Модифицированный метод половинных делений

Постановка задачи и основные определения

Без ограничения общности, рассмотрим задачу безусловной глобальной оптимизации непрерывной функции , заданной на допустимом множестве  в следующей постановке:

. (1)

Положим, что глобальный минимум  принадлежит множеству , причем , а , является многомерным единичным гиперкубом. Далее используются условные обозначения и основные определения, заимствованные из работы [40].

Определим множество -решений задачи (1) следующим образом:

. (2)

Нахождение приближенного решения задачи (1) заключается в поиске хотя бы одной точки множества .

Впервые метод неравномерных покрытий был предложен Ю.Г. Евтушенко в 1971 году [6]. Кратко опишем его основные положения, поскольку предлагаемый в данной работе модифицированный метод половинных делений во многом основывается на идеях этого метода.

В методе неравномерных покрытий строится последовательность  подмножеств множества  и точек  (). В каждой точке  вычисляется значение функции  и определяется её рекордное значение :

. (3)

На каждом из подмножеств  вводятся *миноранты*, т.е. такие функции , что . Тогда последовательности подмножеств  можно поставить в соответствие совокупность покрывающих подмножеств , определяемых из условий

. (4)

Множество  строится таким образом, что  выполняется неравенство , из чего следует, что глобальный минимум функции  на множестве  не может быть меньше рекорда  более чем на . Таким образом, в процессе поиска глобального минимума подмножество  можно исключить из рассмотрения и продолжить оптимизацию на множестве . Алгоритм оптимизации останавливается, когда допустимое множество оказывается покрытым подмножествами , т.е.

. (5)

Алгоритм неравномерных покрытий, предложенный в работах [6, 40], гарантирует нахождение -оптимального решения, что подтверждается основополагающей теоремой [40].

**Теорема 1.** *Пусть для последовательности  выполняется условие (5). Тогда определяемая из (3) рекордная точка  принадлежит множеству* .

Введем понятие *области притяжения* локального минимума, под которой будем понимать множество

 где (6)

. (7)

Здесь  - значение функции в точке локального минимума, *m* – число локальных минимумов функции в допустимой области.

Очевидно, что допустимая область является объединением областей притяжения локальных минимумов  ( ). На самом деле, правильнее было бы исключить линии раздела областей притяжения, но на практике, при использовании численных методов локальной оптимизации, локальный метод обязательно «скатится» с этой линии к одному из локальных минимумов. Далее будем использовать следующее упрощенное определение области притяжения локального минимума

 (8)

где ,  - радиус области притяжения локального минимума.

Метод половинных делений

Большинство точных методов решения задачи глобальной оптимизации используют идею неравномерного покрытия допустимого множества [40, 41, 45]. Идеяметода половинных делений заключается в организации непропорционального деления допустимого множества, например, единичного гиперкуба на параллелепипеды, меньшего размера. Чаще всего реализуется дихотомическое деление параллелепипедов по наибольшему ребру [40, 41, 42], возможен вариант непропорционального деления параллелепипеда по его главной диагонали [45].

Пусть допустимое множество  делится на прямоугольные параллелепипеды  с центрами  и главными диагоналями , где  - радиус параллелепипеда  (). Предположим, что функция  принадлежит классу  и для нее выполняется условие Липшица

. (9)

Тогда в качестве миноранты [4] можно рассматривать функцию

. (10)

Миноранта (10) является вогнутой функцией и ее минимум вычисляется по формуле , а, следовательно, параллелепипед  можно «отбраковать», если .

Алгоритм работает следующим образом. В процессе расчетов поддерживается список параллелепипедов {*Xi*}, первоначально состоящий из одного параллелепипеда *X*. Далее выбирается некоторый элемент списка — *Xi* и определяется множество *Zi* по формуле (4). Если , то *Xi* удаляется из списка и становится элементом последовательности покрывающих множеств {*Zi*}. В противном случае он подвергается дроблению на более мелкие параллелепипеды, которые замещают в списке параллелепипед *Хi*. В их центрах вычисляется значение функции и уточняется рекорд. Алгоритм останавливается, если не остается параллелепипедов, которые можно подвергнуть дальнейшему делению.

Выбор критического параллелепипеда

Сохранив схему двоичного деления параллелепипедов [41], изменим правило выбора «критического» параллелепипеда, подвергающегося дальнейшему двоичному делению. Произвольный выбор «критического» параллелепипеда [41], или ориентация на параллелепипед, имеющий минимальное текущее значение функции в точке  [42], может привести к чрезмерно детальному «изучению» окрестностей локальных минимумов функции и «отсрочить» нахождение её нового рекорда , т.е. увеличить трудоемкость алгоритма.

В работе [43] для одномерных функций предложена одна из самых эффективных стратегий многоэкстремальной оптимизации, основанная на использовании приближенного апостериорного распределения вероятностей расположения глобального экстремума, формируемого в процессе испытаний функции, что реализует более сбалансированную организацию поиска глобального минимума функции. Данная стратегия настолько эффективна, что её часто переносят с одномерного случая на случай оптимизации функций многих переменных. Например, в работе [44] многомерная задача оптимизации, с помощью кривых Пеано, редуцируется к одномерной задаче глобальной оптимизации, в которой используется характеристическая схема оптимизации Стронгина Р. Г.. Аналогичные идеи используются в диагональном методе глобальной оптимизации Я.Д. Сергеева [45]. Попытаемся использовать эту стратегию и для метода половинных делений.

Если в процессе двоичных делений соединять центры параллелепипедов предшественников и потомков, то получится семейство непрерывных непересекающихся линий – траекторий перемещения центров параллелепипедов (рисунок 8), для которых можно вычислять характеристики соответствующих участков кривых по аналогии с одномерным случаем [43]. Оценки характеристик, полученных для участков траекторий перемещения центров параллелепипедов, припишем соответствующим параллелепипедам  с центрами в точках .

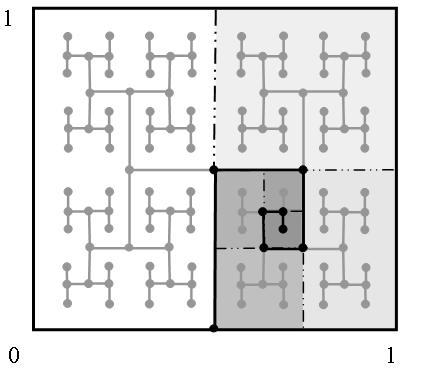


Рисунок 8 – Траектория перемещения центров параллелепипедов

Таким образом, каждый параллелепипед  свяжем с набором параметров . Здесь  - центр параллелепипеда,  - значение функции в точке ,  - радиус параллелепипеда, - расстояние между текущим параллелепипедом и его предшественником. – характеристика i-го параллелепипеда, которую в соответствии с [43] можно вычислить по формуле:

,

где – значение функции, вычисленное в предшествующем параллелепипеде в точке ; *w* – оценка константы Липшица. Последняя, адаптивно вычисляется в процессе вычислений функции:

 ,

где – коэффициент.

Из наборов характеристик  сформируем упорядоченный в направлении убывания характеристики  список , где .

Теперь в методе половинных делений в качестве «критического» параллелепипеда можно выбрать первый параллелепипед из списка , т.е. параллелепипед, для которого выполняется условие: .

Следует отметить, что в списке  на любом этапе работы модифицированного алгоритма половинных делений (МАПД) содержатся параллелепипеды, ещё не подвергнутые двоичному делению. Естественно, что «критические» параллелепипеды каждые раз удаляются из списка. Работа МАПД заканчивается, когда список  пуст.

Очевидно, что МАПД находит  – приближенное решение задачи (1), поскольку для него выполняются условия теоремы 1, так как модификация метода половинных делений касается только стратегии выбора «критического» параллелепипеда.

Двухфазный алгоритм метода половинных делений (ДАМПД)

Как любой алгоритм ГО, алгоритм половинных делений имеет экспоненциальный характер роста сложности в зависимости от числа переменных. Одной из эффективных стратегий совершенствования алгоритмов глобальной оптимизации является использование *локальной техники* [46], когда стратегия глобального поиска удачно сочетается с методами локальной оптимизации.

В этом случае этап глобальной оптимизации можно производить достаточно грубо, поскольку основной задачей этого этапа является определение областей притяжения локальных экстремумов функции, внутри которых можно использовать алгоритмы локальной оптимизации. При ограниченном количестве минимумов функции  области притяжения  имеют значительные размеры. Предположим, что относительно оптимизируемой функции известно количество локальных минимумов –  (включая глобальный) и размеры областей притяжения: . Введем понятие гарантированного радиуса области притяжения минимумов функции, под которым будем понимать .

Дополнительная информация о размерах областей притяжения локальных минимумов функции позволяет построить двухфазную схему алгоритма глобальной оптимизации, в которой выделим фазы глобальной и локальной оптимизации (рисунок 9).

В ДАМПД фаза глобальной оптимизации реализуется с помощью модифицированного алгоритма, с той лишь разницей, что двоичное деление параллелепипедов в МАПД происходит до достижения заданных размеров их радиусов . Величина  определяется размерами гарантированного радиуса  области притяжения глобального минимума. На рисунке 9 пунктирными линиями отмечены гарантированные области притяжения минимумов функции. Заштрихованные прямоугольники являются прямоугольниками, «отбракованными» с использованием константы Липшица. В итоге, исходя из условия , фаза глобальной оптимизации завершилась, породив, для приведенного на рисунке 2 примера, 24 прямоугольника заданного радиуса.

В фазе локальной оптимизации из точек, принадлежащих областям притяжения локальных минимумов, организуется поиск минимумов функции с помощью локальных алгоритмов оптимизации. В нашем случае использовался метод деформированных многогранников.

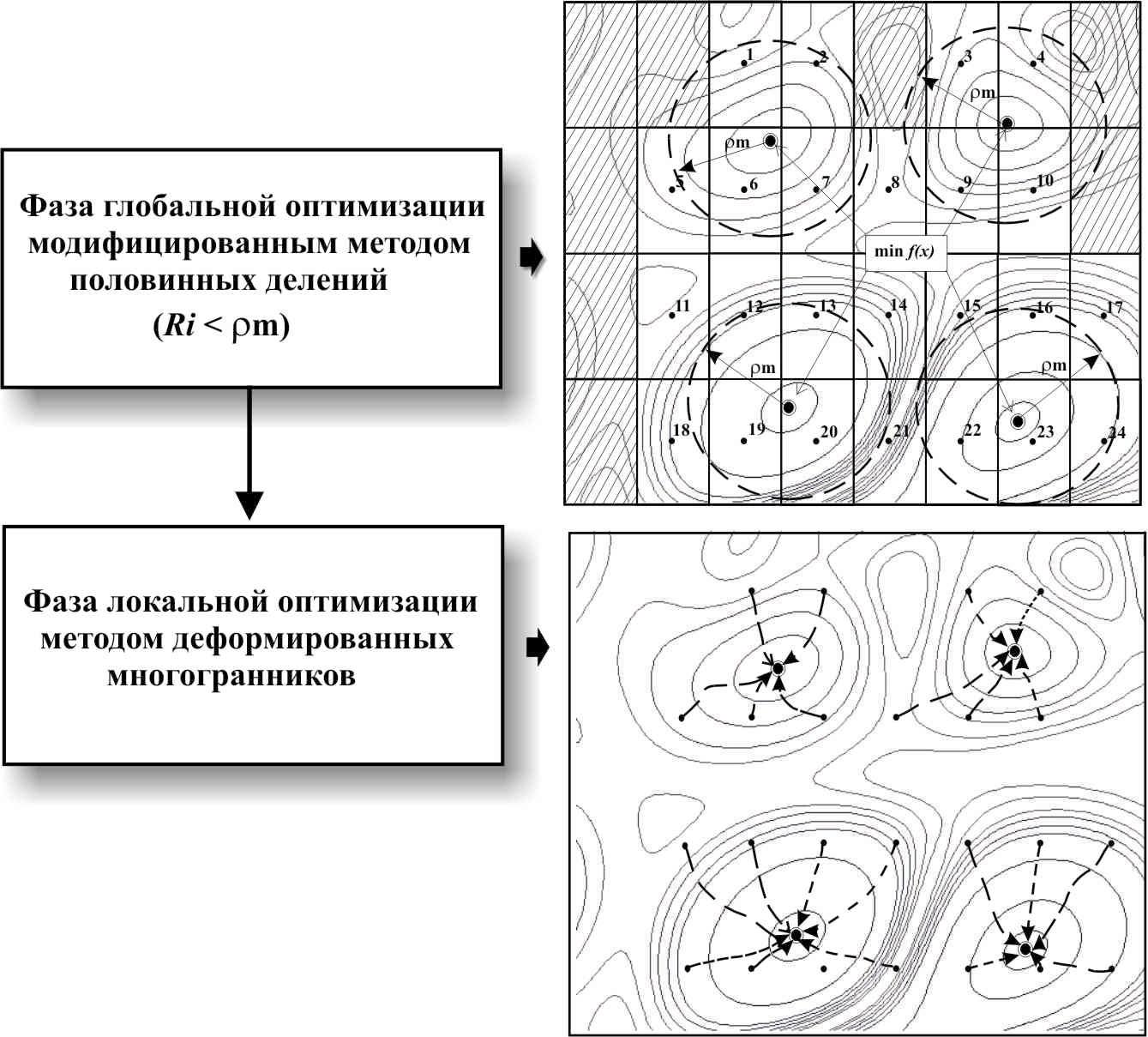


Рисунок 9 – Двухфазный алгоритм метода половинных делений

Такая схема организации процедуры поиска глобального минимума функции существенно сокращает трудоемкость фазы глобальной оптимизации.

Алгоритм построения множества точек начальных приближений для алгоритма локальной оптимизации.

Для больших размерностей вектора независимых переменных оптимизируемой функции в фазе глобальной оптимизации порождается значительное количество параллелепипедов заданного размера. В этом случае объем вычислений в фазе локальной оптимизации становится чрезмерно большим. Рассмотрим следующий адаптивный алгоритм формирования списка точек начальных приближений областей притяжения локальных минимумов оптимизируемой функции, осуществляющий сжатие количества точек начальных приближений, используемых во второй фазе алгоритма.

Будем считать, что область притяжения  определена, если она содержит, по крайней мере, одну точку из множества .

Пусть *r* – количество зон притяжения минимума функции;  – гарантированный радиус области притяжения минимума функции, обеспечивающий нахождение всех локальных минимумов.

Сформируем список областей притяжения локальных минимумов , элементами которого являются структуры , где  - координаты «представителя» *i*-й области притяжения, имеющей наилучшую достигнутую для этой области оценку . Вектор  условно считается центром *i*-ой области притяжения.

Первоначально список *V* – пуст. По мере вычислений функции он наполняется элементами, но в конце этапа глобального поиска не может содержать больше *m* элементов (*m* – заданный размер списка, ). Размер списка *m* зависит от свойств оптимизируемой функции и выбирается из соображений попадания в него зоны притяжения глобального минимума функции. Элементы списка *V* упорядочены таким образом, что .

Пусть в процессе работы алгоритма глобальной оптимизации произведено очередное испытание . Эволюция содержимого списка *V* происходит по следующим простым правилам:

1. Проверяется принадлежность центра очередного параллелепипеда  окрестностям одной из имеющихся областей притяжения  (*l* – текущий размер списка).
   1. Если выполняется условие

. (12)

1.1.1. То при  содержимое элемента  обновляется:

, , где  - центры параллелепипедов «попавших» в область . Далее выполняется действие 1.1.3.

1.1.2. Иначе уточняется только значение . Выполняется действие 1.1.3.

1.1.3. Список *V* упорядочивается в порядке возрастания .

1.2. Если условие (12) не выполняется, то проверяем правило 2.

1. Определяется возможность включения нового элемента в список *V*.

2.1. Если , то элемент  записывается в голову списка *V*.

2.2. Если , то элемент  записывается между  и  элементами списка.

2.3. Если , то элемент  записывается в конец списка *V*.

1. При превышении предельного числа элементов списка, из списка исключается последний элемент списка.

Предложенный алгоритм является эвристическим, однако, вычислительные эксперименты с тестовыми функциями показали его высокую эффективность. Более того, как правило, в первой сотне элементов списка содержится начальное приближение глобального минимума функции. На рисунке 10 представлены элементы списка областей притяжения. Нумерация соответствует весу элемента. Из рисунка видно, что каждый элемент списка сформировался, агрегируя большое количество точек вычисления функции. Например, элемент  агрегировал 9 точек,  - 8, и т.д.

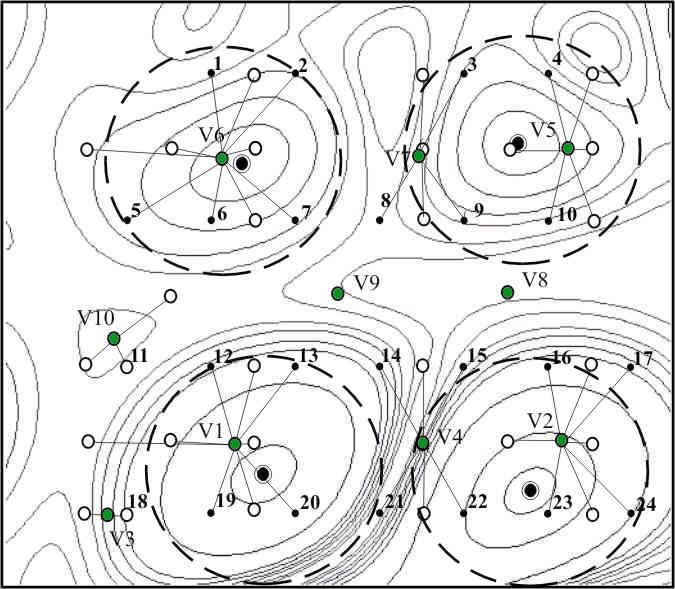


Рисунок 10 – Формирование списка областей притяжения

Параллельная версия двухфазного алгоритма метода половинных делений.

Очевидно, что **з**атратывремени на решение задачи глобальной оптимизации можно, в известной степени, сократить за счет распараллеливания вычислительного процесса. Однако архитектура современных кластеров с распределенной памятью значительно затрудняет распараллеливание алгоритмов глобальной оптимизации, поскольку невольно приходится заниматься организацией эффективной передачи данных между процессорами, а не непосредственным распараллеливанием вычислений. При этом обычно используется «тяжеловесный», но естественный для подобных систем стандарт MPI. Как показала практика, наилучшие результаты по критериям оценки качества распараллеливания вычислений (ускорению и эффективности) получаются, когда удается свести к минимуму обмен информацией между процессорами.

Для двухфазного алгоритма метода половинных делений на каждой фазе работы алгоритма целесообразно назначать разное количество процессоров. Причем для первой фазы глобальной оптимизации за каждым процессором можно закрепить один параллелепипед из списка , построенного МАПД. Очевидно, что параллельные вычисления будут эффективны, если параллелепипеды будут иметь одинаковый радиус. В этом случае, если не учитывать «отбраковку» параллелепипедов по критерию Липшица, на каждом процессоре будет произведено приблизительно одинаковое количество вычислений до достижения заданного уровня разбиения параллелепипедов на части. Этого несложно добиться, если допустимую область разбивать на  равных частей.

Алгоритм глобальной оптимизации ММПД состоит из трёх этапов. В сущности, ещё есть предварительный этап настройки тестовой функции. Реализация алгоритма в рамках нотации ГСП представлена на рисунке 10.

В вершине 1 устанавливаются значения параметров теста GKLS и его инициализация [47]. В первый этап алгоритма входят вершины 2-5. В вершине 2 вычисляется первое значение функции в точке x[i]=0,5, i=0..N, N - размерность пространства и в характеристический список *D* добавляется первый гиперкуб, охватывающий все пространство. В этой же вершине с помощью последовательного варианта алгоритма двоичного деления, описанного выше, формируется начальный характеристический список *D* с числом элементов не менее числа процессоров. При начальном делении областей прореживание выключено и на каждом шаге мы точно можем подсчитать количество элементов и состав списка. Очевидно из самого алгоритма половинного деления, что через 2n шагов список будет состоять из областей равного размера. Это уравнивает начальные условия для последующего параллельного деления.

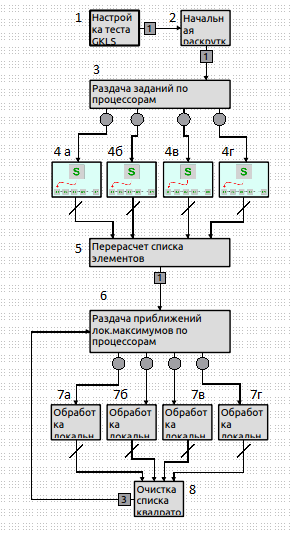


Рисунок 10 – Граф-модель алгоритма ГО ММПД

Первый этап алгоритма осуществляет параллельное половинное деление и глобальный поиск в вершинах 4а-4г. На рисунке 10 приведён вариант алгоритма для 4х процессоров, но алгоритм легко масштабируется на произвольное число процессоров. В вершине 3 происходит раздача заданий по процессорам. Каждый процесс получает один параллелепипед. Вершины 4а-4г имеют одинаковые агрегаты и запускаются каждая на своём процессоре. На каждом из процессоров с помощью МАПД реализуется покрытие частных параллелепипедов (в нашем примере ), причем их дробление производится до достижения определенных размеров. При этом формируются списки начальных приближений зон притяжения локальных минимумов. В вершине 5 эти списки объединяются.

На рисунке 11 представлен базовый модуль деления списка параллелепипедов. Вершины 1-5 реализуют цикл деления параллелепипедов, вершина 7 служит для развязки условия определения режима работы с прореживанием или без прореживания.

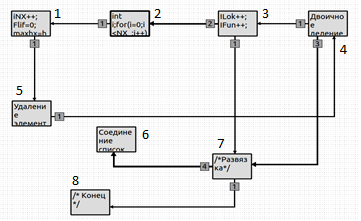


Рисунок 11 – Агрегат “Базовый модуль деления списка параллелепипедов”

Второй этап осуществляет поиск локального минимума из точек определённых во втором этапе. Этап состоит из вершин 6-8. В вершине 6 мастер ветвь раздает начальные точки для локального поиска. В вершинах 7а-7г запускается локальный поиск максимального значения функции с помощью метода деформированных многогранников. В вершине 9 происходит подсчёт количества найденных минимумов. Выходом из второго этапа служит опустошение списка начальных областей или выполнение необходимого минимума запусков локального поиска. Если имеются непроверенные области локальных минимумов, то перед завершением программы они удаляются.