

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования «Нижегородский государственный университет
им. Н.И. Лобачевского»

Институт информационных технологий, математики и механики

**Кафедра математического обеспечения и суперкомпьютерных
технологий**

ОТЧЁТ

по научно-исследовательской практике

Название работы

Выполнил: студент группы М0813-2
_____ Соврасов В.В.

Подпись

Научный руководитель:

доцент, к.ф.м.н.

_____ Баркалов К.А.

Подпись

Нижний Новгород
2016

Содержание

1. Введение	2
2. Алгоритм глобального поиска	2
2.1. Сравнение методов оптимизации	3
2.2. Класс тестовых задач GKLS	4
3. Многоуровневая схема редукции размерности с помощью разверток	4
4. Сравнение различных типов множественных разверток	6
5. Применение локального поиска для ускорения сходимости АГП	7
6. Смешанный алгоритм глобального поиска и его эффективная реализация	8
Список литературы	11

Введение

Задачи нелинейной глобальной оптимизации встречаются в различных прикладных областях и традиционно считаются одними из самых трудоёмких среди оптимизационных задач. Их сложность экспоненциально растёт в зависимости от размерности пространства поиска, поэтому для решения существенно многомерных задач требуются суперкомпьютерные вычисления.

В настоящее время на кафедре МОиСТ активно ведётся разработка программной системы для глобальной оптимизации функций многих вещественных переменных EхаMin. Эта система включает в себя последние теоретические разработки, сделанные на кафедре в этой сфере, в том числе и блочную многошаговую схему редукции размерности [1]. Отличительной чертой системы является то, что, она может работать как на CPU, так на разных типах ускорителей вычислений с высокой степенью параллельности (XeonPhi, GPU Nvidia) [2, 3].

В данной работе будут описаны некоторые улучшения, внесённые в систему, и предварительные исследования, проведённые перед их внедрением.

Алгоритм глобального поиска

Для дальнейшего изложения потребуется описание метода глобальной оптимизации, используемого в системе EхаMin. Многомерные задачи сводятся к одномерным с помощью различных схем редукции размерности, поэтому можно рассматривать минимизацию одномерной функции $f(x)$, $x \in [0, 1]$, удовлетворяющей условию Гёльдера.

Рассматриваемый алгоритм решения данной задачи предполагает построение последовательности точек x_k , в которых вычисляются значения минимизируемой функции $z_k = f(x_k)$. Процесс вычисления значения функции (включающий в себя построение образа $y_k = y(x_k)$) будем называть испытанием, а пару (x_k, z_k) — результатом испытания. Множество пар $\{(x_k, z_k)\}$, $1 \leq k \leq n$ составляет поисковую информацию, накопленную методом после проведения n шагов.

На первой итерации метода испытание проводится в произвольной внутренней точке x_1 интервала $[0; 1]$. Пусть выполнено $k \geq 1$ итераций метода, в процессе которых были проведены испытания в k точках x_i , $1 \leq i \leq k$. Тогда точка x^{k+1} поисковых испытаний следующей $(k + 1)$ -ой итерации определяются в соответствии с правилами:

Шаг 1. Перенумеровать точки множества $X_k = \{x^1, \dots, x^k\} \cup \{0\} \cup \{1\}$, которое включает в себя граничные точки интервала $[0, 1]$, а также точки предшествующих испытаний, нижними индексами в порядке увеличения значений координаты, т.е.

$$0 = x_0 < x_1 < \dots < x_{k+1} = 1$$

Шаг 2. Полагая $z_i = f(x_i)$, $1 \leq i \leq k$, вычислить величины

$$\mu = \max_{1 \leq i \leq k} \frac{|z_i - z_{i-1}|}{\Delta_i}, M = \begin{cases} r\mu, \mu > 0 \\ 1, \mu = 0 \end{cases} \quad (2.1)$$

где r является заданным параметром метода, а $\Delta_i = (x_i - x_{i-1})^{\frac{1}{N}}$.

Шаг 3. Для каждого интервала (x_{i-1}, x_i) , $1 \leq i \leq k+1$, вычислить характеристику в соответствии с формулами

$$R(1) = 2\Delta_1 - 4\frac{z_1}{M}, R(k+1) = 2\Delta_{k+1} - 4\frac{z_k}{M} \quad (2.2)$$

$$R(i) = \Delta_i + \frac{(z_i - z_{i-1})^2}{M^2\Delta_i} - 2\frac{z_i + z_{i-1}}{M}, 1 < i < k+1 \quad (2.3)$$

Шаг 4. Выбрать наибольшую характеристику:

$$t = \arg \max_{1 \leq i \leq k+1} R(i) \quad (2.4)$$

Шаг 5. Провести очередное испытание в точке x_{k+1} , вычисленной по формулам

$$x_{k+1} = \frac{x_t + x_{t-1}}{2}, t = 1, t = k+1$$

$$x_{k+1} = \frac{x_t + x_{t-1}}{2} - \text{sign}(z_t - z_{t-1}) \frac{1}{2r} \left[\frac{|z_t - z_{t-1}|}{\mu} \right]^N, 1 < t < k+1 \quad (2.5)$$

Алгоритм прекращает работу, если выполняется условие $\Delta_t \leq \varepsilon$, где $\varepsilon > 0$ есть заданная точность. В качестве оценки глобально-оптимального решения задачи выбираются значения

$$f_k^* = \min_{1 \leq i \leq k} f(x_i), x_k^* = \arg \min_{1 \leq i \leq k} f(x_i) \quad (2.6)$$

Подробнее метод и теорема о его сходимости описаны в [4].

Сравнение методов оптимизации

Существует несколько критериев оптимальности алгоритмов поиска (минимаксный, критерий одношаговой оптимальности), но большинстве случаев представляет интерес сравнение методов по среднему результату, достижимому на конкретном подклассе липшицевых функций. Достоинством такого подхода является то, что средний показатель можно оценить по конечной случайной выборке задач, используя методы математической статистики.

В качестве оценки эффективности алгоритма будем использовать, операционную характеристику, которая определяется множеством точек на плоскости (K, P) , где K – среднее число поисковых испытаний, предшествующих выполнению условия останова при минимизации функции из данного класса, а P – статистическая вероятность того, что к моменту останова глобальный экстремум будет найден с заданной точностью. Если при выбранном K операционная характеристика одного метода лежит выше характеристики другого, то это

значит, что при фиксированных затратах на поиск первый метод найдёт решение с большей статистической вероятностью. Если же зафиксировать некоторое значение P , и характеристика одного метода лежит левее характеристики другого, то первый метод требует меньше затрат на достижение той же надёжности.

Класс тестовых задач GKLS

Для сравнения алгоритмов глобального поиска в смысле операционной характеристики требуется иметь некоторое множество тестовых задач. Генератор задач GKLS, описанный в [5] позволяет получить такое множество задач с заранее известными свойствами. В данной работе используются два класса, сгенерированные GKLS: 4d Simple и 5d Simple, параметры которых также описаны в [5]. Функции рассматриваемых классов являются непрерывно дифференцируемыми и имеют 10 локальных минимумов, один из которых является глобальным.

Многоуровневая схема редукции размерности с помощью разверток

Одна из постановок задачи глобальной оптимизации звучит следующим образом: найти глобальный минимум N -мерной функции $\phi(y)$ в гиперинтервале $D = \{y \in R^N : a_i \leq x_i \leq b_i, 1 \leq i \leq N\}$. Для построения оценки глобального минимума по конечному количеству вычислений значения функции требуется, чтобы $\phi(y)$ удовлетворяла условию Липшица.

$$\phi(y^*) = \min\{\phi(y) : y \in D\} \quad (3.7)$$

$$|\phi(y_1) - \phi(y_2)| \leq L\|y_1 - y_2\|, y_1, y_2 \in D, 0 < L < \infty \quad (3.8)$$

Классической схемой редукции размерности для алгоритмов глобальной оптимизации является использование разверток — кривых, заполняющих пространство [4].

$$\{y \in R^N : -2^{-1} \leq y_i \leq 2^{-1}, 1 \leq i \leq N\} = \{y(x) : 0 \leq x \leq 1\}$$

Такое отображение позволяет свести задачу в многомерном пространстве к решению одномерной ценой ухудшения её свойств. В частности, одномерная функция $\phi(y(x))$ является не Липшицевой, а Гёльдеровой:

$$|\phi(y(x_1)) - \phi(y(x_2))| \leq H|x_1 - x_2|^{\frac{1}{N}}, x_1, x_2 \in [0, 1]$$

где константа Гельдера H связана с константой Липшица L соотношением

$$H = 4Ld\sqrt{N}, d = \max\{b_i - a_i : 1 \leq i \leq N\}$$

Теоретически с помощью этой схемы можно решить задачу любой размерности, однако на ЭВМ развертка строится с помощью конечноразрядной арифметики, из-за чего начиная

с некоторого N^* построение разветки невозможно (значение N^* зависит от максимального количества значащих разрядов в арифметике с плавающей точкой). Понять почему это происходит нетрудно, обратившись, например к [4].

Чтобы преодолеть эту проблему профессором В. П. Гергелем была предложена следующая идея: использовать композицию разверток меньшей размерности для построения отображения $z(x) : [0; 1] \rightarrow D \in R^N$. Поясним эту схему на примере редукции размерности в четырёхмерной задаче. Пусть $y_2(x)$ — двухмертная развертка (отображает отрезок в прямоугольник), тогда рассмотрим функцию $\psi(x_1, x_2) = \phi(y_2(x_1), y_2(x_2))$. К $\psi(x_1, x_2)$ можно также применить редукцию размерности с помощью развертки. Таким образом, задав точку $x^* \in [0; 1]$, вычислив $y_2(x^*) = (x_1, x_2)$ и пару векторов $(y_2(x_1), y_2(x_2))$, получим четырёхмерную точку. Из инъективности $y_2(x)$ следует инъективность $z(x)$.

Проблемой этого метода является выяснение свойств функции $\phi(z(x))$ и возможности использования одномерного метода Стронгина с гёльдеровой метрикой для оптимизации $\phi(z(x))$. Чтобы не тратить время на теоретическое исследование, были проведены численные эксперименты с целью оценить возможность применения многоуровневой развёртки в четырёхмерном случае.

Прежде всего, рассмотрим линии уровня функции $\psi(x_1, x_2) = \phi(y_2(x_1), y_2(x_2))$ при $\phi(t) = \sum_{i=1}^4 (t_i - 0.5)^2$. Как видно из рис. 3.1, линии уровня имеют довольно сложную струк-

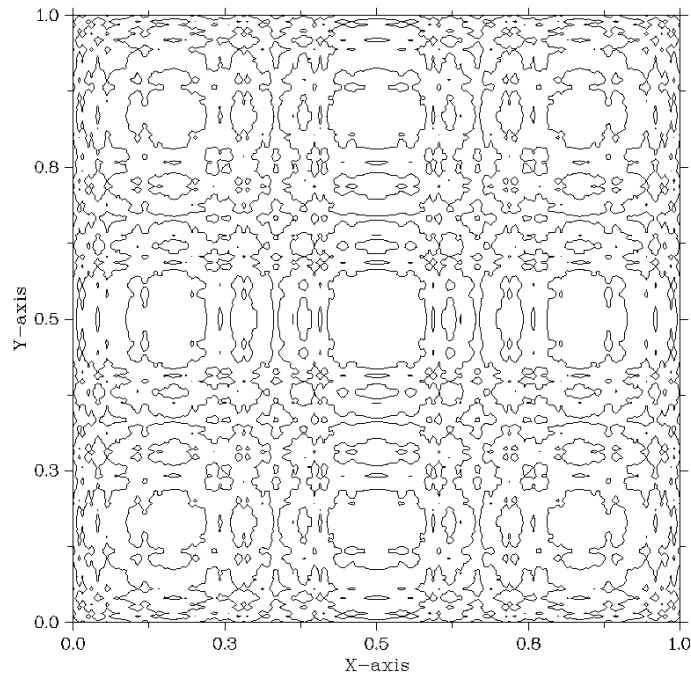


Рис. 3.1: Линии уровня функции $\psi(x_1, x_2) = \phi(y_2(x_1), y_2(x_2))$

туру, что говорит о возможных сложностях применения одномерного метода с разверткой.

Далее был проведён более масштабный вычислительный эксперимент: с помощью многоуровневой развёртки решались 100 задач из класса GKLS Simple 4d. На рис. 3.2 приведены

операционные характеристики метода с простой и многоуровневой развёртками. При этом были зафиксированы следующие параметры алгоритма: надёжность $r = 4.5$, плотность построения всех развёрток 12, критерий остановки попадание точки, поставленной методом в квадрат со стороной $\varepsilon = 10^{-2}$, центром которого является решение задачи. Как видно из

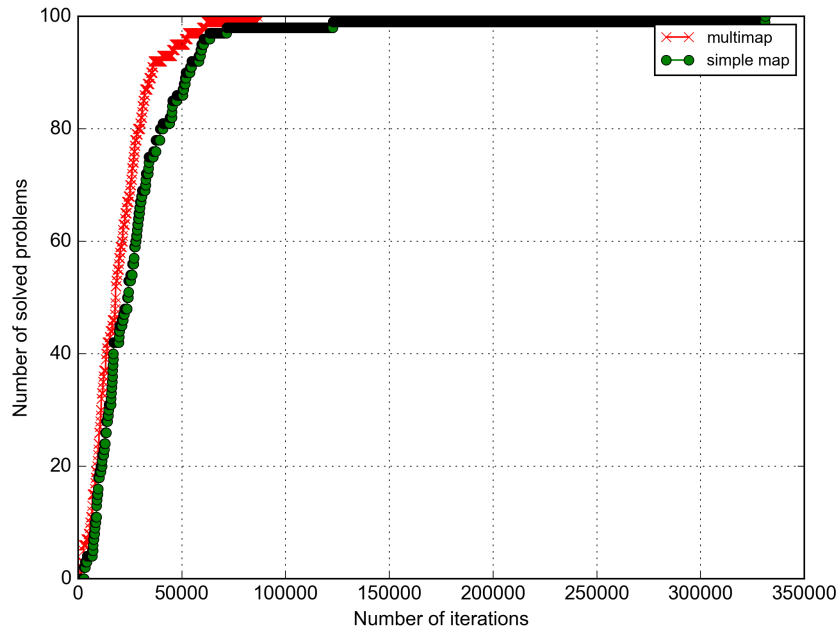


Рис. 3.2: Операционная характеристика метода с различными развёртками на классе GKLS Simple 4d

графика, операционная характеристика метода с многоуровневой развёрткой лежит выше, чем аналогичная кривая метода с простой развёрткой. Кроме того, метод с многоуровневой развёрткой заметно раньше вышел на стопроцентную надёжность на решаемой выборке задач. Исходя из первых экспериментов можно сказать, что при данных параметрах алгоритма многоуровневая развёртка лучше, но при уменьшении точности алгоритма ε до значения 10^{-3} выполнения критерия остановки в случае использования многоуровневой развёртки с выбранной плотностью построения внутренних развёрток не происходит. Если плотность увеличить до 16, то остановка также не будет происходить. Исходя из этого можно делать вывод, что использование на практике многоуровневых развёрток с большой долей вероятности невозможно из-за описанного дефекта сходимости.

Сравнение различных типов множественных разверток

Одним из недостатков использования метода редукции размерности с помощью развёрток является потеря информации о близости точек в многомерном пространстве при решении одномерной задачи. На рис. 4.3 представлена развёртка двухмерного параболоида вращения. Как видно из графика, единственный глобальный уровень расщепился на три.

Это происходит из-за того, что развертка может возвращаться в окрестность некоторых то-

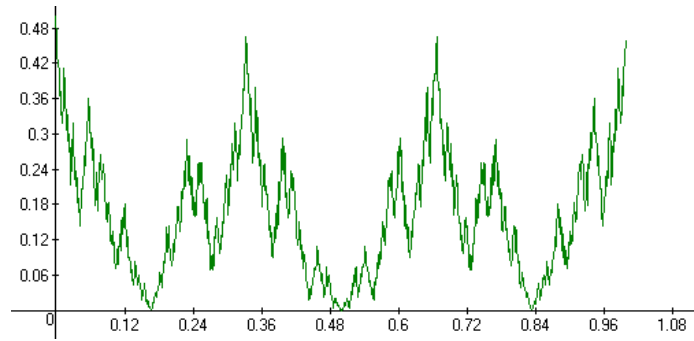


Рис. 4.3: Одномерная функция после подстановки развертки в $\phi(y) = y_1^2 + y_2^2$

чек при разных (далёких друг от друга) значениях одномерного параметра. Чтобы частично компенсировать потерю информации о многомерной окрестности точки, были предложены множественные развёртки [6, 7].

Применение локального поиска для ускорения сходимости АГП

Методы локального поиска могут применяться в сочетании с глобальными алгоритмами для улучшения полученных решений или текущих оценок оптимума. В первом случае локальный метод стартует из точки, найденной глобальным методом, и уточняет решение практически до любой нужной точности. Это позволяет избежать чрезмерных затрат на поиск решения с высокой точностью глобальным методом.

Во втором случае локальный метод используется для ускорения обнаружения локальных оптимумов. Информационно-статистический метод Стронгина позволяет обновлять свою поисковую информацию из любых посторонних источников, в том числе из точек испытаний, полученных от локального метода. Как только глобальный метод находит новую оценку оптимума, из этой точки стартует локальный метод и все или часть испытаний, проведённых им добавляется в поисковую информацию, далее глобальный метод продолжает работу. Каких-либо теоретических исследований подобной схемы не проводилось, поэтому её эффективность проверялась экспериментально.

В качестве метода локальной оптимизации был выбран метод Хука-Дживса [8]. Он прост в реализации и для его работы не требуется знать значений производных оптимизируемой функции.

Были проведены две серии экспериментов, соответствующих следующим схемам добавления точек, полученных локальным методом в поисковую информацию:

- добавление единственной точки, к которой сошёлся локальный метод;
- добавление всех промежуточных точек.

Эксперименты проводились на классах GKLS 4d Simple и GKLS 5d Simple, параметры метода были заданы такие же, как в разделе 3. Из рис. 5.4, 5.5 можно сделать вывод, что ва-

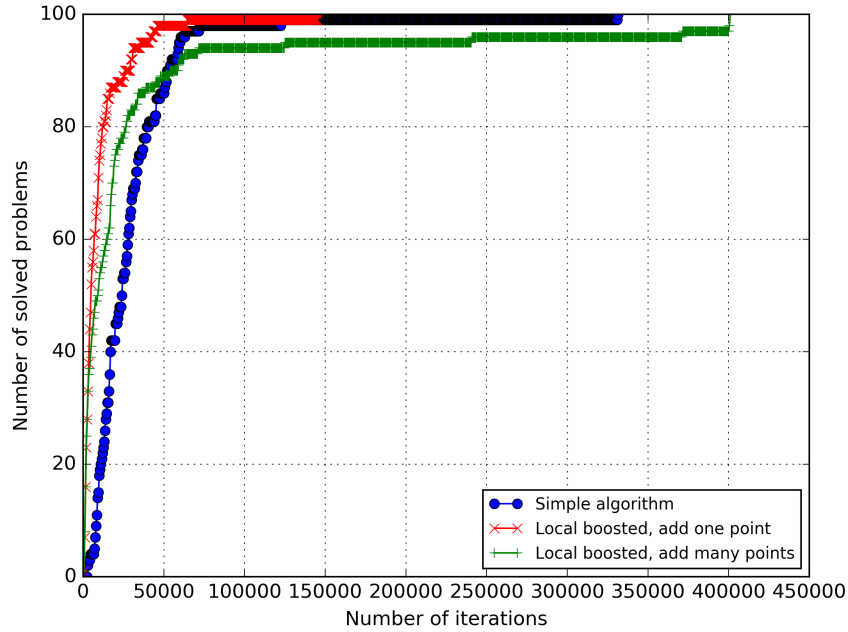


Рис. 5.4: Операционные характеристики на классе GKLS 4d Simple при различных вариантах использования локального метода

риант с использованием только одной лучшей точки, полученной локальным методом, оказался наилучшим. В обоих случаях он заметно ускорил сходимость глобального алгоритма.

Смешанный алгоритм глобального поиска и его эффективная реализация

Ещё одной модификацией метода Стронгина, позволяющей в процессе оптимизации лучше учитывать данные о локальных оптимумах, найденных в процессе поиска, является смешанный алгоритм Стронгина-Маркина [9]. Наряду с характеристикой интервала $R(i)$ (2.3) можно рассматривать $R^*(i)$, которая будет более чувствительна к наличию в интервале текущего найденного минимума функции x_k^* :

$$R^*(i) = \frac{R(i)}{\sqrt{(z_i - z^*)(z_{i-1} - z^*)/\mu + 1.5^{-\alpha}}}$$

где $f(x_k^*) = z^*$, а $\alpha \in [1; 30]$ — степень локальности. Чем она больше, тем более высокая характеристика у интервала, содержащего x_k^* , по сравнению с остальными.

Смешанный алгоритм состоит в следующем: в процессе работы метода каждые S итераций интервал для последующего разбиения выбирается по характеристикам $R^*(i)$. S — параметр смешивания. Такой подход позволяет существенно ускорить сходимость метода.

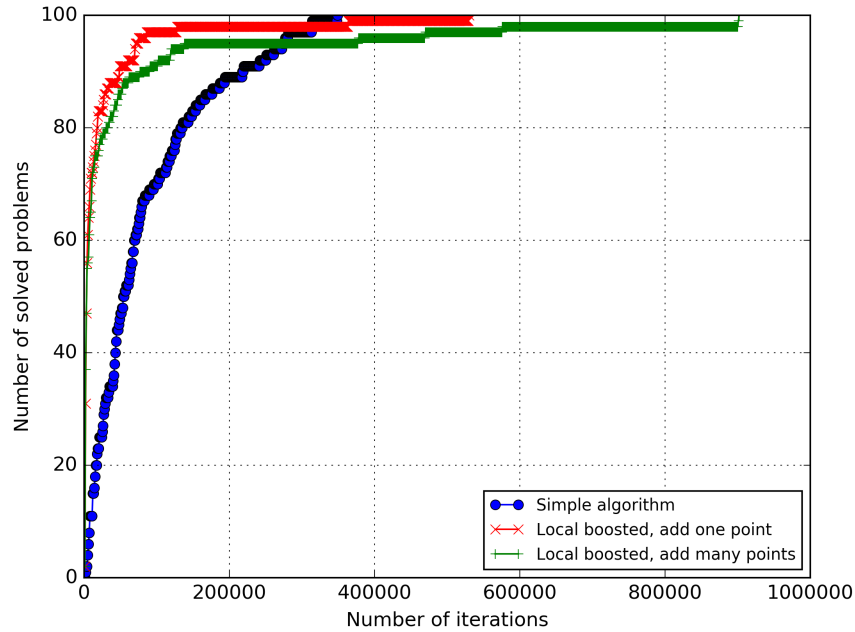


Рис. 5.5: Операционные характеристики на классе GKLS 5d Simple при различных вариантах использования локального метода

На рис. 6.6 приведены операционные характеристики чисто глобального и смешанного алгоритма на классе GKLS 4d Simple. Параметр смешивания S равен 5, $\alpha = 15$, остальные параметры метода были заданы такие же, как в разделе 3.

Из-за того, что интервал имеет сразу две характеристики, появляется проблема эффективной реализации смешанного алгоритма. Если интервал имеет одну характеристику, то для выбора максимальной достаточно организовать приоритетную очередь характеристик [10]. Причём перезаполнение такой очереди необходимо не на каждой итерации: в большинстве случаев характеристики интервалов не меняются, достаточно удалить разбиваемый интервал и вставить в очередь два новых. Такая организация работы метода позволяет существенно сократить объём вычислений. При наличии у интервала двух характеристик можно организовать две связанные очереди. В этом случае необходимо предусмотреть процедуру синхронизации двух очередей.

Перечислим операции, при которых необходима синхронизация:

- вставка интервала сразу в обе очереди;
- удаление интервала из какой-либо очереди.

Синхронизация достигается путём введения перекрёстных ссылок между элементами очередей. На рис. 6.7 приведена схема связанных очередей. Элемент очереди представляет собой совокупность ключа (**LocalR** или **R**), указателя на интервал (**pInterval**) и указателя на элемент связанной очереди, соответствующий тому же интервалу (**pLinkedElement**). Опишем подробнее алгоритмы вставки и удаления элементов.

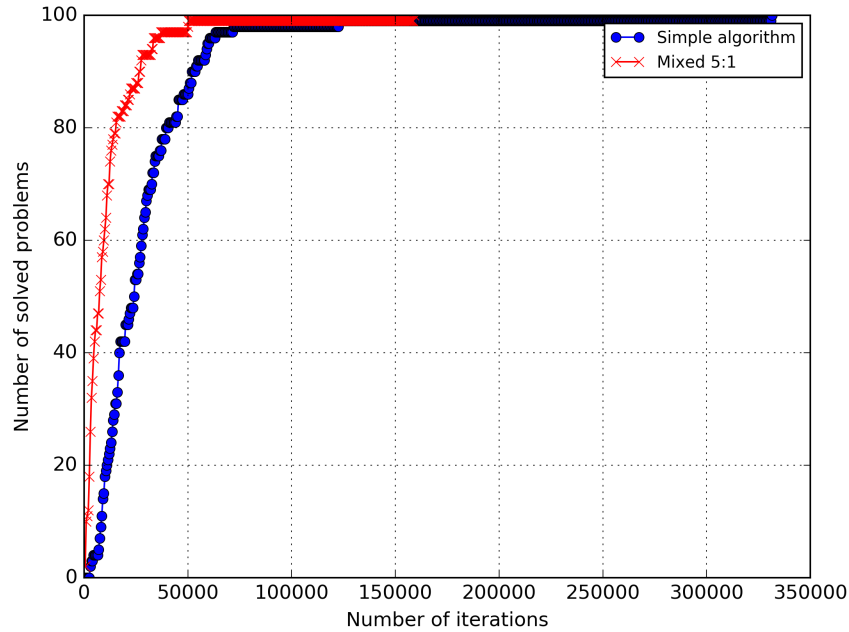


Рис. 6.6: Операционные характеристики обычного и смешанного АГП на классе GKLS 4d Simple

Вставка элемента в пару связанных очередей:

1. Попытаться вставить элемент в очередь глобальных характеристик (он может быть не вставлен, если имеет слишком низкий приоритет).
2. Попытаться вставить элемент в очередь локальных характеристик (он может быть не вставлен, если имеет слишком низкий приоритет).
3. Если элемент интервал вставлен в обе очереди, то выставить перекрёстные ссылки.

Удаление элемента с минимальным ключом:

1. Удалить элемент с минимальным ключом из очереди, запомнить указатель **pLinkedElement**.
2. Если **pLinkedElement** ненулевой, то вызвать процедуру удаления элемента, на который указывает **pLinkedElement** в структуре данных, хранящей связанную очередь.

Последний момент, который надо учесть при реализации: вставка или удаление элемента очереди приводит к тому, что необходимо восстановить её внутреннюю структуру. Если очередь хранится в куче, то восстанавливается свойство кучеобразности. Во время этого процесса требуется производить попарные перестановки элементов, а значит, необходимо обновлять ссылки на эти элементы в связанной очереди.

Стоит заметить, что внесённые модификации (в основном эта работа со ссылками) не увеличивают асимптотическую сложность выполнения операций вставки и удаления по сравнению с единственной очередью.

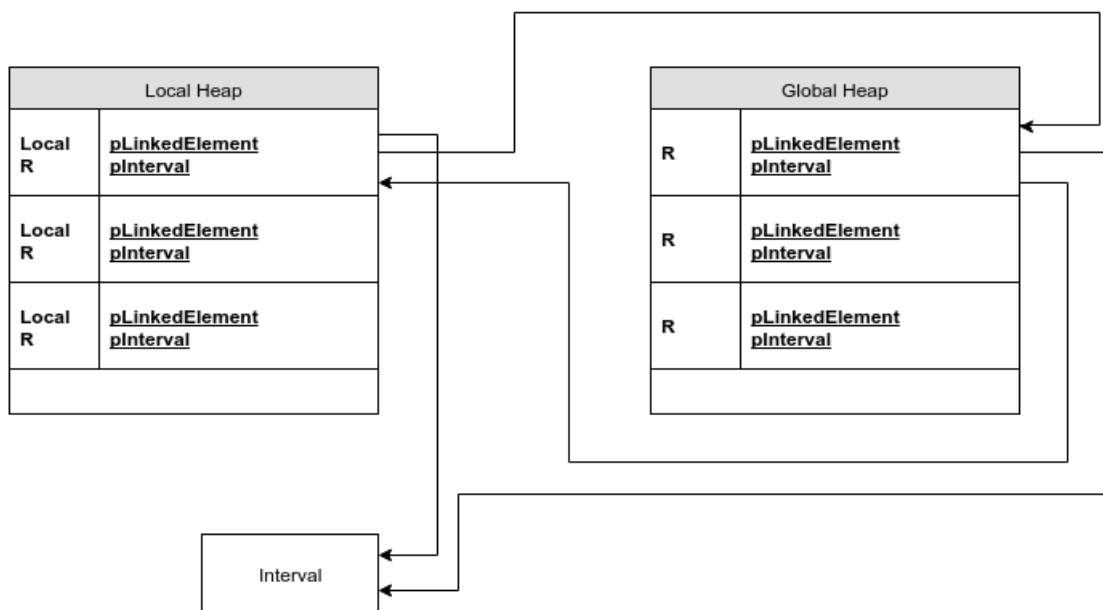


Рис. 6.7: Схема устройства связанных очередей

Список литературы

- [1] В.П. Гергель А.В. Сысоев К.А. Баркалов. «Блочная многошаговая схема параллельного решения задач многомерной глобальной оптимизации». В: *Материалы XIV Международной конференции "Высокопроизводительные параллельные вычисления на кластерных системах 10-12 ноября, ПНИПУ, Пермь. 2014*, 425—432.
- [2] Сысоев А.В. Баркалов К.А. Гергель В.П. Лебедев И.Г.. «MPI-реализация блочной многошаговой схемы параллельного решения задач глобальной оптимизации». В: *Суперкомпьютерные дни в России: Труды международной конференции (28-29 сентября 2015 г., г. Москва)*. М.: Изд-во МГУ, 2015, с. 411—419.
- [3] В.В. Соврасов А.В. Сысоев К.А. Баркалов И.Г. Лебедев. «Реализация параллельного алгоритма поиска глобального экстремума функции на Intel Xeon Phi». В: *Вычислительные методы и программирование 17 (2016)*, с. 101—110.
- [4] Стронгин Р. Г.. «Численные методы в многоэкстремальных задачах (информационно-статистические алгоритмы)». «Наука», М., 1978, 240 стр.
- [5] Сергеев Я.Д. Квасов Д.Е.. *Исследование методов глобальной оптимизации при помощи генератора классов тестовых функций*. Н.Новгород: Изд-во ННГУ, 2011.
- [6] Р. Г. Стронгин. «Параллельная многоэкстремальная оптимизация с использованием множества разверток». В: *Ж. вычисл. матем. и матем. физ.* 31:8 (1991), 1173—1185.
- [7] К. А. Баркалов Р. Г. Стронгин В. П. Гергель. «Параллельные методы решения задач глобальной оптимизации». В: *Известия высших учебных заведений. Приборостроение* № 10 / том 52 (2009), с. 25—33.
- [8] Химмельблау Д. М.. *Прикладное нелинейное программирование*. Издательство МИР, Москва, 1975.
- [9] Р. Г. Стронгин Д. Л. Маркин. «Метод решения многоэкстремальных задач с невыпуклыми ограничениями, использующий априорную информацию об оценках оптимума». В: *Ж. вычисл. матем. и матем. физ.* 27:1 (1987), 52—62.
- [10] N. Santoro T. Strothotte M. D. Atkinson J.R. Sack. «Min-Max Heaps and Generalized Priority Queues». В: *Communications of the ACM* 29 (1986), с. 996—1000.