u^k\}$ *сверхлинейная.*

Доказательство. Для каждого k положим

$$\omega^k = (J_k - \tilde{J}_k)(u^{k+1} - u^k).$$

Тогда условие (44) влечет за собой (35). С учетом (43), требуемый результат немедленно следует из предложения 3. ■

Предположение о том, что для любого достаточно большого k найдется матрица $\tilde{J}_k \in \partial\Phi(u^k)$, удовлетворяющая (44), эквивалентно условию (42). Если точка $\bar{u}$ является CD-регулярным решением уравнения (1), то теорему 4 можно применить с $\bar{\Delta} = \partial\Phi(\bar{u})$, причем в этом случае (43) выполняется автоматически вне зависимости от выбора $\tilde{J}_k \in \partial\Phi(u^k)$, в следствие полунепрерывности дифференциала Кларка сверху. Также теорему 4 можно применять с $\bar{\Delta} = \partial_B\Phi(\bar{u})$, предполагая BD-регулярность решения $\bar{u}$.

Наконец, перейдем к априорному локальному анализу полугладкого метода Джозефи–Ньютона, т. е. к установлению достаточных условий для сходимости этого метода.

Теорема 5. *Пусть отображение* $\Phi\colon \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$ *полугладко в точке* $\bar{u} \in \mathbb{R}^\nu$. *Пусть* $\bar{u}$ *является решением уравнения* (1), *сильно регулярным относительно замкнутого множества* $\bar{\Delta} \subset \partial\Phi(\bar{u})$. *Пусть* Δ *— точечно-множественное отображение из* $\mathbb{R}^\nu$ *во множество подмножеств* $\mathbb{R}^{\nu\times\nu}$ *такое, что*

$$\Delta(u) \subset \partial\Phi(u) \quad \forall\, u \in \mathbb{R}^\nu, \tag{45}$$

и для любого $\varepsilon > 0$ *существует окрестность* O *точки* $\bar{u}$ *такая, что*

$$\operatorname{dist}(J,\, \bar{\Delta}) < \varepsilon \quad \forall\, J \in \Delta(u),\ \forall\, u \in O. \tag{46}$$

Тогда найдется число $\delta > 0$ *такое, что для любого начального приближения* $u^0 \in \mathbb{R}^\nu$, *достаточно близкого к решению* $\bar{u}$, *при всех* $k = 0,\, 1,\, \ldots$ *и при произвольном выборе матрицы* $J_k \in \Delta(u^k)$ *существует единственное решение* u^{k+1} *подзадачи полугладкого метода Джозефи–Ньютона* (27), *удовлетворяющее условию*

$$\|u^{k+1} - u^k\| \leqslant \delta; \tag{47}$$

при этом последовательность $\{u^k\}$ *сходится к* $\bar{u}$, *и скорость сходимости является сверхлинейной. Более того, скорость сходимости является квадратичной, если отображение* Φ *сильно полугладко в точке* $\bar{u}$.

Доказательство. Определим $\varepsilon > 0$, $\bar{\ell} > 0$, $\tilde{U}$, U и V в соответствии с предложением 2 при $\Delta = \bar{\Delta}$. Пусть, кроме того, окрестность $\tilde{U}$ такова, что условие (46) выполняется с $O = \tilde{U}$ и с указанным ε.

Тогда по предложению 2 для любых $u^k \in \tilde{U}$ и $J_k \in \Delta(u^k)$ и всякого $r \in V$ обобщенное уравнение

$$\Phi(u^k) + J_k(u - u^k) + N(u) \ni r \tag{48}$$

имеет в U единственное решение $u(r)$, которое зависит от r на V липшицевым образом с константой $\bar{\ell}$. В частности, обобщенное уравнение (27) имеет в U единственное решение $u^{k+1} = u(0)$.

Определив r^k согласно (38) и используя (45) и полугладкость отображения Φ в точке $\bar{u}$ заключаем, что выполнено (39), и

$$0 \in \Phi(\bar{u}) + N(\bar{u}) = \Phi(u^k) + J_k(\bar{u} - u^k) + N(\bar{u}) - r^k.$$

При необходимости сужая окрестность $\tilde{U}$, в силу (39) получаем, что $r^k \in V$, если $u^k \in \tilde{U}$, а значит, точка $\bar{u}$ является единственным решением уравнения (48) с $r = r^k$, т. е. $\bar{u} = u(r^k)$. Тогда

$$\|u^{k+1} - \bar{u}\| \leqslant \|u(r^k) - u(0)\| \leqslant \bar{\ell}\|r^k\| = o(\|u^k - \bar{u}\|), \tag{49}$$

где последняя оценка следует из (39).

Далее, из (49) следует, что для любого $q \in (0,\, 1)$ существует число $\delta > 0$ такое, что $B(\bar{u},\, \delta/2) \subset \tilde{U}$, $B(\bar{u},\, 3\delta/2) \subset U$, и для любого $u^k \in B(\bar{u},\, \delta/2)$ выполняется неравенство

$$\|u^{k+1} - \bar{u}\| \leqslant q\|u^k - \bar{u}\|, \tag{50}$$

откуда следует, что $u^{k+1} \in B(\bar{u},\, \delta/2)$. Тогда

$$\|u^{k+1} - u^k\| \leqslant \|u^{k+1} - \bar{u}\| + \|u^k - \bar{u}\| < \frac{\delta}{2} + \frac{\delta}{2} = \delta,$$

и значит, u^{k+1} является решением (27), удовлетворяющим (47). Более того, для любой точки $u^{k+1} \in \mathbb{R}^n$, удовлетворяющей (47), справедливы соотношения

$$\|u^{k+1} - \bar{u}\| \leqslant \|u^{k+1} - u^k\| + \|u^k - \bar{u}\| \leqslant \delta + \frac{\delta}{2} = \frac{3\delta}{2},$$

и значит, $u^{k+1} \in U$, откуда следует, что u^{k+1} является решением обобщенного уравнения (27) тогда и только тогда, когда $u^{k+1} = u(0)$. Таким образом, $u(0)$ — единственное решение обобщенного уравнения (27), удовлетворяющее условию (47).

Таким образом, из включения $u^0 \in B(\bar{u},\, \delta/2)$ следует, что последовательность $\{u^k\}$ однозначно определена (при условии, что способ выбора матрицы $J_k \in \Delta(u^k)$ для всех k является фиксированным) и целиком содержится в $B(\bar{u},\, \delta/2)$. Тогда соотношение (50) влечет за собой сходимость этой последовательности к точке $\bar{u}$. Оценки скорости сходимости следуют из предложения 3. ∎

Замечание 1. Теорема 5 обобщает результат о сходимости полугладкого метода Ньютона для обычных нелинейных уравнений [75, 78]. В самом деле, в качестве точечно-множественного отображения $\Delta(\cdot)$ можно взять $\partial_B\Phi(\cdot)$ или $\partial\Phi(\cdot)$. В первом случае теорема 5 применима с $\bar{\Delta} = \partial_B\Phi(\bar{u})$ в предположении BD-регулярности решения $\bar{u}$, в то время как во втором случае ее можно применить с $\bar{\Delta} = \partial\Phi(\bar{u})$, предполагая CD-регулярность решения $\bar{u}$. Конечно, возможны и другие варианты выбора $\Delta(\cdot)$ (к примеру, связанные со специальной структурой конкретной задачи).

Замечание 2. Для классического метода Джозефи–Ньютона в [17] был доказан более тонкий результат о сходимости, в котором предполагается полуустойчивость и хемиустойчивость (англ. *hemistability*) искомого решения. Комбинация этих двух свойств, вообще говоря, слабее сильной регулярности. Заметим, однако, что в отличие от результатов теоремы 5 локальная единственность решения подзадачи метода не имеет места в этих более слабых предположениях.

Глава 3

Методы оптимизации для задач с липшицевыми производными

В данной главе предлагается ряд численных методов для решения задач оптимизации с лишпшицевыми производными. Для этих методов устанавливаются результаты о локальной сходимости и скорости сходимости в различных предположениях. В качестве инструментов анализа при этом выступают итерационные схемы, представленные в предыдущей главе.

Глава состоит из трех разделов. Первый раздел посвящен так называемому *методу множителей* или *методу модифицированных функций Лагранжа*. Во втором разделе будет рассмотрен *метод множителей с линеаризованными ограничениями*. Нужно отметить, что итерации как первого, так и второго из этих методов не используют вторые производные целевой функции и ограничений задачи, в связи с чем указанные методы являются естественными кандидатами на роль методов решения задач с липшицевыми производными. Однако имеющиеся в литературе теории локальной сходимости этих методов предполагают существование вторых производных у целевой функции и ограничений решаемой задачи. В настоящей главе эти теории будут обобщены на задачи оптимизации с липшицевыми производными. Заметим, что такое обобщение зачастую является весьма нетривиальным.

Наконец, в третьем разделе главы речь пойдет о *полугладком методе последовательного квадратичного программирования*. В контексте этого метода нас будут интересовать прежде всего условия, характеризующие его прямую сверхлинейную сходимость.

3.1. Метод модифицированных функций Лагранжа

Первыми работами по методу модифицированных функций Лагранжа или методу множителей являются [40] и [74]. Впоследствии метод модифицированных функций изучался в [13, 14, 33]. Другими ключевыми работами, посвященными этому методу, являются [11, 14, 22]. Отметим, что метод модифицированных функций Лагранжа лежит в основе таких успешных солверов, как LANCELOT [62] и ALGENCAN [10]. Теория локальной и глобальной сходимости этого метода до сих пор остается предметом активных исследований, см., например, [11, 33] и ссылки на литературу в этих работах. Традиционно локальная сходимость метода модифицированных функций изучалась в предположении о двукратной дифференцируемости целевой функции и ограничений решаемой задачи. В данном разделе существующие результаты о локальной сходимости этого метода распространяются на задачи оптимизации с липшицевыми производными. Кроме того, получаются новые результаты о его локальной сходимости в одном лишь предположении о некртичности множителя.

3.1.1. Сходимость при достаточном условии второго порядка оптимальности

Будем рассматривать задачу оптимизации

$$\begin{gathered} f(x) \to \min, \\ h(x) = 0,\ g(x) \leqslant 0, \end{gathered} \tag{1}$$

в которой целевая функция $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ и отображения $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ и $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ дифференцируемы. Напомним, что система Каруша–Куна–Таккера задачи (1) имеет вид

$$\begin{gathered} \frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) = 0, \\ h(x) = 0, \quad \mu \geqslant 0, \quad g(x) \leqslant 0, \quad \langle \mu,\, g(x) \rangle = 0, \end{gathered} \tag{2}$$

где $L\colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ — функция Лагранжа задачи (1),

$$L(x,\, \lambda,\, \mu) = f(x) + \langle \lambda,\, h(x) \rangle + \langle \mu,\, g(x) \rangle. \tag{3}$$

Заметим, что система Каруша–Куна–Таккера (2) эквивалентна обобщенному уравнению

$$\Phi(u) + N(u) \ni 0, \tag{4}$$

в котором $u = (x,\, \lambda,\, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, отображение $\Phi\colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \mapsto \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ задано равенством

$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu),\, h(x),\, -g(x) \right), \tag{5}$$

а в качестве точечно-множественного отображения N из $\mathbb{R}^\nu$ во множество подмножеств $\mathbb{R}^\nu$ используется

$$N(\cdot) = N_Q(\cdot), \quad Q = \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+, \tag{6}$$

где $N_Q(u)$ — нормальный конус ко множеству Q в точке u.

Модифицированной функцией Лагранжа $L_\sigma \colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \mapsto \mathbb{R}$ задачи (1) называется функция

$$L_\sigma(x,\, \lambda,\, \mu) = f(x) + \frac{\sigma}{2}(\|\lambda + h(x)/\sigma\|_2^2 + \|\max\{0,\, \mu + g(x)/\sigma\}\|_2^2),$$

где максимум берется покомпонентно. Параметр σ в определении модифицированной функции Лагранжа называется *обратным параметром штрафа*.

Итерация метода модифицированных функций Лагранжа состоит в следующем. По текущему двойственному приближению $(\lambda^k,\, \mu^k) \in \mathbb{R}^l \times \mathbb{R}^m$, текущему значению обратного параметра штрафа $\sigma_k > 0$, и текущему значению *параметра неточности* $\tau_k \geqslant 0$ очередное приближение $(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ генерируется следующим образом: точка x^{k+1} должна удовлетворять условию

$$\left\| \frac{\partial L_{\sigma_k}}{\partial x}(x^{k+1},\, \lambda^k,\, \mu^k) \right\| \leqslant \tau_k, \tag{7}$$

а пара $(\lambda^{k+1},\, \mu^{k+1})$ вычисляется по формулам

$$\lambda^{k+1} = \lambda^k + h(x^{k+1})/\sigma_k, \quad \mu^{k+1} = \max\{0,\, \mu^k + g(x^{k+1})/\sigma_k\}. \tag{8}$$

Первый результат о сходимости метода модифицированных функций из представленных в данном пункте касается точной версии метода, которая соответствует выбору $\tau_k = 0$ для всех k. Таким образом, *точный* метод модифицированных функций генерирует прямое приближение x^{k+1} как стационарную точку задачи оптимизации без ограничений

$$L_{\sigma_k}(x,\, \lambda^k,\, \mu^k) \to \min, \quad x \in \mathbb{R}^n, \tag{9}$$

и имеют место равенства

$$\begin{aligned} 0 = \frac{\partial L_{\sigma_k}}{\partial x}(x^{k+1},\, \lambda^k,\, \mu^k) &= f'(x^{k+1}) + (h'(x^{k+1}))^{\mathrm{T}}(\lambda^k + h(x^{k+1})/\sigma_k) + \\ &+ \sum_{i=1}^m \max\{0,\, g_i(x^{k+1})/\sigma_k + \mu_i^k\} g_i'(x^{k+1}) = \\ &= f'(x^{k+1}) + (h'(x^{k+1}))^{\mathrm{T}} \lambda^{k+1} + (g'(x^{k+1}))^{\mathrm{T}} \mu^{k+1} = \frac{\partial L}{\partial x}(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1}), \end{aligned} \tag{10}$$

$$0 = h(x^{k+1}) - \sigma_k(\lambda^{k+1} - \lambda^k),$$

$$0 = \max\{-\mu^{k+1},\, g(x^{k+1})/\sigma_k - (\mu^{k+1} - \mu^k)\} = -\min\{\mu^{k+1},\, -g(x^{k+1})/\sigma_k + (\mu^{k+1} - \mu^k)\}.$$

Отсюда следует, что итерация точного метода модифицированных функций может быть записана в виде (2.2) с $\mathcal{A}\colon (\mathbb{R}_+ \setminus \{0\}) \times \mathbb{R}^\nu \times \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$,

$$\mathcal{A}(\sigma,\, \tilde{u},\, u) = \left(\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu),\, h(x) - \sigma(\lambda - \tilde{\lambda}),\, -g(x) + \sigma(\mu - \tilde{\mu})\right),$$

где $\nu = n+l+m$, $\tilde{u} = (\tilde{x},\, \tilde{\lambda},\, \tilde{\mu})$, а точечно-множественное отображение N определено согласно (6). Легко видеть, что $\mathcal{A}(\sigma,\, \tilde{u},\, \tilde{u}) = \Phi(\tilde{u})$, и

$$\Phi(u) - \mathcal{A}(\sigma,\, \tilde{u},\, u) = \left(0,\, \sigma(\lambda - \tilde{\lambda}),\, -\sigma(\mu - \tilde{\mu})\right),$$

а значит,

$$\|(\Phi(u^1) - \mathcal{A}(\sigma,\, \tilde{u},\, u^1)) - (\Phi(u^2) - \mathcal{A}(\sigma,\, \tilde{u},\, u^2))\| = \sigma\|(\lambda^1 - \lambda^2,\, \mu^1 - \mu^2)\|,$$

откуда следует выполнение условия (2.7) с $\omega(\sigma,\, \tilde{u},\, u^1,\, u^2) = \sigma$ и $u^1 = (x^1,\, \lambda^1,\, \mu^1)$, $u^2 = (x^2,\, \lambda^2,\, \mu^2)$ для любых $\sigma > 0$ и $x^1,\, x^2 \in \mathbb{R}^n$, $\lambda^1,\, \lambda^2 \in \mathbb{R}^l$, $\mu^1,\, \mu^2 \in \mathbb{R}^m$. Тогда результат о локальной сходимости и скорости сходимости будет напрямую следовать из теоремы 2.1, если предположить сильную метрическую регулярность решения $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ системы ККТ (2).

Как уже отмечалось, согласно [42], сильная метрическая регулярность решения обобщенного уравнения (4) следует из CD-регулярности (см. с. 58) этого решения. Займемся получением достаточных условий для CD-регулярности в контексте систем Каруша–Куна–Таккера.

Прежде всего напомним, что выполнение в точке $\bar{x} \in \mathbb{R}^n$ *условия линейной независимости* состоит в линейной независимости градиентов $h_j'(\bar{x})$, $j = 1,\, \ldots,\, l$ и $g_i'(\bar{x})$, $i \in A(\bar{x})$, где множество $A(\bar{x})$ было введено в (1.11) и представляет собой множество индексов активных ограничений-неравенств в точке $\bar{x}$. Помимо $A(\bar{x})$, ниже будут использоваться множества $A_+(\bar{x},\, \bar{\mu})$ и $A_0(\bar{x},\, \bar{\mu})$, которые были введены в (1.11) для любых $\bar{x} \in \mathbb{R}^n$ и $\bar{\mu} \in \mathbb{R}^m$.

Для задач оптимизации, целевая функция и ограничения которых являются дважды непрерывно дифференцируемыми, характеризация CD-регулярности (которая для таких задач эквивалентна сильной регулярности) была получена в [81] (достаточность) и в [20] (необходимость). Оказывается, что из этих результатов несложно получить условия, гарантирующие CD-регулярность в случае задач с липшицевыми производными.

Предложение 1. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображения* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *и* $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ *дифференцируемы в точке* $\bar{x} \in \mathbb{R}^n$. *Пусть* $\bar{x}$ *— стационарная точка задачи* (1), *и пусть*

$(\bar\lambda,\, \bar\mu) \in \mathbb{R}^l \times \mathbb{R}^m$ — отвечающий ей множитель Лагранжа. Пусть $H \in \mathbb{R}^{n\times n}$ — произвольная матрица, и

$$J = \begin{pmatrix} H & (h'(\bar x))^{\mathrm T} & (g'(\bar x))^{\mathrm T} \\ h'(\bar x) & 0 & 0 \\ -g'(\bar x) & 0 & 0 \end{pmatrix}. \tag{11}$$

Тогда, если в точке $\bar x$ выполнено условие линейной независимости, а множитель $(\bar\lambda,\, \bar\mu)$ удовлетворяет условию

$$\langle H\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in C_+(\bar x,\, \bar\mu) \setminus \{0\}, \tag{12}$$

где

$$C_+(\bar x,\, \bar\mu) = \{\xi \in \mathbb{R}^n \mid h'(\bar x)\xi = 0,\ g'_{A_+(\bar x,\, \bar\mu)}(\bar x)\xi = 0\},$$

то тройка $\bar u = (\bar x,\, \bar\lambda,\, \bar\mu)$ является сильно регулярным решением обобщенного уравнения

$$\Phi(\bar u) + J(u - \bar u) + N(u) \ni 0 \tag{13}$$

с $\Phi(\cdot)$ и $N(\cdot)$, определенными согласно (5) и, соответственно, (6).

Более того, выполнение условия линейной независимости необходимо для сильной регулярности решения $\bar u$, а выполнение условия (12) необходимо для сильной регулярности решения $\bar u$, если $\bar x$ является локальным решением следующей задачи квадратичного программирования

$$\begin{gathered} \langle f'(\bar x),\, x - \bar x\rangle + \tfrac12 \langle H(x - \bar x),\, x - \bar x\rangle \to \min, \\ h'(\bar x)(x - \bar x) = 0,\ g'_{A(\bar x)}(\bar x)(x - \bar x) \leqslant 0. \end{gathered} \tag{14}$$

Доказательство. Задача (14) локально (вблизи $\bar x$) эквивалентна задаче

$$\begin{gathered} \langle f'(\bar x),\, x - \bar x\rangle + \tfrac12 \langle H(x - \bar x),\, x - \bar x\rangle \to \min, \\ h(\bar x) + h'(\bar x)(x - \bar x) = 0,\ g(\bar x) + g'(\bar x)(x - \bar x) \leqslant 0. \end{gathered} \tag{15}$$

Легко видеть, что система ККТ задачи (15) может быть записана в виде обобщенного уравнения (13) с отображениями $\Phi(\cdot)$ и $N(\cdot)$, заданными согласно (5) и (6), соответственно, и с J, определенной согласно (11). Более того, если $\bar x$ — стационарная точка задачи (1), а $(\bar\lambda,\, \bar\mu)$ — отвечающий ей множитель Лагранжа, то то же самое справедливо и по отношению к задаче (15), и наоборот. При этом множества активных в точке $\bar x$ ограничений-неравенств для указанных двух задач совпадают. Выполнение в точке $\bar x$ условия линейной независимости для них означает одно и то же, а условие (12) совпадает с так называемым сильным достаточным условием второго порядка для задачи (15). С учетом сказанного, требуемое утверждение следует из результатов работ [81] и [20], применяемых к задаче (15) (это можно сделать, поскольку задача (15) является задачей квадратичного программирования и, конечно же, удовлетворяет условиям гладкости, которые требуются в [20, 81]). ■

Замечание 1. Как следует из теоремы 1.1, для произвольной тройки $u = (x,\, \lambda,\, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ справедливо равенство

$$\partial\Phi(u) = \left\{ \left(\begin{array}{ccc} H & (h'(x))^{\mathrm{T}} & (g'(x))^{\mathrm{T}} \\ h'(x) & 0 & 0 \\ -g'(x) & 0 & 0 \end{array} \right) \middle|\, H \in \partial_x \frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) \right\}. \qquad (16)$$

Из (16) и предложения 1 немедленно вытекает следующее: если в стационарной точке задачи (1) выполняется условие линейной независимости, а множитель $(\bar{\lambda},\, \bar{\mu})$, отвечающий этой точке, удовлетворяет сильному достаточному условию второго порядка оптимальности

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \quad \langle H\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in C_+(\bar{x},\, \bar{\mu}) \setminus \{0\}, \qquad (17)$$

то тройка $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu})$ является CD-регулярным решением обобщенного уравнения (4) с $\Phi(\cdot)$ и $N(\cdot)$ определенными согласно (5) и, соответственно, (6).

Для задач с ограничениями равенствами CD-регулярность решения $(\bar{x},\, \bar{\lambda})$ системы Каруша–Куна–Таккера эквивалентна комбинации условия регулярности ограничений

$$\operatorname{rank} h'(\bar{x}) = l$$

и требования, чтобы ни для какой матрицы $H \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda})$ не нашлось вектора $\xi \in \ker h'(\bar{x}) \setminus \{0\}$ такого, что $H\xi \in \operatorname{im}(h'(\bar{x}))^{\mathrm{T}}$. Последнее свойство, вообще говоря, сильнее, чем некритичность множителя $\bar{\lambda}$, а для задач, целевая функция и ограничения которых являются дважды непрерывно дифференцируемыми, эквивалентно некритичности (см. замечание 1.5).

Собирая воедино все сказанное выше, получаем следующий результат.

Теорема 1. *Пусть функция $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ и отображения $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ и $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ являются дифференцируемыми в окрестности точки $\bar{x} \in \mathbb{R}^n$, и их производные являются локально липшицевыми в этой точке. Пусть $\bar{x}$ — стационарная точка задачи (1), удовлетворяющая условию линейной независимости, и пусть сильное достаточное условие второго порядка выполняется для единственного отвечающего этой точке множителя Лагранжа $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$.*

Тогда существуют числа $\bar{\sigma} > 0$ и $\delta > 0$ такие, что для любого начального приближения $(x^0,\, \lambda^0,\, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, достаточно близкого к $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, и любой последовательности

$\{\sigma_k\} \subset (0,\, \bar{\sigma}]$ существует единственная последовательность $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ такая, что для каждого $k = 0,\, 1,\, \ldots$ точка x^{k+1} является стационарной точкой задачи (9), *пара $(\lambda^{k+1},\, \mu^{k+1})$ удовлетворяет соотношениям* (8), *и выполнено неравенство*

$$\|(x^{k+1} - x^k,\, \lambda^{k+1} - \lambda^k,\, \mu^{k+1} - \mu^k)\| \leqslant \delta; \tag{18}$$

эта последовательность сходится к $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, и скорость сходимости линейная. Более того, скорость сходимости является сверхлинейной, если $\sigma_k \to 0$, и квадратичной, если $\sigma_k = O(\|(x^k - \bar{x},\, \lambda^k - \bar{\lambda},\, \mu^k - \bar{\mu})\|)$.

Эта теорема представляет собой первый результат о локальной сходимости метода модифицированных функций Лагранжа без предположения о двукратной дифференцируемости целевой функции и ограничений решаемой задачи.

Однако, следует отметить, что условие линейной независимости и сильное достаточное условие второго порядка являются весьма обременительными. Оказывается, можно получить гораздо более тонкий результат о сходимости метода множителей, воспользовавшись теоремой 2.2.

Прежде всего, отметим, что теорема 2.2 может быть использована для анализа неточной версии метода модифицированных функций, в связи с тем что отображение $\mathcal{A}$ в этой теореме может быть многозначным.

Займемся проверкой предположений теоремы 2.2 для метода модифицированных функций. Пусть отображения Φ и N определены согласно (5) и (6), соответственно. Полуустойчивость решения $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu})$ соответствующего обобщенного уравнения (4) допускает следующую характеризацию, которую легко получить из [58, теорема 8.11].

Предложение 2. *Пусть функция $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ и отображения $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ и $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ дифференцируемы в окрестности точки $\bar{x} \in \mathbb{R}^n$, и их производные локально липшицевы в этой точке. Пусть $\bar{x}$ — стационарная точка задачи* (1), *и пусть $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ — отвечающий ей множитель Лагранжа.*

Тогда точка $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu})$ является полуустойчивым решением уравнения (4) *с отображениями Φ и N, определенными согласно* (5) *и* (6), *соответственно, тогда и только тогда, когда не существует нетривиальной тройки $(\xi,\, \eta,\, \zeta) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, удовлетворяющей системе*

$$\begin{gathered} d + (h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \quad h'(\bar{x})\xi = 0, \quad g'_{A_+(\bar{x},\,\bar{\mu})}(\bar{x})\xi = 0, \\ \zeta_{A_0(\bar{x},\,\bar{\mu})} \geqslant 0, \quad g'_{A_0(\bar{x},\,\bar{\mu})}(\bar{x})\xi \leqslant 0, \quad \zeta_i\langle g_i'(\bar{x}),\, \xi\rangle = 0,\ i \in A_0(\bar{x},\, \bar{\mu}), \\ \zeta_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = 0 \end{gathered} \tag{19}$$

с некоторым вектором $d \in C_x\frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})(\xi)$.

Из предложения 2 следует, что полуустойчивость эквивалентна комбинации из двух свойств: строгого условия Мангасариана–Фромовица, которое состоит в единственности множителя $(\bar{\lambda}, \bar{\mu})$, отвечающего стационарной точке $\bar{x}$, и некритичности множителя, понятие которой было определено в пункте 1.1.3, где также было отмечено, что некритичность следует из достаточного условия второго порядка

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) \quad \langle H\xi, \xi\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{20}$$

где $C(\bar{x})$ — критический конус задачи (1) в точке $\bar{x}$,

$$C(\bar{x}) = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leqslant 0,\ \langle f'(\bar{x}), \xi\rangle \leqslant 0\} = \\ = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0,\ g'_{A_0(\bar{x}, \bar{\mu})}(\bar{x})\xi \leqslant 0\}. \tag{21}$$

В то же время, в отличие от дважды дифференцируемого случая, достаточное условие второго порядка не является необходимым для полуустойчивости решения системы ККТ $(\bar{x}, \bar{\lambda}, \bar{\mu})$, даже если стационарная точка $\bar{x}$ является локальным решением задачи (1) (это демонстрируется примером из замечания 1.5).

Предположим, что значение параметра неточности τ_k вычисляется по текущему приближению к решению системы ККТ, т. е.

$$\tau_k = \tau(x^k, \lambda^k, \mu^k), \tag{22}$$

где $\tau\colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \mapsto \mathbb{R}_+$ — некоторая функция. Тогда итерацию метода модифицированных функций, задаваемую соотношениями (7), (8), можно записать в виде (2.2) с $\mathcal{A}\colon (\mathbb{R}_+ \setminus \{0\}) \times \mathbb{R}^\nu \times \mathbb{R}^\nu \mapsto 2^{\mathbb{R}^\nu}$,

$$\mathcal{A}(\sigma, \tilde{u}, u) = \left(\frac{\partial L}{\partial x}(x, \lambda, \mu) + B(0, \tau(\tilde{x}, \tilde{\lambda}, \tilde{\mu})),\ h(x) - \sigma(\lambda - \tilde{\lambda}),\ -g(x) + \sigma(\mu - \tilde{\mu})\right). \tag{23}$$

Предположим также, что функция τ удовлетворяет условию

$$\tau(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \leqslant \theta(\tilde{x}, \tilde{\lambda}, \tilde{\mu})\|(\tilde{x} - \bar{x}, \tilde{\lambda} - \bar{\lambda}, \tilde{\mu} - \bar{\mu}))\| \quad \forall\, (\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m,$$

где $\theta\colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \mapsto \mathbb{R}_+$ есть некоторая функция такая, что $\theta(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \to 0$ при $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$. В этом случае для отображения $\mathcal{A}$, определяемого согласно (23), предположение 2) теоремы 2.2 выполняется с $\Pi = (0, \bar{\sigma}]$, где $\bar{\sigma} > 0$ — любое достаточно малое число, и с

$$\omega(\sigma, \tilde{u}, u) = \max\{\sigma, \theta(\tilde{x}, \tilde{\lambda}, \tilde{\mu})\}.$$

На практике в качестве функции τ с нужными свойствами можно использовать функции, основанные на невязке системы ККТ (2). Для задач с липшицевыми производными подойдет любая функция τ такая, что

$$\tau(\tilde{x},\, \tilde{\lambda},\, \tilde{\mu}) = o\left(\left\|\left(\frac{\partial L}{\partial x}(\tilde{x},\, \tilde{\lambda},\, \tilde{\mu}),\, h(\tilde{x}),\, \min\{\tilde{\mu},\, -g(\tilde{x})\}\right)\right\|\right). \tag{24}$$

Проверить предположение 3) теоремы 2.2 для метода модифицированных функций гораздо сложнее. Начнем с доказательства нескольких вспомогательных утверждений.

Предложение 3. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображения* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *и* $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ *непрерывны вблизи точки* $\bar{x} \in \mathbb{R}^n$*, и пусть* $\bar{x}$ *является строгим локальным решением задачи* (1).

Тогда существует число $\bar{\delta} > 0$ *такое, что для любого* $\delta \in (0,\, \bar{\delta}]$ *и любого* $M > 0$ *найдется число* $\sigma(\delta,\, M) > 0$ *такое, что для любого* $\sigma \in (0,\, \sigma(\delta,\, M)]$ *и любой пары* $(\tilde{\lambda},\, \tilde{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$*, удовлетворяющей неравенству* $\|(\tilde{\lambda},\, \tilde{\mu})\| \leqslant M$*, всякое глобальное решение* x *задачи*

$$L_\sigma(x,\, \tilde{\lambda},\, \tilde{\mu}) \to \min, \quad x \in B(\bar{x},\, \bar{\delta}), \tag{25}$$

удовлетворяет оценке $\|x - \bar{x}\| \leqslant \delta$.

Это предложение следует из более общих результатов в [86, пример 1.21]. Для задач с ограничениями-равенствами оно было доказано в [14, теорема 2.2].

Предложение 4. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображения* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *и* $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ *дифференцируемы в окрестности точки* $\bar{x} \in \mathbb{R}^n$*, и их производные непрерывны в этой точке. Пусть* $\bar{x}$ *— стационарная точка задачи* (1)*, удовлетворяющая строгому условию Мангасариана–Фромовица, и пусть при этом* $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ *— единственный отвечающий ей множитель Лагранжа.*

Тогда для любого $\varepsilon > 0$ *найдется число* $\delta(\varepsilon) > 0$ *такое, что для всех* $\sigma > 0$ *и всех* $(\tilde{\lambda},\, \tilde{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ *таких, что* $\sigma|\tilde{\mu}_i| \leqslant \delta(\varepsilon)$ *для каждого* $i \in \{1,\, \ldots,\, m\} \setminus A(\bar{x})$*, для всякой стационарной точки* x *задачи*

$$L_\sigma(x,\, \tilde{\lambda},\, \tilde{\mu}) \to \min, \quad x \in \mathbb{R}^n, \tag{26}$$

такой, что $\|x - \bar{x}\| \leqslant \delta(\varepsilon)$*, векторы*

$$\lambda = \tilde{\lambda} + h(x)/\sigma, \quad \mu = \max\{0,\, \tilde{\mu} + g(x)/\sigma\} \tag{27}$$

удовлетворяют неравенству $\|(\lambda - \bar{\lambda},\, \mu - \bar{\mu})\| \leqslant \varepsilon$.

Доказательство. Предположим, что для некоторых последовательностей $\{\sigma_k\} \subset \mathbb{R}_+ \setminus \{0\}$,

$\{x^k\} \subset \mathbb{R}^n$, $\{(\tilde{\lambda}^k,\, \tilde{\mu}^k)\} \subset \mathbb{R}^l \times \mathbb{R}^m$ выполнено следующее: для каждого k точка x^k является стационарной точкой задачи (26) при $\sigma = \sigma_k$ и $(\tilde{\lambda},\, \tilde{\mu}) = (\tilde{\lambda}^k,\, \tilde{\mu}^k)$, и

$$\{x^k\} \to \bar{x}, \quad \left\{\sigma_k \tilde{\mu}_i^k\right\} \to 0 \quad \text{при } k \to \infty \quad \forall\, i \in \{1,\, \ldots,\, m\} \setminus A(\bar{x}). \tag{28}$$

Рассмотрим последовательность $\{(\lambda^k,\, \mu^k)\}$, задаваемую соотношениями (27) при $\sigma = \sigma_k$, $x = x^k$ и $(\tilde{\lambda},\, \tilde{\mu}) = (\tilde{\lambda}^k,\, \tilde{\mu}^k)$. Аналогично (10) имеем: для всех k выполняется равенство

$$\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) = 0, \tag{29}$$

и следовательно, в силу сходимости последовательности $\{x^k\}$ к $\bar{x}$ всякая предельная точка $(\hat{\lambda},\, \hat{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ последовательности $\{(\lambda^k,\, \mu^k)\}$ удовлетворяет равенству

$$\frac{\partial L}{\partial x}(\bar{x},\, \hat{\lambda},\, \hat{\mu}) = 0.$$

Из (27) напрямую следует, что

$$\mu^k \geqslant 0 \quad \forall\, k. \tag{30}$$

Значит, для любого $i \in \{1,\, \ldots,\, m\} \setminus A(\bar{x})$, поскольку $g_i(\bar{x}) < 0$, из (27) и (28) следует:

$$\mu_i^k = \max\{0,\, \tilde{\mu}_i^k + g_i(x^k)/\sigma_k\} = \frac{1}{\sigma_k} \max\left\{0,\, \sigma_k \tilde{\mu}_i^k + g_i(x^k)\right\} = 0 \tag{31}$$

при всех достаточно больших k. Пользуясь соотношениями (30) и (31), заключаем, что $\hat{\mu} \geqslant 0$, и что $\hat{\mu}_i = 0$ для всех $i \in \{1,\, \ldots,\, m\} \setminus A(\bar{x})$. Следовательно, пара $(\hat{\lambda},\, \hat{\mu})$ является множителем Лагранжа, отвечающим точке $\bar{x}$, а значит, в силу строгого условия Мангасариана–Фромовица, $(\hat{\lambda},\, \hat{\mu}) = (\bar{\lambda},\, \bar{\mu})$.

Остается доказать, что последовательность $\{(\lambda^k,\, \mu^k)\}$ ограничена: тогда из проведенного рассуждения будет вытекать, что эта последовательность сходится к $(\bar{\lambda},\, \bar{\mu})$. Предположим, что это не так. Тогда без ограничения общности можно считать, что $t_k = \|(\lambda^k,\, \mu^k)\| \to \infty$ при $k \to \infty$, и что последовательность $\{(\lambda^k,\, \mu^k)/t_k\}$ сходится к некоторому вектору $(\eta,\, \zeta) \in \mathbb{R}^l \times \mathbb{R}^m$, $\|(\eta,\, \zeta)\| = 1$. Из (29) получаем, что при всех достаточно больших k имеет место равенство

$$\frac{1}{t_k} f'(x^k) + (h'(x^k))^{\mathrm{T}} \frac{\lambda^k}{t_k} + (g'(x^k))^{\mathrm{T}} \frac{\mu^k}{t_k} = 0.$$

Тогда переход к пределу при $k \to \infty$ дает соотношение

$$(h'(\bar{x}))^{\mathrm{T}} \eta + (g'(\bar{x}))^{\mathrm{T}} \zeta = 0.$$

Кроме того, из (30) и (31) непосредственно следует, что $\zeta \geqslant 0$, и $\zeta_i = 0$ для всех $i \in \{1,\, \ldots,\, m\} \setminus A(\bar{x})$. Однако существование такого вектора $(\eta,\, \zeta)$ противоречит условию Мангасариана–Фромовица. ■

Лемма 1. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображения* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *и* $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ *дифференцируемы в окрестности точки* $\bar{x} \in \mathbb{R}^n$, *и их производные локально липшицевы в этой точке. Пусть* $\bar{x}$ *— стационарная точка задачи* (1), *и пусть* $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ *— отвечающий ей множитель Лагранжа.*

Тогда для любого направления ξ *из конуса* $C(\bar{x})$, *введенного в* (21), *любых последовательностей* $\{t_k\} \to 0+$ *и* $\{\xi^k\} \to \xi$ *и любой последовательности* $\{(\lambda^k, \mu^k)\} \subset \mathcal{M}(\bar{x})$, *сходящейся к* $(\bar{\lambda}, \bar{\mu})$, *следующие пределы существуют или не существуют одновременно и в случае существования равны:*

$$\lim_{k\to\infty} \frac{L(\bar{x}+t_k\xi, \bar{\lambda}, \bar{\mu}) - L(\bar{x}, \bar{\lambda}, \bar{\mu})}{t_k^2} = \lim_{k\to\infty} \frac{L(\bar{x}+t_k\xi^k, \lambda^k, \mu^k) - L(\bar{x}, \lambda^k, \mu^k)}{t_k^2}.$$

Доказательство. Для любого k справедливо равенство

$$\begin{gathered}(L(\bar{x}+t_k\xi^k, \lambda^k, \mu^k) - L(\bar{x}, \lambda^k, \mu^k)) - (L(\bar{x}+t_k\xi, \bar{\lambda}, \bar{\mu}) - L(\bar{x}, \bar{\lambda}, \bar{\mu})) = \\ = L(\bar{x}+t_k\xi^k, \lambda^k, \mu^k) - L(\bar{x}+t_k\xi, \lambda^k, \mu^k) + L(\bar{x}+t_k\xi, \lambda^k, \mu^k) - L(\bar{x}+t_k\xi, \bar{\lambda}, \bar{\mu}),\end{gathered} \tag{32}$$

поскольку $L(\bar{x}, \lambda^k, \mu^k) = L(\bar{x}, \bar{\lambda}, \bar{\mu}) = f(\bar{x})$, в следствие того, что пары $(\bar{\lambda}, \bar{\mu})$ и (λ^k, μ^k) принадлежат множеству $\mathcal{M}(\bar{x})$.

По теореме о среднем существует число $\theta_k \in [0, 1]$ такое, что

$$\begin{gathered}L(\bar{x}+t_k\xi^k, \lambda^k, \mu^k) - L(\bar{x}+t_k\xi, \lambda^k, \mu^k) = \\ = L(\bar{x}+t_k\xi^k, \lambda^k, \mu^k) - L(\bar{x}+t_k\xi, \lambda^k, \mu^k) - \left\langle \frac{\partial L}{\partial x}(\bar{x}, \lambda^k, \mu^k), t_k(\xi^k-\xi) \right\rangle = \\ = \left\langle \frac{\partial L}{\partial x}(\bar{x}+t_k(\theta_k\xi^k+(1-\theta_k)\xi), \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k, \mu^k), t_k(\xi^k-\xi) \right\rangle = \\ = O(t_k^2\|\xi^k-\xi\|) = o(t_k^2),\end{gathered} \tag{33}$$

где при выводе предпоследней оценки мы воспользовались локальной липшицевостью производных функции f и отображений h и g в точке $\bar{x}$.

Кроме того, поскольку $\xi \in C(\bar{x})$, используя (21) и теорему о среднем, можно заключить, что

$$\begin{gathered}L(\bar{x}+t_k\xi, \lambda^k, \mu^k) - L(\bar{x}+t_k\xi, \bar{\lambda}, \bar{\mu}) = \langle \lambda^k-\bar{\lambda}, h(\bar{x}+t_k\xi)\rangle + \langle \mu^k-\bar{\mu}, g(\bar{x}+t_k\xi)\rangle = \\ = \langle \lambda^k-\bar{\lambda}, h(\bar{x}+t_k\xi) - h(\bar{x}) - t_kh'(\bar{x})\xi\rangle + \langle \mu^k-\bar{\mu}, g(\bar{x}+t_k\xi) - g(\bar{x}) - t_kg'(\bar{x})\xi\rangle = \\ = \langle \lambda^k-\bar{\lambda}, t_k(h'(\bar{x}+\theta_kt_k\xi) - h'(\bar{x}))\xi\rangle + \langle \mu^k-\bar{\mu}, t_k(g'(\bar{x}+\theta_kt_k\xi) - g'(\bar{x}))\xi\rangle = \\ == O(t_k^2\|\lambda^k-\bar{\lambda}\|) + O(t_k^2\|\mu^k-\bar{\mu}\|) = o(t_k^2).\end{gathered} \tag{34}$$

Из (33) и (34) получаем, что выражение в левой части (32) представляет собой $o(t_k^2)$, откуда следует требуемое. ∎

Следующий результат представляет собой далеко идущее обобщение леммы Финслера–Дебре [25, 35].

Лемма 2. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображения* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *и* $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ *дифференцируемы в окрестности точки* $\bar{x} \in \mathbb{R}^n$*, и их производные локально липшицевы в этой точке. Пусть* $\bar{x}$ *— стационарная точка задачи* (1)*, и пусть* $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ *— отвечающий ей множитель Лагранжа такой, что выполнено условие*

$$\liminf_{t \to 0+} \frac{L(\bar{x} + t\xi,\, \bar{\lambda},\, \bar{\mu}) - L(\bar{x},\, \bar{\lambda},\, \bar{\mu})}{t^2/2} > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}. \tag{35}$$

Тогда найдутся $\bar{\sigma} > 0$*,* $\bar{\delta} > 0$*,* $\bar{\gamma} > 0$ *и* $\bar{\varepsilon} > 0$ *такие, что для всех* $\sigma \in (0, \bar{\sigma}]$

$$L(x,\, \lambda,\, \mu) - L(\bar{x},\, \lambda,\, \mu) + \frac{1}{2\sigma}\|h(x)\|_2^2 + \\ + \frac{1}{2\sigma} \sum_{i \in A_+(\bar{x},\, \bar{\mu})} (g_i(x))^2 + \frac{1}{2\sigma} \sum_{i \in A_0(\bar{x},\, \bar{\mu})} (\max\{0,\, g_i(x)\})^2 \geqslant \bar{\gamma}\|x - \bar{x}\|^2$$

для всех $x \in B(\bar{x}, \bar{\delta})$ *и всех* $(\lambda, \mu) \in \mathcal{M}(\bar{x})$ *таких, что* $\|(\lambda - \bar{\lambda},\, \mu - \bar{\mu})\| \leqslant \bar{\varepsilon}$*.*

Доказательство. От противного, предположим, что существуют последовательность чисел $\{\sigma_k\} \subset \mathbb{R}_+ \setminus \{0\}$, стремящаяся к нулю, последовательность точек $\{x^k\} \subset \mathbb{R}^n \setminus \{\bar{x}\}$, сходящаяся к $\bar{x}$, и последовательность отвечающих точке $\bar{x}$ множителей Лагранжа $\{(\lambda^k,\, \mu^k)\} \subset \mathcal{M}(\bar{x})$, сходящаяся к множителю $(\bar{\lambda},\, \bar{\mu})$, такие, что

$$L(x^k,\, \lambda^k,\, \mu^k) - L(\bar{x},\, \lambda^k,\, \mu^k) + \frac{1}{2\sigma_k}\|h(x^k)\|_2^2 + \\ + \frac{1}{2\sigma_k} \sum_{i \in A_+(\bar{x},\, \bar{\mu})} (g_i(x^k))^2 + \frac{1}{2\sigma_k} \sum_{i \in A_0(\bar{x},\, \bar{\mu})} \big(\max\{0,\, g_i(x^k)\}\big)^2 \leqslant o(\|x^k - \bar{x}\|^2). \tag{36}$$

Для каждого k положим $t_k = \|x^k - \bar{x}\|$ и $\xi^k = (x^k - \bar{x})/t_k$. Без ограничения общности можно считать, что последовательность $\{\xi^k\}$ сходится к некоторому вектору $\xi \in \mathbb{R}^n \setminus \{0\}$. Разделив (36) на t_k^2/σ_k и перейдя к пределу при $k \to \infty$, получим, что

$$\|h'(\bar{x})\xi\|^2 + \|g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})\xi\|^2 + \|\max\{0,\, g'_{A_0(\bar{x},\, \bar{\mu})}(\bar{x})\xi\}\|^2 = 0,$$

а значит, $\xi \in C(\bar{x}) \setminus \{0\}$ (см. (21)).

С другой стороны, поскольку последние три слагаемых в левой части формулы (36) неотрицательны,

$$L(\bar{x} + t_k\xi^k,\, \lambda^k,\, \mu^k) - L(\bar{x},\, \lambda^k,\, \mu^k) \leqslant o(t_k^2).$$

Разделив это соотношение на t_k^2 и воспользовавшись леммой 1, приходим к противоречию с условием (35). ∎

Теперь докажем следующий факт.

Предложение 5. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображения* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *and* $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ *дифференцируемы в окрестности точки* $\bar{x} \in \mathbb{R}^n$*, и их производные локально липшицевы в этой точке* $\bar{x}$*. Пусть* $\bar{x}$ *— стационарная точка задачи* (1)*, и пусть* $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ *— отвечающий ей множитель Лагранжа, удовлетворяющий условию* (35)*.*

Тогда найдутся числа $\bar{\sigma} > 0$ *и* $\bar{\delta} > 0$ *такие, что выполнено следующее:*

а) *существуют числа* $\bar{\gamma} > 0$ *и* $\bar{\varepsilon} > 0$ *такие, что для любого* $\sigma \in (0,\, \bar{\sigma}]$

$$\begin{gathered} L_\sigma(x,\, \lambda,\, \mu) \geqslant L_\sigma(\bar{x},\, \lambda,\, \mu) + \bar{\gamma}\|x - \bar{x}\|^2 \\ \forall\, x \in B(\bar{x},\, \bar{\delta}),\ \forall\, (\lambda,\, \mu) \in \mathcal{M}(\bar{x})\colon \|(\lambda - \bar{\lambda},\, \mu - \bar{\mu})\| \leqslant \bar{\varepsilon}; \end{gathered} \tag{37}$$

б) *для любого* $\delta \in (0,\, \bar{\delta}]$ *существует* $\varepsilon(\delta) > 0$ *такое, что для любого* $\sigma \in (0,\, \bar{\sigma}]$ *и любой пары* $(\tilde{\lambda},\, \tilde{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$*, удовлетворяющей неравенству* $\|(\tilde{\lambda} - \bar{\lambda},\, \tilde{\mu} - \bar{\mu})\| \leqslant \varepsilon(\delta)$*, для любого глобального решения* x *задачи* (25) *выполнено:* $\|x - \bar{x}\| \leqslant \delta$*.*

Доказательство. Сперва докажем часть а). Формулировка и доказательство этой части являются обобщениями таковых из [33, предложение 3.1] на случай задач с липшицевыми производными.

Для любого фиксированного $\sigma > 0$, любого $x \in \mathbb{R}^n$ и любой пары $(\lambda,\, \mu) \in \mathbb{R}^l \times \mathbb{R}^m$ справедливо равенство

$$\begin{multlined} L_\sigma(x,\, \lambda,\, \mu) = L(x,\, \lambda,\, \mu) + \frac{\sigma}{2}\|\lambda\|_2^2 + \frac{1}{2\sigma}\|h(x)\|_2^2 + \\ + \frac{\sigma}{2}\|\max\{0,\, \mu + g(x)/\sigma\}\|_2^2 - \langle \mu,\, g(x)\rangle. \end{multlined} \tag{38}$$

Если точка $x \in \mathbb{R}^n$ достаточно близка к $\bar{x}$, и если пара $(\lambda,\, \mu)$ достаточно близка к $(\bar{\lambda},\, \bar{\mu})$, для любого $i \in A_+(\bar{x},\, \bar{\mu})$ имеем: $\sigma\mu_i > -g_i(x)$, и

$$\frac{\sigma}{2}\left(\max\{0,\, \mu_i + g_i(x)/\sigma\}\right)^2 - \mu_i g_i(x) = \frac{\sigma}{2}\mu_i^2 + \frac{1}{2\sigma}(g_i(x))^2, \tag{39}$$

Это соотношение также выполняется для всякого $i \in A_0(\bar{x}, \bar{\mu})$ такого, что $\sigma\mu_i \geqslant -g_i(x)$, откуда следует, что

$$\frac{\sigma}{2}\left(\max\{0, \mu_i + g_i(x)/\sigma\}\right)^2 - \mu_i g_i(x) \geqslant \frac{\sigma}{2}\mu_i^2 + \frac{1}{2\sigma}\left(\max\{0, g_i(x)\}\right)^2. \tag{40}$$

С другой стороны, если $\mu \geqslant 0$ для любого $i \in A_0(\bar{x}, \bar{\mu})$ такого, что $\sigma\mu_i < -g_i(x)$, выполнено неравенство $g_i(x) < 0$, и значит,

$$\frac{\sigma}{2}\left(\max\{0, \mu_i + g_i(x)/\sigma\}\right)^2 - \mu_i g_i(x) = -\mu_i g_i(x) \geqslant \frac{\sigma}{2}\mu_i^2 = \frac{\sigma}{2}\mu_i^2 + \frac{1}{2\sigma}\left(\max\{0, g_i(x)\}\right)^2. \tag{41}$$

Наконец, для $i \in \{1, \ldots, m\} \setminus A(\bar{x})$ в том случае, если $(\lambda, \mu) \in \mathcal{M}(\bar{x})$, имеет место равенство $\mu_i = 0$, и следовательно, для любого x, достаточно близкого к $\bar{x}$,

$$\frac{\sigma}{2}\left(\max\{0, \mu_i + g_i(x)/\sigma\}\right)^2 - \mu_i g_i(x) = 0. \tag{42}$$

В то же время, для любого $i \in A(\bar{x})$ и любого $\mu \geqslant 0$ верно равенство

$$\frac{\sigma}{2}\left(\max\{0, \mu_i + g_i(\bar{x})/\sigma\}\right)^2 - \mu_i g_i(\bar{x}) = \frac{\sigma}{2}\mu_i^2. \tag{43}$$

Из (38)–(43) получаем, что при любом фиксированном $\sigma > 0$ для всех $x \in \mathbb{R}^n$, достаточно близких к $\bar{x}$, и всех $(\lambda, \mu) \in \mathcal{M}(\bar{x})$, достаточно близких к $(\bar{\lambda}, \bar{\mu})$, выполнена цепочка соотношений

$$\begin{aligned}
L_\sigma(x, \lambda, \mu) - L_\sigma(\bar{x}, \lambda, \mu) = L(x, \lambda, \mu) - L(\bar{x}, \lambda, \mu) + \frac{1}{2\sigma}\|h(x)\|_2^2 +& \\
+ \frac{\sigma}{2}\|\max\{0, \mu + g(x)/\sigma\}\|_2^2 - \langle \mu, g(x)\rangle -& \\
- \left(\frac{1}{2\sigma}\|h(\bar{x})\|_2^2 + \frac{\sigma}{2}\|\max\{0, \mu + g(\bar{x})/\sigma\}\|_2^2 - \langle \mu, g(\bar{x})\rangle\right) \geqslant& \\
\geqslant L(x, \lambda, \mu) - L(\bar{x}, \lambda, \mu) + \frac{1}{2\sigma}\|h(x)\|_2^2& \\
+ \frac{1}{2\sigma}\sum_{i \in A_+(\bar{x}, \bar{\mu})}(g_i(x))^2 + \frac{1}{2\sigma}\sum_{i \in A_0(\bar{x}, \bar{\mu})}\left(\max\{0, g_i(x)\}\right)^2.&
\end{aligned} \tag{44}$$

Тогда, объединяя (44) с леммой 2, получаем существование чисел $\bar{\sigma} > 0$, $\bar{\delta} > 0$ и $\bar{\varepsilon} > 0$ таких, что

$$L_{\bar{\sigma}}(x, \lambda, \mu) \geqslant L_{\bar{\sigma}}(\bar{x}, \lambda, \mu) + \bar{\gamma}\|x - \bar{x}\|^2$$
$$\forall\, x \in B(\bar{x}, \bar{\delta}),\ \forall\,(\lambda, \mu) \in \mathcal{M}(\bar{x})\colon \|(\lambda - \bar{\lambda}, \mu - \bar{\mu})\| \leqslant \bar{\varepsilon}.$$

Выполнение соотношения (37) для $\sigma \in (0, \bar{\sigma}]$ следует из того, что при фиксированных $x \in \mathbb{R}^n$ и $(\lambda, \mu) \in \mathcal{M}(\bar{x})$ разность $L_\sigma(x, \lambda, \mu) - L_\sigma(\bar{x}, \lambda, \mu)$ не возрастает по $\sigma > 0$, в чем легко убедиться непосредственной проверкой.

Теперь докажем часть б). Прежде всего, заметим, что из свойства, выполнение которого было доказано в части а), следует, что $\bar{x}$ — строгое локальное решение задачи (1), а значит, выполнены условия предложения 3. Пусть $\bar{\sigma} > 0$ и $\bar{\delta} > 0$ выбраны согласно части a) и таким образом, что с $\bar{\delta}$ выполняется утверждение предложения 3.

Предположим, что найдется $\delta_0 \in (0, \bar{\delta}]$ и последовательности $\{\sigma_k\} \subset \mathbb{R}_+$, $\{x^k\} \subset \mathbb{R}^n$ и $\{(\tilde{\lambda}^k, \tilde{\mu}^k)\} \to (\bar{\lambda}, \bar{\mu})$ такие, что при каждом k выполняется неравенство $\sigma_k \leqslant \bar{\sigma}$, x^k является решением задачи (25) при $\sigma = \sigma_k$ и $(\lambda, \mu) = (\tilde{\lambda}^k, \tilde{\mu}^k)$, и

$$\|x^k - \bar{x}\| > \delta_0. \tag{45}$$

Если последовательность $\{\sigma_k\}$ сходится к нулю, то (45) противоречит предложению 3. Таким образом, без ограничения общности можно считать, что последовательность $\{\sigma_k\}$ сходится к некоторому числу $\hat{\sigma} \in (0, \bar{\sigma}]$.

Будем рассматривать задачу (25) как параметрическую, в которой σ и $(\tilde{\lambda}, \tilde{\mu})$ играют роль параметров с базовыми значениями $\hat{\sigma}$ и $(\bar{\lambda}, \bar{\mu})$, соответственно. В силу выбора $\bar{\sigma} > 0$ и $\bar{\delta} > 0$, точка $\bar{x}$ является строгим (глобальным) решением задачи (25) при этих базовых значениях параметра. Тогда, пользуясь известными фактами об устойчивости строгого локального решения задачи оптимизации (см., например, [12, теоерма 3.1]), вновь приходим к противоречию с (45). ∎

На самом деле можно показать, что условия (35) и (37) эквивалентны.

Замечание 2. Выражение в левой части формулы (35) представляет собой нижнюю вторую производную отображения $L(\cdot, \bar{\lambda}, \bar{\mu})$ в точке $\bar{x}$ по направлению ξ (напомним, что $\frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0$). Условие (35) следует из достаточного условия второго порядка (20). В самом деле, в силу компактности множества $\partial_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})$, достаточное условие второго порядка (20) влечет за собой существование числа $\gamma > 0$ такого, что

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) \quad \langle H\xi, \xi\rangle \geqslant \gamma\|\xi\|^2 \quad \forall\, \xi \in C(\bar{x}).$$

Рассмотрим произвольный вектор $\xi \in C(\bar{x}) \setminus \{0\}$ и произвольную последовательность $\{t_k\} \to 0+$. Применяя соответствующую теорему о среднем (см., например, [41, теорема 2.3]), получаем:

$$\frac{L(\bar{x} + t_k\xi, \bar{\lambda}, \bar{\mu}) - L(\bar{x}, \bar{\lambda}, \bar{\mu})}{t_k^2/2} = \langle H_k\xi, \xi\rangle, \tag{46}$$

где $H_k \in \partial_x \frac{\partial L}{\partial x}(\bar{x} + \tilde{t}_k\xi, \bar{\lambda}, \bar{\mu})$ при некотором $\tilde{t}_k \in (0, t_k)$. Т. к. дифференциал Кларка отображения в точке ограничен константой, с которой данное отображение является в этой точке локально липшицевым, последовательность $\{H_k\}$ ограничена, и при этом, в силу полунепрерывности дифференциала Кларка сверху, любая ее предельная точка лежит во множестве

$\partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})$. Следовательно, последовательность величин в левой части равенства (46) ограничена, и любая ее предельная точка больше либо равна $\gamma\|\xi\|^2$. Значит, выполнено условие (35).

Как видно из примера 1 ниже, обратная импликация не имеет места.

Замечание 3. Как легко видеть, условие (35) эквивалентно следующему условию квадратичного роста функции Лагранжа относительно критического конуса: существуют числа $\gamma > 0$ и $\rho > 0$ такие, что

$$L(\bar{x} + \xi,\, \bar{\lambda},\, \bar{\mu}) \geqslant L(\bar{x},\, \bar{\lambda},\, \bar{\mu}) + \gamma\|\xi\|^2 \quad \forall\, \xi \in C(\bar{x}) \cap B(0,\, \rho).$$

Более того, при выполнении строгого условия Мангасариана–Фромовица, условие (35) следует из обычного условия квадратичного роста для задачи (1) в точке $\bar{x}$, состоящего в существовании чисел $\gamma > 0$ и $\rho > 0$ таких, что

$$f(x) \geqslant f(\bar{x}) + \gamma\|x - \bar{x}\|^2 \quad \forall\, x \in B(\bar{x},\, \rho)\colon h(x) = 0,\, g(x) \leqslant 0. \tag{47}$$

Чтобы убедиться в этом, рассмотрим произвольный вектор $\xi \in C(\bar{x}) \setminus \{0\}$ и произвольную последовательность $\{t_k\} \to 0+$. Вспомним, что строгое условие Мангасариана–Фромовица эквивалентно выполнению обычного условия Мангасариана–Фромовица в точке $\bar{x}$ для системы ограничений

$$h(x) = 0, \quad g_{A_+(\bar{x},\, \bar{\mu})}(x) = 0, \quad g_{A_0(\bar{x},\, \bar{\mu})}(x) \leqslant 0.$$

Следовательно, в силу (21), вектор ξ является (внутренним) касательным вектором к множеству, задаваемому этой системой ограничений в точке $\bar{x}$ (см., к примеру, [19, следствие 2.91]), и, в частности, найдется последовательность $\{\xi^k\} \subset \mathbb{R}^n$, сходящаяся к ξ и такая, что при каждом k выполняются соотношения

$$h(\bar{x} + t_k\xi^k) = 0, \quad g_{A_+(\bar{x},\, \bar{\mu})}(\bar{x} + t_k\xi^k) = 0, \quad g_{A_0(\bar{x},\, \bar{\mu})}(\bar{x} + t_k\xi^k) \leqslant 0.$$

Так как $\bar{\mu}_i = 0$ для всех $i \in \{1,\, \ldots,\, m\} \setminus A_+(\bar{x},\, \bar{\mu})$, с учетом (47) для всех достаточно больших k имеем:

$$L(\bar{x} + t_k\xi^k,\, \bar{\lambda},\, \bar{\mu}) - L(\bar{x},\, \bar{\lambda},\, \bar{\mu}) = f(\bar{x} + t_k\xi^k) - f(\bar{x}) \geqslant \gamma\|\xi^k\|^2 t_k^2.$$

Применяя лемму 1, получаем (35).

Согласно [59, теорема 1], достаточное условие второго порядка (20) влечет за собой условие квадратичного роста (47) с некоторыми $\gamma > 0$ и $\rho > 0$. Заметим, однако, что, в отличие от дважды дифференцируемого случая (см., например, [19, теорема 3.70]), обратная импликация не имеет места даже для задач с ограничениями-равенствами, даже если в рассматриваемой стационарной точке выполняется условие регулярности ограничений, и даже при некритичности отвечающего этой точке множителя Лагранжа. Это демонстрирует следующий пример.

Пример 1. Пусть $n = 2$, $l = 1$, $m = 0$, $f(x) = (x_1 + x_2 - \sqrt{x_1^2 + x_2^2})^2/2$, и $h(x) = x_1 + x_2$. Тогда точка $\bar{x} = (0,\, 0)$ является единственным решением задачи (1), а $\bar{\lambda} = 0$ — единственным отвечающим этой точке множителем Лагранжа, который, как легко проверить, является некритическим.

Рассмотрим последовательность $\{x^k\} \subset \mathbb{R}^2$ такую, что $\{x^k\} \to 0$, и $x_1^k = x_2^k > 0$ при всех k. Отображение $\frac{\partial L}{\partial x}(\cdot,\, \bar{\lambda})$ является дифференцируемым во всех точках такой последовательности, и

$$\frac{\partial^2 L}{\partial x^2}(x^k,\, \bar{\lambda}) = \frac{\sqrt{2}-1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{\sqrt{2}-1}{2}\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Следовательно, матрица в правой части последнего равенства (обозначим ее через H) принадлежит множеству $\partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda})$. В то же время, для $\xi = (1,\, -1) \in \ker h'(\bar{x})$ имеем:

$$\langle H\xi,\, \xi \rangle = -2(\sqrt{2}-1) < 0.$$

Таким образом, достаточное условие второго порядка (20) не выполняется, несмотря на очевидное выполнение условия квадратичного роста.

Соединяя часть б) предложения 5 с предложением 4 и принимая во внимание замечание 3, видим, что точный (а значит, и неточный) метод модифицированных функций удовлетворяет предположению 3) теоремы 2.2 при выполнении строгого условия Мангасариана–Фромовица и условия квадратичного роста.

Теорема 2. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображения* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *и* $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ *дифференцируемы в окрестности точки* $\bar{x} \in \mathbb{R}^n$, *и их производные локально липшицевы в этой точке. Пусть* $\bar{x}$ *— стационарная точка задачи* (1), *удовлетворяющая строгому условию Мангасариана–Фромовица, и пусть единственный отвечающий этой точке множитель Лагранжа* $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ *является некритическим. Пусть также существуют числа* $\gamma > 0$ *и* $\rho > 0$ *такие, что выполнено условие* (47). *Наконец, пусть функция* $\tau\colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \mapsto \mathbb{R}_+$ *удовлетворяет оценке* $\tau(x,\, \lambda,\, \mu) = o(\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu}))\|)$.

Тогда найдется $\bar{\sigma} > 0$ *и* $\delta > 0$ *такие, что для любого начального приближения* $(x^0,\, \lambda^0,\, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, *достаточно близкого к* $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, *и любой последовательности* $\{\sigma_k\} \subset (0,\, \bar{\sigma}]$ *существует последовательность* $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, *удовлетворяющая при каждом* $k = 0,\, 1,\, \ldots$ *условию* (7) *с* τ_k, *определяемым согласно* (22), *и условиям* (8) и (18); *любая такая последовательность сходится к* $(\bar{x}, \bar{\lambda}, \bar{\mu})$*, и скорость сходимости является линейной. Более того, скорость сходимости является сверхлинейной, если* $\sigma_k \to 0+$*. Если же* $\tau_k = O(\|(x^k - \bar{x}, \lambda^k - \bar{\lambda}, \mu^k - \bar{\mu})\|^2)$*, а* $\sigma_k = O(\|(x^k - \bar{x}, \lambda^k - \bar{\lambda}, \mu^k - \bar{\mu})\|)$*, то скорость сходимости квадратичная.*

Для задач с ограничениями-равенствами, целевая функция и ограничения которых дважды непрерывно дифференцируемы, полуустойчивость эквивалентна сильной метрической регулярности, которая в свою очередь эквивалентна комбинации из условия регулярности ограничений и некритичности множителя. Поэтому для таких задач последняя теорема не дает ничего нового по сравнению с теоремой 2.1, за исключением возможности решать подзадачи неточно. Однако, в отсутствие двукратной дифференцируемости, даже для задач с ограничениями-равенствами полуустойчивость, вообще говоря, слабее сильной метрической регулярности. Действительно, для всякого $t > 0$, система Лагранжа задачи из примера 1 с возмущенным ограничением $x_1 + x_2 = t$ имеет два решения $(x, \lambda) = ((t, 0), 0)$ и $(x, \lambda) = ((0, t), 0)$, и как следствие, решение $(\bar{x}, \bar{\lambda})$ системы Лагранжа невозмущенной задачи не может быть сильно метрически регулярным. Таким образом, теорема 2.2 содержательна и в случае задач с ограничениями-равенствами.

Наиболее тонкий результат о локальной сходимости метода модифицированных функций Лагранжа для задач, целевая функция и ограничения которых являются дважды непрерывно дифференцируемыми, был получен в [33]. А именно, в этой работе было доказано, что метод модифицированных функций обладает локальной сходимостью при выполнении одного лишь достаточного условия второго порядка. В оставшейся части данного пункта мы распространим этот результат на задачи с липшицевыми производными.

Заметим, что, согласно следствию 1.1, предположение 1) теоремы 2.3 с $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ и с заменой множества $\bar{U}$ его подмножеством $\{\bar{x}\} \times \mathcal{M}(\bar{x})$ эквивалентно некритичности множителя $(\bar{\lambda}, \bar{\mu})$. В свою очередь, как уже отмечалось, некритичность множителя следует из достаточного условия второго порядка (20).

Как и прежде, будем предполагать, что значение параметра неточности τ_k в неточном методе модифицированных функций выбирается в соответствии с (22). При этом предположим, что функция τ удовлетворяет условию

$$\tau(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \leqslant \theta(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \operatorname{dist}((\tilde{x}, \tilde{\lambda}, \tilde{\mu}), \{\bar{x}\} \times \mathcal{M}(\bar{x})) \quad \forall\, (\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$$

с некоторой функцией $\theta\colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \mapsto \mathbb{R}_+$ такой, что $\theta(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \to 0$ при $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$. Тогда для любого $c > 0$ предположение 2) теоремы 2.3 для метода модифицированных функций выполняется с $\Pi = (0, \bar{\sigma}]$ и $\omega(\sigma, \tilde{u}, u) = c\sigma + \theta(\tilde{x}, \tilde{\lambda}, \tilde{\mu})$, где число $\bar{\sigma} > 0$

достаточно мало. Заметим, что, как и раньше, функцию τ с нужными свойствами можно выбирать согласно (24).

Наконец, при выполнении достаточного условия второго порядка, используя часть a) предложения 5, замечание 2, и рассуждая аналогично [33, следствие 3.2, предложение 3.3], несложно проверить, что предположение 3) теоремы 2.3 (с определенным выше Π) выполняется для любого $c > 0$ и любого достаточно малого $\bar{\sigma} > 0$.

Теорема 3. *Пусть функция $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ и отображения $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ и $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ дифференцируемы в окрестности точки $\bar{x} \in \mathbb{R}^n$, и их производные локально липшицевы в этой точке $\bar{x}$. Пусть $\bar{x}$ — стационарная точка задачи (1), а $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ — отвечающий ей множитель Лагранжа. Пусть функция $\tau\colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \mapsto \mathbb{R}_+$ удовлетворяет оценке*

$$\tau(x, \lambda, \mu) = o(\operatorname{dist}((x, \lambda, \mu), \{\bar{x}\} \times \mathcal{M}(\bar{x}))). \tag{48}$$

Тогда для любого $c > 0$ существует число $\bar{\sigma} > 0$ такое, что для любого начального приближения $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, достаточно близкого к $(\bar{x}, \bar{\lambda}, \bar{\mu})$, и любой последовательности $\{\sigma_k\} \subset (0, \bar{\sigma}]$ существует последовательность $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, удовлетворяющая при всех $k = 0, 1, \ldots$ условию (7) с τ_k, определяемым согласно (22), условиям (8) и

$$\|(x^{k+1} - x^k, \lambda^{k+1} - \lambda^k, \mu^{k+1} - \mu^k)\| \leqslant c(\|x^k - \bar{x}\| + \operatorname{dist}((\lambda^k, \mu^k), \mathcal{M}(\bar{x}))); \tag{49}$$

любая такая последовательность сходится к $(\bar{x}, \lambda^*, \mu^*)$, где $(\lambda^*, \mu^*) \in \mathcal{M}(\bar{x})$, и скорости сходимости последовательности $\{(x^k, \lambda^k, \mu^k)\}$ к $(\bar{x}, \lambda^*, \mu^*)$ и последовательности $\{\|x^k - \bar{x}\| + \operatorname{dist}((\lambda^k, \mu^k), \mathcal{M}(\bar{x}))\}$ к нулю линейные. Более того, они являются сверхлинейными, если $\sigma_k \to 0+$, и квадратичными, если $\tau_k = O((\operatorname{dist}((x^k, \lambda^k, \mu^k), \{\bar{x}\} \times \mathcal{M}(\bar{x})))^2)$, и $\sigma_k = O(\|x^k - \bar{x}\| + \operatorname{dist}((\lambda^k, \mu^k), \mathcal{M}(\bar{x})))$.

Кроме того, для любого $\varepsilon > 0$ справедливо неравенство $\|(\lambda^* - \bar{\lambda}, \mu^* - \bar{\mu})\| < \varepsilon$, если начальное приближение (x^0, λ^0, μ^0) достаточно близко к $(\bar{x}, \bar{\lambda}, \bar{\mu})$.

Отметим, что при выводе теоремы 2.3 мы предполагали, что норма на произведении $\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ задана как сумма норм на пространствах $\mathbb{R}^n$ и $\mathbb{R}^l \times \mathbb{R}^m$. Также следует заметить, что при применении теоремы 2.3 нужно выбирать $\varepsilon_0 > 0$ таким образом, чтобы любая последовательность, генерируемая соответствующей итерационной схемой, не покидала окрестность тройки $(\bar{x}, \bar{\lambda}, \bar{\mu})$, в которой условие локализации (2.16) эквивалентно (49).

Условие локализации (49) в теореме 3 является более сильным (и, в каком-то смысле, «менее практичным»), чем условие (18) в теоремах 1 и 2. Заметим, однако, что согласно следствию 1.1, некритичность множителя $(\bar{\lambda},\, \bar{\mu})$ эквивалентна выполнению оценки расстояния

$$\|x-\bar{x}\| + \mathrm{dist}((\lambda,\, \mu),\, \mathcal{M}(\bar{x})) \leqslant \ell \left\| \begin{pmatrix} \dfrac{\partial L}{\partial x}(x,\, \lambda,\, \mu) \\ h(x) \\ \min\{\mu,\, -g(x)\} \end{pmatrix} \right\| \tag{50}$$

для всех $(x,\, \lambda,\, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, достаточно близких к $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$. Поэтому для контроля параметра σ_k можно использовать вычислимую невязку системы ККТ (2), фигурирующую в правой части формулы (50).

3.1.2. Сходимость при некритичности множителя

В настоящем пункте для задач оптимизации с ограничениями-равенствами будут получены теоремы о локальной сходимости метода модифицированных функций при единственном предположении о близости начального двойственного приближения к строго некритическому множителю Лагранжа (напомним, что строгая некритичность была определена в замечании 1.3). Здесь рассматриваются задачи с липшицевыми производными, но даже в дважды дифференцируемом случае эти теоремы являются первыми результатами о локальной сходимости метода модифицированных функций в предположениях, более слабых, чем достаточное условие второго порядка. В конце пункта будет показано, что для задач со смешанными ограничениями полученные результаты в общем случае не имеют места, но часть из них справедлива при выполнении условия строгой дополнительности (т. е. если $A_0(\bar{x}) = \varnothing$).

Итак, рассмотрим задачу оптимизации (1) с $m = 0$, т. е. задачу

$$f(x) \to \min, \quad h(x) = 0. \tag{51}$$

Функция Лагранжа $L\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}^n$ такой задачи имеет вид

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

а система Каруша–Куна–Таккера совпадает с так называемой *системой Лагранжа*

$$\frac{\partial L}{\partial x}(x,\, \lambda) = 0, \quad h(x) = 0. \tag{52}$$

Таким образом, для задачи (51) множество $\mathcal{M}(\bar{x})$ в любой точке $\bar{x} \in \mathbb{R}^n$ совпадает со множеством всех $\lambda \in \mathbb{R}^l$ таких, что пара $(\bar{x},\, \lambda)$ удовлетворяет системе (52). Заметим также, что система (52) может быть записана в виде

$$\Phi(u) = 0, \tag{53}$$

где $u = (x,\, \lambda)$, а отображение $\Phi\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}^n \times \mathbb{R}^l$ задано по правилу

$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x,\, \lambda),\, h(x) \right). \tag{54}$$

Следует отметить, что такое задание отображения Φ соответствует формуле (5), поэтому мы не говорим о переопределении этого отображения. В свою очередь уравнение (53) эквивалентно обобщенному уравнению (4) с отображением Φ, заданным согласно (54), и отображением $N\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto 2^{\mathbb{R}^n \times \mathbb{R}^l}$, тождественно равным множеству $\{0\}$.

Далее, модифицированная функция Лагранжа для задачи (51) имеет вид

$$L_\sigma(x,\, \lambda) = f(x) + \frac{\sigma}{2}\|\lambda + h(x)/\sigma\|_2^2,$$

где $\sigma > 0$. Итерация метода модифицированных функций Лагранжа для задачи (51) состоит в следующем: по текущему приближению $(x^k,\, \lambda^k)$ к решению системы Лагранжа (52), текущему значению обратного параметра штрафа $\sigma_k > 0$ и текущему значению параметра неточности $\tau_k \geqslant 0$ очередное прямое приближение x^{k+1} ищется как точка, удовлетворяющая условию

$$\left\| \frac{\partial L_{\sigma_k}}{\partial x}(x^{k+1},\, \lambda^k) \right\| \leqslant \tau_k, \tag{55}$$

а очередное двойственное приближение λ^{k+1} вычисляется по формуле

$$\lambda^{k+1} = \lambda^k + h(x^{k+1})/\sigma_k. \tag{56}$$

Из (55), (56) видно, что точная версия метода, соответствующая выбору $\tau_k = 0$ для всех k, генерирует прямое приближение x^{k+1} как стационарную точку задачи

$$L_{\sigma_k}(x,\, \lambda^k) \to \min, \quad x \in \mathbb{R}^n.$$

При этом аналогично (10) имеет место равенство

$$\frac{\partial L}{\partial x}(x^{k+1},\, \lambda^{k+1}) = 0.$$

Отсюда следует, что итерация точного метода модифицированных функций состоит в решении системы уравнений

$$\Phi_{\sigma_k}(\lambda^k,\, u) = 0,$$

где $\Phi_\sigma\colon \mathbb{R}^l \times (\mathbb{R}^n \times \mathbb{R}^l) \mapsto \mathbb{R}^n \times \mathbb{R}^l$ есть семейство отображений индексируемое числовым параметром $\sigma \geqslant 0$ и заданное по правилу

$$\Phi_\sigma(\tilde{\lambda},\, u) = \left(\frac{\partial L}{\partial x}(x,\, \lambda),\, h(x) - \sigma(\lambda - \tilde{\lambda}) \right). \tag{57}$$

Будем предполагать, что значение обратного параметра штрафа σ_k и значение параметра неточности τ_k выбираются по текущему приближению, т. е.

$$\sigma_k = \sigma(x^k,\, \lambda^k), \quad \tau_k = \tau(x^k,\, \lambda^k), \tag{58}$$

где $\sigma\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}_+$ и $\tau\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}_+$ — некоторые функции. От функции σ мы будем требовать, чтобы она принимала нулевое значение тогда и только тогда, когда текущее приближение является решением системы Лагранжа (52). При этом будем считать, что в таком случае метод модифицированных функций Лагранжа останавливается, и итерация (55), (56) не осуществляется, а точки генерируемой траектории для всех последующих значений k равны этому решению.

Наконец, дополним итерацию метода модифицированных функций условием локализации. А именно, будем требовать, чтобы очередное приближение $(x^{k+1},\, \lambda^{k+1})$ в дополнение к условиям (55), (56) удовлетворяло неравенству

$$\|(x^{k+1} - x^k,\, \lambda^{k+1} - \lambda^k)\| \leqslant c\, \mathrm{dist}((x^k,\, \lambda^k),\, \bar{U}), \tag{59}$$

где $c > 0$ — фиксированное число, а $\bar{U}$ — множество решений системы Лагранжа (52) или, эквивалентно, уравнения (53) с оператором Φ, заданным согласно (54).

С учетом вышесказанного, при любом $c > 0$ итерация метода модифицированных функций (55), (56), снабженная условием локализации (59), эквивалентна итерации схемы (2.17) с отображением $\mathcal{A}\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto 2^{\mathbb{R}^n \times \mathbb{R}^l}$, заданным по правилу

$$\mathcal{A}(\tilde{u},\, u) = \left(\frac{\partial L}{\partial x}(x,\, \lambda) + B(0,\, \tau(\tilde{x},\, \tilde{\lambda})),\, h(x) - \sigma(\tilde{x},\, \tilde{\lambda})(\lambda - \tilde{\lambda})\right), \tag{60}$$

где $\tilde{u} = (\tilde{x},\, \tilde{\lambda})$, и отображением $N\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto 2^{\mathbb{R}^n \times \mathbb{R}^l}$, тождественно равным множеству $\{0\}$. Заметим, что отображение $\mathcal{A}$ не зависит от параметра π, т. е. речь идет о непараметрическом варианте схемы (2.17).

Пусть $(\bar{x},\, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ — решение системы Лагранжа (52). В силу следствия (1), строгая некритичность множителя $\bar{\lambda}$ влечет за собой выполнение предположения 1) теоремы 2.3 с $\bar{u} = (\bar{x},\, \bar{\lambda})$.

Предположим, что функция $\tau(\cdot)$ удовлетворяет условию

$$\tau(x,\, \lambda) = o(\mathrm{dist}((x,\, \lambda),\, \{\bar{x}\} \times \mathcal{M}(\bar{x}))) \tag{61}$$

при $(x,\, \lambda) \to (\bar{x},\, \bar{\lambda})$. В таком случае предположение 2) теоремы 2.3 выполняется для отображения $\mathcal{A}$, определенного согласно (60), если функция $\sigma(\cdot)$ принимает достаточно малые значения, когда ее аргумент $(x,\,\lambda)$ близок к точке $(\bar{x},\,\bar{\lambda})$, и если множество $\bar{U}$ локально вблизи точки $(\bar{x},\,\bar{\lambda})$ совпадает с множеством $\{\bar{x}\}\times\mathcal{M}(\bar{x})$. Последнее имеет место, если множитель $\bar{\lambda}$ является некритическим, поскольку, согласно следствию 1.1, некритичность множителя $\bar{\lambda}$ эквивалентна оценке расстояния (1.46), которая для задач с ограничениями-равенствами состоит в существовании числа $\ell>0$ такого, что для любой пары $(x,\,\lambda)\in\mathbb{R}^n\times\mathbb{R}^l$, достаточно близкой к $(\bar{x},\,\bar{\lambda})$, выполняется неравенство

$$\|x-\bar{x}\|+\mathrm{dist}(\lambda,\,\mathcal{M}(\bar{x}))\leqslant\ell\left\|\left(\begin{array}{c}\dfrac{\partial L}{\partial x}(x,\,\lambda,\,\mu)\\ h(x)\}\end{array}\right)\right\|. \tag{62}$$

Отметим, что функция τ удовлетворяет условию (61), если

$$\tau(x,\,\lambda)=o(\rho(x,\,\lambda))$$

при $(x,\,\lambda)\to(\bar{x},\,\bar{\lambda})$, где $\rho\colon\mathbb{R}^n\times\mathbb{R}^l\mapsto\mathbb{R}_+$ есть невязка системы Лагранжа (52),

$$\rho(x,\,\lambda)=\left\|\left(\frac{\partial L}{\partial x}(x,\,\lambda),\,h(x)\right)\right\|. \tag{63}$$

Таким образом, результаты о локальной сходимости и скорости сходимости будут следовать из теоремы 2.3, если установить, что строгая некритичность гарантирует выполнение ее третьего предположения. Ниже мы проделаем это для двух специальных правил управления обратным параметром штрафа (т. е. двух видов функции $\sigma(\cdot)$). Но сначала докажем следующие две леммы.

Лемма 3. *Пусть $H\in\mathbb{R}^{n\times n}$, $B\in\mathbb{R}^{l\times n}$, и пусть*

$$H\xi\notin\mathrm{im}\,B^{\mathrm{T}}\quad\forall\,\xi\in\ker B\setminus\{0\}. \tag{64}$$

Тогда для любого $M>0$ найдется число $\gamma>0$ такое, что соотношение

$$\left\|\left(\tilde{H}+t(B+\Omega)^{\mathrm{T}}\tilde{B}\right)\xi\right\|\geqslant\gamma\|\xi\|\quad\forall\,\xi\in\mathbb{R}^n$$

выполняется для любой матрицы $\tilde{H}\in\mathbb{R}^{n\times n}$, достаточно близкой к H, любой матрицы $\tilde{B}\in\mathbb{R}^{l\times n}$, достаточно близкой B, любого достаточно большого по модулю числа $t\in\mathbb{R}$ и любой матрицы $\Omega\in\mathbb{R}^{l\times n}$, удовлетворяющей неравенству $\|\Omega\|\leqslant M/|t|$.

Доказательство. От противного: предположим, что для некоторого $M>0$ найдутся последовательности матриц $\{H_k\}\subset\mathbb{R}^{n\times n}$, $\{B_k\}\subset\mathbb{R}^{l\times n}$ и $\{\Omega_k\}\subset\mathbb{R}^{l\times n}$, а также последовательности $\{t_k\}\subset\mathbb{R}$ и $\{\xi^k\}\subset\mathbb{R}^n\setminus\{0\}$ такие, что $\{H_k\}\to H$, $\{B_k\}\to B$, $|t_k|\to\infty$, $\|\Omega_k\|\leqslant M/|t_k|$ для всех k, и

$$H_k\xi^k+t_k(B+\Omega_k)^{\mathrm{T}}B_k\xi^k=o(\|\xi^k\|) \tag{65}$$

при $k \to \infty$. Без ограничения общности можно считать, что $\|\xi^k\| = 1$ для всех k и что $\{\xi^k\} \to \xi \neq 0$. Тогда оценка (65) означает существование последовательности $\{w^k\} \subset \mathbb{R}^n$ такой, что $\{w_k\} \to 0$, и

$$H_k\xi^k + t_k(B + \Omega_k)^{\mathrm{T}} B_k\xi^k = w_k \tag{66}$$

при всех k. Как следствие, поскольку

$$B^{\mathrm{T}} B_k\xi^k = -\frac{1}{t_k} H_k\xi^k - \Omega_k^{\mathrm{T}} B_k\xi^k + \frac{1}{t_k} w^k \to 0$$

при $k \to \infty$, $B^{\mathrm{T}} B\xi = 0$, а значит, $\xi \in \ker B$.

С другой стороны, оценка (66) влечет за собой выполнение равенства

$$H_k\xi^k + t_k\Omega_k^{\mathrm{T}} B_k\xi^k - w^k = -t_k B^{\mathrm{T}} B_k\xi^k \in \operatorname{im} B^{\mathrm{T}}$$

для всех k, где второе слагаемое в левой части стремится к нулю при $k \to \infty$, так как последовательность $\{t_k\Omega_k\}$ ограничена, а $\{B_k\xi^k\} \to B\xi = 0$. Следовательно, $H\xi \in \operatorname{im} B^{\mathrm{T}}$ в силу замкнутости множества $\operatorname{im} B^{\mathrm{T}}$, и мы приходим к противоречию с (64). ■

Лемма 4. *Пусть функция $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ и отображение $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ дифференцируемы в точке $\bar{x}$, и их производные локально липшицевы в этой точке. Пусть $\bar{x}$ — стационарная точка задачи* (51), *и пусть отвечающий ей множитель Лагранжа $\bar{\lambda} \in \mathcal{M}(\bar{x})$ является строго некритическим.*

Тогда для любого $M > 0$ найдется число $\gamma > 0$ такое, что для любого достаточно малого $\sigma > 0$, любого $\lambda \in \mathbb{R}^l$, достаточно близкого к $\bar{\lambda}$, и любого $x \in \mathbb{R}^n$, удовлетворяющего неравенству $\|x - \bar{x}\| \leqslant \sigma M$, выполняется соотношение

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(x,\, \lambda) \quad \left\| \left(H + \frac{1}{\sigma}(h'(x))^{\mathrm{T}} h'(x) \right) \xi \right\| \geqslant \gamma\|\xi\| \quad \forall\, \xi \in \mathbb{R}^n.$$

Доказательство. Предположим, что для некоторого $M > 0$ найдутся последовательности $\{\sigma_k\} \subset \mathbb{R}_+ \setminus \{0\}$, $\{(x^k,\, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$, $\{H_k\} \subset \mathbb{R}^{n\times n}$ и $\{\xi^k\} \subset \mathbb{R}^n$ такие, что $\|x^k - \bar{x}\| \leqslant \sigma_k M$ и $H_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k)$ для всех k, $\{\sigma_k\} \to 0$, $\{\lambda^k\} \to \bar{\lambda}$, и

$$\left(H_k + \frac{1}{\sigma_k}(h'(\bar{x}) + (h'(x^k) - h'(\bar{x})))^{\mathrm{T}} h'(x^k) \right) \xi^k = \left(H_k + \frac{1}{\sigma_k}(h'(x^k))^{\mathrm{T}} h'(x^k) \right) \xi^k = o(\|\xi^k\|)$$

при $k \to \infty$. Поскольку отображения f' и h' являются локально липшицевыми в точке $\bar{x}$, отображение $\frac{\partial L}{\partial x}(\cdot,\, \lambda^k)$ является локально липшицевым в точке x^k для всех достаточно больших k, и более того, т. к. последовательность $\{\lambda^k\}$ ограничена, соответствующую константу Липшица можно взять одинаковой для всех таких k. Тогда, поскольку норма всех матриц, принадлежащих дифференциалу Кларка отображения в точке, ограничена константой Липшица, с которой это отображение локально липшицево в этой точке, последовательность $\{H_k\}$ ограничена, а значит, без ограничения общности можно считать ее сходящейся к некоторой матрице $H \in \mathbb{R}^{n\times n}$. Используя лемму 1.5 и полунепрерывность дифференциала Кларка сверху, заключаем, что $H \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda})$.

Кроме того, в силу локальной липшицевости отображения h' в точке $\bar{x}$,

$$\|h'(x^k) - h'(\bar{x})\| = O(\|x^k - \bar{x}\|) = O(\sigma_k),$$

откуда, в частности, следует ограниченность последовательности $\{(h'(x^k) - h'(\bar{x}))/\sigma_k\}$.

Тем самым мы приходим к противоречию с леммой 3, применяемой с $B = h'(\bar{x})$, $\tilde{H} = H_k$, $\tilde{B} = h'(x^k)$, $\Omega = (h'(x^k) - h'(\bar{x}))$ и $t = 1/\sigma_k$. ■

Из леммы 4, в частности, следует, что, если вектор $\lambda \in \mathbb{R}^l$ достаточно близок к строго некритическому множителю $\bar{\lambda}$, то для любого достаточно малого $\sigma > 0$ найдется окрестность точки $\bar{x}$ такая, что

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(x,\, \lambda) \quad \det\left(H + \frac{1}{\sigma}(h'(x))^{\mathrm{T}} h'(x)\right) \neq 0 \tag{67}$$

для всех x из этой окрестности. Следующий простой пример демонстрирует, что, вообще говоря, эта окрестность существенно зависит от σ.

Пример 2. Пусть $n = l = 1$, $h(x) = x^2/2$, и пусть $f\colon \mathbb{R} \mapsto \mathbb{R}$ — произвольная функция, дважды непрерывно дифференцируемая в точке $\bar{x} = 0$ и такая, что $f'(0) = 0$. Тогда $\mathcal{M}(0) = \mathbb{R}$, и любой множитель $\bar{\lambda} \in \mathcal{M}(0) \setminus \{-f''(0)\}$ является (строго) некритическим.

Зафиксируем произвольный множитель $\bar{\lambda} < -f''(0)$ и произвольную последовательность $\{x^k\} \subset \mathbb{R}$, сходящуюся к точке 0 и такую, что $f''(x^k) + \bar{\lambda} < 0$ при всех k. Положим $\sigma_k = -(x^k)^2/(f''(x^k) + \bar{\lambda}) > 0$. Ясно, что последовательность $\{\sigma_k\} \to 0$, но при этом для всех k

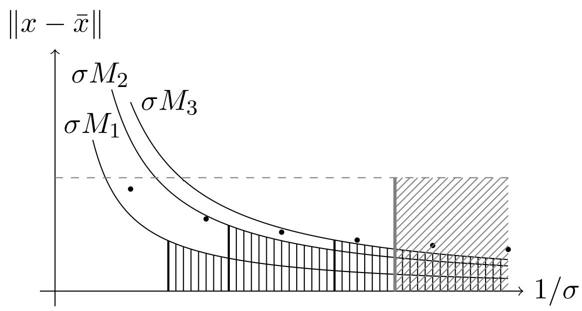

Рис. 3.1. Области невырожденности.

условие (67) нарушается при $\lambda = \bar{\lambda}$, $\sigma = \sigma_k$ и $x = x^k$. Следовательно, радиус окрестности, в которой выполнено условие (67), нельзя выбрать одинаковым для всех достаточно малых $\sigma > 0$, даже если $\lambda = \bar{\lambda}$.

В то же время, как легко показать от противного, если множитель $\bar{\lambda}$ удовлетворяет достаточному условию второго порядка (20), т. е. если

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \quad \langle H\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\},$$

то условие (67) выполняется для всех достаточно малых $\sigma > 0$ и для всех $(x,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$, достаточно близких к $(\bar{x},\, \bar{\lambda})$. Ситуация проиллюстрирована на рисунке 3.1. Черными точками показана последовательность из примера 2. Вертикальной штриховкой обозначены области невырожденности (т. е. области, в которых выполнено условие (67), если вектор λ достаточно близок к множителю $\bar{\lambda}$), которые получаются в результате применения леммы 4 с тремя разными значениями M: $M_1 < M_2 < M_3$. Наконец, наклонной штриховкой показана прямоугольная область невырожденности, которая существовала бы, если бы множитель $\bar{\lambda}$ удовлетворял достаточному условию второго порядка (20).

Рассмотрим вариант метода модифицированных функций, в котором параметр штрафа определяется согласно (58) с использованием функции $\sigma(\cdot) = \sigma_\theta(\cdot)$ вида

$$\sigma_\theta(x,\, \lambda) = (\rho(x,\, \lambda))^\theta, \tag{68}$$

где $\rho(\cdot)$ — невязка системы Лагранжа, определенная в (63), а $\theta \in (0,\, 1]$ — фиксированное число.

Замечание 4. Для любого $\sigma > 0$, любого $\tilde{\lambda} \in \mathbb{R}^l$ и любой пары $u = (x,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$ такой, что отображения f' и h' локально липшицевы в точке x, для отображения Φ_σ, определенного согласно (57), выполнено соотношение

$$\partial_u \Phi_\sigma(\tilde{\lambda},\, u) = \left\{ \left(\begin{array}{cc} H & (h'(x))^{\mathrm{T}} \\ h'(x) & -\sigma I \end{array} \right) \middle|\; H \in \partial_x \frac{\partial L}{\partial x}(x,\, \lambda) \right\},$$

где символом I обозначена единичная матрица размера $l \times l$. Это следует из теоремы 1.1.

Пользуясь леммой 4, докажем следующее утверждение.

Следствие 1. *Пусть выполнены предположения леммы 4. Пусть $c > 0$ и $\theta \in (0,\, 1]$ — произвольные числа, и пусть функция $\sigma_\theta(\cdot)$ определена согласно* (68).

Тогда все матрицы во множестве $\partial_u \Phi_{\sigma_\theta(\tilde{x},\,\tilde{\lambda})}(\tilde{\lambda},\, u)$ *не вырождены, если пара* $(\tilde{x},\, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ *достаточно близка к* $(\bar{x},\, \bar{\lambda})$, $\tilde{x} \neq \bar{x}$ *и/или* $\tilde{\lambda} \notin \mathcal{M}(\bar{x})$, *и если* $u = (x,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$ *удовлетворяет неравенству*

$$\|(x - \tilde{x},\, \lambda - \tilde{\lambda})\| \leqslant c(\operatorname{dist}((\tilde{x},\, \tilde{\lambda}),\, \{\bar{x}\} \times \mathcal{M}(\bar{x}))). \tag{69}$$

Доказательство. Зафиксируем произвольные числа $c > 0$ и $\theta \in (0,\, 1]$. В силу оценки расстояния (62), которая выполняется при (строгой) некритичности, для любой пары $(\tilde{x},\, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$, достаточно близкой к $(\bar{x},\, \bar{\lambda})$ и такой, что $x \neq \bar{x}$ и/или $\lambda \notin \mathcal{M}(\bar{x})$, величина $\rho(\tilde{x},\, \tilde{\lambda})$ положительна, а значит, $\sigma_\theta(\tilde{x},\, \tilde{\lambda}) > 0$. Кроме того, $\sigma_\theta(\tilde{x},\, \tilde{\lambda}) \to 0$ при $(\tilde{x},\, \tilde{\lambda}) \to (\bar{x},\, \bar{\lambda})$.

Далее, вновь пользуясь оценкой расстояния (62), получаем, что для любой пары $(x,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$, удовлетворяющей (69), выполняется оценка

$$\|x - \bar{x}\| \leqslant \|x - \tilde{x}\| + \|\tilde{x} - \bar{x}\| = O(\rho(\tilde{x},\, \tilde{\lambda})) = O\left(\sigma_\theta(\tilde{x},\, \tilde{\lambda})(\rho(\tilde{x},\, \tilde{\lambda}))^{1-\theta}\right) = O(\sigma_\theta(\tilde{x},\, \tilde{\lambda}))$$

при $(\tilde{x},\, \tilde{\lambda}) \to (\bar{x},\, \bar{\lambda})$. Наконец, из (69) следует, что $\lambda \to \bar{\lambda}$ при $(\tilde{x},\, \tilde{\lambda}) \to (\bar{x},\, \bar{\lambda})$.

Тогда, применяя лемму 4, заключаем, что если пара $(\tilde{x},\, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ достаточно близка к $(\bar{x},\, \bar{\lambda})$, и при этом $\tilde{x} \neq \bar{x}$ и/или $\tilde{\lambda} \notin \mathcal{M}(\bar{x})$, для любой точки $u = (x,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$, удовлетворяющей (69), матрица

$$H + \frac{1}{\sigma_\theta(\tilde{x},\, \tilde{\lambda})}(h'(x))^{\mathrm{T}} h'(x)$$

является невырожденной для всех $H \in \partial_x \frac{\partial L}{\partial x}(x,\, \lambda)$. В силу замечания 4, отсюда следует, что всякая матрица из множества $\partial_u \Phi_{\sigma_\theta(\tilde{x},\,\tilde{\lambda})}(\tilde{\lambda},\, u)$ имеет невырожденную подматрицу $(-\sigma_\theta(\tilde{x},\, \tilde{\lambda})I)$ с невырожденным дополнением Шура, а значит, является невырожденной (см., например, [87, предложение 3.9]). Таким образом, все матрицы из $\partial_u \Phi_{\sigma_\theta(\tilde{x},\,\tilde{\lambda})}(\tilde{\lambda},\, u)$ невырождены. ∎

Для заданного $c > 0$ определим функцию $\delta_c \colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}_+$ по правилу

$$\delta_c(x,\, \lambda) = c(\operatorname{dist}((x,\, \lambda),\, \{\bar{x}\} \times \mathcal{M}(\bar{x}))). \tag{70}$$

Через $\pi(\lambda)$ будем обозначать ортогональную проекцию вектора $\lambda \in \mathbb{R}^l$ на аффинное множество $\mathcal{M}(\bar{x})$.

Лемма 5. *Пусть функция* $f \colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображение* $h \colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *дифференцируемы в некоторой окрестности точки* $\bar{x}$, *и их производные локально липшицевы в этой точке.*

Пусть $\bar{x}$ *— стационарная точка задачи* (51)*, и пусть отвечающий ей множитель Лагранжа* $\bar{\lambda} \in \mathcal{M}(\bar{x})$ *является некритическим.*

Тогда для любых $c > 0$, $\theta \in (0,\, 1]$ *и* $\gamma \in (0,\, 1)$ *найдется число* $\varepsilon > 0$ *такое, что для функции* $\sigma_\theta(\cdot)$*, определенной согласно* (68)*, и любой пары* $(\tilde{x},\, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$*, удовлетворяющей неравенству*

$$\|(\tilde{x} - \bar{x},\, \tilde{\lambda} - \bar{\lambda})\| \leqslant \varepsilon, \tag{71}$$

соотношение

$$\|\Phi_{\sigma_\theta(\tilde{x},\tilde{\lambda})}(\tilde{\lambda},\, u) - \Phi_{\sigma_\theta(\tilde{x},\tilde{\lambda})}(\tilde{\lambda},\, (\bar{x},\, \pi(\tilde{\lambda})))\| \geqslant \gamma\sigma_\theta(\tilde{x},\, \tilde{\lambda})\|(x - \bar{x},\, \lambda - \pi(\tilde{\lambda}))\| \tag{72}$$

выполняется в любой точке $u = (x,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$*, удовлетворяющей оценке*

$$\|(x - \tilde{x},\, \lambda - \tilde{\lambda})\| \leqslant \delta_c(\tilde{x},\, \tilde{\lambda}),$$

где функция $\delta_c(\cdot)$ *определена согласно* (70)*.*

Доказательство. Предположим противное: пусть существуют числа $c > 0$, $\theta \in (0,\, 1]$, $\gamma \in (0,\, 1)$ и последовательности $\{(\tilde{x}^k,\, \tilde{\lambda}^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$ и $\{(x^k,\, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$ такие, что $\{(\tilde{x}^k,\, \tilde{\lambda}^k)\} \to (\bar{x},\, \bar{\lambda})$, и для каждого k выполнены включение $(x^k,\, \lambda^k) \in B((\tilde{x}^k,\, \tilde{\lambda}^k),\, \delta_k)$ и неравенство

$$\|\Phi_{\sigma_k}(\tilde{\lambda}^k,\, (x^k,\, \lambda^k)) - \Phi_{\sigma_k}(\tilde{\lambda}^k,\, (\bar{x},\, \pi(\tilde{\lambda}^k)))\| < \gamma\sigma_k\|(x^k - \bar{x},\, \lambda^k - \pi(\tilde{\lambda}^k))\|, \tag{73}$$

где $\sigma_k = \sigma_\theta(\tilde{x}^k,\, \tilde{\lambda}^k)$, $\delta_k = \delta_c(\tilde{x}^k,\, \tilde{\lambda}^k)$.

Положим $t_k = \|(x^k - \bar{x},\, \lambda^k - \pi(\tilde{\lambda}^k))\|$. Из неравенства (73) следует, что $\sigma_k > 0$ и $t_k > 0$. Также заметим, что $\sigma_k \to 0$ при $k \to \infty$, и что

$$\|\Phi_{\sigma_k}(\tilde{\lambda}^k,\, (x^k,\, \lambda^k)) - \Phi_{\sigma_k}(\tilde{\lambda}^k,\, (\bar{x},\, \pi(\tilde{\lambda}^k)))\| = \left\|\left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k),\, h(x^k) - \sigma_k(\lambda^k - \pi(\tilde{\lambda}^k))\right)\right\|.$$

Следовательно, (73) влечет за собой оценки

$$\left\|\frac{\partial L}{\partial x}(x^k,\, \lambda^k)\right\| < \gamma\sigma_k t_k = o(t_k) \tag{74}$$

и

$$\|h(x^k) - h(\bar{x}) - \sigma_k(\lambda^k - \pi(\tilde{\lambda}^k))\| < \gamma\sigma_k t_k = o(t_k) \tag{75}$$

при $k \to \infty$.

Положим $\xi^k = (x^k - \bar{x})/t_k$ и $\eta^k = (\lambda^k - \pi(\tilde{\lambda}^k))/t_k$. Без ограничения общности можно считать, что последовательность $\{(\xi^k,\, \eta^k)\}$ сходится к некоторому вектору $(\xi,\, \eta) \in \mathbb{R}^n \times \mathbb{R}^l$ такому, что

$$\|(\xi,\, \eta)\| = 1. \tag{76}$$

Из (75) немедленно следует, что

$$\xi \in \ker h'(\bar{x}). \tag{77}$$

Более того,

$$\begin{aligned}\frac{\partial L}{\partial x}(x^k,\, \lambda^k) = \frac{\partial L}{\partial x}(x^k,\, \pi(\tilde{\lambda}^k)) - \frac{\partial L}{\partial x}(\bar{x},\, \pi(\tilde{\lambda}^k)) + \frac{\partial L}{\partial x}(x^k,\, \lambda^k) - \frac{\partial L}{\partial x}(x^k,\, \pi(\tilde{\lambda}^k)) = \\ = \frac{\partial L}{\partial x}(x^k,\, \bar{\lambda}) - \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda}) + (h'(x^k) - h'(\bar{x}))^{\mathrm{T}}(\pi(\tilde{\lambda}^k) - \bar{\lambda}) + (h'(x^k))^{\mathrm{T}}(\lambda^k - \pi(\tilde{\lambda}^k)) = \\ = \frac{\partial L}{\partial x}(\bar{x} + t_k\xi_k,\, \bar{\lambda}) - \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda}) + t_k(h'(x^k))^{\mathrm{T}}\eta^k + o(t_k\|\eta^k\|) = \\ = \frac{\partial L}{\partial x}(\bar{x} + t_k\xi,\, \bar{\lambda}) - \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda}) + t_k(h'(x^k))^{\mathrm{T}}\eta^k + O(t_k\|\xi^k - \xi\|) + o(t_k\|\eta^k\|),\end{aligned} \tag{78}$$

где в последнем переходе учтено, что в сделанных предположениях отображение $\frac{\partial L}{\partial x}(\cdot,\, \bar{\lambda})$ является локально липшицевым в точке $\bar{x}$. Соединяя (74) и (78), получаем существование вектора $d \in C_x\frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda})(\xi)$, удовлетворяющего равенству

$$d + (h'(\bar{x}))^{\mathrm{T}}\eta = 0. \tag{79}$$

Поскольку множитель $\bar{\lambda}$ некритический, соотношения (77) и (79) влекут за собой равенство $\xi = 0$, и, в частности, $\{\xi^k\} \to 0$.

Кроме того, из (75) следует, что

$$\eta^k = \frac{1}{\sigma_k t_k}(h(x^k) - h(\bar{x})) + \zeta^k, \tag{80}$$

где $\zeta^k \in \mathbb{R}^l$ удовлетворяет неравенству $\|\zeta^k\| \leqslant \gamma$. Заметим также, что, в силу $(x^k,\, \lambda^k) \in B((\tilde{x}^k,\, \tilde{\lambda}^k),\, \delta_k)$, из (70) следует, что

$$\|x^k - \bar{x}\| \leqslant \|x^k - \tilde{x}^k\| + \|\tilde{x}^k - \bar{x}\| \leqslant (c+1)(\mathrm{dist}((\tilde{x}^k,\, \tilde{\lambda}^k),\, \{\bar{x}\} \times \mathcal{M}(\bar{x}))).$$

Вновь пользуясь оценкой расстояния (62), заключаем, что

$$\|x^k - \bar{x}\| = O(\rho(\tilde{x}^k,\, \tilde{\lambda}^k)). \tag{81}$$

Обозначим через P ортогональный проектор на подпространство $(\mathrm{im}\, h'(\bar{x}))^{\perp}$ в $\mathbb{R}^l$. Применяя P к обеим частям формулы (80), пользуясь теоремой о среднем и принимая во внимание локальную липшицевость отображения h' в точке $\bar{x}$, а также (68) и (81), получаем, что

$$\|P\eta^k\| \leqslant \frac{1}{\sigma_k} \sup_{\tau\in[0,\,1]} \|P(h'(\bar{x}+\tau(x^k-\bar{x}))-h'(\bar{x}))\|\|\xi^k\| + \|P\zeta^k\| =$$

$$= \|P\zeta^k\| + O\left(\frac{\|x^k-\bar{x}\|\|\xi^k\|}{\sigma_k}\right) = \|P\zeta^k\| + O((\rho(\tilde{x}^k,\,\tilde{\lambda}^k))^{1-\theta}\|\xi^k\|). \quad (82)$$

Так как $\|\zeta^k\| \leqslant \gamma$ для всех k, без ограничения общности можно считать, что последовательность $\{\zeta^k\}$ сходится к некоторому вектору $\zeta \in \mathbb{R}^l$, который удовлетворяет неравенству $\|\zeta\| \leqslant \gamma$. Тогда, поскольку $\{\xi^k\} \to 0$, осуществляя предельный переход в (82), получаем:

$$\|P\eta\| \leqslant \|P\zeta\| \leqslant \|\zeta\| \leqslant \gamma < 1. \quad (83)$$

Однако, так как $\xi = 0$, имеем $d \in C_x\frac{\partial L}{\partial x}(\bar{x},\,\bar{\lambda})(0)$, а значит, $d = 0$. Тогда из (79) следует, что $\eta \in \ker(h'(\bar{x}))^{\mathrm{T}} = (\operatorname{im} h'(\bar{x}))^{\perp}$, а значит, $P\eta = \eta$, и, в силу (83), $\|\eta\| < 1$. Поскольку $\xi = 0$, это противоречит (76). ■

Теперь можно доказать, что подзадача (55), (56) точного (а значит, и неточного) метода множителей имеет решение, удовлетворяющее условию локализации, если параметр штрафа выбирается согласно (58) с функцией $\sigma(\cdot) = \sigma_\theta(\cdot)$ вида (68) и если текущая точка находится вблизи решения $(\bar{x},\,\bar{\lambda})$ системы Лагранжа (52) такого, что множитель $\bar{\lambda}$ является строго некритическим.

Для $\delta > 0$, $\sigma \geqslant 0$, $\tilde{x} \in \mathbb{R}^n$ и $\tilde{\lambda} \in \mathbb{R}^l$ через $U_\delta(\sigma,\,\tilde{x},\,\tilde{\lambda})$ обозначим множество решений задачи оптимизации

$$\|\Phi_\sigma(\tilde{\lambda},\,u)\|^2 \to \min, \quad \|(x-\tilde{x},\,\lambda-\tilde{\lambda})\| \leqslant \delta, \quad (84)$$

относительно $u = (x,\,\lambda) \in \mathbb{R}^n \times \mathbb{R}^l$. Ясно, что это множество не пусто.

Предложение 6. *Пусть выполнены предположения леммы 4.*

Тогда для любого $c > 3$, любого $\theta \in (0,\,1]$ и любого $(\tilde{x},\,\tilde{\lambda}) \in \mathbb{R}^n\times\mathbb{R}^l$, достаточно близкого к $(\bar{x},\,\bar{\lambda})$, уравнение

$$\Phi_{\sigma_\theta(\tilde{x},\,\tilde{\lambda})}(\tilde{\lambda},\,u) = 0, \quad (85)$$

где функция $\sigma_\theta(\cdot)$ определена в (68), имеет решение $u = (x,\,\lambda) \in \mathbb{R}^n \times \mathbb{R}^l$, удовлетворяющее условию (69).

Доказательство. Прежде всего заметим, что, если $\tilde{x} = \bar{x}$ и $\tilde{\lambda} \in \mathcal{M}(\bar{x})$, то требуемое утверждение выполняется очевидным образом: в качестве u можно взять $(\tilde{x},\,\tilde{\lambda})$. Поэтому в остальной части доказательства будем полагать, что $\tilde{x} \neq \bar{x}$ и/или $\tilde{\lambda} \notin \mathcal{M}(\bar{x})$.

Зафиксируем произвольное число $\gamma \in (2/(c-1),\, 1)$. Из леммы 5 следует, что существует число $\varepsilon > 0$ такое, что для всех $(\tilde{x},\, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$, удовлетворяющих условию (71), неравенство (72) выполнено для всех $u = (x,\, \lambda)$ из множества $U_{\delta_c(\tilde{x},\, \tilde{\lambda})}(\sigma_\theta(\tilde{x},\, \tilde{\lambda}),\, \tilde{x},\, \tilde{\lambda})$, где функция $\delta_c(\cdot)$ определена в (70). Согласно следствию 1, при необходимости уменьшая ε, мы можем гарантировать, что множество $\partial_u \Phi_{\sigma_\theta(\tilde{x},\, \tilde{\lambda})}(\tilde{\lambda},\, u)$ не содержит вырожденных матриц. Покажем, что для любого $(\tilde{x},\, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$, удовлетворяющего неравенству (71) с указанным ε, любая пара $u = (x,\, \lambda) \in U_{\delta_c(\tilde{x},\, \tilde{\lambda})}(\sigma_\theta(\tilde{x},\, \tilde{\lambda}),\, \tilde{x},\, \tilde{\lambda})$ является решением уравнения (85). При этом из определения множества $U_{\delta_c(\tilde{x},\, \tilde{\lambda})}(\sigma_\theta(\tilde{x},\, \tilde{\lambda}),\, \tilde{x},\, \tilde{\lambda})$ будет сразу следовать, что любая такая пара удовлетворяет условию (69).

Если $\|(x - \tilde{x},\, \lambda - \tilde{\lambda})\| = \delta_c(\tilde{x},\, \tilde{\lambda})$, то из (70) следует, что

$$\|(x - \bar{x},\, \lambda - \pi(\tilde{\lambda}))\| \geqslant \|(x - \tilde{x},\, \lambda - \tilde{\lambda})\| - \|(\tilde{x} - \bar{x},\, \tilde{\lambda} - \pi(\tilde{\lambda}))\| =$$
$$= (c-1)\|(\tilde{x} - \bar{x},\, \tilde{\lambda} - \pi(\tilde{\lambda}))\| \geqslant (c-1)\|\tilde{\lambda} - \pi(\tilde{\lambda})\| = (c-1)\operatorname{dist}(\tilde{\lambda},\, \mathcal{M}(\bar{x})).$$

Тогда, используя (72), получаем

$$\|\Phi_{\sigma_\theta(\tilde{x},\, \tilde{\lambda})}(\tilde{\lambda},\, u) - \Phi_{\sigma_\theta(\tilde{x},\, \tilde{\lambda})}(\tilde{\lambda},\, (\bar{x},\, \pi(\tilde{\lambda})))\| \geqslant \gamma\sigma_\theta(\tilde{x},\, \tilde{\lambda})\|(x - \bar{x},\, \lambda - \pi(\tilde{\lambda}))\| \geqslant$$
$$\geqslant \gamma(c-1)\sigma_\theta(\tilde{x},\, \tilde{\lambda})\operatorname{dist}(\tilde{\lambda},\, \mathcal{M}(\bar{x})) > 2\sigma_\theta(\tilde{x},\, \tilde{\lambda})\operatorname{dist}(\tilde{\lambda},\, \mathcal{M}(\bar{x})), \quad (86)$$

где учтен выбор γ.

С другой стороны, в силу (70), $\|(\bar{x} - \tilde{x},\, \pi(\tilde{\lambda}) - \tilde{\lambda})\| \leqslant \delta_c(\tilde{x},\, \tilde{\lambda})$, и, поскольку точка u является решением задачи (84) с $\sigma = \sigma_\theta(\tilde{x},\, \tilde{\lambda})$ и $\delta = \delta_c(\tilde{x},\, \tilde{\lambda})$, имеем:

$$\|\Phi_{\sigma_\theta(\tilde{x},\, \tilde{\lambda})}(\tilde{\lambda},\, u) - \Phi_{\sigma_\theta(\tilde{x},\, \tilde{\lambda})}(\tilde{\lambda},\, (\bar{x},\, \pi(\tilde{\lambda})))\| \leqslant \|\Phi_{\sigma_\theta(\tilde{x},\, \tilde{\lambda})}(\tilde{\lambda},\, u)\| + \|\Phi_{\sigma_\theta(\tilde{x},\, \tilde{\lambda})}(\tilde{\lambda},\, (\bar{x},\, \pi(\tilde{\lambda})))\| \leqslant$$
$$\leqslant 2\|\Phi_{\sigma_\theta(\tilde{x},\, \tilde{\lambda})}(\tilde{\lambda},\, (\bar{x},\, \pi(\tilde{\lambda})))\| = 2\sigma_\theta(\tilde{x},\, \tilde{\lambda})\|\tilde{\lambda} - \pi(\tilde{\lambda})\| = 2\sigma_\theta(\tilde{x},\, \tilde{\lambda})\operatorname{dist}(\tilde{\lambda},\, \mathcal{M}(\bar{x})),$$

что противоречит (86).

Следовательно, $\|(x - \tilde{x},\, \lambda - \tilde{\lambda})\| < \delta_c(\tilde{x},\, \tilde{\lambda})$, и значит, u является локальным решением задачи оптимизации без ограничений с целевой функцией, совпадающей с целевой функцией задачи (84). Согласно [21, предложение 2.3.2], отсюда следует, что

$$0 \in \partial_u \left(\|\Phi_{\sigma_\theta(\tilde{x},\, \tilde{\lambda})}(\tilde{\lambda},\, u)\|^2\right),$$

и, в соответствии с формулой для дифференциала Кларка суперпозиции [21, теорема 2.6.6], последнее влечет за собой существование матрицы $\mathcal{J} \in \partial_u \Phi_{\sigma_\theta(\tilde{x},\, \tilde{\lambda})}(\tilde{\lambda},\, u)$ такой, что

$$\mathcal{J}^{\mathrm{T}} \Phi_{\sigma_\theta(\tilde{x},\, \tilde{\lambda})}(\tilde{\lambda},\, u) = 0.$$

Но, согласно выбору ε, матрица $\mathcal{J}$ невырождена, а значит, точка u является решением уравнения (85). ■

С учетом предложения 6, а также представленных выше обсуждений, касающихся проверки предположений 1) и 2) теоремы 2.3, справедлив следующий результат.

Теорема 4. *Пусть выполнены предположения леммы 4, и пусть функция* $\tau\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}_+$ *удовлетворяет условию* (61).

Тогда для любых $c > 3$ *и* $\theta \in (0,\, 1]$ *справедливо следующее: для любого начального приближения* $(x^0,\, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^l$*, достаточно близкого к* $(\bar{x},\, \bar{\lambda})$*, существует последовательность* $\{(x^k,\, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$*, генерируемая методом множителей с* σ_k *и* τ_k*, вычисляемыми согласно* (58), (68)*, и удовлетворяющая условию*

$$\|(x^{k+1} - x^k,\, \lambda^{k+1} - \lambda^k)\| \leqslant c\,\mathrm{dist}((x^k,\, \lambda^k),\, \bar{U}) \tag{87}$$

для всех k*; любая такая последовательность сходится к некоторой паре* $(\bar{x},\, \lambda^*)$*, где* $\lambda^* \in \mathcal{M}(\bar{x})$*, и скорости сходимости последовательности* $\{(x^k,\, \lambda^k)\}$ *к* $(\bar{x},\, \lambda^*)$ *и последовательности* $\{\mathrm{dist}((x^k,\, \lambda^k),\, \{\bar{x}\}\times\mathcal{M}(\bar{x}))\}$ *к нулю являются сверхлинейными. Кроме того, для любого* $\varepsilon > 0$ *справедливо неравенство* $\|\lambda^* - \bar{\lambda}\| < \varepsilon$*, если начальное приближение* $(x^0,\, \lambda^0)$ *достаточно близко к* $(\bar{x},\, \bar{\lambda})$.

Теперь обратимся к варианту метода модифицированных функций, в котором обратный параметр штрафа является фиксированным:

$$\sigma(x,\, \lambda) = \sigma \quad \forall\, (x,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l,$$

где $\sigma > 0$. Начнем с доказательства двух лемм алгебраического характера.

Лемма 6. *Пусть выполнены предположения леммы 3.*

Тогда для любого $M > 0$ *и любого* $\varepsilon > 0$ *справедливо следующее: для любой матрицы* $\tilde{H} \in \mathbb{R}^{n\times n}$*, достаточно близкой к* H*, любой матрицы* $\tilde{B} \in \mathbb{R}^{l\times n}$*, достаточно близкой к* B*, любого достаточно большого по модулю числа* t *и всякой матрицы* $\Omega \in \mathbb{R}^{l\times n}$*, удовлетворяющей неравенству* $\|\Omega\| \leqslant M/|t|$*, матрица* $\tilde{H} + t(B+\Omega)^{\mathrm{T}}\tilde{B}$ *не вырождена, и при этом*

$$\left\|\left(\tilde{H} + t(B+\Omega)^{\mathrm{T}}\tilde{B}\right)^{-1}(B+\Omega)^{\mathrm{T}}\right\| \leqslant \varepsilon. \tag{88}$$

Доказательство. Зафиксируем произвольные $M > 0$ и $\varepsilon > 0$. Невырожденность матрицы

$\tilde{H} + t(B+\Omega)^{\mathrm{T}}\tilde{B}$ напрямую следует из леммы 3. Таким образом, остается показать, что (возможно, уменьшая расстояние между $\tilde{H}$ и H и между $\tilde{B}$ и B, а также увеличивая $|t|$) можно гарантировать выполнение (88).

От противного: предположим существование последовательностей $\{H_k\} \subset \mathbb{R}^{n\times n}$, $\{B_k\} \subset \mathbb{R}^{l\times n}$, $\{\Omega_k\} \subset \mathbb{R}^{l\times n}$, $\{t_k\} \subset \mathbb{R}$ и $\{\eta^k\} \subset \mathbb{R}^n$ таких, что $\{H_k\} \to H$, $\{B_k\} \to B$, $|t_k| \to \infty$, $\|\Omega_k\| \leqslant M/|t_k|$, $\|\eta^k\| = 1$, и $\det(H_k + t_k(B+\Omega_k)^{\mathrm{T}}B_k) \neq 0$ для всех k, и что вектор

$$\xi^k = (H_k + t_k(B+\Omega_k)^{\mathrm{T}}B_k)^{-1}(B+\Omega_k)^{\mathrm{T}}\eta^k \tag{89}$$

удовлетворяет неравенству

$$\|\xi^k\| > \varepsilon \tag{90}$$

при всех k. Из (89) получаем:

$$(B+\Omega_k)^{\mathrm{T}}\eta^k = H_k\xi^k + t_k(B+\Omega_k)^{\mathrm{T}}B_k\xi^k. \tag{91}$$

Заметим, что из неравенства (90) следует ограниченность последовательности $\{\eta^k/\|\xi^k\|\}$. Кроме того, без ограничения общности можно полагать, что последовательность $\{\xi^k/\|\xi^k\|\}$ сходится к некоторому вектору $\xi \in \mathbb{R}^n$ такому, что $\|\xi\| = 1$. Тогда, разделив обе части соотношения (91) на $t_k\|\xi^k\|$ и перейдя к пределу при $k \to \infty$, получим, что $B^{\mathrm{T}}B\xi = 0$, а значит, $\xi \in \ker B$.

Далее, в силу (91), для всех k выполняется равенство

$$H_k\frac{\xi^k}{\|\xi^k\|} - \Omega_k^{\mathrm{T}}\frac{\eta^k}{\|\xi^k\|} + t_k\Omega_k^{\mathrm{T}}B_k\frac{\xi^k}{\|\xi^k\|} = \frac{1}{\|\xi^k\|}B^{\mathrm{T}}(\eta^k - t_kB_k\xi^k) \in \operatorname{im} B^{\mathrm{T}}.$$

Второе слагаемое в левой части этого равенства стремится к нулю, поскольку $\{\|\Omega_k\|\} \to 0$, в то время как последовательность $\{\eta^k/\|\xi^k\|\}$ ограничена. Более того, третье слагаемое в левой части указанного равенства также стремится к нулю, т. к. последовательность $\{t_k\Omega_k\}$ ограничена, а $\{B_k\xi^k/\|\xi^k\|\} \to B\xi = 0$. Тогда из замкнутости множества $\operatorname{im} B^{\mathrm{T}}$ следует включение $H\xi \in \operatorname{im} B^{\mathrm{T}}$, которое противоречит (64). ∎

Лемма 7. *Пусть, в дополнение к предположениям леммы 3, матрица H симметрична.*

Тогда для любых $M > 0$ и $\varepsilon > 0$ справедливо следующее: для любой матрицы $\tilde{H} \in \mathbb{R}^{n\times n}$, достаточно близкой к H, любого достаточно большого по модулю числа t и всякой матрицы $\Omega \in \mathbb{R}^{l\times n}$, удовлетворяющей неравенству $\|\Omega\| \leqslant M/|t|$, матрица $\tilde{H} + t(B+\Omega)^{\mathrm{T}}\tilde{B}$ не вырождена, и при этом

$$\left\| t(B+\Omega)\left(\tilde{H} + t(B+\Omega)^{\mathrm{T}}(B+\Omega)\right)^{-1}(B+\Omega)^{\mathrm{T}} \right\| \leqslant 1+\varepsilon. \tag{92}$$

Доказательство. Как и при доказательстве предыдущей леммы, невырожденность матрицы $\tilde{H}+t(B+\Omega)^{\mathrm{T}}(B+\Omega)$ является прямым следствием леммы 3. Если при этом оценка (92) не имеет места, то существуют последовательность симметричных матриц $\{H_k\}$ размера $n\times n$ и последовательности $\{\Omega_k\} \in \mathbb{R}^{l\times n}$, $\{t_k\} \in \mathbb{R}$, $\{\eta^k\} \subset \mathbb{R}^n$ такие, что $\{H_k\} \to H$, $|t_k| \to \infty$, и для всех k выполняются соотношения $\|\Omega_k\| \leqslant M/|t_k|$, $\|\eta^k\| = 1$, $\det(H_k+t_k(B+\Omega_k)^{\mathrm{T}}(B+\Omega_k)) \neq 0$, и

$$\left\|t_k(B+\Omega_k)\left(H_k+t_k(B+\Omega_k)^{\mathrm{T}}(B+\Omega_k)\right)^{-1}(B+\Omega_k)^{\mathrm{T}}\eta^k\right\| > 1+\varepsilon. \tag{93}$$

Для каждого k положим

$$\begin{aligned} W_k = (B+\Omega_k)\left(H_k+t_k(B+\Omega_k)^{\mathrm{T}}(B+\Omega_k)\right)^{-1} = \\ = \left(\left(H_k+t_k(B+\Omega_k)^{\mathrm{T}}(B+\Omega_k)\right)^{-1}(B+\Omega_k)^{\mathrm{T}}\right)^{\mathrm{T}}, \end{aligned} \tag{94}$$

где во втором равенстве учтена симметричность матрицы H_k. Из 6 следует, что $\{W_k\} \to 0$.

Далее, для каждого k вектор η^k можно разложить в сумму

$$\eta^k = \eta_1^k + \eta_2^k,$$

где $\eta_1^k \in \ker B^{\mathrm{T}} = (\operatorname{im} B)^{\perp}$ and $\eta_2^k \in \operatorname{im} B$. Заметим, что $t_kW_k(B+\Omega_k)^{\mathrm{T}}\eta_1^k = W_k(t_k\Omega_k^{\mathrm{T}})\eta_1^k$, и, поскольку последовательности $\{\eta_1^k\}$ и $\{t_k\Omega_k\}$ ограничены, а последовательность $\{W_k\}$ стремится к нулю, $\{t_kW_k(B+\Omega_k)^{\mathrm{T}}\eta_1^k\} \to 0$. С другой стороны, так как $\eta_2^k \in \operatorname{im} B$, можно указать вектор $\xi_2^k \in \mathbb{R}^n$ такой, что $B\xi_2^k = \eta_2^k$, и при этом последовательность $\{\xi_2^k\}$ ограничена. Следовательно, с учетом (94),

$$\begin{aligned} \left\|t_kW_k(B+\Omega_k)^{\mathrm{T}}\eta_2^k\right\| &= \left\|W_k(t_k(B+\Omega_k)^{\mathrm{T}})B\xi_2^k\right\| \leqslant \\ &\leqslant \left\|W_k(H_k+t_k(B+\Omega_k)^{\mathrm{T}}(B+\Omega_k))\xi_2^k\right\| + \left\|W_k(H_k+t_k(B+\Omega_k)^{\mathrm{T}}\Omega_k)\xi_2^k\right\| = \\ &= \left\|(B+\Omega_k)\xi_2^k\right\| + \left\|W_k(H_k+t_k(B+\Omega_k)^{\mathrm{T}}\Omega_k)\xi_2^k\right\| \leqslant \\ &\leqslant \|\eta_2^k\| + \|\Omega_k\xi_2^k\| + \left\|W_k(H_k+t_k(B+\Omega_k)^{\mathrm{T}}\Omega_k)\xi_2^k\right\| \leqslant \\ &\leqslant 1 + \|\Omega_k\xi_2^k\| + \left\|W_k(H_k+t_k(B+\Omega_k)^{\mathrm{T}}\Omega_k)\xi_2^k\right\|. \end{aligned}$$

Последние два слагаемых в правой части последней формулы стремятся к нулю, поскольку последовательности $\{\xi_2^k\}$ and $\{H_k+t_k(B+\Omega_k)^{\mathrm{T}}\Omega_k\}$ ограничены, в то время как $\{\Omega_k\} \to 0$ и $\{W_k\} \to 0$. Поэтому

$$\limsup_{k\to\infty}\left\|t_kW_k(B+\Omega_k)^{\mathrm{T}}\eta^k\right\| \leqslant \lim_{k\to\infty}\left\|t_kW_k(B+\Omega_k)^{\mathrm{T}}\eta_1^k\right\| + \limsup_{k\to\infty}\left\|t_kW_k(B+\Omega_k)^{\mathrm{T}}\eta_2^k\right\| \leqslant 1,$$

что противоречит (93). ∎

Предложение 7. *Пусть выполнены предположения леммы 4.*

Тогда для любого $c > 2$ выполнено следующее: для любого $\sigma > 0$ существует окрестность точки $(\bar{x},\, \bar{\lambda})$ такая, что, если точка $(\tilde{x},\, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ лежит в этой окрестности, уравнение

$$\Phi_\sigma(\tilde{\lambda},\, u) = 0 \tag{95}$$

имеет решение $u = (x,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$, удовлетворяющее условию (69).

Доказательство. Для любого $\sigma > 0$ точка $\bar{u} = (\bar{x},\, \bar{\lambda})$ является решением уравнения

$$\Phi_\sigma(\bar{\lambda},\, u) = 0.$$

Кроме того, из замечания 4 и леммы 4 следует, что, если число σ достаточно мало, любая матрица во множестве $\partial_u \Phi_\sigma(\bar{\lambda},\, \bar{u})$ имеет невырожденную подматрицу с невырожденным дополнением Шура, и поэтому любая матрица из $\partial_u \Phi_\sigma(\bar{\lambda},\, \bar{u})$ является невырожденной. Тогда теорема Кларка об обратной функции [21, теорема 7.1.1] гарантирует, что для любого такого σ найдутся окрестность U_σ точки $\bar{u}$ и окрестность V_σ нуля такие, что для любого $r \in V_\sigma$ уравнение

$$\Phi_\sigma(\bar{\lambda},\, u) = r \tag{96}$$

имеет в U_σ единственное решение $u_\sigma(r)$, и при этом функция $u_\sigma(\cdot)\colon V_\sigma \mapsto U_\sigma$ удовлетворяет условию Липшица с некоторой константой ℓ_σ.

Определим величину $r(\tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ формулой

$$r(\tilde{\lambda}) = \begin{pmatrix} 0 \\ -\sigma(\tilde{\lambda} - \bar{\lambda}) \end{pmatrix}.$$

Если вектор $\tilde{\lambda} \in \mathbb{R}^l$ достаточно близок к $\bar{\lambda}$, вектор $r(\tilde{\lambda})$ лежит в V_σ, и значит, уравнение (96) с $r = r(\tilde{\lambda})$ имеет в U_σ единственное решение $u_\sigma(r(\tilde{\lambda}))$. Заметим, что это решение

$$u = u_\sigma(r(\tilde{\lambda})) \tag{97}$$

удовлетворяет (95). Более того, поскольку $u_\sigma(0) = \bar{u}$,

$$\|u - \bar{u}\| = \|u_\sigma(r(\tilde{\lambda})) - u_\sigma(0)\| \leqslant \ell_\sigma \|r(\tilde{\lambda})\| = \ell_\sigma \sigma \|\tilde{\lambda} - \bar{\lambda}\|. \tag{98}$$

Теперь покажем, что для любого достаточно малого $\sigma > 0$ существует окрестность точки $(\bar{x},\, \bar{\lambda})$ такая, что для любого вектора $\tilde{u} = (\tilde{x},\, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ из этой окрестности, точка u, определенная формулой, удовлетворяет оценке (69). Предположим, что это не так. Тогда найдутся $c > 2$, $M > 0$ и последовательности $\{\sigma_k\} \subset \mathbb{R}_+ \setminus \{0\}$ и $\{\tilde{u}^k\} \subset \mathbb{R}^n \times \mathbb{R}^l$, $\tilde{u}^k = (\tilde{x}^k,\, \tilde{\lambda}^k)$, такие, что $\sigma_k \to 0$, $\{\tilde{u}^k\} \to \bar{u}$, при всех k справедливо неравенство $\ell_{\sigma_k}\|\tilde{u}^k - \bar{u}\| \leqslant M$, и при этом $u^k = u_{\sigma_k}(r(\tilde{\lambda}^k))$ не удовлетворяет (69), т.е.

$$\|u^k - \tilde{u}^k\| > c\,\mathrm{dist}(\tilde{u}^k,\, \{\bar{x}\} \times \mathcal{M}(\bar{x})). \tag{99}$$

Тогда из (98) следует, что для всех k выполняется неравенство

$$\|x^k - \bar{x}\| \leqslant \sigma_k M. \tag{100}$$

Кроме того, принимая во внимание, что $\Phi_{\sigma_k}(\tilde{\lambda}^k,\, u^k) = 0$, и $\Phi(\bar{x},\, \pi(\tilde{\lambda}^k)) = 0$, можем записать

$$\Phi_{\sigma_k}(\tilde{\lambda}^k,\, u^k) - \Phi_{\sigma_k}(\tilde{\lambda}^k,\, (\bar{x},\, \pi(\tilde{\lambda}^k))) = \left(0,\, -\sigma_k(\tilde{\lambda}^k - \pi(\tilde{\lambda}^k))\right).$$

Применяя теорему о среднем (см., например, [32, предложение 7.1.16]) и привлекая замечание 4, получаем существование точек $u^{k,\,i}$ из отрезка, соединяющего u^k и $(\bar{x},\, \pi(\tilde{\lambda}^k))$, чисел $\alpha_{k,\,i} \geqslant 0$ и матриц $H_{k,\,i} \in \partial_x \frac{\partial L}{\partial x}(x^{k,\,i},\, \lambda^{k,\,i})$, $i = 1,\, \ldots,\, n$, таких, что $\sum_{i=1}^n \alpha_{k,\,i} = 1$, и

$$\begin{pmatrix} \displaystyle\sum_{i=1}^n \alpha_{k,\,i}\, H_{k,\,i} & \left(\displaystyle\sum_{i=1}^n \alpha_{k,\,i}\, h'(x^{k,\,i})\right)^{\mathrm{T}} \\ \displaystyle\sum_{i=1}^n \alpha_{k,\,i}\, h'(x^{k,\,i}) & -\sigma_k I \end{pmatrix} \begin{pmatrix} x^k - \bar{x} \\ \lambda^k - \pi(\tilde{\lambda}^k) \end{pmatrix} = \begin{pmatrix} 0 \\ -\sigma_k(\tilde{\lambda}^k - \pi(\tilde{\lambda}^k)) \end{pmatrix}$$

для всех достаточно больших k. Так как $\{\sigma_k\} \to 0$, $\{u^k\} \to \bar{u}$, и выполнено (100), из леммы 4 следует, что для всех достаточно больших k матрица в левой части последней формулы является невырожденной (как матрица, содержащая невырожденную подматрицу с невырожденным дополнением Шура). Тогда

$$\begin{pmatrix} x^k - \bar{x} \\ \lambda^k - \pi(\tilde{\lambda}^k) \end{pmatrix} = \begin{pmatrix} H_k & B_k^{\mathrm{T}} \\ B_k & -\sigma_k I \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ -\sigma_k(\tilde{\lambda}^k - \pi(\tilde{\lambda}^k)) \end{pmatrix},$$

где через H_k и B_k обозначены $\sum_{i=1}^n \alpha_{k,\,i} H_{k,\,i}$ и $\sum_{i=1}^n \alpha_{k,\,i} h'(x^{k,\,i})$ соответственно. Выражая обратную матрицу в последней формуле через дополнение Шура подматрицы $-\sigma_k I$ (см., например, [14, раздел 1.2]), получаем, что

$$\begin{pmatrix} x^k - \bar{x} \\ \lambda^k - \pi(\tilde{\lambda}^k) \end{pmatrix} = \begin{pmatrix} \left(H_k + \dfrac{1}{\sigma_k} B_k^{\mathrm{T}} B_k\right)^{-1} B_k^{\mathrm{T}} \\ -I + \dfrac{1}{\sigma_k} B_k \left(H_k + \dfrac{1}{\sigma_k} B_k^{\mathrm{T}} B_k\right)^{-1} B_k^{\mathrm{T}} \end{pmatrix} \left(\tilde{\lambda}^k - \pi(\tilde{\lambda}^k)\right). \tag{101}$$

Для каждого $i = 1,\, \ldots,\, n$, поскольку отображения f' и h' локально липшицевы в точке $\bar{x}$, отображение $\frac{\partial L}{\partial x}(\cdot,\, \lambda^{k,\,i})$ локально липшицево в точке $x^{k,\,i}$ для всех достаточно больших k, и более того, в силу ограниченности последовательности $\{\lambda^{k,\,i}\}$, соответствующую константу Липшица можно выбрать одинаковой для всех таких k. Поскольку норма всех матриц дифференциала Кларка отображения в точке ограничена константой Липшица, с которой это отображение локально липшицево в этой точке, отсюда следует, что последовательности $\{H_{k,\,i}\}$ ограничены, и поэтому без ограничения общности можно считать, что они сходятся к некоторым матрицам $H_i \in \mathbb{R}^{n\times n}$ при $k\to\infty$. Тогда, пользуясь леммой 1.5 и полунепрерывностью дифференциала Кларка сверху, заключаем, что $H_i \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\,\bar{\lambda})$. Кроме того, без ограничения общности можно считать, что $\alpha_{k,\,i}$ стремятся к некоторым числам $\alpha_i \geqslant 0$. Тогда $\sum_{i=1}^{n}\alpha_i = 1$, и, полагая $H = \sum_{i=1}^{n}\alpha_i H_i$, имеем: $\{H_k\}\to H$ и $H \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\,\bar{\lambda})$.

Далее, пользуясь локальной липшицевостью отображения h' в точке $\bar{x}$ и неравенством (100), получаем, что

$$\|B_k - h'(\bar{x})\| = \left\|\sum_{i=1}^{n}\alpha_{k,\,i}(h'(x^{k,\,i}) - h'(\bar{x}))\right\| \leqslant \sum_{i=1}^{n}\alpha_{k,\,i}\|h'(x^{k,\,i}) - h'(\bar{x})\| =$$
$$= O\left(\sum_{i=1}^{n}\alpha_{k,\,i}\|x^{k,\,i} - \bar{x}\|\right) = O\left(\sum_{i=1}^{n}\alpha_{k,\,i}\|x^{k} - \bar{x}\|\right) = O(\sigma_k)$$

при $k\to\infty$. Теперь, применяя лемму 6 с $B = h'(\bar{x})$, $\tilde{H} = J_k$, $\tilde{B} = B_k$, $\Omega = B_k - h'(\bar{x})$, и $t = 1/\sigma_k$, заключаем, что

$$\left\{\left(H_k + \frac{1}{\sigma_k}B_k^{\mathrm{T}}B_k\right)^{-1}B_k^{\mathrm{T}}\right\}\to 0. \tag{102}$$

Наконец, учитывая симметричность матриц H и H_k для всех k и используя лемму 7 с B, $\tilde{H}$, Ω и t такими же, как и при применении леммы 6, получаем, что

$$\left\|\frac{1}{\sigma_k}B_k\left(H_k + \frac{1}{\sigma_k}B_k^{\mathrm{T}}B_k\right)^{-1}B_k^{\mathrm{T}}\right\|\to 1. \tag{103}$$

Объединяя (101) с (102) и (103), заключаем, что для любого фиксированного $\tilde{c} > 1$ неравенства

$$\|u^k - (\bar{x},\,\pi(\tilde{\lambda}^k))\| \leqslant \tilde{c}\|\tilde{\lambda}^k - \pi(\tilde{\lambda}^k)\| \leqslant \tilde{c}\|\tilde{u}^k - (\bar{x},\,\pi(\tilde{\lambda}^k))\|$$

выполняются для всех достаточно больших k. Тогда из оценки расстояния (62) следует, что для всех достаточно больших k справедливы соотношения

$$\|u^k - \tilde{u}^k\| \leqslant \|u^k - (\bar{x},\,\pi(\tilde{\lambda}^k))\| + \|\tilde{u}^k - (\bar{x},\,\pi(\tilde{\lambda}^k))\| \leqslant (1+\tilde{c})\,\mathrm{dist}((\tilde{x}^k,\,\tilde{\lambda}^k),\,\{\bar{x}\}\times\mathcal{M}(\bar{x})).$$

И мы приходим к противоречию с (99). ■

Отметим, что, если функция f и отображение h дважды непрерывно дифференцируемы в точке $\bar{x}$, можно показать, что предложение 7 остается верным, если заменить условие $c > 2$ на $c > 1$ в его формулировке.

Тем самым, в сделанных предположениях для метода модифицированных функций с фиксированным параметром штрафа выполняется предположение 3) теоремы 2.3. С учетом представленных выше обсуждений, касающихся выполнения для метода модифицированных функций предположений 1) и 2) этой теоремы, получаем следующий результат.

Теорема 5. *Пусть выполнены условия леммы 4, и пусть функция* $\tau\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}_+$ *удовлетворяет условию* (61).

Тогда для любого $c > 2$ *существует число* $\bar{\sigma} > 0$ *такое, что для всякого* $\sigma \in (0, \bar{\sigma})$ *справедливо следующее: для любого начального приближения* $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^l$*, достаточно близкого к* $(\bar{x}, \bar{\lambda})$*, существует последовательность* $\{(x^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$*, генерируемая методом модифицированных функций с* $\sigma_k = \sigma$ *для всех* k*, и с* τ_k*, вычисляемым согласно* (58)*, которая удовлетворяет условию* (18) *при всех* k*; любая такая последовательность сходится к* $(\bar{x}, \lambda^*)$*, где* $\lambda^* \in \mathcal{M}(\bar{x})$*, и при этом скорости сходимости последовательности* $\{(x^k, \lambda^k)\}$ *к* $(\bar{x}, \lambda^*)$ *и последовательности* $\{\mathrm{dist}((x^k, \lambda^k), \{\bar{x}\} \times \mathcal{M}(\bar{x}))\}$ *к нулю линейные. Кроме того, для любого* $\varepsilon > 0$ *выполняется неравенство* $\|\lambda^* - \bar{\lambda}\| < \varepsilon$*, если начальное приближение* (x^0, λ^0) *достаточно близко* $(\bar{x}, \bar{\lambda})$*.*

Вернемся к рассмотрению задачи (1), т. е. задачи со смешанными ограничениями. Сначала покажем, что без дополнительных предположений анализ, проведенный в предшествующей части настоящего пункта, на случай таких задач обобщить нельзя.

Будем рассматривать метод модифицированных функций Лагранжа (7), (8), в котором значение обратного параметра штрафа σ_k и значение параметра неточности τ_k выбираются по текущему приближению, т. е.

$$\sigma_k = \sigma(x^k, \lambda^k, \mu^k), \quad \tau_k = \tau(x^k, \lambda^k, \mu^k),$$

где $\sigma\colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \mapsto \mathbb{R}_+$ и $\tau\colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \mapsto \mathbb{R}_+$ — некоторые функции, причем функция $\sigma(\cdot)$ равна нулю тогда и только тогда, когда ее аргумент является решением системы ККТ (2). Будем считать, что если в текущем приближении $u^k = (x^k, \lambda^k, \mu^k)$ функция $\sigma(\cdot)$ равна нулю, то метод останавливается, и $u^{k+j} = u^k$ для всех $j \geqslant 1$. Если дополнить такой метод условием локализации (49), то он будет эквивалентен итерационной схеме (2.17) с отображением $\mathcal{A}\colon \mathbb{R}^\nu \times \mathbb{R}^\nu \mapsto 2^{\mathbb{R}^\nu}$, $\nu = n + l + m$, вида

$$\mathcal{A}(\tilde{u}, u) = \biggl(\frac{\partial L}{\partial x}(x, \lambda, \mu) + B(0, \tau(\tilde{x}, \tilde{\lambda}, \tilde{\mu})),$$
$$h(x) - \sigma(\tilde{x}, \tilde{\lambda}, \tilde{\mu})(\lambda - \tilde{\lambda}), -g(x) + \sigma(\tilde{x}, \tilde{\lambda}, \tilde{\mu})(\mu - \tilde{\mu})\biggr)$$

и отображением N, заданным согласно (6).

Как уже отмечалось выше, система Каруша–Куна–Таккера (2) эквивалентна обобщенному уравнению (4) с отображениями Φ и N, заданными согласно (5) и (6) соответственно. При этом, если тройка $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ является решением системы ККТ и множитель $(\bar{\lambda}, \bar{\mu})$ является некритическим, то, в силу следствия 1.1, для $\bar{u}$ как для решения этого обобщенного уравнения выполняется предположение 1) теоремы 2.3.

Проверка предположения 2) теоремы 2.3 для рассматриваемого метода в том случае, если множитель $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ является некритическим, а функция τ удовлетворяет условию (48), также не составляет труда. Однако, как показывает следующий пример, предположение 3) теоремы 2.3 в случае задач со смешанными ограничениями для метода модифицированных функций, вообще говоря, не выполняется.

Пример 3. Пусть $n = 1$, $l = 0$, $m = 2$, $f(x) = -x^2/2$, $g(x) = (-x,\, x^3/6)$. Тогда $\bar{x} = 0$ — единственная стационарная точка задачи (1), и $\mathcal{M}(\bar{x}) = \{\mu \in \mathbb{R}^2 \mid \mu_1 = 0,\, \mu_2 \geqslant 0\}$. Множитель $\bar{\mu} = 0$ является (строго) некритическим.

Легко видеть, что итерация (7), (8) применительно к рассматриваемой задаче эквивалентна решению следующей системы

$$\begin{gathered} -x - \mu_1 + \frac{1}{2}x^2\mu_2 = 0, \\ \mu_1 \geqslant 0, \quad x + \sigma(\mu_1 - \tilde{\mu}_1) \geqslant 0, \quad \mu_1\,(x + \sigma(\mu_1 - \tilde{\mu}_1)) = 0, \\ \mu_2 \geqslant 0, \quad \frac{1}{6}x^3 - \sigma(\mu_2 - \tilde{\mu}_2) \leqslant 0, \quad \mu_2\left(\frac{1}{6}x^3 - \sigma(\mu_2 - \tilde{\mu}_2)\right) = 0 \end{gathered} \tag{104}$$

относительно (x, μ), где $\tilde{\mu}$ — текущее двойственное приближение, а σ — текущее значение обратного параметра штрафа.

Пусть $\tilde{\mu}_1 > 0$ и $\tilde{\mu}_2 = 0$.

1. Если $\mu_1 = \mu_2 = 0$, то из первого соотношения в (104) следует, что $x = 0$. Но тогда вторая строка в (104) противоречит предположению о том, что $\tilde{\mu}_1 > 0$.

2. Если $\mu_1 > 0$, $\mu_2 = 0$, то первое соотношение в (104) влечет за собой равенство $x = -\mu_1$, и поэтому, в силу соотношений во второй строке в (104),

$$-\mu_1(1 - \sigma) = \sigma\tilde{\mu}_1 > 0,$$

что не может выполняться, если $\sigma \leqslant 1$.

3. Если $\mu_1 = 0$, $\mu_2 > 0$, то из первого соотношения в (104) следует, что либо $x = 0$, либо $x\mu_2 = 2$. В первом случае вторая строка в (104) дает выполнение неравенства

$-\sigma\tilde{\mu}_1 \geqslant 0$, которое противоречит предположению $\tilde{\mu}_1 > 0$. С другой стороны, второй случай невозможен, если пара $(x,\, \mu)$ достаточно близка к $(\bar{x},\, \bar{\mu})$.

4. Наконец, если $\mu_1 > 0$ и $\mu_2 > 0$, то из третьей строки в (104) получаем, что $\mu_2 = x^3/(6\sigma)$, и следовательно, $x > 0$. Более того, из первой строки в (104) следует, что $\mu_1 = -x + x^2\mu_2/2$, а значит, $\mu_1 \leqslant -x + x^2/2 < 0$, если $\mu_2 \leqslant 1$ и $x \in (0,\, 2)$.

Таким образом, в любой окрестности множителя $\bar{\mu}$ существует вектор $\tilde{\mu}$ такой, что система (104) не имеет решений в некоторой фиксированной окрестности точки $(\bar{x},\, \bar{\lambda})$, если параметр $\sigma > 0$ достаточно мал. Следовательно, предположение 3) теоремы 2.3 не может выполняться ни с каким $c > 0$ для точного метода модифицированных функций ни в случае, если значение обратного параметра штрафа σ_k выбирается таким образом, что оно стремится к нулю, когда текущее приближение стремится к $(\bar{x},\, \bar{\mu})$, ни в случае, если обратный параметр штрафа зафиксирован на достаточно малом уровне.

Однако, в примере 3 нарушено условие строгой дополнительности: $\bar{\mu} = 0$ (на самом деле, $A_0(\bar{x},\, \mu) \neq 0$ для любого $\mu \in \mathcal{M}(\bar{x})$ в этом примере). Если предположить выполнение условия строгой дополнительности $\bar{\mu}_A > 0$, то ситуация из примера 3 оказывается невозможной.

При выполнении условия строгой дополнительности система ККТ (2) локально (вблизи $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$) эквивалентна системе уравнений

$$f'(x) + (h'(x))^{\mathrm{T}}\lambda + (g'_A(x))^{\mathrm{T}}\mu_A = 0, \quad h(x) = 0, \quad g_A(x) = 0, \tag{105}$$

с дополнительным уравнением $\mu_{\{1,\, \ldots,\, m\}\setminus A} = 0$. Система (105) представляет собой систему Лагранжа задачи оптимизации

$$\begin{gathered} f(x) \to \min, \\ h(x) = 0,\ g_A(x) = 0. \end{gathered} \tag{106}$$

При выполнении условия строгой дополнительности множитель $(\bar{\lambda},\, \bar{\mu}) \in \mathcal{M}(\bar{x})$ является (строго) некритическим тогда и только тогда, когда множитель $(\bar{\lambda},\, \bar{\mu}_A)$, отвечающий $\bar{x}$ в задаче (106), строго некритический. Поэтому, если множитель $(\bar{\lambda},\, \bar{\mu})$ строго некритический, то для метода модифицированных функций, применяемого к задаче (106), справедливы установленные ранее результаты о разрешимости подзадач. В случае фиксированного параметра штрафа отсюда можно получить разрешимость подзадач метода модифицированных функций, применяемого к задаче (1), и обобщить теорему 5 на задачи со смешанными ограничениями.

3.2. Метод множителей с линеаризованными ограничениями

Метод множителей с линеаризованными ограничениями [37, 68, 79] традиционно вводится для задачи вида (1), в которой $m = n$ и $g(x) = -x$, $x \in \mathbb{R}^n$, т. е. для задачи с ограничениями равенствами и условием неотрицательности переменных:

$$f(x) \to \min, \quad h(x) = 0, \; x \geqslant 0. \tag{107}$$

По текущему приближению $(x^k, \lambda^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ и значению параметра штрафа $s_k \geqslant 0$ очередное прямое приближение $x^{k+1} \in \mathbb{R}^n$ вычисляется как стационарная точка задачи оптимизации

$$f(x) + \langle \lambda^k, h(x) \rangle + \frac{s_k}{2} \|h(x)\|_2^2 \to \min, \quad h(x^k) + h'(x^k)(x - x^k) = 0, \; x \geqslant 0, \tag{108}$$

а очередная точка $(\lambda^{k+1}, \mu^{k+1}) \in \mathbb{R}^l \times \mathbb{R}^n$ двойственной траектории выбирается таким образом, чтобы пара $(\lambda^{k+1} - \lambda^k, \mu^{k+1})$ была отвечающим x^{k+1} множителем Лагранжа.

Именно этот метод лежит в основе чрезвычайно успешного пакета MINOS [69]. Отметим, что, если $s_k > 0$, целевая функция задачи (108) представляет собой модифицированную функцию Лагранжа с $\sigma_k = 1/s_k$, включающую в себя только ограничения-равенства задачи (107).

В случае двукратной дифференцируемости функции f и отображения h наиболее тонкий известный результат о локальной сходимости данного метода получен в [51], где установлена локальная сверхлинейная сходимость в предположении о единственности множителя и при выполнении достаточного условия второго порядка. В конце настоящего раздела этот результат будет не только перенесен на задачи оптимизации с липшицевыми производными, но и усилен: сверхлинейная оценка скорости сходимости будет заменена квадратичной.

При анализе локальной сходимости метода множителей с линеаризованными ограничениями естественно считать, что параметр штрафа в (108) не меняется: $s_k = s \geqslant 0$ для всех k (см. обсуждение этого вопроса в [37]; в оригинальной работе [79] полагалось $s_k = 0$ для всех k).

Система ККТ задачи (108) имеет вид

$$f'(x) + (h'(x))^{\mathrm{T}}(\lambda + \lambda^k) - \mu + s(h'(x))^{\mathrm{T}} h(x) = 0,$$

$$h(x^k) + h'(x^k)(x - x^k) = 0, \quad \mu \geqslant 0, \quad x \geqslant 0, \quad \langle \mu, x \rangle = 0.$$

Отсюда и из правила для λ^{k+1} следует, что метод множителей с линеаризованными ограничениями представляет собой частный случай схемы (2.2) при $\nu = n+l+n$, $\mathcal{A}\colon \mathbb{R}^\nu \times \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$,

$$\mathcal{A}(\tilde{u},\, u) = \left(f'(x) + (h'(x))^{\mathrm{T}}\lambda - \mu + s(h'(x))^{\mathrm{T}}h(x),\, h(\tilde{x}) + h'(\tilde{x})(x-\tilde{x}),\, x\right), \tag{109}$$

и при $N(\cdot)$, определенном согласно (6).

Для того чтобы применить теорему 2.1 в данном случае, переопределим отображение Φ следующим образом:

$$\Phi(u) = \left(f'(x) + (h'(x))^{\mathrm{T}}\lambda - \mu + s(h'(x))^{\mathrm{T}}h(x),\, h(x),\, x\right). \tag{110}$$

При этом система ККТ задачи (107) остается эквивалентной обобщенному уравнению (4) с отображением $N(\cdot)$, определенным согласно (6), для любого фиксированного s. Это отражает хорошо известный факт, что в условиях оптимальности вместо обычной функции Лагранжа можно использовать модифицированную. С другой стороны, такое обобщенное уравнение совпадает с обобщенным уравнением, отвечающим системе ККТ задачи

$$f(x) + \frac{s}{2}\|h(x)\|_2^2 \to \min, \quad h(x) = 0,\ x \geqslant 0. \tag{111}$$

Если тройка $(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ является решением системы ККТ задачи (107), удовлетворяющим условию линейной независимости и сильному достаточному условию второго порядка (17), то, как легко видеть, эти же свойства справедливы и для задачи (111). Поэтому, в силу сказанного в предыдущем разделе, указанные условия гарантируют сильную метрическую регулярность решения $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ обобщенного уравнения (4), в котором Φ определено согласно (110), а N — согласно (6).

Далее, легко видеть, что $\mathcal{A}(\tilde{u},\, \tilde{u}) = \Phi(\tilde{u})$, и

$$\begin{aligned}\|(\Phi(u^1) - \mathcal{A}(\tilde{u},\, u^1)) - (\Phi(u^2) - \mathcal{A}(\tilde{u},\, u^2))\| \leqslant \|(h'(x^1) - h'(\tilde{x}))^{\mathrm{T}}(\lambda^1 - \lambda^2)\| +\\ + \|(h'(x^1) - h'(x^2))^{\mathrm{T}}(\lambda^2 - \tilde{\lambda})\| + \|h(x^1) - h(x^2) - h'(\tilde{x})(x^1 - x^2)\| \leqslant\\ \leqslant \omega(\tilde{u},\, u^1,\, u^2)\|u^1 - u^2\|,\end{aligned}$$

где

$$\omega(\tilde{u},\, u^1,\, u^2) = O\left(\sup_{t\in[0,\,1]} \|h'(tx^1 + (1-t)x^2) - h'(\tilde{x})\| + \|\lambda^2 - \tilde{\lambda}\|\right). \tag{112}$$

Для того чтобы завершить проверку предположения 2) теоремы 2.1, остается заметить, что $\omega(\tilde{u},\, u^1,\, u^2) \to 0$ при $\tilde{u},\, u^1,\, u^2 \to \bar{u}$ в силу непрерывности отображения h' в точке $\bar{x}$.

Кроме того, из (112) и локальной липшицевости отображения h' в точке $\bar{x}$ следует, что

$$\omega(\tilde{u},\, \tilde{u},\, u) = O(\|u - \tilde{u}\|). \tag{113}$$

Если $\{u^k\} \subset \mathbb{R}^\nu$ сходится к $\bar{u}$ (сверх)линейно, то $\|u^{k+1} - u^k\| = O(\|u^k - \bar{u}\|)$, и соотношение (113) влечет оценку $\omega(u^k,\, u^k,\, u^{k+1}) = O(\|u^k - \bar{u}\|)$.

Теорема 6. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображение* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *дифференцируемы в окрестности точки* $\bar{x} \in \mathbb{R}^n$*, причем их производные локально липшицевы в этой точке. Пусть* $\bar{x}$ *является стационарной точкой задачи* (107)*, удовлетворяющей условию линейной независимости, и пусть сильное достаточное условие второго порядка* (17) *выполняется для единственного отвечающего этой стационарной точке множителя Лагранжа* $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^n$.

Тогда для любого $s \geqslant 0$ *существует число* $\delta > 0$ *такое, что для любого начального приближения* $(x^0,\, \lambda^0,\, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$*, достаточно близкого к* $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$*, существует единственная последовательность* $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ *такая, что для каждого* $k = 0,\, 1,\, \ldots$ *точка* x^{k+1} *является стационарной в задаче* (108)*, пара* $(\lambda^{k+1} - \lambda^k,\, \mu^{k+1})$ *является отвечающим ей множителем Лагранжа, и выполнено неравенство* (18)*; эта последовательность сходится к* $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$*, и скорость сходимости квадратичная.*

Как и в случае метода модифицированных функций Лагранжа, локальную сходимость метода множителей с линеаризованными ограничениями можно установить в более слабых предположениях, чем условие линейной независимости и сильное достаточное условие второго порядка, если воспользоваться теоремой 2.2. Теорема, полученная на этом пути, не будет, однако, гарантировать единственность траектории метода.

Пусть отображения Φ и $\mathcal{A}$ определены согласно (110) и, соответственно, (109). Аналогично обсуждению после формулы (111) можно показать, что строгое условие Мангасариана–Фромовица и достаточное условие второго порядка (53) гарантируют полуустойчивость тройки $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ как решения обобщенного уравнения (4) с указанным отображением Φ и отображением N, заданным формулой (6).

Кроме того, в силу теоремы о среднем,

$$\|\Phi(u) - \mathcal{A}(\tilde{u},\, u)\| = O\left(\sup_{t \in [0,\,1]} \|h'(tx + (1-t)\tilde{x}) - h'(\tilde{x})\|\right) \|u - \tilde{u}\|.$$

Поэтому предположение 2) теоремы 2.2 выполняется, и остается проверить предположение 3).

Предложение 8. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображение* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *дифференцируемы в окрестности точки* $\bar{x} \in \mathbb{R}^n$*, и их производные непрерывны в этой точке. Пусть*

$\bar{x}$ — стационарная точка задачи (107)*, удовлетворяющая строгому условию Мангасариана–Фромовица, и пусть достаточное условие второго порядка* (53) *выполняется для единственного отвечающего ей множителя Лагранжа $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^n$.*

Тогда для любого $s \geqslant 0$ и любого $\varepsilon > 0$ существует число $\tilde{\varepsilon} > 0$ такое, что для любой пары $(\tilde{x},\, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$, удовлетворяющей неравенству $\|(\tilde{x} - \bar{x},\, \tilde{\lambda} - \bar{\lambda})\| \leqslant \tilde{\varepsilon}$, существует стационарная точка x задачи оптимизации

$$f(x) + \langle h(x),\, \tilde{\lambda}\rangle + \frac{s}{2}\|h(x)\|_2^2 \to \min \quad h(\tilde{x}) + h'(\tilde{x})(x - \tilde{x}) = 0,\ x \geqslant 0, \tag{114}$$

удовлетворяющая условию $\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\| \leqslant \varepsilon$ с произвольной парой $(\lambda,\, \mu) \in \mathbb{R}^l \times \mathbb{R}^n$ такой, что $(\lambda - \tilde{\lambda},\, \mu)$ — множитель Лагранжа, отвечающий точке x.

Доказательство. Задачу (114) можно рассматривать как параметрическую, в которой $\tilde{x}$ и $\tilde{\lambda}$ играют роль параметров с базовыми значениями $\bar{x}$ и $\bar{\lambda}$ соответственно. Непосредственно проверяется, что $\bar{x}$ — стационарная точка задачи

$$f(x) + \langle h(x),\, \bar{\lambda}\rangle + \frac{s}{2}\|h(x)\|_2^2 \to \min \quad h(\bar{x}) + h'(\bar{x})(x - \bar{x}) = 0,\ x \geqslant 0, \tag{115}$$

т. е. задачи (114) при этих базовых значениях параметров. Кроме того, легко видеть, что точка $\bar{x}$ как стационарная точка задачи (115) удовлетворяет строгому условию Мангасариана–Фромовица, и что единственный отвечающий ей множитель Лагранжа равен $(0,\, \bar{\mu})$. Пусть $\bar{L}\colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n \mapsto \mathbb{R}$ — функция Лагранжа задачи (115), т. е.

$$\bar{L}(x,\, \lambda,\, \mu) = f(x) + \langle h(x),\, \bar{\lambda}\rangle + \frac{s}{2}\|h(x)\|_2^2 + \langle \lambda,\, h(\bar{x}) + h'(\bar{x})(x - \bar{x})\rangle - \langle \mu,\, x\rangle.$$

Имеем

$$\partial_x \frac{\partial \bar{L}}{\partial x}(\bar{x},\, 0,\, \bar{\mu}) = \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) + s(h'(\bar{x}))^{\mathrm{T}} h'(\bar{x}).$$

Поскольку критические конусы задач (107) и (115) совпадают, из последнего равенства следует, что в стационарной точке $\bar{x}$ задачи (115) для отвечающего ей множителя $(0,\, \bar{\mu})$ выполнено достаточное условие второго порядка. И, в частности, точка $\bar{x}$ является строгим локальным решением задачи (115). Поскольку строгое условие Мангасариана–Фромовица влечет за собой выполнение обычного условия Мангасариана–Фромовица, по теореме устойчивости Робинсона [80] следует, что для любой пары $(\tilde{x},\, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$, достаточно близкой к $(\bar{x},\, \bar{\lambda})$, задача (114) имеет локальное решение $x(\tilde{x},\, \tilde{\lambda})$ такое, что $x(\tilde{x},\, \tilde{\lambda}) \to \bar{x}$ при $(\tilde{x},\, \tilde{\lambda}) \to (\bar{x},\, \bar{\lambda})$. Кроме того, так как условие Мангасариана–Фромовица устойчиво к малым возмущениям (см., например, [19, замечание 2.88]), для всех пар $(\tilde{x}, \tilde{\lambda})$, достаточно близких к $(\bar{x}, \bar{\lambda})$, локальное решение $x = x(\tilde{x}, \tilde{\lambda})$ задачи (114) удовлетворяет условию Мангасариана–Фромовица, и следовательно, является стационарной точкой этой задачи (см., к примеру, [19, теорема 3.9]).

Наконец, рассуждая аналогично тому, как это делалось при доказательстве предложения 4, можно убедиться в том, что при любом выборе пары $(\lambda, \mu) \in \mathbb{R}^l \times \mathbb{R}^n$ такой, что $(\lambda - \tilde{\lambda}, \mu)$ — множитель Лагранжа, отвечающий точке $x(\tilde{x}, \tilde{\lambda})$, выполнено $(\lambda, \mu) \to (\bar{\lambda}, \bar{\mu})$ при $(\tilde{x}, \tilde{\lambda}) \to (\bar{x}, \bar{\lambda})$. ■

Применяя теорему 2.2 и предложение 8, получаем следующее.

Теорема 7. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображение* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *дифференцируемы в окрестности точки* $\bar{x} \in \mathbb{R}^n$, *и их производные локально липшицевы в этой точке. Пусть* $\bar{x}$ *— стационарная точка задачи* (107), *удовлетворяющая строгому условию Мангасариана–Фромовица, и пусть достаточное условие второго порядка* (53) *выполняется для единственного отвечающего этой стационарной точке множителя Лагранжа* $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^n$.

Тогда для любого $s \geqslant 0$ *существует число* $\delta > 0$ *такое, что для любого начального приближения* $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$, *достаточно близкого к* $(\bar{x}, \bar{\lambda}, \bar{\mu})$, *существует последовательность* $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ *такая, что для каждого* $k = 0, 1, \ldots$ *точка* x^{k+1} *является стационарной в задаче* (108), *пара* $(\lambda^{k+1} - \lambda^k, \mu^{k+1})$ *является отвечающим ей множителем Лагранжа, и выполнено неравенство* (18); *любая такая последовательность сходится к* $(\bar{x}, \bar{\lambda}, \bar{\mu})$, *и скорость сходимости является квадратичной.*

3.3. Полугладкий метод последовательного квадратичного программирования

Данный раздел посвящен установлению необходимых и достаточных условий для прямой сверхлинейной сходимости метода последовательного квадратичного программирования [16] для задачи (1) в случае, когда отображения f', h' и g' являются полугладкими. Метод последовательного квадратичного программирования является оптимизационным частным случаем метода Джозефи–Ньютона, представленного в разделе 2.2, и состоит в следующем. По текущему приближению (x^k, λ^k, μ^k) к решению системы Каруша-Куна-Таккера (2) и текущей симметричной итерационной матрице $H_k \in \mathbb{R}^{n \times n}$ очередное прямое приближение $x^{k+1} \in \mathbb{R}^n$

вычисляется как стационарная точка задачи

$$\begin{gathered}\langle f'(x^k),\, x - x^k\rangle + \frac{1}{2}\langle H_k(x - x^k),\, x - x^k\rangle \to \min,\\ h(x^k) + h'(x^k)(x - x^k) = 0,\ g(x^k) + g'(x^k)(x - x^k) \leqslant 0,\end{gathered} \tag{116}$$

а очередное двойственное приближение $(\lambda^{k+1},\, \mu^{k+1}) \in \mathbb{R}^l \times \mathbb{R}^m_+$ — как отвечающий ей множитель Лагранжа.

Полугладким методом последовательного квадратичного программирования будем называть метод последовательного квадратичного программирования, в котором итерационная матрица выбирается по правилу

$$H_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k).$$

Методы такого типа рассматривались, например, в [39, 76].

Система ККТ задачи (116) имеет вид

$$\begin{gathered}f'(x^k) + H_k(x - x^k) + (h'(x^k))^{\mathrm{T}}\lambda + (g'(x^k))^{\mathrm{T}}\mu = 0,\\ h(x^k) + h'(x^k)(x - x^k) = 0,\\ \mu \geqslant 0, \quad g(x^k) + g'(x^k)(x - x^k) \leqslant 0, \quad \langle \mu,\, g(x^k) + g'(x^k)(x - x^k)\rangle = 0.\end{gathered} \tag{117}$$

Легко видеть, что решение такой системы эквивалентно осуществлению итерации метода Джозефи–Ньютона (2.27) с $u^k = (x^k,\, \lambda^k,\, \mu^k)$, отображениями $\Phi(\cdot)$ и $N(\cdot)$, определенными согласно (5) и (6) соответственно, и с

$$J_k = \begin{pmatrix} H_k & (h'(x^k))^{\mathrm{T}} & (g'(x^k))^{\mathrm{T}} \\ -h'(x^k) & 0 & 0 \\ -g'(x^k) & 0 & 0 \end{pmatrix}. \tag{118}$$

Полугладкость отображений f', h' и g' в точке $\bar{x}$ влечет за собой полугладкость отображения Φ в точке $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu})$. Кроме того, принимая во внимание предложение 1 и замечание 1, мы можем применить теорему 2.5 с $\bar{\Delta} = \partial\Phi(\bar{x})$ и $\Delta(\cdot) = \partial\Phi(\cdot)$, если точка $\bar{x}$ и множитель $(\bar{\lambda},\, \bar{\mu})$ удовлетворяют условию линейной независимости и сильному достаточному условию второго порядка (17). Отсюда немедленно следует результат о локальной сходимости и скорости сходимости полугладкого метода последовательного квадратичного программирования.

Теорема 8. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображения* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *и* $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ *дифференцируемы в окрестности точки* $\bar{x} \in \mathbb{R}^n$*, и их производные полугладки в этой точке. Пусть* $\bar{x}$ *— стационарная точка задачи* (1)*, удовлетворяющая условию линейной независимости, и пусть сильное достаточное условие второго порядка* (17) *выполняется для единственного отвечающего ей множителя Лагранжа* $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$.

Тогда существует число $\delta > 0$ такое, что для любого начального приближения $(x^0,\, \lambda^0,\, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, достаточно близкого к $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, для каждого $k = 0, 1, \ldots$ и при любом выборе матрицы H_k из множества $\partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$ существует единственная стационарная точка x^{k+1} задачи (116) и единственный отвечающий ей множитель Лагранжа $(\lambda^{k+1},\, \mu^{k+1})$ такие, что

$$\|(x^{k+1} - x^k,\, \lambda^{k+1} - \lambda^k,\, \mu^{k+1} - \mu^k)\| \leqslant \delta. \tag{119}$$

При этом последовательность $\{(x^k,\, \lambda^k,\, \mu^k)\}$ сходится к $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, и скорость сходимости сверхлинейная. Более того, скорость сходимости является квадратичной, если отображения f', h' и g' сильно полугладки в точке $\bar{x}$.

Теорема 8 повторяет основной результат [39], который был получен там путем непосредственного и весьма «громоздкого» анализа. Результаты о локальной сходимости полугладкого метода Джозефи–Ньютона, полученные в разделе 2.2, позволяют вывести тот же результат проще и элегантнее. Отметим, что локальная сходимость полугладкого метода последовательного квадратичного программирования изучалась также в [76], но в более сильных предположениях, включающих в себя условие строгой дополнительности.

Как хорошо известно (см., например, [18, упражнение 14.8]), сверхлинейная или даже квадратичная сходимость прямо-двойственной траектории не влечет за собой даже линейную сходимость прямой траектории. В то же время, как правило, именно прямая сходимость представляет наибольший интерес. В связи с этим займемся получением необходимых и достаточных условий прямой сверхлинейной сходимости метода последовательного квадратичного программирования применительно к задачам, производные целевой функции и ограничений которых являются полугладкими, предполагая сходимость прямо-двойственной траектории алгоритма.

Будем рассматривать метод последовательного квадратичного программирования как частный случай возмущенного варианта полугладкого метода последовательного квадратичного программирования. Этот возмущенный вариант состоит в следующем. По текущему приближению $(x^k,\, \lambda^k,\, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ к решению системы Каруша–Куна–Таккера (2) очередное приближение $(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1})$ находится из условий

$$\begin{gathered}
\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) + W_k(x^{k+1} - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) + \omega_1^k = 0,\\
h(x^k) + h'(x^k)(x^{k+1} - x^k) + \omega_2^k = 0,\\
\mu^{k+1} \geqslant 0,\ g(x^k) + g'(x^k)(x^{k+1} - x^k) + \omega_3^k \leqslant 0,\\
\langle \mu^{k+1}, g(x^k) + g'(x^k)(x^{k+1} - x^k) + \omega_3^k \rangle = 0
\end{gathered} \tag{120}$$

с некоторой матрицей $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$ и возмущениями $\omega_1^k \in \mathbb{R}^n$, $\omega_2^k \in \mathbb{R}^l$ и $\omega_3^k \in \mathbb{R}^m$.

В дальнейшем нам понадобится следующая лемма.

Лемма 8. *Пусть отображение* $\Phi\colon \mathbb{R}^p \times \mathbb{R}^q \mapsto \mathbb{R}^r$ *имеет вид*

$$\Phi(x,\, y) = K(x)y + b(x), \tag{121}$$

где $K\colon \mathbb{R}^p \mapsto \mathbb{R}^{r\times q}$ *и* $b\colon \mathbb{R}^p \mapsto \mathbb{R}^r$ *— заданные отображения, полугладкие в точке* $\bar{x} \in \mathbb{R}^p$. *Пусть последовательность* $\{(x^k,\, y^k)\} \subset \mathbb{R}^p \times \mathbb{R}^q$ *сходится к точке* $(\bar{x},\, \bar{y})$, *где* $\bar{y} \in \mathbb{R}^q$.

Тогда для любой последовательности матриц $\{W_k\} \subset \mathbb{R}^{r\times p}$ *такой, что*

$$W_k \in \partial_x \Phi(x^k,\, y^k) \quad \forall\, k,$$

выполняется оценка

$$\Phi(x^k,\, y^k) - \Phi(\bar{x},\, y^k) - W_k(x^k - \bar{x}) = o(\|x^k - \bar{x}\|).$$

Доказательство. Применяя лемму 1.5 с $y_1^k = y^k$ и $y_2^k = \bar{y}$ при любом k, получаем существование последовательности матриц $\{\bar{W}_k\} \subset \mathbb{R}^{r\times p}$, такой что $\bar{W}_k \in \partial_x \Phi(x^k,\, \bar{y})$ для всех достаточно больших k и $W_k - \bar{W}_k = O(\|y^k - \bar{y}\|)$. Используя (121) и полугладкость отображений K и b в точке $\bar{x}$, выводим оценку

$$\begin{aligned}
&\|\Phi(x^k,\, y^k) - \Phi(\bar{x},\, y^k) - W_k(x^k - \bar{x})\| \leqslant \\
&\qquad \leqslant \|(\Phi(x^k,\, y^k) - \Phi(x^k,\, \bar{y})) - (\Phi(\bar{x},\, y^k) - \Phi(\bar{x}, \bar{y}))\| + \|(W_k - \bar{W}_k)(x^k - \bar{x})\| + \\
&\qquad\qquad + \|\Phi(x^k,\, \bar{y}) - \Phi(\bar{x},\, \bar{y}) - \bar{W}_k(x^k - \bar{x})\| = \|(K(x^k) - K(\bar{x}))(y^k - \bar{y})\| = \\
&+ O(\|x^k - \bar{x}\|\|y^k - \bar{y}\|) + o(\|x^k - \bar{x}\|) = O(\|x^k - \bar{x}\|\|y^k - \bar{y}\|) + o(\|x^k - \bar{x}\|) = o(\|x^k - \bar{x}\|).
\end{aligned}$$

■

Сначала установим ряд необходимых условий прямой сверхлинейной сходимости алгоритма, задаваемого соотношениями (120).

Предложение 9. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображения* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *и* $g\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *дифференцируемы в окрестности точки* $\bar{x} \in \mathbb{R}^n$, *причем их производные полугладки в этой точке. Пусть* $\bar{x}$ *является стационарной точкой задачи* (1), *и пусть* $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ *— отвечающий ей множитель Лагранжа. Предположим, что последовательность* $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ *сходится* $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, *и для всех достаточно больших* k *тройка*

$(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1})$ удовлетворяет системе (120) с некоторой матрицей $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k)$ и возмущениями $\omega_1^k \in \mathbb{R}^n$, $\omega_2^k \in \mathbb{R}^l$ и $\omega_3^k \in \mathbb{R}^m$.

Если скорость сходимости последовательности $\{x^k\}$ сверхлинейна, то

$$\frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - W_k(x^{k+1} - x^k) = o(\|x^k - \bar{x}\|), \tag{122}$$

$$\omega_2^k = o(\|x^k - \bar{x}\|), \tag{123}$$

$$(\omega_3^k)_{A_+(\bar{x},\bar{\mu})} = o(\|x^k - \bar{x}\|). \tag{124}$$

Если, кроме того,

$$\{(\omega_3^k)_{\{1,\,\dots,\,m\}\setminus A(\bar{x})}\} \to 0 \text{ при } k \to \infty, \tag{125}$$

то

$$\pi_{C(\bar{x})}(-\omega_1^k) = o(\|x^k - \bar{x}\|). \tag{126}$$

Доказательство. Пусть $\{\tilde{W}_k\}$ — произвольная последовательность матриц, такая что $\tilde{W}_k \in \partial_x \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k)$ для любого k. Легко видеть, что, поскольку $\{(x^k,\, \lambda^k,\, \mu^k)\}$ сходится к $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ (а значит, $\{(\lambda^k,\, \mu^k)\}$ ограничена), и производные функции f и отображений h и g локально липшицевы в точке $\bar{x}$, найдется окрестность U точки $\bar{x}$ и $\ell > 0$ такие, что для всех k отображение $\frac{\partial L}{\partial x}(\cdot,\, \lambda^k,\, \mu^k)$ удовлетворяет условию Липшица на U с константой ℓ, и как x^k, так и x^{k+1} принадлежат U для всех достаточно больших k. Отсюда следует, что $\|W_k\| \leqslant \ell$ и $\|\tilde{W}_k\| \leqslant \ell$ для всех таких k (поскольку матрицы из дифференциала Кларка локально липшицевого отображения ограничены константой Липшица). Таким образом, последовательности $\{W_k\}$ и $\{\tilde{W}_k\}$ ограничены. Тогда, привлекая лемму 8 (с $p = r = n$, $q = l+m$, $K(x) = ((h'(x))^{\mathrm{T}},\, (g'(x))^{\mathrm{T}})$, $b(x) = f'(x)$, $y^k = (\lambda^k,\, \mu^k)$, $\bar{y} = (\bar{\lambda},\, \bar{\mu})$) и пользуясь сверхлинейной сходимостью последовательности $\{x^k\}$ к $\bar{x}$, получаем

$$\begin{aligned}
&\frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - W_k(x^{k+1} - x^k) = \\
&\qquad = \left(\frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k, \mu^k) - \tilde{W}_k(x^{k+1} - \bar{x})\right) - \\
&\qquad - \left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k, \mu^k) - W_k(x^k - \bar{x})\right) - (W_k - \tilde{W}_k)(x^{k+1} - \bar{x}) = \\
&\qquad = o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|) + O(\|x^{k+1} - \bar{x}\|) = o(\|x^k - \bar{x}\|),
\end{aligned} \tag{127}$$

т. е. (122) выполняется.

Далее, из (120), опираясь на сверхлинейную сходимость $\{x^k\}$ к $\bar{x}$, ограниченность последовательности $\{W_k\}$, лемму 8 и локальную липшицевость производных отображений h и g в точке $\bar{x}$, получаем

$$-\omega_1^k = \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) + W_k(x^{k+1} - x^k) +$$
$$+ (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) =$$
$$= (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) +$$
$$+ (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) +$$
$$+ \left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k, \mu^k) - W_k(x^k - \bar{x})\right) + W_k(x^{k+1} - \bar{x}) =$$
$$= ((h'(x^k))^{\mathrm{T}} - (h'(\bar{x}))^{\mathrm{T}})(\lambda^{k+1} - \lambda^k) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) +$$
$$+ ((g'(x^k))^{\mathrm{T}} - (g'(\bar{x}))^{\mathrm{T}})(\mu^{k+1} - \mu^k) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}) + o(\|x^k - \bar{x}\|) =$$
$$= (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}) + o(\|x^k - \bar{x}\|), \quad (128)$$

и

$$\omega_2^k = -h(x^k) - h'(x^k)(x^{k+1} - x^k) =$$
$$= -h(x^k) + h(\bar{x}) + h'(\bar{x})(x^k - \bar{x}) + (h'(x^k) - h'(\bar{x}))(x^k - \bar{x}) - h'(x^k)(x^{k+1} - \bar{x}) =$$
$$= o(\|x^k - \bar{x}\|).$$

Поскольку $\bar{\mu}_{A_+(\bar{x},\bar{\mu})} > 0$, для достаточно больших k выполнено $\mu^k_{A_+(\bar{x},\bar{\mu})} > 0$. Тогда из последнего равенства в (120) следует

$$(\omega_3^k)_{A_+(\bar{x},\bar{\mu})} = -g_{A_+(\bar{x},\bar{\mu})}(x^k) - g'_{A_+(\bar{x},\bar{\mu})}(x^k)(x^{k+1} - x^k) =$$
$$= -g_{A_+(\bar{x},\bar{\mu})}(x^k) + g_{A_+}(\bar{x}) + g'_{A_+(\bar{x},\bar{\mu})}(\bar{x})(x^k - \bar{x}) =$$
$$+ (g'_{A_+(\bar{x},\bar{\mu})}(x^k) - g'_{A_+(\bar{x},\bar{\mu})}(\bar{x}))(x^k - \bar{x}) - g'_{A_+(\bar{x},\bar{\mu})}(x^k)(x^{k+1} - \bar{x}) = o(\|x^k - \bar{x}\|).$$

Для всех k положим

$$\tilde{\omega}_1^k = (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}). \quad (129)$$

Тогда (128) влечет оценку

$$\omega_1^k + \tilde{\omega}_1^k = o(\|x^k - \bar{x}\|). \quad (130)$$

Если выполнено (125), то, поскольку $\{g_{\{1,\ldots,m\}\setminus A(\bar{x})}(x^k)\} \to g_{\{1,\ldots,m\}\setminus A(\bar{x})}(\bar{x}) < 0$, из последнего равенства в (120) следует, что $\mu^k_{\{1,\ldots,m\}\setminus A(\bar{x})} = 0$ при всех достаточно больших k. Принимая это во внимание, из (129) получаем, что для всех таких k, для любого $\xi \in C(\bar{x})$ справедливы соотношения

$$\langle \tilde{\omega}_1^k,\, \xi \rangle = \langle \lambda^{k+1} - \bar{\lambda},\, h'(\bar{x})\xi \rangle + \langle \mu^{k+1} - \bar{\mu},\, g'(\bar{x})\xi \rangle = \langle \mu^k_{A_0(\bar{x},\bar{\mu})},\, g'_{A_0(\bar{x},\bar{\mu})}(\bar{x})\xi \rangle \leqslant 0,$$

где также было использовано неравенство $\mu^{k+1} \geqslant 0$. Таким образом, $\tilde{\omega}_1^k \in (C(\bar{x}))^\circ$. Значит, согласно (132), $\pi_{C(\bar{x})}(\tilde{\omega}_1^k) = 0$. В совокупности с (130) и тем фактом, что $\pi_{C(\bar{x})}(\cdot)$ является неразжимающим, это дает (126). ■

Перейдем к получению достаточных условий.

При доказательстве следующего утверждения окажутся полезными следующие два свойства евклидовой проекции. Если $C \subset \mathbb{R}^s$ — выпуклый замкнутый конус, то

$$\pi_C(z - \pi_C(z)) = 0 \quad \forall\, z \in \mathbb{R}^s, \tag{131}$$

и

$$\{z \in \mathbb{R}^s \mid \pi_C(z) = 0\} = C^\circ, \tag{132}$$

где $C^\circ = \{z \in \mathbb{R}^s \mid \langle z,\, v\rangle \leqslant 0\, \forall\, v \in C\}$ – конус, отрицательно сопряженный с C.

Предложение 10. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображения* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *и* $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ *дифференцируемы в окрестности точки* $\bar{x} \in \mathbb{R}^n$*, причем их производные полугладки в этой точке. Пусть* $\bar{x}$ *является стационарной точкой задачи* (1)*, и пусть* $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ *— отвечающий этой точке множитель Лагранжа, для которого выполняется достаточное условие второго порядка* (20)*.*

Тогда для всех $(x,\, \lambda,\, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$*, достаточно близких к* $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$*, справедлива оценка*

$$\|x - \bar{x}\| = O\left(\left\|\begin{pmatrix} \pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x, \lambda, \mu)\right) \\ h(x) \\ \min\{\mu,\, -g(x)\}\end{pmatrix}\right\|\right) \tag{133}$$

.

Доказательство. От противного, предположим, что (133) не выполняется. Тогда найдутся последовательность $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, сходящаяся к $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, и последовательность $t_k \to +\infty$ такие, что для всех k

$$\|x^k - \bar{x}\| > t_k \left\|\begin{pmatrix} \pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k)\right) \\ h(x^k) \\ \min\{\mu^k, -g(x^k)\}\end{pmatrix}\right\|.$$

Последнее неравенство эквивалентно соотношениям

$$\pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k)\right) = o(\|x^k - \bar{x}\|), \tag{134}$$

$$h(x^k) = o(\|x^k - \bar{x}\|), \tag{135}$$

$$\min\{\mu^k,\ -g(x^k)\} = o(\|x^k - \bar{x}\|). \tag{136}$$

Из (135) следует, что

$$0 = h(\bar{x}) + h'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) = h'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|). \tag{137}$$

Кроме того, поскольку $g_{A_+(\bar{x},\,\bar{\mu})}(\bar{x}) = 0 < \bar{\mu}_{A_+(\bar{x},\,\bar{\mu})}$, из (136) получаем, что для всех достаточно больших k

$$\begin{aligned} 0 = \min\{\mu^k_{A_+(\bar{x},\,\bar{\mu})},\ -g_{A_+(\bar{x},\,\bar{\mu})}(x^k)\} + o(\|x^k - \bar{x}\|) = -g_{A_+(\bar{x},\,\bar{\mu})}(x^k) + o(\|x^k - \bar{x}\|) = \\ = -g_{A_+(\bar{x},\,\bar{\mu})}(\bar{x}) - g'_{A_+(\bar{x},\,\bar{\mu})}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) = \\ = -g'_{A_+(\bar{x},\,\bar{\mu})}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \end{aligned} \tag{138}$$

и аналогично, так как $g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(\bar{x}) < 0 = \bar{\mu}_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}$,

$$\begin{aligned} 0 = \min\{\mu^k_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})},\ -g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(x^k)\} + o(\|x^k - \bar{x}\|) = \\ = \mu^k_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} + o(\|x^k - \bar{x}\|). \end{aligned} \tag{139}$$

Поскольку число возможных разбиений множества $A_0(\bar{x},\,\bar{\mu})$ конечно, переходя при необходимости к подпоследовательности, можно считать, что существуют множества индексов I_1 и I_2, такие что $I_1 \cup I_2 = A_0(\bar{x},\,\bar{\mu})$, $I_1 \cap I_2 = \emptyset$, и для всех k выполнено

$$\mu^k_{I_1} \geqslant -g_{I_1}(x^k), \quad \mu^k_{I_2} < -g_{I_2}(x^k). \tag{140}$$

Тогда, в силу (136), имеем

$$\begin{aligned} 0 = \min\{\mu^k_{I_1},\ -g_{I_1}(x^k)\} + o(\|x^k - \bar{x}\|) = -g_{I_1}(x^k) + o(\|x^k - \bar{x}\|) = \\ = -g_{I_1}(\bar{x}) - g'_{I_1}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) = -g'_{I_1}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \end{aligned} \tag{141}$$

$$0 = \min\{\mu^k_{I_2},\ -g_{I_2}(x^k)\} + o(\|x^k - \bar{x}\|) = \mu^k_{I_2} + o(\|x^k - \bar{x}\|). \tag{142}$$

Наконец, из (140) следует

$$\begin{aligned} -\mu^k_{I_2} > g_{I_2}(x^k) = g_{I_2}(\bar{x}) + g'_{I_2}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) = \\ = g'_{I_2}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \end{aligned}$$

и значит, с учетом (142),

$$g'_{I_2}(\bar{x})(x^k - \bar{x}) \leqslant o(\|x^k - \bar{x}\|). \qquad (143)$$

Без ограничения общности можно считать, что $x^k \neq \bar{x}$ для всех k и что последовательность $\{(x^k - \bar{x})/\|x^k - \bar{x}\|\}$ сходится к некоторому $\xi \in \mathbb{R}^n \setminus \{0\}$ ($\|\xi\| = 1$). Тогда, используя (137), (138), (141) и (143), получаем, что $\xi \in C(\bar{x}) \setminus \{0\}$.

Далее, привлекая (131) и (134), заключаем, что

$$0 = \pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) - \pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)\right)\right) =$$
$$= \pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + o(\|x^k - \bar{x}\|)\right),$$

и значит, в силу (132) и леммы 8,

$$(C(\bar{x}))^\circ \ni \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + o(\|x^k - \bar{x}\|) =$$
$$= \left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(\bar{x},\, \lambda^k,\, \mu^k)\right) + \left(\frac{\partial L}{\partial x}(\bar{x},\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\right) +$$
$$+ o(\|x^k - \bar{x}\|) =$$
$$= W_k(x^k - \bar{x}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) + o(\|x^k - \bar{x}\|), \quad (144)$$

где $\{W_k\}$ является произвольной последовательностью матриц, такой что $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$ для всех k.

По лемме 1.5 найдется последовательность $\{\bar{W}_k\}$, такая что $\bar{W}_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \bar{\lambda},\, \bar{\mu})$ для достаточно больших k и что $W_k - \bar{W}_k = O(\|(\lambda^k - \bar{\lambda},\, \mu^k - \bar{\mu})\|)$. Тогда, поскольку последовательность $\{\bar{W}_k\}$ ограничена (в силу локальной липшицевости $\frac{\partial L}{\partial x}(\cdot,\, \bar{\lambda},\, \bar{\mu})$ в точке $\bar{x}$) и поскольку $\{(\lambda^k,\, \mu^k)\}$ сходится к $(\bar{\lambda},\, \bar{\mu})$, переходя, если нужно, к подпоследовательности, можно считать, что и $\{\bar{W}_k\}$, и $\{W_k\}$ сходятся к некоторой матрице $\bar{W} \in \mathbb{R}^{n \times n}$. Так как дифференциал Кларка полунепрерывен сверху, $\bar{W} \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})$. Учитывая (139) и (142), включение $\xi \in C(\bar{x})$ и равенства $g'_{I_1}(\bar{x})\xi = 0$ (см. (141)) и $\bar{\mu}_{\{1,\, \ldots,\, m\} \setminus A_+(\bar{x},\, \bar{\mu})} = 0$, из (144) получаем

$$0 \geqslant \langle W_k(x^k - \bar{x}),\, \xi\rangle + \langle \lambda^k - \bar{\lambda},\, h'(\bar{x})\xi\rangle + \langle \mu^k - \bar{\mu},\, g'(\bar{x})\xi\rangle + o(\|x^k - \bar{x}\|) =$$
$$= \langle W_k(x^k - \bar{x}),\, \xi\rangle + \langle \mu^k_{I_2 \cup (\{1,\, \ldots,\, m\} \setminus A(\bar{x}))},\, g'_{I_2 \cup (\{1,\, \ldots,\, m\} \setminus A(\bar{x}))}(\bar{x})\xi\rangle + o(\|x^k - \bar{x}\|) =$$
$$= \langle W_k(x^k - \bar{x}),\, \xi\rangle + o(\|x^k - \bar{x}\|).$$

Разделив обе части полученного соотношения на $\|x^k - \bar{x}\|$ и перейдя к пределу, заключаем что

$$\langle \bar{W}\xi, \xi\rangle \leqslant 0,$$

что противоречит достаточному условию второго порядка (20), поскольку $\xi \in C(\bar{x}) \setminus \{0\}$. ■

Равенство (133) представляет собой оценку расстояния между «прямой» частью решения системы Каруша-Куна-Таккера и соответствующей компонентой точек, близких к этому решению. Прямо-двойственная оценка расстояния для задач с липшицевыми производными была получена в разделе 1.3.

Теорема 9. *Пусть в дополнение к условиям предложения 9, множитель Лагранжа* $(\bar{\lambda}, \bar{\mu})$*, отвечающий стационарной точке* $\bar{x}$ *задачи* (1)*, удовлетворяет достаточному условию второго порядка* (20)*.*

Если выполнено (125) *и*

$$\pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - W_k(x^{k+1} - x^k) - \omega_1^k\right) =$$
$$= o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \quad (145)$$

$$\omega_2^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \quad (146)$$

$$(\omega_3^k)_{A(\bar{x})} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \quad (147)$$

то скорость сходимости $\{x^k\}$ *сверхлинейна.*

Доказательство. Пользуясь сходимостью последовательности $\{(x^k, \lambda^k, \mu^k)\}$ к точке $(\bar{x}, \bar{\lambda}, \bar{\mu})$ и локальной липшицевостью отображений h' и g' в точке $\bar{x}$, из первого и второго равенств в (120) получаем соотношения

$$\begin{aligned}
\frac{\partial L}{\partial x}(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) &= \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) \\
&\quad + (h'(x^{k+1}))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^{k+1}))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&= \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) \\
&\quad + \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&\quad + o(\|x^{k+1} - x^k\|) \\
&= \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - W_k(x^{k+1} - x^k) - \omega_1^k \\
&\quad + o(\|x^{k+1} - x^k\|) \quad (148)
\end{aligned}$$

и

$$h(x^{k+1}) = h(x^k) + h'(x^k)(x^{k+1} - x^k) + o(\|x^{k+1} - x^k\|) = -\omega_2^k + o(\|x^{k+1} - x^k\|). \quad (149)$$

Поскольку последовательность $\{(\omega_3^k)_{\{1,\ldots,m\}\setminus A(\bar{x})}\} \to 0$ (в силу (125)), а последовательность $\{g_{\{1,\ldots,m\}\setminus A(\bar{x})}(x^k)\} \to g_{\{1,\ldots,m\}\setminus A(\bar{x})}(\bar{x}) < 0$, для всех достаточно больших k справедливо равенство $\mu^{k+1}_{\{1,\ldots,m\}\setminus A(\bar{x})} = 0$. Следовательно,

$$\min\{\mu^{k+1}_{\{1,\ldots,m\}\setminus A(\bar{x})},\ -g_{\{1,\ldots,m\}\setminus A(\bar{x})}(x^{k+1})\} = 0. \tag{150}$$

Заметим, что последнее равенство в (120) может быть записано в форме

$$\min\{\mu^{k+1},\ -g(x^k) - g'(x^k)(x^{k+1} - x^k) - \omega_3^k\} = 0.$$

Для каждого $i \in A(\bar{x})$, пользуясь последним равенством, а также свойством

$$|\min\{a,\, b\} - \min\{a,\, c\}| \leqslant |b - c| \quad \forall\, a,\, b,\, c \in \mathbb{R},$$

выводим оценку

$$\begin{aligned}
&\|\min\{\mu^{k+1}_{A(\bar{x})},\ -g_{A(\bar{x})}(x^{k+1})\}\| = \\
&\qquad = \|\min\{\mu^{k+1}_{A(\bar{x})},\ -g_{A(\bar{x})}(x^k) - g'_{A(\bar{x})}(x^k)(x^{k+1} - x^k) + o(\|x^{k+1} - x^k\|)\} - \\
&\qquad\quad - \min\{\mu^{k+1}_{A(\bar{x})},\ -g_{A(\bar{x})}(x^k) - g'_{A(\bar{x})}(x^k)(x^{k+1} - x^k) - (\omega_3^k)_{A(\bar{x})}\}\| \leqslant \\
&\qquad\qquad \leqslant \|(\omega_3^k)_{A(\bar{x})}\| + o(\|x^{k+1} - x^k\|).
\end{aligned} \tag{151}$$

Учитывая предложение 10, соотношения (145)–(149) и (150)–(151), заключаем, что

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

т. е. существует последовательность $\{t_k\}$, состоящая из положительных чисел, сходящаяся к нулю и такая, что

$$\|x^{k+1} - \bar{x}\| \leqslant t_k(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|)$$

при всех достаточно больших k. Отсюда следует, что

$$(1 - t_k)\|x^{k+1} - \bar{x}\| \leqslant t_k\|x^k - \bar{x}\|,$$

и значит, для всех достаточно больших k

$$\|x^{k+1} - \bar{x}\| \leqslant \frac{t_k}{1 - t_k}\|x^k - \bar{x}\|,$$

т. е.

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

что и требовалось. ∎

Замечание 5. Условие (145) следует из (122) и (126). Поэтому, в силу предложения 9, оно является необходимым для прямой сверхлинейной сходимости возмущенного полугладкого SQP (120) (при выполнении (125)).

Замечание 6. В теореме 9 достаточное условие второго порядка (20) может быть заменено следующим секвенциальным условием второго порядка

$$\liminf_{k\to\infty} \max_{W\in\partial_x \frac{\partial L}{\partial x}(x^k,\,\lambda^k,\,\mu^k)} \langle W\xi,\,\xi\rangle > 0 \quad \forall\,\xi \in C(\bar{x}) \setminus \{0\} \tag{152}$$

(пользуясь леммой 1.5, легко убедиться, что (152) следует из (20)). Для этого понадобится секвенциальный вариант оценки (133), который легко установить по аналогии с доказательством предложения 10.

Применим проведенный анализ прямой сверхлинейной сходимости возмущенного полугладкого SQP (120) для того, чтобы получить необходимые и достаточные условия прямой сверхлинейной сходимости метода последовательного квадратичного программирования. Метод последовательного квадратичного программирования представляет собой частный случай схемы (120) с

$$\omega_1^k = (H_k - W_k)(x^{k+1} - x^k), \quad \omega_2^k = 0, \quad \omega_3^k = 0,$$

где $\{W_k\}$ является произвольной последовательностью матриц, такой что $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\,\lambda^k,\,\mu^k)$ для любого k. Из предложения 9, теоремы 9, а также замечаний 5 и 6 вытекает следующая теорема.

Теорема 10. *Пусть функция* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ *и отображения* $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *и* $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ *дифференцируемы в окрестности точки* $\bar{x} \in \mathbb{R}^n$*, причем их производные полугладки в этой точке. Пусть* $\bar{x}$ *является стационарной точкой задачи* (1)*, и пусть* $(\bar{\lambda},\,\bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ *— отвечающий этой точке множитель Лагранжа. Предположим, что последовательность* $\{(x^k,\,\lambda^k,\,\mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$*, сгенерированная методом последовательного квадратичного программирования, сходится* $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$*.*

Тогда, если скорость сходимости последовательности $\{x^k\}$ *сверхлинейная, то выполняется соотношение*

$$\pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^{k+1},\lambda^k,\mu^k) - \frac{\partial L}{\partial x}(x^k,\lambda^k,\mu^k) - H_k(x^{k+1}-x^k)\right) = o(\|x^{k+1}-x^k\|). \tag{153}$$

Обратно, при выполнении условия (152) *соотношение* (153) *влечет за собой сверхлинейную скорость сходимости последовательности* $\{x^k\}$*.*

Таким образом, при выполнении (152) прямая сверхлинейная сходимость метода последовательного квадратичного программирования характеризуется условием (153). Это условие представляет собой естественное обобщение условия Денниса-Морé [70, теорема 18.5] на случай задач, производные целевой функции и ограничений которых являются полугладкими.

Обычное условие Денниса-Морé для гладкого случая наталкивает на мысль о том, чтобы заменить (153) следующим условием

$$\max_{W \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)} \|\pi_{C(\bar{x})}((W - H_k)(x^{k+1} - x^k))\| = o(\|x^{k+1} - x^k\|). \tag{154}$$

Последнее условие использовалось в [49], где было показано, что (154) является необходимым, а при выполнении (152) и достаточным для прямой сверхлинейной сходимости метода последовательного квадратичного программирования в случае, когда ограничения-неравенства отсутствуют. Если f, h и g дважды непрерывно-дифференцируемы в окрестности точки $\bar{x}$, то, как легко убедиться, применяя теорему о среднем, условия (153) и (154) эквивалентны. В полугладком случае то, как соотносятся эти два условия, на данный момент до конца не известно. Из предложения 9 легко следует, что (154) необходимо для прямой сверхлинейной сходимости. Поэтому оно следует из (153) при выполнении (152). Обратное включение, возможно, не выполняется, но привести пример его нарушения оказывается сложно. А именно, ниже будет показано, что при выполнении достаточно разумного дополнительного предположения условие (154) достаточно для прямой сверхлинейной сходимости метода последовательного квадратичного программирования и поэтому влечет за собой (153).

Более точно, дополнительное предположение состоит в том, что множества

$$A_k = \{i = 1,\, \ldots,\, m \mid g_i(x^k) + g_i'(x^k)(x^{k+1} - x^k) = 0\} \tag{155}$$

индексов активных ограничений-неравенств траектории метода последовательного квадратичного программирования стабилизируются, т. е. $A_k = A$ для некоторого $A \subset \{1,\, \ldots,\, m\}$ и всех достаточно больших k. В силу последнего равенства в (117), по непрерывности при всех больших k справедливы включения

$$A_+(\bar{x},\, \bar{\mu}) \subset A_k \subset A(\bar{x}). \tag{156}$$

Поэтому условие стабилизации автоматически выполняется, если двойственная траектория $\{(\lambda^k,\, \mu^k)\}$ сходится к множителю $(\bar{\lambda},\, \bar{\mu})$, удовлетворяющему условию строгой дополнительности, т. е. если $\bar{\mu}_{A(\bar{x})} > 0$ (в этом случае $A_+(\bar{x},\, \bar{\mu}) = A(\bar{x})$). В отсутствие строгой дополнительности, стабилизационное свойство, вообще говоря, может нарушаться. Тем не менее, его выполнение представляется естественным и типичным. Заметим, что если траектория не стабилизируется, едва ли можно ожидать сходимость ее двойственной части.

Следующий результат обобщает теорему [49, теорема 2.2] в части достаточности на случай смешанных ограничений.

Теорема 11. *Пусть* $f\colon \mathbb{R}^n \mapsto \mathbb{R}$, $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *и* $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ *дифференцируемы в окрестности точки* $\bar{x} \in \mathbb{R}^n$*, причем их производные полугладки в этой точке. Пусть* $\bar{x}$ *является стационарной точкой задачи* (1)*, и пусть* $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ *является отвечающим ей множителем Лагранжа. Пусть последовательность* $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$*, сгенерированная методом последовательного квадратичного программирования, сходится к* $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$*. Предположим также, что выполнены* (152) *и* (154)*, и что существует множество индексов* $A \subset \{1,\, \ldots,\, m\}$*, такое что* $A_k = A$ *для всех достаточно больших* k*, где множества* A_k *заданы согласно* (155)*.*

Тогда скорость сходимости последовательности $\{x^k\}$ *сверхлинейна.*

Доказательство. Определим множество

$$\tilde{C}(\bar{x}) = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_A(\bar{x})\xi = 0,\ g'_{A(\bar{x})\setminus A}(\bar{x})\xi \leqslant 0\}.$$

По лемме Хоффмана (см., например, [32, лемма 3.2.3])

$$\begin{gathered}\mathrm{dist}(x^{k+1} - \bar{x},\, \tilde{C}(\bar{x})) = O(\|h'(\bar{x})(x^{k+1} - \bar{x})\| + \|g'_A(\bar{x})(x^{k+1} - \bar{x})\| + \\ + \|\max\{0,\, g'_{A(\bar{x})\setminus A}(\bar{x})(x^{k+1} - \bar{x})\}\|). \quad (157)\end{gathered}$$

Из второй строки в (117) и локальной липшицевости производной отображения h в точке $\bar{x}$ имеем

$$\begin{gathered}h'(\bar{x})(x^{k+1} - \bar{x}) = h'(\bar{x})(x^{k+1} - x^k) + h'(\bar{x})(x^k - \bar{x}) - h(x^k) - h'(x^k)(x^{k+1} - x^k) = \\ = -(h'(x^k) - h'(\bar{x}))(x^{k+1} - x^k) - (h(x^k) - h(\bar{x}) - h'(\bar{x})(x^k - \bar{x})) = o(\|x^k - \bar{x}\|). \quad (158)\end{gathered}$$

Для любого достаточно большого k справедливо равенство $g_A(x^k) + g'_A(x^k)(x^{k+1} - x^k) = 0$, и аналогично (158)

$$g'_A(\bar{x})(x^{k+1} - \bar{x}) = o(\|x^k - \bar{x}\|). \quad (159)$$

Наконец, если $i \in A(\bar{x}) \setminus A$ и $\langle g'_i(\bar{x}),\, x^{k+1} - \bar{x}\rangle > 0$, то с учетом последней строки в (117) и локальной липшицевости производной отображения g в точке $\bar{x}$

$$\max\{0,\, \langle g_i'(\bar{x}),\, x^{k+1}-\bar{x}\rangle\} = \langle g_i'(\bar{x}),\, x^{k+1}-\bar{x}\rangle =$$
$$= \langle g_i'(\bar{x}),\, x^{k+1}-x^k\rangle + \langle g_i'(\bar{x}),\, x^k-\bar{x}\rangle \leqslant \langle g_i'(\bar{x}),\, x^{k+1}-x^k\rangle + \langle g_i'(\bar{x}),\, x^k-\bar{x}\rangle -$$
$$- g_i(x^k) - \langle g_i'(x^k),\, x^{k+1}-x^k\rangle = -\langle g_i'(x^k)-g_i'(\bar{x}),\, x^{k+1}-x^k\rangle -$$
$$-\,(g_i(x^k)-g_i(\bar{x})-\langle g_i'(\bar{x}),\, x^k-\bar{x}\rangle) = o(\|x^k-\bar{x}\|). \quad (160)$$

Из соотношений (157)–(160) следует, что

$$\mathrm{dist}(x^{k+1}-\bar{x},\, \tilde{C}(\bar{x})) = o(\|x^k-\bar{x}\|).$$

Последнее равенство означает, что для всех k существует вектор $\xi^k \in \tilde{C}(\bar{x})$, такой что

$$x^{k+1}-\bar{x} = \xi^k + o(\|x^k-\bar{x}\|). \quad (161)$$

Из первой строки в (117) и из полугладкости отображений f', h' и g' в точке $\bar{x}$, с учетом леммы 8 и сходимости последовательности $\{(\lambda^k,\, \mu^k)\}$ к $(\bar{\lambda},\, \bar{\mu})$, для любых матриц $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$ справедлива цепочка соотношений

$$-\,H_k(x^{k+1}-x^k) = \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + (h'(x^k))^T(\lambda^{k+1}-\lambda^k) + (g'(x^k))^T(\mu^{k+1}-\mu^k) =$$
$$= \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(\bar{x},\, \lambda^k,\, \mu^k) - W_k(x^k-\bar{x}) +$$
$$+\,\frac{\partial L}{\partial x}(\bar{x},\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) + (h'(x^k))^T(\lambda^{k+1}-\lambda^k) +$$
$$+\,(g'(x^k))^T(\mu^{k+1}-\mu^k) + W_k(x^k-\bar{x}) =$$
$$= W_k(x^k-\bar{x}) + (h'(\bar{x}))^T(\lambda^k-\bar{\lambda}) + (g'(\bar{x}))^T(\mu^k-\bar{\mu}) +$$
$$+\,(h'(\bar{x}))^T(\lambda^{k+1}-\lambda^k) + (g'(\bar{x}))^T(\mu^{k+1}-\mu^k) + o(\|x^k-\bar{x}\|) =$$
$$= W_k(x^k-\bar{x}) + (h'(\bar{x}))^T(\lambda^{k+1}-\bar{\lambda}) + (g'(\bar{x}))^T(\mu^{k+1}-\bar{\mu}) + o(\|x^k-\bar{x}\|).$$

Значит,

$$W_k(x^{k+1}-\bar{x}) = (W_k-H_k)(x^{k+1}-x^k) - (h'(\bar{x}))^T(\lambda^{k+1}-\bar{\lambda}) -$$
$$-\,(g'(\bar{x}))^T(\mu^{k+1}-\bar{\mu}) + o(\|x^k-\bar{x}\|). \quad (162)$$

Из определения множества A следует, что для всех достаточно больших k выполняется неравенство $g_{\{1,\, \ldots,\, m\}\setminus A}(x^k) + g'_{\{1,\, \ldots,\, m\}\setminus A}(x^k)(x^{k+1}-x^k) < 0$. Тогда, с учетом последней строки в (117),

$$\mu^{k+1}_{\{1,\, \ldots,\, m\}\setminus A} = 0 \quad (163)$$

для всех таких k. Кроме того, в силу (156), $\tilde{C}(\bar{x}) \subset C(\bar{x})$, а значит, (154) остается выполненным, если заменить в нем $C(\bar{x})$ на $\tilde{C}(\bar{x})$. Тогда, используя (162), (163), и тот факт, что $\langle x,\, \xi\rangle \leqslant \langle \pi_{\tilde{C}(\bar{x})}(x),\, \xi\rangle$ для всех $x \in \mathbb{R}^n$ и всех $\xi \in \tilde{C}(\bar{x})$ (см. (131) и (132)), получаем

$$\begin{aligned}\langle W_k\xi^k,\, \xi^k\rangle = \langle W_k(x^{k+1}-\bar{x}),\, \xi^k\rangle + o(\|x^k-\bar{x}\|\|\xi^k\|) = \langle (W_k-H_k)(x^{k+1}-x^k),\, \xi^k\rangle - \\ - \langle (h'(\bar{x}))^T(\lambda^{k+1}-\bar{\lambda}) + (g'(\bar{x}))^T(\mu^{k+1}-\bar{\mu}),\, \xi^k\rangle + o(\|x^k-\bar{x}\|\|\xi^k\|) \leqslant \\ \leqslant \langle \pi_{\tilde{C}(\bar{x})}((W_k-H_k)(x^{k+1}-x^k)),\, \xi^k\rangle + o(\|x^k-\bar{x}\|\|\xi^k\|) = \\ = o((\|x^{k+1}-x^k\| + \|x^k-\bar{x}\|)\|\xi^k\|).\end{aligned} \tag{164}$$

Из (152) и включения $\tilde{C}(\bar{x}) \subset C(\bar{x})$ следует, что существуют $\gamma > 0$ и последовательность матриц $\{W_k\}$, такие что $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$ и для всех достаточно больших k

$$\langle W_k\xi^k,\, \xi^k\rangle \geqslant \gamma\|\xi^k\|^2.$$

Тогда (164) влечет за собой оценку

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

а значит, с учетом (161),

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

Повторяя рассуждение, завершающее доказательство теоремы 9, получаем, что последовательность $\{x^k\}$ сходится со сверхлинейной скоростью. ∎

Из доказательства теоремы 11 ясно, что для всякой последовательности непустых множеств $\Delta_k \subset \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$ эта теорема останется верной при замене условия (154) следующим условием

$$\sup_{W\in\Delta_k} \|\pi_{C(\bar{x})}((W-H_k)(x^{k+1}-x^k))\| = o(\|x^{k+1}-x^k\|),$$

если при этом секвенциальное достаточное условие второго порядка (152) заменить условием

$$\liminf_{k\to\infty} \sup_{W\in\Delta_k} \langle W\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}. \tag{165}$$

Отметим, что при выполнении достаточного условия второго порядка (20) условие (165) выполняется для любой последовательности непустых множеств $\Delta_k \subset \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$, и, в частности, можно гарантировать прямую сверхлинейную сходимость траектории метода последовательного квадратичного программирования, если эта траектория стабилизируется и выполнено соотношение

$$\min_{W\in\partial_x \frac{\partial L}{\partial x}(x^k,\lambda^k,\mu^k)} \|\pi_{C(\bar{x})}((W-H_k)(x^{k+1}-x^k))\| = o(\|x^{k+1}-x^k\|).$$

Заключение

В работе представлена единая теория локальной сходимости ряда численных методов оптимизации применительно к задачам с липшицевыми производными. Для этого были предложены абстрактные итерационные схемы решения обобщенных уравнений, а также разработана соответствующая теория.

Основные результаты диссертационной работы заключаются в следующем.

1. Получена оценка расстояния до множества решений системы Каруша–Куна–Таккера задачи оптимизации с липшицевыми производными.

2. Предложены абстрактные итерационные схемы для решения обобщенных уравнений. Установлена их локальная сходимость в различных предположениях о регулярности решения обобщенного уравнения: при сильной метрической регулярности, при полуустойчивости и при верхней липшицевой устойчивости.

3. Установлена локальная сходимость и скорость сходимости метода модифицированных функций Лагранжа при выполнении достаточного условия второго порядка применительно к задачам с липшицевыми производными. При этом для задач с ограничениями-равенствами также получены результаты о локальной сходимости метода модифицированных функций при условии строгой некритичности множителя, которое является более слабым, чем достаточное условие второго порядка.

4. Установлена локальная квадратичная сходимость метода множителей с линеаризованными ограничениями при выполнении строгого условия Мангасариана–Фромовица и достаточного условия второго порядка приментиельно к задачам с липшицевыми производными.

5. Получены необходимые и достаточные условия прямой локальной сверхлинейной сходимости полугладкого метода последовательного квадратичного программирования.

Литература

1. Дарьина А. Н., Измаилов А. Ф., Солодов М. В. Смешанные комплементарные задачи: регулярность, оценки расстояния до решения и ньютоновские методы // Журнал вычислительной математики и математической физики. — 2004. — Т. 44, № 1. — С. 51–69.

2. Измаилов А. Ф., Куренной А. С. Частный дифференциал Кларка и другие обобщенные дифференциалы // Теоретические и прикладные задачи нелинейного анализа. — М. : ВЦ РАН, 2010. — С. 77–90.

3. Измаилов А. Ф., Куренной А. С. Частный дифференциал Кларка и проекция полного дифференциала Кларка // XVIII международная научная конференция студентов, аспирантов и молодых ученых «Ломоносов-2011», секция «Вычислительная математика и математическая кибернетика»: Сб. тезисов / МГУ им. М. В. Ломоносова. — М. : МАКС Пресс, 2011. — 11–15 апреля. — С. 38.

4. Измаилов А. Ф., Куренной А. С. Абстрактная ньютоновская схема для нахождения неизолированных решений негладких обобщенных уравнений // Тихоновские чтения: Научная конференция / МГУ им. М. В. Ломоносова. — Т. 1. — М. : МАКС Пресс, 2012. — 23–31 октября. — С. 41.

5. Измаилов А. Ф., Куренной А. С. Методы множителей для задач оптимизации с липшицевыми производными // Журнал вычислительной математики и математической физики. — 2012. — Т. 52, № 12. — С. 2140–2148.

6. Измаилов А. Ф., Куренной А. С. Об условиях регулярности для комплементарных задач // XIX международная научная конференция студентов, аспирантов и молодых ученых «Ломоносов-2012», секция «Вычислительная математика и математическая кибернетика»: Сб. тезисов / МГУ им. М. В. Ломоносова. — М. : МАКС Пресс, 2012. — 9–13 апреля. — С. 61–62.

7. Измаилов А. Ф., Куренной А. С., Солодов М. В. Метод Джозефи–Ньютона для полугладких обобщенных уравнений // Ломоносовские чтения: Научная конференция, посвященная 300-летию со дня рождения М. В. Ломоносова / МГУ им. М. В. Ломоносова. — М. : МАКС Пресс, 2011. — С. 24.

8. Прасолов В. В. Задачи и теоремы линейной алгебры. — М. : Наука, 1996.

9. Федерер Г. Геометрическая теория меры. — М. : Наука, 1987.

10. ALGENCAN. — `http://www.ime.usp.br/egbirgin/tango/`.

11. Andreani R., Birgin E. G., Martínez J. M., Schuverdt M. L. On augmented Lagrangian methods with general lower-level constraints // SIAM Journal on Optimization. — 2007. — Vol. 18. — P. 1286–1309.

12. Arutyunov A. V., Izmailov A. F. Sensitivity analysis for cone-constrained optimization problems under the relaxed constraint qualifications // Mathematics of Operations Research. — 2005. — Vol. 30. — P. 333–353.

13. Bertsekas D. P. Multiplier methods: a survey // Automatica. — 1976. — Vol. 12. — P. 133–145.

14. Bertsekas D. P. Constrained optimization and Lagrange multiplier methods. — New York, USA : Academic Press, 1982.

15. Algorithms for complementarity problems and generalized equations : PhD Thesis : 95–14 / Computer Sciences Department, University of Wisconsin ; Executor: S. C. Billups. — Madison : 1995.

16. Boggs P. T., Tolle J. W. Sequential quadratic programming // Acta numerica. — 1995. — Vol. 4, no. 1. — P. 1–51.

17. Bonnans J. F. Local analysis of Newton-type methods for variational inequalities and nonlinear programming // Applied Mathematics and Optimization. — 1994. — Vol. 29. — P. 161–186.

18. Bonnans J. F., Gilbert J. C., Lemaréchal C., Sagastizábal C. A. Numerical Optimization: Theoretical and Practical Aspects. — Berlin, Heidelberg: Springer, 2006.

19. Bonnans J. F., Shapiro A. Perturbation Analysis of Optimization Problems. — New York, USA : Springer-Verlag, 2000.

20. Bonnans J. F., Sulem A. Pseudopower expansion of solutions of generalized equations and constrained optimization // Mathematical Programming. — 1995. — Vol. 70. — P. 123–148.

21. Clarke F. H. Optimization and Nonsmooth Analysis. — New York, USA : John Wiley & Sons, 1983.

22. Conn A. R., Gould N., Sartenaer A., Toint P. L. Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints // SIAM Journal on Optimization. — 1996. — Vol. 6. — P. 674–703.

23. Daryina A. N., Izmailov A. F., Solodov M. V. A class of active-set Newton methods for mixed complementarity problems // SIAM Journal on Optimization. — 2005. — Vol. 15, no. 2. — P. 409–429.

24. De Luca T., Facchinei F., Kanzow C. A theoretical and numerical comparison of some semismooth algorithms for complementarity problems // Computational Optimization and Applications. — 2000. — Vol. 16, no. 2. — P. 173–205.

25. Debreu G. Definite and semidefinite quadratic forms // Econometrica. — 1952. — Vol. 20. — P. 295–300.

26. Dontchev A., Rockafellar R. Characterizations of Lipschitzian stability in nonlinear programming // Mathematical Programming with Data Perturbations. — 1997. — P. 65–82.

27. Dontchev A. L., Rockafellar R. T. Implicit Functions and Solution Mappings. — New York, USA : Springer, 2009.

28. Dontchev A. L., Rockafellar R. T. Newton's method for generalized equations: a sequential implicit function theorem // Mathematical Programming. — 2010. — Vol. 123. — P. 139–159.

29. Fabian M., Preiss D. On the Clarke's generalized Jacobian // Proceedings of the 14th Winter School on Abstract Analysis / Rendiconti del Circollo Matematico di Palermo. — Vol. 14 of Serie II. — 1987. — P. 305–307.

30. Facchinei F., Fischer A., Kanzow C. A semismooth Newton method for variational inequalities: The case of box constraints // Complementarity and Variational Problems: State of the Art / Ed. by M. C. Ferris, J.-S. Pang. — Philadelphia : SIAM, 1997. — P. 76–90.

31. Facchinei F., Fischer A., Kanzow C. On the accurate identification of active constraints // SIAM Journal on Optimization. — 1999. — P. 14–32.

32. Facchinei F., Pang J.-S. Finite-Dimensional Variational Inequalities and Complementarity Problems. — New York, USA : Springer-Verlag, 2003.

33. Fernández D., Solodov M. V. Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition // SIAM Journal on Optimization. — 2012. — Vol. 22. — P. 384–407.

34. Ferris M. C., Kanzow C., Munson T. S. Feasible descent algorithms for mixed complementarity problems // Mathematical Programming. — 1999. — Vol. 86. — P. 475–497.

35. Finsler P. Über das vorkommen definiter und semidefiniter formen und scharen quadratischer formen // Commentarii Mathematici Helvetici. — 1937. — Vol. 94. — P. 188–192.

36. Fischer A. Local behavior of an iterative framework for generalized equations with nonisolated solutions // Mathematical Programming. — 2002. — Vol. 94. — P. 91–124.

37. Friedlander M. P., Saunders M. A. A globally convergent linearly constrained Lagrangian method for nonlinear optimization // SIAM Journal on Optimization. — 2005. — Vol. 15. — P. 863–897.

38. Hager W., Gowda M. Stability in the presence of degeneracy and error estimation // Mathematical Programming. — 1999. — Vol. 85. — P. 181–192.

39. Han J., Sun D. Superlinear convergence of approximate Newton methods for LC^1 optimization problems without strict complementarity // Recent Advances in Nonsmooth Optimization / Ed. by D.-Z. Du. — World Scientific Publishing, 1995. — P. 141–158.

40. Hestenes M. R. Multiplier and gradient methods // Journal of Optimization Theory and Applications. — 1969. — Vol. 4. — P. 303–320.

41. Hiriart-Urruty J.-B., Strodiot J. J., Nguyen V. H. Generalized Hessian matrix and second-order optimality conditions for problems with $\mathrm{C}^{1,1}$-data // Applied Mathematics and Optimization. — 1984. — Vol. 11. — P. 43–56.

42. Izmailov A. F. Strongly regular nonsmooth generalized equations // Mathematical Programming. — 2013.

43. Izmailov A. F., Kurennoy A. S. Abstract Newtonian frameworks and their applications // SIAM Journal on Optimization. — 2013. — Vol. 23, no. 4. — P. 2369–2396.

44. Izmailov A. F., Kurennoy A. S. Multiplier methods for optimization problems with Lipschitzian derivatives // VII Московская международная конференция по исследованию операций (ORM2013): Труды. — Т. 1. — М. : МАКС Пресс, 2013. — С. 64–66.

45. Izmailov A. F., Kurennoy A. S. On regularity conditions for complementarity problems // Computational Optimization and Applications. — 2013. — DOI: 10.1007/s10589-013-9604-1.

46. Izmailov A. F., Kurennoy A. S., Solodov M. V. The Josephy–Newton method for semismooth generalized equations and semismooth SQP for optimization // Set-Valued and Variational Analysis. — 2013. — Vol. 21. — P. 17–45.

47. Izmailov A. F., Kurennoy A. S., Solodov M. V. A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems // Mathematical Programming. — 2013. — Vol. 142. — P. 591–604.

48. Izmailov A. F., Kurennoy A. S., Solodov M. V. Local convergence of the method of multipliers for variational and optimization problems under the sole noncriticality assumption // Computational Optimization and Applications. To appear.

49. Izmailov A. F., Pogosyan A. L., Solodov M. V. Semismooth Newton method for the lifted reformulation of mathematical programs with complementarity constraints // Computational Optimization and Applications. — 2012. — Vol. 51, no. 1. — P. 199–221.

50. Izmailov A. F., Solodov M. V. Karush–Kuhn–Tucker systems: regularity conditions, error bounds, and a class of Newton-type methods // Mathematical Programming. — 2003. — Vol. 95. — P. 631–650.

51. Izmailov A. F., Solodov M. V. Solution sensitivity for Karush–Kuhn–Tucker systems with nonunique Lagrange multipliers // Optimization. — 2010. — Vol. 95. — P. 747–775.

52. Izmailov A. F., Solodov M. V. Stabilized SQP revisited // Mathematical Programming. — 2012. — Vol. 133. — P. 93–120.

53. Newton's method for generalized equations : Technical Summary Report : 1965 / Mathematics Research Center, University of Wisconsin ; Executor: N. H. Josephy. — Madison, Wisconsin : 1979.

54. Quasi-Newton methods for generalized equations : Technical Summary Report : 1966 / Mathematics Research Center, University of Wisconsin ; Executor: N. H. Josephy. — Madison, Wisconsin : 1979.

55. Kanzow C., Fukishima M. Solving box constrained variational inequalities by using the natural residual with D-gap function globalization // Operations Research Letters. — 1998. — Vol. 86. — P. 45–51.

56. Klatte D. Nonlinear optimization problems under data perturbations // Modern Methods of Optimization / Ed. by Werner Krabs, Jochem Zowe. — Springer Berlin Heidelberg, 1992. — Vol. 378 of Lecture Notes in Economics and Mathematical Systems. — P. 204–235.

57. Klatte D. Upper Lipschitz behavior of solutions to perturbed $C^{1,1}$ programs // Mathematical Programming. — 2000. — Vol. 88, no. 2. — P. 285–311.

58. Klatte D., Kummer B. Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications. — Dordrecht : Kluwer Academic Publishers, 2002.

59. Klatte D., Tammer K. On the second order sufficient conditions to perturbed $C^{1,1}$ optimization problems // Optimization. — 1988. — Vol. 19. — P. 169–180.

60. Kummer B. Newton's method for non-differentiable functions // Mathematical research. — 1988. — Vol. 45. — P. 114–125.

61. Kummer B. Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis // Lecture Notes in Economics and Mathematical Systems. — 1992. — Vol. 382. — P. 171–194.

62. LANCELOT. — `http://www.cse.scitech.ac.uk/nag/lancelot/lancelot.shtml`.

63. Levy A. B. Errata in implicit multifunction theorems for the sensitivity analysis of variational conditions // Mathematical Programming. — 1999. — Vol. 86, no. 2. — P. 439–441.

64. Luo Z.-Q., Pang J.-S., Ralph D. Mathematical programs with equilibrium constraints. 1996 // Cambridge University Press. — 1996.

65. Mangasarian O. L. A finite Newton method for classification // Optimization Methods and Software. — 2002. — Vol. 17, no. 5. — P. 913–929.

66. Mordukhovich B. S. Stability theory for parametric generalized equations and variational inequalities via nonsmooth analysis // Transactions of the American Mathematical Society. — 1994. — Vol. 343. — P. 609–657.

67. Mordukhovich B. S. Variational Analysis and Generalized Differentiation. — Berlin, Germany : Springer, 2006.

68. Murtagh B. A., Saunders M. A. A projected Lagrangian algorithm and its implementation for sparse nonlinear constraints // Mathematical Programming Study. — 1982. — Vol. 16. — P. 84–117.

69. MINOS 5.0 user's guide : Technical Report SOL : 83.20 / Stanford University ; Executor: B. A. Murtagh, M. A. Saunders : 1983.

70. Nocedal J., Wright S. J. Numerical Optimization. — Second edition. — New York : Springer, 2006.

71. Outrata J., Kocvara M., Zowe J. Nonsmooth approach to optimization problems with equilibrium constraints: theory, applications and numerical results. — Springer, 1998. — Vol. 28.

72. Pang J.-S., Gabriel S. A. NE/SQP: A robust algorithm for the nonlinear complementarity problem // Mathematical Programming. — 1993. — Vol. 60, no. 1–3. — P. 295–337.

73. Pang J.-S., Qi L. Nonsmooth equations: motivation and algorithms // SIAM Journal on Optimization. — 1993. — Vol. 3. — P. 443–465.

74. Powell M. J. D. A method for nonlinear constraints in minimization problems // Optimization. — London and New York : Academic Press, 1969. — P. 283–298.

75. Qi L. Convergence analysis of some algorithms for solving nonsmooth equations // Mathematics of operations research. — 1993. — Vol. 18, no. 1. — P. 227–244.

76. Qi L. Superlinearly convergent approximate Newton methods for LC^1 optimization problems // Mathematical Programming. — 1994. — Vol. 64. — P. 277–294.

77. Qi L., Jiang H. Semismooth Karush–Kuhn–Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations // Mathematics of Operations Research. — 1997. — Vol. 22. — P. 301–325.

78. Qi L., Sun J. A nonsmooth version of newton's method // Mathematical programming. — 1993. — Vol. 58, no. 1-3. — P. 353–367.

79. Robinson S. M. A quadratically convergent algorithm for general nonlinear programming problems // Mathematical Programming. — 1972. — Vol. 3. — P. 145–156.

80. Robinson S. M. Stability theory for systems of inequalities, Part II. Differentiable nonlinear systems // SIAM Journal on Numerical Analysis. — 1976. — Vol. 13. — P. 497–513.

81. Robinson S. M. Strongly regular generalized equations // Mathematics of Operations Research. — 1980. — Vol. 5. — P. 43–62.

82. Robinson S. M. Generalized equations and their solutions. part ii: applications to nonlinear programming // Mathematical Programming Study. — 1982. — Vol. 19. — P. 200–221.

83. Robinson S. M. Newton's method for a class of nonsmooth functions // Set-Valued Analysis. — 1994. — Vol. 2. — P. 291–305.

84. Rockafellar R. T. Computational schemes for large-scale problems in extended linear-quadratic programming // Mathematical Programming. — 1990. — Vol. 48, no. 1–3. — P. 447–474.

85. Rockafellar R. T., Wets R. J. B. Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time // SIAM Journal on Control and Optimization. — 1990. — Vol. 28, no. 4. — P. 810–822.

86. Rockafellar R. T., Wets R. J.-B. Variational Analysis. — Berlin : Springer-Verlag, 1998.

87. Serre D. Matrices: Theory and applications. — 2nd edition. — Springer, 2010. — Vol. 216.

88. Stein O. Lifting mathematical programs with complementarity constraints // Mathematical programming. — 2012. — Vol. 131, no. 1-2. — P. 71–94.

89. von Heusinger A., Kanzow C. SC^1 optimization reformulations of the generalized Nash equilibrium problem // Optimisation Methods and Software. — 2008. — Vol. 23, no. 6. — P. 953–973.

90. von Heusinger A., Kanzow C. Optimization reformulations of the generalized Nash equilibrium problem using Nikaido-Isoda-type functions // Computational Optimization and Applications. — 2009. — Vol. 43, no. 3. — P. 353–377.

91. Wild E. W. Optimization-based Machine Learning and Data Mining. — ProQuest, 2008.