

© 2020 г. С.Ю. ГОРОДЕЦКИЙ, канд. физ.-мат. наук (gorosyu@gmail.com)
(Нижегородский государственный университет им. Н.И. Лобачевского)

ДИАГОНАЛЬНОЕ ОБОБЩЕНИЕ МЕТОДА DIRECT НА ЗАДАЧИ С ОГРАНИЧЕНИЯМИ¹

Метод DIRECT решает задачи липшицевой глобальной оптимизации в гиперинтервале при неограниченном диапазоне значений констант Липшица. Предложено расширение принципов DIRECT на задачи с многоэкстремальными ограничениями при использовании в гиперинтервалах сразу двух измерений функций на концах выбираемых главных диагоналей. Представлены вычислительные иллюстрации, включая решение задачи с разрывами. Выполнен анализ сходимости.

Ключевые слова: глобальная оптимизация, липшицевы функции, метод DIRECT, многоэкстремальные ограничения, функции с разрывами, двухточечная диагональная схема, вычислительные эксперименты.

DOI: 10.31857/S0005231020080073

1. Введение

Статья продолжает исследования, начатые в [1], и посвящена разработке специального DIRECT-подобного метода условной глобальной оптимизации для задач

$$(1) \quad Q^* = \min Q(x), \quad x \in X,$$

$$(2) \quad X = \{x \in D = [a, b] \subset \mathbb{R}^N : g(x) \leq 0\},$$

где векторы a и b конечны, целевая функция Q и функция ограничения g могут сложно зависеть от x и считаются многоэкстремальными. Конкретные требования на них будут указаны далее, но считаем, что минимум в (1)–(2) существует в обычном или расширенном смысле согласно [1], где под расширенной понимается трактовка (1)–(2) при $X = \emptyset$ как задачи минимизации невязки $g(x)$ на D . Если ограничений неравенств в (2) несколько, то в (1)–(2) они заменяются одним ограничением с функцией вида $g(x) = \max\{g_1(x), \dots, g_m(x)\}$. Предполагается алгоритмическая форма задания функций и существенность времени их вычисления. Это делает оправданным разработку для (1)–(2) методов, основанных на планировании новых точек испытаний, исходя из принятых требований оптимальности к их размещению.

Потребность в таких методах возникает в различных областях, включая разработку систем управления. Действительно, формальное применение современных методов синтеза оптимальных регуляторов по заданным (как правило, линейным) целевым выходам (например, [2–6]) могут еще не полностью удовлетворять конструктора-разработчика, так как его еще часто интересует уменьшение или ограничение содержательных нелинейных показателей качества (таких как время окончательного вхождения откликов нелинейной системы в желаемую окрестность входного задающего сигнала, величины

¹ Работа выполнена при поддержке научно-образовательного математического центра “Математика технологий будущего”.

перерегулирований и т.п.). Оптимальная настройка таких критериев может быть выполнена численно в форме решения задач вида (1)–(2) по оставшимся свободным параметрам первичного алгоритма синтеза регулятора. Заметим, что нелинейные критерии в таких задачах часто имеют разрывы либо за счет эффекта бифуркаций, либо из-за самого вида критерия (например, разрывы характерны для указанного выше критерия “времени вхождения”). В [7] использование прямой численной оптимизации продемонстрировано на модельном примере, причем методов локальной оптимизации оказалось недостаточно.

Применение к таким задачам различных методов липшицевой глобальной оптимизации, использующих оценки констант Липшица по результатам измерений функций (не претендуя на общность, укажем лишь несколько методов из [8–12]), вызовет неограниченный рост этих оценок при появлении разрывов в окрестностях глобальных минимумов. Это приведет к вырождению методов в почти равномерный перебор.

Задачами указанного выше типа еще в период 1978–90 гг. занималась группа в отделе Ю.И. Неймарка в НИИ Прикладной математики и кибернетики при ГГУ (ныне ННГУ) им. Н.И. Лобачевского. Это были крупные хозяйственные работы по заказу ряда предприятий. Для решения задач вида (1)–(2), связанных с оптимальной настройкой следящих систем по 4–5 основным параметрам, применялись методы условной глобальной оптимизации с адаптивными стохастическими моделями, разработанные в [13] для некоторых подклассов кусочно-непрерывных функций. Методы основывались на общей концепции, первоначально предложенной Ю.И. Неймарком в [14, гл. VIII, § 7], первый вариант реализации которой представлен в [15]. Методы обладали сходимостью за счет всюду плотного в пределе размещения точек испытаний. Однако в [13] для этих методов были получены аналитические оценки относительной плотности размещения испытаний, доказывающие существенную неравномерность такого размещения. Недостатком этих методов являлась невозможность их точной вычислительной реализации, а наличие лишь приближенной.

Методы с несколько похожим поведением могут быть получены на основе совершенно другого подхода, лежащего в основе оригинального метода DIRECT (от **D**ividing **R**ectangles), предложенного в [16] 1993 г. Этот метод, хотя и относится к липшицевой оптимизации, в действительности применим и к задачам с наличием разрывов. Исходная версия метода построена в [16] для задач на гиперинтервале D без функциональных ограничений

$$(3) \quad f^* = f(x^*) = \min f(x), \quad x \in D = [a, b] \subset \mathbb{R}^N.$$

Целевая функция предполагается липшицевой с константой $L^f = L$. Однако значение L считается неизвестным, не оценивается и может быть любым числом из диапазона $[0, \infty)$. DIRECT принципиально отличается от других методов липшицевой оптимизации именно отказом от возможного оценивания и использования конкретных значений константы L . Он относится к компонентным методам, использует адаптивное разбиение начального множества D на компоненты-гиперинтервалы $D_i = [a^i, b^i]$, в центральных точках c^i которых проводятся измерения функции f .

Кратко поясним принцип работы. При известном значении константы Липшица $L^f = L$ легко вычислить нижнюю оценку значений f для каждого гиперинтервала с центром в c^i

$$f_i^-(L) = f^-(L, D_i) = f_i - L d_i/2, \quad f_i = f(c^i),$$

где $d_i = \text{diam}(D_i) = \|b^i - a^i\|$. В этом случае наиболее приоритетным для новых измерений следует считать гиперинтервал D_t с наименьшим значением нижней оценки функции. Поскольку в методе DIRECT значение $L^f = L$ неизвестно, он использует иное правило отбора делимых гиперинтервалов D_t . Оно основано на двух принципах. Первый — недоминируемость

$$(4) \quad \exists L \in [0, \infty) : f^-(L, D_t) = \min \left\{ f^-(L, D_i) : i = 1, \dots, M_k \right\},$$

где M_k — количество гиперинтервалов после k итераций, т.е. недоминируемым является гиперинтервал, являющийся “лучшим” хотя бы при одном из значений L . Второй — достаточная улучшаемость достигнутого минимального значения целевой функции. Это значит, что хотя бы при некоторых значениях L из (4) и заданных $\eta_k > 0$ должно выполняться условие

$$(5) \quad f^-(L, D_t) \leq f_*^k - \eta_k, \quad f_*^k = \min \left\{ f(x^i) : i = 1, \dots, n_k \right\},$$

где x^i — точки выполненных измерений, n_k — их число (в оригинальном DIRECT всегда $n_k = M_k$).

Гиперинтервалы, удовлетворяющие условиям (4) и (5), называют потенциально оптимальными. Именно такие гиперинтервалы DIRECT выделяет для деления на очередной итерации. В [16] найдено простое геометрическое представление, позволяющее легко находить множество гиперинтервалов D_t , удовлетворяющих (4)–(5). Если сопоставить гиперинтервалам D_i точки на плоскости $(d/2, f)$ с координатами $(d_i/2, f_i)$, где $d_i = \|b^i - a^i\|$, $f_i = f(c^i)$, и добавить к ним точку $(0, f_*^k - \eta_k)$, то множеству точек $(d_t/2, f_t)$ потенциально оптимальных гиперинтервалов D_t соответствуют вершины правой-нижней части границы выпуклой линейной оболочки множества всех точек $(d_i/2, f_i)$, включая дополнительную. Для вычислительно эффективного выделения $(d_t/2, f_t)$ и, следовательно, нахождения делимых гиперинтервалов D_t используют правило Грэхема [17].

Для дробления выбранных гиперинтервалов метод применяет схему их деления на три равные части по большему ребру (см. описание в [18]). С использованием центральной схемы измерений при делении одного D_t достаточно провести только два новых измерения $f(x)$ в центрах крайних из трех новых гиперинтервалов, поскольку в центре среднего результат уже известен.

Заметим, что сама схема деления на три в рекурсивной реализации использовалась в глобальной оптимизации для задач (1)–(2) и (3) значительно раньше, еще в программной разработке Compromiss Solver ВЦ АН СССР 80-х гг. XX в. (предтече метода деления на три является метод деления на два, который был опубликован в [19]).

К настоящему времени известны различные обобщения DIRECT. Обширная библиография приведена в [8]. Укажем на некоторые. Оригинальный DIRECT относительно медленно уточняет найденные оценки решения, в [20]

предложена достаточно удачная локально-ориентированная модификация. В [21] построен, а также описан в [8] более эффективный вариант метода [16] для задач (3), работающий по диагональной схеме деления на три (в каждом гиперинтервале проводится не одно, а два измерения на концах специально ориентируемых главных диагоналей, образующих вместе эффективную диагональную кривую, построение которой было ранее предложено в [22]). В [21] кроме двухточечной диагональной схемы измерений применен также специальный механизм балансировки локальной и глобальной стратегий поиска за счет искусственных усечений множеств потенциально оптимальных гиперинтервалов, что улучшило сходимость к глобальному решению. В [23] получено нетривиальное обобщение метода DIRECT на задачи с вычислимой липшицевой производной, константа Липшица которой неизвестна и может изменяться на промежутке $[0, \infty)$; метод построен для задач (3) с размерностью $N = 1$. В [24] метод из [23] распространен на многомерные задачи с использованием специальной нецентральной одноточечной схемы разбиения множества D в задаче (3). Также в 2001 г. получен концептуально близкий к DIRECT метод для задач (1)–(2) с несколькими ограничениями, второе издание 2009 г. этой публикации приведено в [18]. Представленный в [18] метод в отличие от [1] использует специально сконструированный на случай наличия ограничений подход к выделению делимых гиперинтервалов. Принцип их выделения существенно отличается от оригинального DIRECT.

Заметный интерес к DIRECT-подобным методам привел к проведению в [25] представительного экспериментального сравнительного исследования нескольких методов [16, 20, 21] этого семейства в сопоставлении с группой эвристических методов, включая эволюционно-генетические алгоритмы. Тестирование проведено на задачах без функциональных ограничений размерности $N = 5$, представляющих простой и сложный классы из 100 тестовых функций GKLS генератора (его описание приведено, например, в [8]). Результаты, представленные в [25], показали, что при достаточно большом числе измерений n_k решение с наибольшей надежностью обеспечивается именно диагональной реализацией DIRECT, представленной в [21], и особенно на классе сложных тестовых задач. Это актуализирует разработку диагональной версии одноточечного варианта метода [1] в задачах с функциональными ограничениями.

В [1] предложено два подхода к обобщению метода DIRECT на задачи с ограничениями. Оба существенно отличаются от представленного в [18]. Первый подход сводит исходную задачу с ограничениями к последовательности задач без ограничений с перестраиваемой целевой функцией. Второй основан на непосредственном распространении принципов построения DIRECT на задачи с ограничениями. В обоих подходах в [1] применена схема деления на три с измерением функций в центральных точках (далее измерения функций задачи называем испытанием). Данная статья обобщает второй из подходов, предложенных в [1], на двухточечную диагональную схему измерений [22], использованную в [21]. При этом испытания задачи в гиперинтервалах D_t проводятся в двух точках a^t, b^t , расположенных на концах одной из главных диагоналей. Ориентация этой диагонали определяется специальным образом согласно безызбыточной стратегии разбиения, предложенной в [22] (описа-

на также в [8, 21]). Прежде чем перейти к формальной постановке задачи, поясним основы подхода и терминологию, использованную в [1] и далее применяемую.

Поскольку в задачах с ограничением (1)–(2) добавляется фактор допустимости или недопустимости точек измерений, возможна ситуация пустоты допустимого множества X . При $X \neq \emptyset$ целью является определение элементов x^* из множества X^* глобальных минимумов задачи. В случае пустоты ($X = \emptyset$), следуя [1], неявно трактуем решение в расширенном смысле, как определение глобального минимума x^* невязки в ограничениях (предполагается что этот минимум существует):

$$g^* = g(x^*) = \min g(x), \quad x \in D = [a, b] \subset \mathbb{R}^N.$$

Множество таких минимумов по-прежнему обозначаем через X^* .

Далее удобно трактовать эту задачу как задачу с добавленным функциональным ограничением $\tilde{g}(x) \leq 0$, заданным фиктивной функцией $\tilde{g}(x) \equiv -1$.

Заметим, что в (1)–(2) при $X \neq \emptyset$ и наличии допустимых и недопустимых подмножеств точек различия значений $Q(x)$ на подмножествах с нарушением ограничений не оказывают влияния на X^* — множество глобальных минимумов задачи. Однако наличие слишком малых значений Q в недопустимых областях может существенно повлиять на процесс численного решения. Поэтому целевую функцию в (1) при $X \neq \emptyset$ целесообразно заменить функцией вида

$$f(x) = \max\{Q(x); Q^* - \xi\}, \quad \xi \geq 0.$$

Поскольку значение в условном глобальном минимуме Q^* неизвестно, заменим его текущей оценкой Q_k^* . В результате метод с учетом ограничений далее будет применен не к решению задачи в форме (1)–(2), а к специально модифицированной, возникающей из указанных выше замен

$$(6) \quad \min f_k(x), \quad x \in X_k,$$

$$(7) \quad X_k = \{x \in D = [a, b] \subset \mathbb{R}^N : g_k(x) \leq 0\},$$

где k — номер итерации. Ее точный вид приведен в разделе 2. Чтобы не усложнять обозначения, далее будем опускать индексы k у f_k и g_k , записывая их как f и g . Предположения о функциях далее примем по отношению к задаче (6)–(7). Они будут отличаться от требований в [1].

В [1] предположения о функциях следующие: f и g липшицевы в евклидовой норме с неизвестными константами Липшица: $L^f = L \in [0, \infty)$ и $L^g \in [0, \alpha L)$ (где $\alpha > 0$ — введенный параметр класса задач). При неограниченном L значение L^g также может быть сколько угодно большим. Принцип потенциальной оптимальности при отборе делимых на итерации гиперинтервалов D_t в [1] заменен на модифицированную потенциальную оптимальность, которая состоит в одновременном выполнении для D_t трех основных принципов.

Первый назовем принципом наименьшего нарушения, а именно: из множества \mathfrak{D}^k всех гиперинтервалов текущего разбиения выделим на итерации k сокращенное подмножество $\hat{\mathfrak{D}}^k$ гиперинтервалов, для которых выполнены требования:

$\nexists D', D'' \in \tilde{\mathfrak{D}}^k$, что если с учетом измерений $f' = f(c')$, $f'' = f(c'')$ выполнено:

$$(8) \quad D' \neq D'', \quad \text{diam}(D') = \text{diam}(D''), \quad f' = f'', \quad g' = g(c') > 0, \quad g'' = g(c'') > 0,$$

то $g' \neq g''$ и, кроме того, если гиперинтервал текущего разбиения $D'' \notin \tilde{\mathfrak{D}}^k$, а $D' \in \tilde{\mathfrak{D}}^k$, то из выполнения для них условий (8) должно следовать неравенство $g' < g''$.

Второй принцип — модифицированной недоминируемости. К модифицированно недоминируемым отнесем все $D_t \in \tilde{\mathfrak{D}}^k$, удовлетворяющие следующим условиям

$$\exists L \in [0, \infty):$$

$$(9) \quad f^-(L, D_t) = \min \left\{ f^-(L, D_i) : D_i \in \tilde{\mathfrak{D}}^k, \exists L^g \in [0, \alpha L] : g^-(L^g, D_i) \leq 0 \right\},$$

$$(10) \quad \exists L^g \in [0, \alpha L] : g^-(L^g, D_t) \leq 0.$$

Очевидно, что эти условия — прямое обобщение принципа недоминируемости метода DIRECT, описанного, например, в [8].

Третий принцип — достаточная улучшаемость. Пусть f_*^k — текущая оценка минимального значения в (6)–(7) по результатам испытаний, тогда для D_t из $\tilde{\mathfrak{D}}^k$ при некотором значении L , когда выполнены условия (9)–(10), должно также для заданного η_k соблюдаться неравенство

$$(11) \quad f^-(L, D_t) \leq f_*^k - \eta_k, \quad \eta_k > 0.$$

Здесь нижние оценки функций имеют вид:

$$(12) \quad f_i^-(L) = f^-(L, D_i) = f_i - L d_i/2; \quad g_i^-(L^g) = g^-(L^g, D_i) = g_i - L^g d_i/2.$$

В [1] предложен конструктивный алгоритм выделения гиперинтервалов D_t с указанными свойствами (9)–(11). В примененной для этого технике существенно использована удобная структура нижней оценки значения ограничения при центральном измерении из (12). Важным оказалось то, что знак первого элемента g_i в $g^-(L^g, D_i)$ однозначно определяет допустимость или недопустимость измерения, выполненного в центре D_i .

В двухточечном диагональном обобщении [21] стандартного метода DIRECT (в задачах без ограничений) использован другой вид нижних оценок функции f . В двухточечной схеме целевая функция f в D_i измерена в двух точках a^i, b^i на концах выбранной главной диагонали D_i . Нижнюю оценку функции f строим на этой диагонали. При обычных для DIRECT предположениях о липшицевости $f(x)$ с константой $L^f = L \in [0, \infty)$ получим минимальное значение оценки на диагонали

$$f_i^-(L) = f^-(L, D_i) = \frac{f_i^a + f_i^b}{2} - L \frac{d_i}{2}.$$

Такой вид оценки не подходит для функции $g(x)$ при наличии ограничений, если строить двухточечное обобщение метода из [1]. Действительно, в оценке такого вида для $g(x)$ первым элементом будет $(g_i^a + g_i^b)/2$. Очевидно, знак этой суммы однозначно не выделяет ситуации наличия допустимой точки среди двух испытаний в a^i, b^i .

Поэтому в данной статье применен другой вид нижних оценок, вытекающий из измененных предположений относительно свойств f и g в задаче (6)–(7). Поясним это на примере $g(x)$. Пусть получены измерения g_i^a, g_i^b на концах a^i, b^i главной диагонали длины d_i гиперинтервала D_i . Вычислим значение $L_i^g = |g_i^a - g_i^b|/d_i$. Очевидно, что в D_i должно выполняться неравенство $L^g \geq L_i^g$. Примем значение L^g для D_i в виде $L^g = L_i^g + \Delta L^g$, где ΔL^g будем считать одинаковым для всех гиперинтервалов. Тогда нижняя оценка значений функции g на диагонали от a^i до b^i примет вид

$$(13) \quad g_i^-(\Delta L^g) = g^-(\Delta L^g, D_i) = \frac{g_i^a + g_i^b}{2} - (L_i^g + \Delta L^g) \frac{d_i}{2} = G_i - \Delta L^g \frac{d_i}{2};$$

$$\Delta L^g \geq 0,$$

где $G_i = \min\{g_i^a; g_i^b\}$. При этом знак G_i однозначно определит наличие допустимого испытания на концах используемой диагонали. Это позволит непосредственно обобщить результаты [1] на случай двухточечных диагональных измерений.

2. Предварительные замечания и постановка задачи

2.1. Преобразование исходной задачи

Пусть выполнено k итераций некоторого метода и проведено n_k испытаний (измерений функций задачи) в точках $x^i \in D$ с сохранением в памяти значений Q_i, g_i ($i = 1, \dots, n_k$). Определим рекордное значение Q_k^* в (1)–(2):

$$(14) \quad Q_k^* = \begin{cases} \min\{Q_i : g_i \leq 0, i = 1, \dots, n_k\}, & \text{если } \exists g_i \leq 0, \\ +\infty, & \text{если } \forall i = 1, \dots, n_k : g_i > 0. \end{cases}$$

Если еще не найдено ни одной допустимой точки, значение $Q_k^* = +\infty$, и с учетом расширенной трактовки решения выполним переход от решения (1)–(2) к задаче поиска минимума невязки $g(x)$ на D до тех пор, пока не встретится первая допустимая точка. Для единообразного рассмотрения алгоритмов решения перейдем от исходной формы задачи (1)–(2) к ее измененному представлению в форме (6)–(7), где функции $f_k(x)$ и $g_k(x)$ имеют вид:

$$(15) \quad f_k(x) = \begin{cases} \max\{Q(x); Q_k^* - \xi_k\}, & Q_k^* \neq +\infty, \\ g(x), & Q_k^* = +\infty; \end{cases}$$

$$(16) \quad g_k(x) = \begin{cases} g(x), & Q_k^* \neq +\infty, \\ -1, & Q_k^* = +\infty. \end{cases}$$

При проведении испытаний этой задачи в точках x^i вычисляются и сохраняются значения Q^i и g^i исходной задачи, а величины вспомогательных функций $f_k(x^i)$ и $g_k(x^i)$ восстанавливаются согласно (15), (16) по сохранившимся в памяти результатам измерений исходных функций. Численный метод решает задачу (6)–(7), (15), (16), которая на итерациях сама изменяет свою структуру в зависимости от рекордного значения (14). Такая форма

представления задачи удобна для построения метода. Текущее минимальное значение вспомогательной целевой функции (15) обозначим

$$(17) \quad f_*^k = \min \{ f_k(x^i) : g_k(x^i) \leq 0, i = 1, \dots, n_k \}.$$

Значение f_*^k всегда конечно. Далее, как уже было указано выше, обычно будем опускать у вспомогательных функций f_k и g_k нижний индекс k .

2.2. Использование безызбыточной диагональной стратегии разбиения

Поясним правила размещения точек испытаний при использовании безызбыточной диагональной стратегии разбиения [22] (см. также [8]). Перед началом поиска проводим два начальных испытания в вершинах a и b исходного гиперинтервала D , соответствующих минимальным и максимальным значениям координат. В последующем каждый из делимых гиперинтервалов разделяется на три равные части по первому из его больших ребер. У очередного делимого гиперинтервала D_t перед делением всегда имеется пара вершин a^t и b^t (назовем их активными), в которых испытания уже проведены. Эти вершины лежат на одной из главных диагоналей (далее будем называть ее активной). Эти диагонали могут иметь разные ориентации. Пусть r — длина делимого ребра, а e — нормированный вектор, направленный из вершины a^t вдоль этого ребра так, чтобы $e^\top (b^t - a^t) > 0$. Согласно [22] при делении D_t выбираются две точки u и v , в которых могут проводиться новые испытания:

$$u = a^t + \frac{2}{3} r e; \quad v = b^t - \frac{2}{3} r e.$$

Три новых гиперинтервала порождаются из D_t по следующим парам активных вершин (см. рис. 1): a^t и v , v и u , u и b^t .

В [22] показано, что при описанном способе выбора точек новых измерений некоторые из них могут являться активными и для других смежных гиперинтервалов (см. правую часть рис. 1). Поэтому в некоторых из новых точек u или v испытания задачи могут быть уже проведены ранее и повторно их

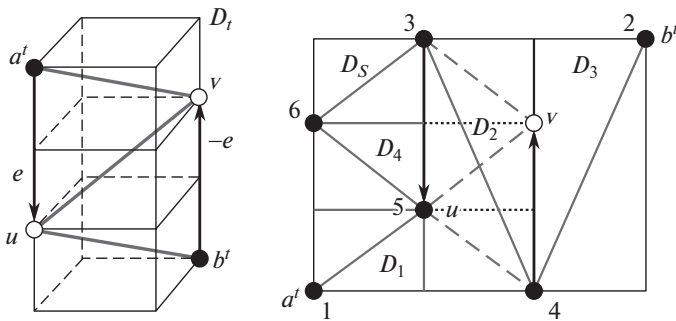


Рис. 1. Слева — выбор новых точек измерений и ориентаций активных диагоналей при делении на три выбранного гиперинтервала по безызбыточной диагональной стратегии в \mathbb{R}^3 ; справа — пример возможного разбиения начального D в \mathbb{R}^2 : при делении D_2 новая точка u совпала с точкой прежнего испытания с номером 5.

выполнять не следует — можно извлечь готовые результаты из памяти, экономя на количестве испытаний. Быстрое выполнение операции поиска для последующего извлечения обеспечено созданием специально организованной структуры хранения результатов испытаний.

2.3. Новая модель поведения функций задачи и диагональное обобщение DIRECT-подобного принципа отбора делимых гиперинтервалов с учетом ограничений

Пусть на итерации k имеется текущее разбиение $\mathfrak{D}^k = \{D_i^k\}_{i=1}^{M_k}$ начального множества D на гиперинтервалы $D_i = D_i^k$, полученное по безызбыточной диагональной стратегии разбиения [8, 22]. В каждом D_i , в вершинах a^i и b^i , оканчивающих активную диагональ, проведены испытания задачи. Пусть $f_i^a = f(a^i)$, $f_i^b = f(b^i)$, $g_i^a = g(a^i)$, $g_i^b = g(b^i)$ — результаты испытаний. Введем обозначения:

$$(18) \quad L_i = L_i^f = \left| f_i^a - f_i^b \right| / d_i, \quad L_i^g = \left| g_i^a - g_i^b \right| / d_i.$$

Примем, что функции f и g , определяемые соотношениями (15) и (16), липшицевы в D с нормой $\|\cdot\|$ и константами $L^f = L$ и L^g , причем в гиперинтервале D_i значения этих констант с использованием (18) представимы в виде:

$$L = L_i + \Delta L, \quad L^g = L_i^g + \Delta L^g.$$

Приращения ΔL и ΔL^g неотрицательны и одинаковы для всех гиперинтервалов разбиения. Таким образом, гиперинтервалам разбиения назначаются разные по величине минимально возможные значения констант и одинаковые приращения к ним. Введем следующие предположения. Считаем, что приращение ΔL неизвестно и его величина может быть сколь угодно большой, т.е. $\Delta L \in [0, \infty)$; значение ΔL^g также неизвестно, но $\Delta L^g \in [0, \alpha \Delta L]$, где $\alpha > 0$ — постоянный параметр метода. Заметим, что указанные выше предположения применительно к функции f впервые были использованы и экспериментально исследованы в [26] при построении методов для задач без ограничений в \mathbb{R}^1 .

Выполнив преобразования, аналогичные (13), получим следующий вид минимальных значений нижних оценок функций f и g на активных диагоналях гиперинтервалов D_i :

$$(19) \quad \begin{aligned} f_i^-(\Delta L) &= f^-(\Delta L, D_i) = F_i - \Delta L d_i / 2, \\ g_i^-(\Delta L^g) &= g^-(\Delta L^g, D_i) = G_i - \Delta L^g d_i / 2, \end{aligned}$$

$$(20) \quad F_i = \min \left\{ f_i^a, f_i^b \right\}, \quad G_i = \min \left\{ g_i^a, g_i^b \right\}.$$

Значения F_i и G_i из (20) назовем базовыми характеристиками функций f и g на гиперинтервале D_i . Следует обратить внимание на то, что нижние оценки (19), (20) двухточечной диагональной схемы отличаются от аналогичных оценок центральной схемы заменой прежних значений f_i , L , g_i , L^g новыми F_i , ΔL , G_i , ΔL^g . Это позволяет доказать ряд фактов и разработать алгоритмы, близкие к полученным в [1].

Введем необходимые определения.

Определение 1. Подмножеством гиперинтервалов с наименьшим нарушением (ограничений) назовем подмножество $\tilde{\mathfrak{D}}^k$ множества всех гиперинтервалов текущего разбиения \mathfrak{D}^k , в которое из каждой группы гиперинтервалов $\{D\} \subseteq \mathfrak{D}^k$ одинакового диаметра, с одинаковыми значениями базовых характеристик F и положительными значениями G базовых характеристик ограничений включены только гиперинтервалы с наименьшим значением G в этой группе.

Проводимое сокращение \mathfrak{D}^k до подмножества $\tilde{\mathfrak{D}}^k$ является значимым при решении преобразованной задачи (6)–(7), когда уже обнаружены допустимые точки, т.е. $Q_k^* \neq +\infty$. Именно в этом случае в силу (15)–(16) может возникнуть значительное количество гиперинтервалов D_i с одинаковыми значениями $F_i = Q_k^* - \xi_k$ и $G_i > 0$.

Определение 2. Гиперинтервал $D_t \in \mathfrak{D}^k$ назовем модифицировано недоминируемым на итерации k , если $D_t \in \tilde{\mathfrak{D}}^k$ и $\exists \Delta L \in [0, \infty)$ такое, что для нижних оценок на активных диагоналях из (19) функций $f_k(x)$, $g_k(x)$ из (15)–(16) выполнено:

$$(21) \quad f^-(\Delta L, D_t) = \min \left\{ f^-(\Delta L, D_i) : D_i \in \tilde{\mathfrak{D}}^k, \exists \Delta L^g \in [0, \alpha \Delta L] : g^-(\Delta L^g, D_i) \leq 0 \right\},$$

$$(22) \quad \exists \Delta L^g \in [0, \alpha \Delta L] : g^-(\Delta L^g, D_t) \leq 0.$$

Определение 3. Гиперинтервал D_t назовем модифицированно потенциально оптимальным (на итерации k), если он модифицированно недоминируемый на этой итерации и хотя бы для некоторых ΔL , при которых для него соблюдены требования (21)–(22), выполняется также условие достаточной улучшаемости с заданным $\eta_k > 0$ и f_*^k из (17):

$$(23) \quad f^-(\Delta L, D_t) \leq f_*^k - \eta_k.$$

Следуя [1], сформулируем простое утверждение, позволяющее записать требования (21)–(22) в эквивалентной, но более простой форме.

Утверждение 1 (об эквивалентности). Условия (21)–(22) модифицированной недоминируемости $D_t \in \tilde{\mathfrak{D}}^k$ в диапазоне $\Delta L \in [\Delta L', \Delta L'']$ эквивалентны следующим условиям: $\exists \Delta L \in [\Delta L', \Delta L'']$ такое, что

$$(24) \quad f^-(\Delta L, D_t) = \min \left\{ f^-(\Delta L, D_i) : D_i \in \tilde{\mathfrak{D}}^k, g^-(\alpha \Delta L, D_i) \leq 0 \right\}; \\ g^-(\alpha \Delta L, D_t) \leq 0.$$

Доказательство утверждения приведено в Приложении.

Заметим, что по отношению к определению 2 утверждение применяем, положив $[\Delta L', \Delta L''] = [0, \infty)$.

Основная задача ставится следующим образом. Для получения диагонального варианта DIRECT-подобного метода условной глобальной оптимизации требуется определить подходы к построению и структуре алгоритмов выде-

ления на итерации k подмножества (обозначим его $\widehat{\mathcal{D}}_k$) всех модифицированно потенциально оптимальных гиперинтервалов с учетом упрощающего утверждения 1. Также необходимо провести анализ сходимости построенного метода и выполнить его экспериментальную апробацию. В данной статье описаны указанные выше подходы к обработке информации в алгоритмах формирования $\widehat{\mathcal{D}}_k$ и их структура, выполнен анализ сходимости, а также представлены вычислительные иллюстрации работы метода на двумерных задачах. Детальное описание и аналитическое обоснование алгоритмов выделения $\widehat{\mathcal{D}}_k$, а также вычислительный эксперимент на задачах разных размерностей не являются целью данной статьи, а составляют предмет отдельной публикации.

3. О построении алгоритмов отбора делимых гиперинтервалов при двухточечной диагональной схеме измерений

Множество $\widehat{\mathcal{D}}_k$ гиперинтервалов, делимых на текущей k -й итерации, является множеством всех модифицированно потенциально оптимальных гиперинтервалов. Поясним правила их выделения и реализующие их алгоритмы, используя иллюстративные примеры.

Заметим, что в обычном методе DIRECT для задач без ограничений отбор делимых гиперинтервалов происходит сравнением на плоскости $(d/2, f)$ их изображающих точек, где первая координата — половина диаметра гиперинтервала, а вторая — значение целевой функции в его центре. При наличии ограничений и использовании одноточечной центральной схемы гиперинтервалы D_i в [1] представлены трехмерными изображающими точками $(d_i/2, f_i, g_i)$, включающими результат вычисления функций f и g в центре D_i . В рассматриваемой двухточечной схеме значения f_i, g_i в центральных точках заменим базовыми характеристиками этих функций F_i, G_i из (20). После такой замены выделение искомого подмножества $\widehat{\mathcal{D}}_k$ может быть выполнено по двухэтапной процедуре, похожей на описанную в [1].

Поясним характер возникающей при этом обработки накопленной поисковой информации. На первом этапе на множестве трехмерных точек $(d_i/2, F_i, G_i)$ выделим подмножества, соответствующие гиперинтервалам D_i^d

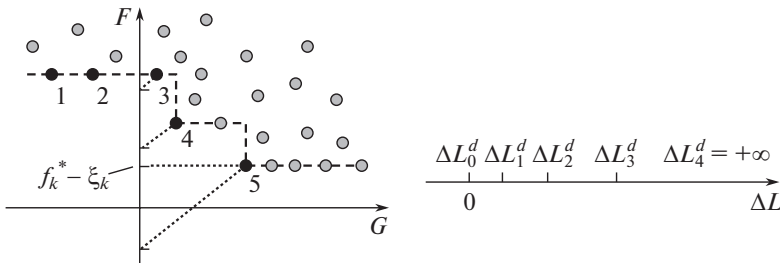


Рис. 2. Слева — набор точек (G_j, F_j) , соответствующих d -слою гиперинтервалов с $\text{diam}(D_j) = d$, черные точки отвечают группам $\{\widehat{D}_s^d\}$ отобранных гиперинтервалов; справа — порождаемые для $\{\widehat{D}_s^d\}$ границы диапазонов $[\Delta L_s^d, \Delta L_{s+1}^d]$ приращений ΔL : значение $s = 0$ соответствует точкам 1 и 2, $s = 1, 2, 3$ соответствуют точкам 3, 4, 5.

одинакового диаметра с $d_i = d$ (назовем такие подмножества d -слоями). В каждом d -слое выполним предварительный отбор гиперинтервалов за счет сравнения соответствующих им точек на плоскости (G, F) при выбранном d . Сравнения выполним так, чтобы выделенные в результате отбора группы гиперинтервалов (обозначим их $\{\widehat{D}_s^d\}$, $s = 0, \dots, S(d)$) удовлетворяли определениям 1 и 2, но только в пределах множества гиперинтервалов своего d -слоя. Проводя рассуждения аналогично [1], но с учетом особенностей диагональных испытаний, можно показать, что в достаточно общей ситуации, представленной на рис. 2 слева, отобранными на плоскости (G, F) будут точки, выделенные черным. Аналогично [1] для групп гиперинтервалов $\{\widehat{D}_s^d\}$, соответствующих выделенным точкам, наборы значений ΔL , при которых для \widehat{D}_s^d в пределах d -слоя выполнены условия (21), (22), образуют диапазоны вида $[\Delta L_s^d, \Delta L_{s+1}^d)$, где при $G_s \leq 0$ значение $\Delta L_s^d = 0$, а при $G_s > 0$ $\Delta L_s^d = 2G_s/(\alpha d)$. В последнем диапазоне $\Delta L_{S(d)+1}^d = +\infty$. Показанные на рис. 2 наклонные точечные линии, имеют коэффициент наклона $1/\alpha$, где α — введенный ранее параметр класса функций ограничений. Прересечения этих линий с осью F порождают указанные значения ΔL_s^d , при которых впервые выполняются условия неположительности $g^-(\alpha \Delta L_s^d, \widehat{D}_s^d) \leq 0$, а ординаты точек пересечения совпадают с $f^-(\Delta L_s^d, \widehat{D}_s^d)$. Легко видеть, что в каждом d -слое будет существовать хотя бы одна точка, соответствующая группе отбираемых гиперинтервалов.

В начале второго этапа происходит объединение множеств значений ΔL_s^d по всем d и по всем s с формированием общего упорядоченного ряда значений:

$$(25) \quad 0 \leq \Delta L_1 < \Delta L_2 < \dots < \Delta L_{m_k} < \Delta L_{m_k+1} = +\infty.$$

Далее, при продолжении второго этапа обработки для каждого промежутка $[\Delta L_j, \Delta L_{j+1})$ приращений ΔL выделяются для различных d -слоев из групп $\{\widehat{D}_s^d\}_{s=0}^{S(d)}$ отобранных гиперинтервалов те, для которых не пусто пересечение:

$$[\Delta L_j, \Delta L_{j+1}) \cap [\Delta L_s^d, \Delta L_{s+1}^d) \neq \emptyset.$$

Обозначим множество таких гиперинтервалов через $\widetilde{\mathfrak{D}}_k^j$, а его элементы через \widetilde{D}_k^j . Каждое из множеств $\widetilde{\mathfrak{D}}_k^j$ отдельно представим набором изображающих точек с координатами $(d/2, F)$.

Заметим, что аналогично геометрической интерпретации правил метода DIRECT, минимальное значение нижней оценки функции $f_k(x)$ на активной диагонали одного из гиперинтервалов \widetilde{D}_k^j , соответствующего некоторой изображающей точке $(d/2, F)$, можно получить как ординату пересечения с осью F прямой, проведенной с коэффициентом наклона ΔL через эту изображающую точку. В оригинальном методе DIRECT сравнение гиперинтервалов происходит по значениям аналогичных нижних оценок для целевой функции при условии, что диапазон изменения константы Липшица есть $[0, +\infty)$. В рассматриваемом случае диапазон изменения ΔL ограничен промежутком $[\Delta L_j, \Delta L_{j+1})$, что требует новых алгоритмов обработки. Их точное описание и обоснование выходят за рамки данной статьи.

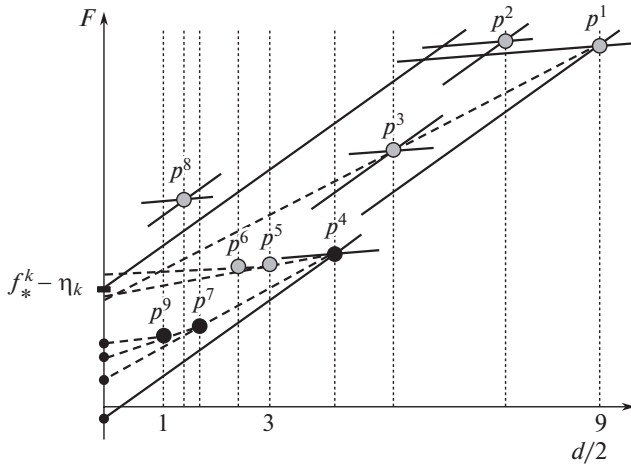


Рис. 3. Пример отбора в множество $\widehat{\mathcal{D}}_k^j$ гиперинтервалов, делимых на k -й итерации, из множества $\widetilde{\mathcal{D}}_k^j$, сформированного для значений ΔL из $[\Delta L_j, \Delta L_{j+1}]$; если гиперинтервалу из $\widetilde{\mathcal{D}}_k^j$ соответствует множество точек $P_k^j = \{p^1, \dots, p^9\}$, то отобранным для деления гиперинтервалам будут отвечать точки $p^9; p^7; p^4$, выделенные черным.

На полученном множестве точек эти алгоритмы должны проводить дополнительный отбор так, чтобы для гиперинтервалов, соответствующих оставшимся точкам, оказались выполнены определения 2 и 3, причем условия из определения 2 должны выполняться для отобранных гиперинтервалов только в пределах множества $\widetilde{\mathcal{D}}_k^j$ и только для значений ΔL из $[\Delta L_j, \Delta L_{j+1}]$. Пример отбора показан на рис. 3. Он иллюстрирует результат выделения подмножества $\widehat{\mathcal{D}}_k^j$ из $\widetilde{\mathcal{D}}_k^j$ на конкретном примере. Вертикальные точечные линии на рис. 3 соответствуют диаметрам возможных существующих d -слоев после 9 итераций на примере исходного множества в виде некоторого четырехмерного гиперинтервала. Границам промежутка $[\Delta L_j, \Delta L_{j+1}]$ на рисунке соответствуют коэффициенты наклона сплошных линий, проведенных через точки. Наклонные штриховые линии отражают процесс отбора. Черным выделены те изображающие точки, которым соответствуют гиперинтервалы, отобранные в этом примере в множество $\widehat{\mathcal{D}}_k^j$ для их последующего включения в полный набор $\widehat{\mathcal{D}}_k$ гиперинтервалов, делимых на итерации k .

Повторяя процедуру выделения $\widehat{\mathcal{D}}_k^j$ для $j = 1, \dots, m_k$ и включая (с устранением повторов) все отобранные таким образом гиперинтервалы в искомое множество $\widehat{\mathcal{D}}_k$, получим полный набор $\widehat{\mathcal{D}}_k$ гиперинтервалов, делимых на шаге. Нетрудно видеть, что полученное $\widehat{\mathcal{D}}_k$ совпадает со множеством всех модифицированно потенциально оптимальных на итерации k гиперинтервалов.

Лемма 1 (о наибольшем гиперинтервале). Построенное множество гиперинтервалов $\widehat{\mathcal{D}}_k$ всегда содержит хотя бы один гиперинтервал наибольшего диаметра из гиперинтервалов текущего разбиения \mathcal{D}^k .

Лемма используется при доказательстве теоремы о сходимости в разделе 4. Доказательства леммы и последующей теоремы приведены в Приложении.

4. Структурное описание диагонального метода с учетом ограничений, анализ сходимости

4.1. Описание построенного метода

На каждой итерации метод должен выделить очередной набор $\widehat{\mathcal{D}}_k$ делимых гиперинтервалов D_t , реализуя принципы, предложенные в разделе 2. Краткое описание структуры возникающей при этом обработки накопленной поисковой информации приведено в разделе 3. Алгоритмы этой обработки являются центральной частью построенного метода и в отличие от [1] учитывают специфику двухточечных диагональных измерений в гиперинтервалах. Из материала разделов 2 и 3 следует, что выполнение алгоритмов обработки зависит от значения ряда величин, а именно: η_k из условия (23); α — параметр, введенный в начале подраздела 2.3 при описании требований к диапазону значений ΔL^g — приращений константы Липшица функции ограничения; а также ξ_k из (15). Значения η_k и ξ_k выберем пропорциональными некоторой величине Δf_k , которая в [1] была названа базовым значением целевой функции решаемой задачи, а именно примем: $\eta_k = \Delta f_k \varepsilon$; $\xi_k = \Delta f_k \delta$. Способ вычисления Δf_k предложен и подробно описан в [1, раздел 3] (в подразделе Новое базовое значение). Единственное отличие, применительно к рассматриваемому диагональному варианту метода, состоит в том, что здесь для вычисления Δf_k вместо набора измерений целевой функции в центрах гиперинтервалов текущего разбиения используется набор имеющихся измерений целевой функции $f_k(x)$ из (15) на концах активных диагоналей этих гиперинтервалов. При вычислении Δf_k каждое такое измерение учитывается только один раз, хотя почти каждая точка измерения входит в несколько гиперинтервалов. Значение Δf_k содержательно можно трактовать как грубую оценку разности $f_k^\mu - f_*^k$, где f_*^k определяется в (17), а f_k^μ — значение, при котором относительная мера множества точек $\{x \in D : f_k(x) \leq f_k^\mu\}$ в D равна заданной величине μ из интервала $(0, 1)$. f_k^μ формируется в [1] приближенно по результатам проведенных испытаний. Принятый в [1] алгоритм вычисления Δf_k предполагает задание еще одного параметра \tilde{n} , который определяет пороговое число измерений. А именно если после очередной k -й итерации число измерений n_k впервые превысит заданное \tilde{n} (обозначим этот номер k через \tilde{k}), то после этой итерации происходит обновление $\Delta f_{\tilde{k}}$, и далее при всех $k > \tilde{k}$ используется $\Delta f_k = \Delta f_{\tilde{k}}$.

Таким образом, имеем набор параметров: μ , \tilde{n} , а также ε , α , δ . Параметры ε и δ всегда неотрицательны и влияют на работу метода следующим образом. Увеличение ε , как правило, увеличивает равномерность размещения точек измерений, а его уменьшение увеличивает тенденцию к делению гиперинтервалов малого диаметра. Уменьшение δ , как правило, приводит к уменьшению доли измерений в недопустимых областях, а его увеличение повышает долю недопустимых измерений. Метод управляет этими тенденциями за счет использования трех наборов значений ε , α , δ . Набор ε_0 , α_0 , δ_0 метод применяет на первой стадии поиска, пока число измерений меньше заданного \tilde{n} , а далее использует два других (ε_1 , α_1 , δ_1 , либо ε_2 , α_2 , δ_2), попеременно применяя их в зависимости от номера итерации k . Если k кратно дополнительному параметру \overline{K} , применяется первый набор, а если не кратно, — второй. Значения

параметров двух последних наборов, как правило, выбираются так, чтобы первый соответствовал стратегии уточнения найденных оценок решения, а второй — поиску в менее исследованных областях. Начальный набор должен способствовать достаточной равномерности размещения начальных точек.

Одноточечный метод, построенный в [1], имеет маркировку ExDIR (от **DIRECT Extention**), его двухточечное диагональное обобщение обозначим как ExDIR-diag.

Поскольку DIRECT-подобные методы не оценивают константы Липшица, останов по точности невозможен, поэтому указывают ресурс по количеству испытаний.

Алгоритм 1 (принципиальное описание метода ExDIR-diag).

1. Задаем параметры $\mu, \tilde{n}, \overline{K}, \varepsilon_s, \alpha_s, \delta_s, (s = 0, 1, 2)$ и n_{\max} — максимальное количество испытаний.

2. Проводим два начальных испытания в вершинах a и b исходного гиперинтервала D из (1), соответствующих минимальным и максимальным значениям координат, полагаем номер итерации $k = 1$, количество испытаний $n_k = 2$ и множество гиперинтервалов $\mathfrak{D}^k = \{D\}$, количество гиперинтервалов $M_k = 1$.

3. С учетом проведенных на k -й итерации испытаний обновляем Q_k^* из (14), при необходимости корректируем текущий вид функций (15), (16) решаемой задачи (6)–(7). Если задача изменилась, пересчитываем значения функций $f_k(x)$ и $g_k(x)$ на концах a^i, b^i активных диагоналей всех гиперинтервалов D_i текущего множества \mathfrak{D}^k гиперинтервалов. При этом функции исходной задачи повторно не вычисляются, используются их ранее сохраненные в памяти значения. При необходимости корректируем f_*^k из (17) и x_*^k — точку испытания, соответствующую значению f_*^k . Если одинаковое рекордное значение f_*^k наблюдается в точках нескольких испытаний, то в качестве x_*^k принимаем последнее. Вычисляем базовое значение Δf_k согласно [1] с учетом замечаний и описания, приведенного выше.

4. Если $n_k \geq n_{\max}$, то останавливаем поиск и выдаем значения Q_k^*, f_*^k, x_*^k в качестве оценки решения. При $n_k < n_{\max}$, в зависимости от превышения числом испытаний n_k порогового значения \tilde{n} и от кратности k значению \overline{K} выбираем, как описано выше, один из трех наборов значений $\varepsilon_s, \alpha_s, \delta_s, (s = 0, 1, 2)$ в качестве параметров $\varepsilon, \alpha, \delta$.

5. Выполняя двухэтапную обработку информации, структурно описанную в разделе 3, из множества \mathfrak{D}^k всех гиперинтервалов разбиения выделяем подмножество $\widehat{\mathfrak{D}}_k$ всех модифицированно потенциально оптимальных.

6. Полагаем $n_{k+1} = n_k$. Для каждого гиперинтервала \widehat{D}_t^k из $\widehat{\mathfrak{D}}_k$ согласно описанию из подраздела 2.2 определяем потенциальные точки новых испытаний u^t и v^t для его деления на три по большему ребру. Для каждой из этих двух точек выполняем поиск в базе проведенных испытаний. В точках u^t и v^t , не найденных в базе, проводим новые испытания, увеличивая при этом счетчик n_{k+1} . Каждый гиперинтервал \widehat{D}_t^k разделяем на три по большему ребру, \widehat{D}_t^k исключаем из множества \mathfrak{D}^k , добавляем в него три новых гиперинтервала. Полагаем $M_{k+1} = M_k + 2$. Принимаем измененное множество \mathfrak{D}^k в качестве \mathfrak{D}^{k+1} . Полагаем $k = k + 1$ и переходим к выполнению п. 3.

4.2. Анализ сходимости

Исследование сходимости метода ExDIR-diag проведем при более слабых предположениях о свойствах функций задачи (1)–(2), чем было принято при построении метода. Как и в [1], решение исходной задачи (1)–(2) понимаем в расширенном смысле. А именно если в (2) допустимое множество $X \neq \emptyset$, то предполагаем существование глобального минимума в (1)–(2), целью решения считаем определение элементов x^* из множества X^* глобальных минимумов задачи: $X^* \subseteq X$, $x^* \in X^*$, $Q(x^*) = Q^*$. В случае пустоты допустимого множества ($X = \emptyset$) неявно трактуем решение для (1)–(2) в расширенном смысле как обеспечение минимума невязки в ограничениях, предполагая, что минимум существует:

$$g^* = \min g(x), \quad x \in D = [a, b] \subset \mathbb{R}^N.$$

При этом целью решения задачи (1)–(2) при $X = \emptyset$ является определение элементов x^* из множества X^* , которое в данном случае понимается как множество глобальных минимумов $g(x)$ на D : $X^* \subseteq D$, $x^* \in X^*$, $g(x^*) = g^*$.

При анализе сходимости не будем предполагать липшицевость функций, но наложим следующие дополнительные требования.

А. Значения функций Q и g ограничены на D .

В. При непустоте допустимого множества X для всякого решения x^* существует открытое подмножество $\chi \subset X$, замыкание которого $\bar{\chi}$ содержит x^* , а функции Q и g непрерывны на этом замыкании. Если $X = \emptyset$, то для всякого расширенного решения x^* существует открытое подмножество $\chi \subset D$, замыкание которого $\bar{\chi}$ содержит x^* , а функция g непрерывна на этом замыкании.

С. Множество решений X^* , понимаемое в обычном или расширенном варианте, замкнуто и является глобально устойчивым (по аналогии с терминологией [27]) в том смысле, что для любой минимизирующей последовательности x_k , т.е. последовательности, удовлетворяющей при $X \neq \emptyset$ условиям $x_k \in X$, $Q(x_k) \rightarrow Q^*$ при $k \rightarrow \infty$, а при $X = \emptyset$ условиям $x_k \in D$, $g(x_k) \rightarrow g^*$ при $k \rightarrow \infty$, выполняется требование $\rho(x_k, X^*) \rightarrow 0$, где

$$\rho(x_k, X^*) = \inf \{\|x_k - x^*\| : x^* \in X^*\}.$$

Теорема 1 (о сходимости). Пусть для задачи с ограничениями неравенствами (1)–(2), решение которой понимается в расширенном смысле, выполняются требования А, В и С, тогда двухточечный диагональный метод ExDIR-diag, расширяющий принципы метода DIRECT на задачи с ограничениями, порождает на множестве поиска D последовательность испытаний со всюду плотным в пределе характером размещения, при этом все предельные точки последовательности оценок решения x_k^* принадлежат множеству решений X^* .

Заметим, что всюду плотный характер размещения точек испытаний не означает их равномерного расположения на множестве поиска D . Хотя для DIRECT-подобных методов, в отличие от методов из [13], до сих пор не получено аналитических оценок относительной плотности размещения испытаний, приведенные в разделе 5 вычислительные иллюстрации показывают существенно неравномерное распределение точек измерений в зависимости от поведения функций Q и g задачи (1)–(2). При достаточно большом коли-

честве измерений наиболее высокая концентрация испытаний наблюдается в окрестностях глобальных минимумов.

5. Вычислительные иллюстрации

В качестве вычислительных иллюстраций работы построенного двухточечного диагонального метода ExDIR-diag приведем примеры размещения испытаний на трех задачах с номерами № 1, № 3 и № 7 из тестового набора, использованного в [1]. Там же представлены постановки и описания этих задач. Здесь укажем лишь общие характеристики: число ограничений неравенств m , число условных локальных минимумов n_{loc} , параметры глобального минимума Q^* и x^* , Q^o — наименьшее значение в локальном минимуме, не совпадающем с глобальным.

В задаче № 1 $m = 3$ и функции ограничений сильно разномасштабные, $n_{loc} = 5$, $Q^* = -1,48968$, $x^* = (0,94248; 0,94526)$, $Q^o = -1,0$; в задаче № 3 $m = 1$, $n_{loc} = 8$, $Q^* = -0,81911$, $x^* = (1,30499; 2,27249)$, $Q^o = -0,81814$, эта задача характеризуется малым отличием в значениях глобального и следующего по значению локального минимума ($Q^o - Q^* = 0,00097$).

Численные эксперименты проведены с использованием двух групп значений параметров, которые для удобства обозначены $E1$ и $E3$. Набор параметров $E1$ включает: $\mu = 0,3$, $\tilde{n} = 100$, $\overline{K} = 2$; $\varepsilon_0 = 0,5$, $\alpha_0 = 1,0$, $\delta_0 = 0,0$; $\varepsilon_1 = 0,0001$, $\alpha_1 = 0,5$, $\delta_1 = 0,1$; $\varepsilon_2 = 0,1$, $\alpha_2 = 2,0$, $\delta_2 = 0,1$. Набор $E3$ отличается изменением части параметров: $\varepsilon_1 = 0,00001$, $\alpha_1 = 0,5$, $\delta_1 = 0,01$; $\varepsilon_2 = 0,1$, $\alpha_2 = 2,0$, $\delta_2 = 0,01$.

Представленные на рис. 4 и рис. 5 результаты тестирования соответствуют следующей постановке эксперимента.

Для задачи вводим желаемую погрешность ϵ определения глобально минимального значения Q^* . Вычисления проводим до первого испытания, в результате которого текущая оценка Q_k^* значения Q^* окажется меньшей или равной $Q^* + \epsilon$. В этот момент определяем n^* — количество проведенных испытаний, k^* — номер итерации и M^* — количество гиперинтервалов. Значения n^* , k^* и M^* указываем наряду с видами размещения испытаний. Допустимое множество на всех рисунках выделено серым. Результаты рис. 4 показывают размещения испытаний в одноточечном методе ExDIR из [1] в сравнении с его диагональным обобщением — ExDIR-diag для задачи № 1. Использована группа параметров $E1$. Размещение выглядит более рациональным у диагонального варианта метода (на рисунке справа). Для достижения точности $\epsilon = 0,002$ по значению функции ему понадобилось провести $n^* = 111$ испытаний (выполнено $k^* = 15$ итераций, $M^* = 157$). У метода ExDIR испытаний $n^* = 219$ за $k^* = 43$ итерации, $M^* = n^*$. Число делимых в среднем на итерации гиперинтервалов для ExDIR-diag оказалось примерно в два раза больше, чем у ExDIR.

В задаче № 3 малое отличие значений глобального и следующего по глубине локального минимума увеличивает число испытаний, необходимых для определения глобального минимума с заданной точностью $\epsilon = 0,00051$. Методу ExDIR потребовалось $n^* = 577$ испытаний (выполнена $k^* = 61$ итерация, $M^* = n^*$), а диагональному методу ExDIR-diag потребовалось $n^* = 369$ испы-

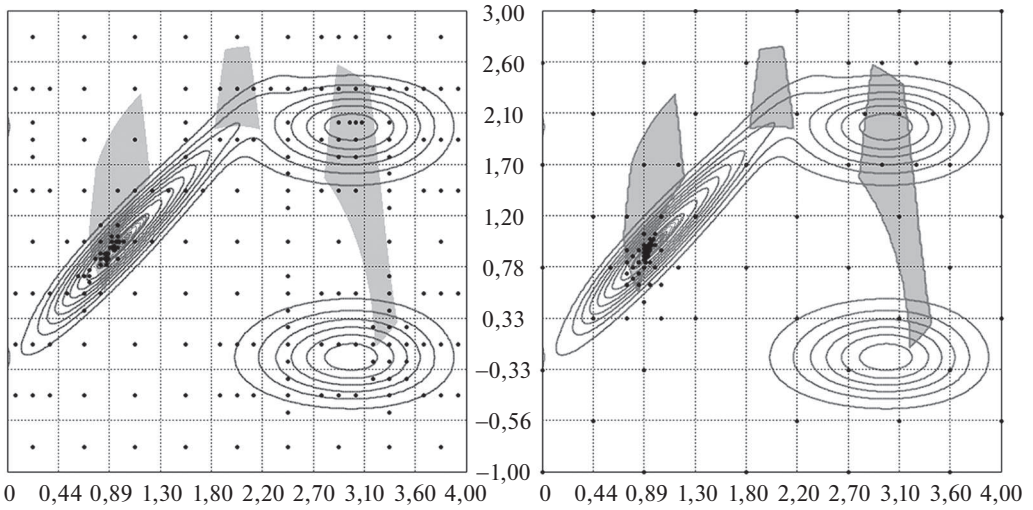


Рис. 4. Размещение испытаний методами ExDIR (слева) и ExDIR-diag (справа) в задаче № 1 до момента определения решения с погрешностью $\epsilon = 0,002$.

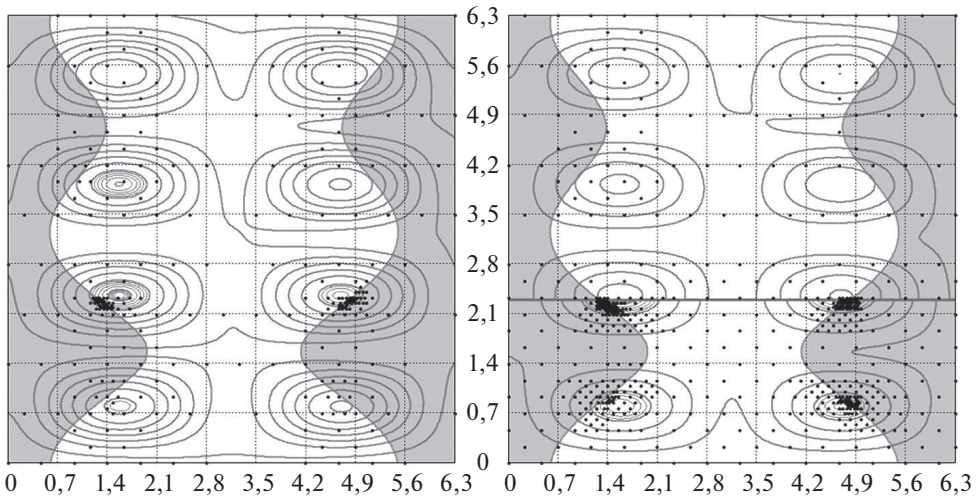


Рис. 5. Размещение испытаний в ExDIR-diag до момента получения решения с погрешностью $\epsilon = 0,00051$ в задаче № 3 без разрыва (слева) и задаче № 7 с разрывом.

таний (выполнено $k^* = 30$ итераций, $M^* = 547$). В обоих методах использовалась группа параметров E3. Размещение испытаний в диагональном методе ExDIR-diag показано на рис. 5 слева.

Дополнительно на рис. 5 справа представлена задача № 7 из [1] для иллюстрации влияния разрывов на поведение метода. Она отличается от задачи № 3 только добавленным разрывом в целевую функцию $Q(x)$ вдоль прямой $x_2 = 2,3$ за счет вычитания из $Q(x)$ значения $\Delta = 1$ на множестве точек с $x_2 \leq 2,3$. Линия разрыва проходит в непосредственной близости от точки глобального минимума с $x_2^* = 2,27249$. Общие характеристики зада-

чи № 7 отличаются от задачи № 3 лишь тем, что для нее $Q^* = -1,81911$ и $Q^o = -1,81814$.

Поскольку метод ExDIR-diag не оценивает константы Липшица, то несмотря на разрывы в областях концентрации точек не происходит вырождения этого метода в равномерный перебор. Видно, что метод ExDIR-diag на задаче с разрывом сохраняет целесообразное поведение. В процессе решения правильное распознавание области глобального минимума происходит достаточно быстро, однако наличие разрывов заметно замедляет уточнение найденного решения. В результате для достижения той же точности $\epsilon = 0,00051$ методу потребовалось $n^* = 707$ испытаний (выполнено $k^* = 53$ итерации, $M^* = 1137$). Вычисления проведены с группой параметров $E1$.

Следует отметить, что представленную в статье версию двухточечного диагонального метода следует рассматривать как базовую, которую можно использовать для его дальнейшего развития.

6. Заключение

Оригинальный метод DIRECT, построенный в [16] для решения задач многоэкстремальной оптимизации без функциональных ограничений, привлекателен тем, что сохраняет способность поиска глобального экстремума при наличии конечных (умеренной величины) разрывов целевой функции в непосредственной близости от решений. Эта особенность связана со специальными принципами отбора делимых гиперинтервалов, входящих в текущее покрытие множества поиска. Эти принципы не используют оценок констант Липшица, хотя липшицевость целевой функции (при неограниченности диапазона возможных значений константы) предполагается.

В данной статье впервые предложено обобщение принципов DIRECT с учетом сразу двух дополнительных факторов: наличия функциональных ограничений и использования испытаний функций в гиперинтервале не в одной центральной точке, а в двух на концах специально ориентируемых главных диагоналей. Для одновременного учета пар измерений потребовалось изменить традиционные для DIRECT предположения о свойствах функций решаемой задачи. Введенные обобщенные принципы отбора можно рассматривать как прямое развитие и расширение принципов построения метода DIRECT. В статье показана лишь структура двухэтапной обработки информации при отборе делимых гиперинтервалов. Детальное описание и обоснование алгоритмов реализации обобщенных принципов требует отдельной публикации.

Проведенный анализ показывает, что сходимость оценок решения построенного метода (как и в методе DIRECT) достигается за счет всюду плотного в пределе размещения точек испытаний, но выполненная численная апробация демонстрирует существенную неравномерность их размещения с преимущественной концентрацией в окрестностях решений, что определяет эффективность метода. Метод рассчитан на использование в прикладных задачах, где значения функций определяются в результате затратных по времени вычислений, а сами функции могут иметь разрывы. Функции качества такого вида могут возникать, например, в нелинейных задачах управления при оптимальной настройке свободных параметров по численно рассчитываемым для замкнутой системы дополнительным нелинейным критериям.

Построенная реализация метода является последовательной, но вычисления функций на наборе гиперинтервалов, выделенных для деления на очередной итерации, можно выполнять параллельно.

ПРИЛОЖЕНИЕ

Доказательство утверждения 1. Если условия (24) выполнены при некотором ΔL из указанного промежутка, то (21)–(22), очевидно, будут справедливы при том же ΔL и при $\Delta L^g = \alpha \Delta L$. Пусть теперь выполнены (21)–(22) и для D_i нашлась соответствующая пара $\widetilde{\Delta L}, \widetilde{\Delta L}^g$. Но поскольку $g^-(\Delta L^g, D_i)$ монотонно убывает с возрастанием ΔL^g , то (21)–(22) тем более будут выполнены при значениях $\widetilde{\Delta L}, \Delta L^g = \alpha \widetilde{\Delta L}$, что и требовалось. Утверждение доказано.

Доказательство леммы 1. На первом этапе обработки поисковой информации на итерации k согласно описанию в разделе 3 в каждом d -слое выделяется хотя бы один модифицированно недоминируемый гиперинтервал \widehat{D}_s^d этого слоя. Последний из выделенных имеет связанный диапазон $[\Delta L_s^d, \Delta L_{s+1}^d]$ с наибольшим номером $s = S_k(d)$, где $\Delta L_{S_k(d)+1}^d = +\infty$. Это верно, в частности, и для d -слоя гиперинтервалов, имеющих на текущей k -й итерации наибольший диаметр $d = d_k^{\max}$.

Далее, при построении разбиения (25) оси приращений ΔL в последнем промежутке $[\Delta L_j, \Delta L_{j+1}]$ с номером $j = m_k$ значение $\Delta L_{m_k+1} = +\infty$. Этому промежутку соответствует множество гиперинтервалов $\mathfrak{D}_k^{m_k}$, которое по построению обязательно включает в себя хотя бы один гиперинтервал \widehat{D}_s^d d -слоя с $d = d_k^{\max}$ (т.е. максимального диаметра), соответствующий диапазону $[\Delta L_s^d, \Delta L_{s+1}^d]$ значений ΔL с номером $s = S_k(d_k^{\max})$. На плоскости сравнения $(d/2, F)$ при сопоставлении гиперинтервалов из $\mathfrak{D}_k^{m_k}$ этому гиперинтервалу будет соответствовать самая правая точка в полученном множестве точек $P_k^{m_k}$. Поскольку сравнение происходит для промежутка $[\Delta L_{m_k}, +\infty)$, точка с наибольшим $d = d_k^{\max}$ при достаточно большом ΔL обязательно будет доминировать остальные и для нее также выполнится условие (17). Следовательно, гиперинтервал наибольшего диаметра будет включен в множество $\widehat{\mathfrak{D}}_k$ гиперинтервалов, делимых на итерации k . Лемма доказана.

Доказательство теоремы 1. В силу предположения о глобальной устойчивости множества решений X^* (требование С), а также требований В к структуре множества глобальных минимумов X^* и функциям задачи, для доказательства теоремы достаточно обосновать всюду плотное в пределе размещение точек испытаний. Действительно, множество поиска D компактно. При выполнении требований А и В всюду плотное в пределе размещение испытаний приведет к появлению подпоследовательностей точек измерений, сходящихся к каждому из решений x^* . Каждое x^* принадлежит замыканию некоторого открытого подмножества χ из X , если $X \neq \emptyset$, или из D в противном случае. При всюду плотном в пределе размещении испытаний сколь угодно близко от x^* найдутся точки, являющиеся центрами открытых шаров, включенных в χ . В каждом из таких шаров в какой-то момент метод разместит точку испытания. Поэтому найдутся подпоследовательности допустимых точек из подмножеств χ , сходящиеся к x^* при неограниченном возрастании

нии числа итераций. Функции Q , g или (при $X = \emptyset$) функция g непрерывны на замыканиях \bar{X} . Таким образом, последовательность рекордных измерений x_k^* будет минимизирующей и в силу требования С ее предельные точки будут являться решениями задачи.

Остается обосновать всюду плотное в пределе размещение испытаний. Приведем лишь схему рассуждения. Нужно поведение следует из леммы 1, которая устанавливает, что в построенном методе на каждой итерации обязательно делится по крайней мере один из гиперинтервалов наибольшего диаметра. Поскольку деление гиперинтервалов происходит на три равные части по наибольшему ребру, это обеспечивает строгое убывание диаметра с некоторым коэффициентом, отделенным от единицы. Этих двух факторов достаточно. Более подробное рассуждение можно построить аналогично доказательству теоремы 5.6 из [8] с учетом леммы 1. Теорема доказана.

СПИСОК ЛИТЕРАТУРЫ

1. *Городецкий С.Ю.* Несколько подходов к обобщению метода DIRECT на задачи с функциональными ограничениями // Математическое моделирование. Оптимальное управление. 2013. № 6 (1). С. 189–215. Нижний Новгород: Вестн. Нижегород. ун-та им. Н.И. Лобачевского.
2. Методы классической и современной теории автоматического управления: Учебник в 3-х т. Т. 2. Синтез регуляторов и теория оптимизации систем автоматического управления / Под ред. Н.Е. Егупова. М.: Изд-во МГТУ им. Н.Э. Баумана, 2000.
3. *Баландин Д.В., Коган М.М.* Синтез законов управления на основе линейных матричных неравенств. М.: Физматлит, 2007.
4. *Александров А.Г.* Методы построения систем автоматического управления. М.: Физматкнига, 2008.
5. *Gershon E., Shaked U., Yaesh I.* H_∞ Control and Estimation of State-multiplicative Linear Systems. London: Springer, 2005.
6. *Баландин Д.В., Коган М.М.* Оптимальное по Парето обобщенное H_2 -оптимальное управление и задачи виброзащиты // АИТ. 2017. № 8. С. 76–90.
Balandin D.V., Kogan M.M. Pareto Optimal Generalized H_2 -control and Vibroprotection Problems // Autom. Remote Control. 2017. V. 78. No. 8. P. 1417–1429.
7. *Городецкий С.Ю., Сорокин А.С.* Построение оптимальных регуляторов по нелинейным критериям качества на примере одной динамической системы // Математическое моделирование. Оптимальное управление. 2012. № 2 (1). С. 165–176. Нижний Новгород: Вестн. Нижегород. ун-та им. Н.И. Лобачевского.
8. *Сергеев Я.Д., Квасов Д.Е.* Диагональные методы глобальной оптимизации. М.: Физматлит, 2008.
9. *Strongin R.G., Sergeyev Ya.D.* Global optimization with non-convex constraints: Sequential and parallel algorithms. Dordrecht: Kluwer Acad. Publishers, 2000.
10. *Евтушенко Ю.Г., Малкова В.У., Станевичюс А.А.* Параллельный поиск глобального экстремума функций многих переменных // Журн. вычисл. матем. и матем. физ. 2009. Т. 49. № 2. С. 255–269.
11. *Стронгин Р.Г., Гергель В.П., Гришагин В.А., Баркалов К.А.* Параллельные вычисления в задачах глобальной оптимизации / Предисл.: В.А. Садовничий. М.: Изд-во МГУ, 2013.

12. *Городецкий С.Ю.* Триангуляционные методы параболоидов в задачах многоэкстремальной оптимизации с ограничениями для класса функций с липшицевыми производными по направлениям // Математическое моделирование. Оптимальное управление. 2012. № 1 (1). С. 144–155. Нижний Новгород: Вестн. Нижегород. ун-та им. Н.И. Лобачевского.
13. *Городецкий С.Ю.* Исследование процедур глобальной оптимизации с адаптивными стохастическими моделями. Дисс. на соискание уч. степ. канд. физ.-мат. наук. Горький: ГГУ, 1984.
14. *Неймарк Ю.И.* Динамические системы и управляемые процессы. М.: Наука, 1978.
15. *Городецкий С.Ю., Неймарк Ю.И.* О поисковых характеристиках алгоритма глобальной оптимизации с адаптивной стохастической моделью // Пробл. случайного поиска. Рига: Зинатне, 1981. С. 83–105.
16. *Jones D.R., Perttunen C.D., Stuckman B.E.* Lipschitzian Optimization without the Lipschitz Constant // J. Optim. Theory Appl. 1993. V. 79. No 1. P. 157–181.
17. *Препарата Ф.Ф., Шеймос М.И.* Вычислительная геометрия. Введение. М.: Мир, 1989.
18. *Jones D.R.* The DIRECT global optimization algorithm / Encyclopedia of optimization. 7 Vols. 2nd revised and expanded ed., ed. by C.A. Floudas, P.M. Pardalos. Springer, 2009. P. 725–735.
19. *Евтушенко Ю.Г., Ратъкин В.А.* Метод половинных делений для глобальной оптимизации функций многих переменных // Изв. АН СССР. Технич. кибернетика. 1987. № 1. С. 119–127.
20. *Gablonsky J.M., Kelley C.T.* A Locally-Biased form of the DIRECT Algorithm // J. Global Optim. 2001. V. 21. No. 1. P. 27–37.
21. *Sergeyev Ya.D., Kvasov D.E.* Global Search Based on Efficient Diagonal Partitions and a Set of Lipschitz Constants // SIAM J. Optim. 2006. V. 16. No. 3. P. 910–937.
22. *Sergeyev Ya.D.* An Efficient Strategy for Adaptive Partition of N-dimensional Intervals in the Framework of Diagonal Algorithms // J. Optim. Theory Appl. 2000. V. 107. No. 1. P. 145–168.
23. *Sergeyev Ya.D., Kvasov D.E.* A Univariate Global Search Working with a Set of Lipschitz Constants for the First Derivative // Optimization Lettes. 2009. No. 3. P. 303–318.
24. *Kvasov D.E., Sergeyev Ya.D.* Lipschitz Gradients for Global Optimization in a One-Point-Based Partitioning Scheme // J. Comput. Appl. Math. 2012. V. 236. P. 4042–4054.
25. *Sergeyev Ya.D., Mukhametzhonov M.S., Kvasov D.E.* On the Efficiency of Nature-Inspired Metaheuristics in Expensive Global Optimization with Limited Budget // Sci. Reports. 8, 453 (2018).
26. *Городецкий С.Ю.* О новой модели поведения целевой функции для диагональной реализации DIRECT-подобных методов // Научное периодическое издание CETERIS PARIBUS. М.: РИЦ ЭФИР, 2016. № 1. С. 4–16.
27. *Карманов В.Г.* Математическое программирование: Уч. пос. М.: Физматлит, 2008.

Статья представлена к публикации членом редколлегии Б.Т. Поляком.

Поступила в редакцию 23.07.2019

После доработки 06.10.2019

Принята к публикации 30.01.2020