Адаптивные параллельные алгоритмы для многомерной многоэкстремальной оптимизации

А.В. Гергель, Д.В. Гнатюк

Нижегородский государственный университет им Н. И. Лобачевского

1. Задачи многомерной многоэкстремальной оптимизации и многошаговая схема редукции размерности

Задача многомерной многоэкстремальной оптимизации может быть определена как проблема поиска наименьшего значения действительной функции $\varphi(y)$

$$\varphi(y^*) = \min\{\varphi(y): y \in D\}, \tag{1}$$

где D есть область поиска, представляющая собой некоторый гиперпараллелепипед N-мерного евклидова пространства.

Многие регулярные поисковые методы решения многомерных оптимизационных задач сводят многомерную задачу (явно или неявно) к системе одномерных подзадач. Один из наиболее общих методов редукции размерности состоит в применении *многошаговой схемой редукции размерности*, согласно которой решение многомерной задачи оптимизации может быть получено посредством решения последовательности «вложенных» одномерных задач (см., например, [1-3]):

$$\min_{y\in D} \varphi(y) = \min_{y_1\in[a_1,b_1]} \dots \min_{y_N\in[a_N,b_N]} \varphi(y_1,\dots y_N) \tag{2}$$

Согласно (2) решение многомерной задачи (1) сводится к решению одномерной задачи:

$$\varphi^* = \min_{y\in D} \varphi(y) = \min_{y_1\in[a_1,b_1]} \tilde{\varphi}_1(y_1), \tag{3}$$

где

$$\tilde{\varphi}_i(y_i) = \varphi_i(y_1,\dots,y_i) = \min_{y_{i+1}\in[a_{i+1},b_{i+1}]} \varphi_{i+1}(y_1,\dots,y_i,y_{i+1}) \quad 1 \le i < N, \tag{4}$$

$$\varphi_N(y_1,\dots,y_N) = \varphi(y_1,\dots,y_N). \tag{5}$$

Правила (3) – (5) определяют множество задач:

$$F_l = \{f_i(y_i, x),\ 1 \le i \le l\}. \tag{6}$$

порождаемых в соответствии с многошаговой схемой редукции. Количество задач в множестве F_l в процессе поиска может изменяться: увеличиваться при переходе к следующей переменной и уменьшаться при завершении решения какой-либо из задач (можно отметить, что при этом количество задач не превышает размерность решаемой задачи N). При этом активной – решаемой – в множестве F_l является только одна задача – это задача с максимальным номером варьируемой переменной.

В работе предлагается обобщенная (адаптивная) многошаговой схемы редукции размерности, в которой предлагается осуществлять одновременное решение всех задач множества F_l. Такое решение может выполняться последовательно (при наличии единственного процессора), тогда для выполнения очередной итерации глобального поиска необходимо выбирать для решения одну из задач множества F_l в соответствии с тем или иным правилом выбора задач. Но решение задач может выполняться и параллельно, если используемый вычислитель является многопроцессорным или многоядерным. Именно параллельный – более общий вариант – будет рассматриваться далее в работе.

Важно отметить, что выполнение итерации глобального поиска для любой задачи множества F_l с номером переменной меньшим, чем N (N есть размерность решаемой задачи), будет приводить к порождению новой одномерной задачи вида (4), и, тем самым, количество задач в семействе F_l может оказаться значительным (десятки и сотни тысяч для сложных задач оптимизации).

Выделим в семействе решаемых задач оптимизации F_l из (6) задачи для последней варьируемой переменной (задачи уровня N) – будем называть задачи подобного вида *терминальными* (нетерминальные задачи семейства будут именоваться далее *структурными*). В соответствии с выдвинутыми в начале работы предположениями именно при решении таких задач выполняется наибольший объем вычислений, т.к. только в этих задачах осуществляется вычисление значений функционалов исходной решаемой задачи оптимизации.

2. Централизованная схема параллельного глобального поиска

Общая схема построения параллельных методов на основе рассмотренных в статье алгоритмов состоит в следующем. Характеристики интервалов, вычисляемые алгоритмами глобального поиска, рассматриваются как некоторая мера важности интервалов на предмет содержания в них искомого решения оптимизационной задачи. Следуя данному пониманию, в параллельных методах после выбора точки проведения испытания для первого процессора в точном соответствии с последовательным алгоритмом, для второго процессора точку испытания выбирается из следующего по важности интервала (т.е. из интервала со следующей по порядку максимальной характеристикой) и т.д.

Рассмотренная выше идея послужила основой для целого направления работ по разработке параллельных информационно-статистических алгоритмов глобального поиска – см., например, [2, 4]. Один из важных результатов в данном направлении состоит в том, что для параллельных алгоритмов оптимизации оказалось возможным формулирование общей схемы характеристической представимости [3-4]. Данная схема может быть обобщена для предлагаемого в работе подхода.

Схема характеристической представимости параллельных алгоритмов многоэкстремальной оптимизации на основе адаптивной многошаговой редукции размерности. Пусть p>1 есть число процессоров, используемых для решения задачи оптимизации. Каждый из процессоров, получив точку $y \in D$, осуществляет независимо и параллельно с другими процессорами вычисление значения функции $\varphi(y)$ из (1). В процессе вычисления значения функции испытание считается незавершенным – тем самым, при выполнении алгоритма будут существовать точки завершенных и незавершенных испытаний.

Далее, пусть проводятся первые p>1 испытаний в некоторых точках области поиска D и количество завершенных испытаний k=0. Тогда выбор точки очередной итерации (*решающее правило алгоритма*) включает выполнение следующего набора действий:

1. **Начало итерации**. Если завершено k>0 испытаний в точках $y^1, y^2,\ldots, y^k$, и реализуются испытания в точках $y^{k+1}, y^{k+2},\ldots, y^{k+p}$, вычислительная схема алгоритма находится в состоянии ожидания. Если же одно из параллельных испытаний завершается, номер k увеличивается на единицу и алгоритм переходит к формированию точки x^{k+p+1} нового k+p+1 испытания.

2. **Вычисление характеристик**. Для всех задач семейства F_l необходимо выполнить правила 2.1 – 2.4.

2.1 **Формирование системы интервалов**. Координаты $x^1, x^2,\ldots, x^{s(j)}$, испытаний (завершенных и незавершенных) вместе с граничными точками a_j и b_j области одномерной задачи оптимизации $f_j(x) \in F_l$ перенумеровываются нижним индексом в порядке возрастания:

$$a_j = x_0 \le x_1 \le x_2 \le \ldots \le x_{s-1} \le x_s = b_j .$$

2.2. **Список допустимых интервалов**. Сформировать список номеров I_s допустимых интервалов $(x_{i-1}, x_i), i \in I_s$, в качестве которых рассматриваются такие интервалы, граничные точки которых являются координатами завершенных интервалов, либо граничными очками области $[a_j, b_j]$.

2.3. **Вычисление характеристик допустимых интервалов**. Каждому допустимому интервалу, ставится в соответствие величина $R(i)$, называемая далее *характеристикой интервала*.

2.4. **Выбор интервала с максимальной характеристикой**. Среди допустимых интервалов определить интервал $(x_{t-1}, x_t), t \in I_s$, которому соответствует максимальная характеристика $R(t)$, т.е.

$$R(t) = \max\{R(i) : i \in I_s\} \tag{7}$$

(данная характеристика $R(f_j)= R(t)$ будет называться также *характеристикой задачи* $f_j(x) \in F_l$).

3. **Вычисление точки нового испытания**. Точка x^{k+p+1} очередной итерации глобального поиска выбирается внутри интервала (x_{t-1}, x_t) для задачи $f_\tau(x) \in F_l$ с максимальной характеристикой $R(f_\tau)$ в соответствии с правилом

$$x^{k+p+1} = s(t) \in (x_{t-1}, x_t)\, , R(f_\tau) = \max\{R(f_j) : f_j \in F_l\} \ . \tag{8}$$

Если выбранная задача является структурной (номер переменной меньше размерности задачи), то порождается новая задача семейства в соответствии с правилами (3)-(5) и выполнение правил начинается с шага 2. Для терминальной задачи в соответствии с выбранной точкой x^{k+p+1} формируется вектор варьируемых переменных y^{k+p+1} для очередного испытания. Для вычисления значений оптимизируемой функции данный вектор в асинхронном режиме отсылается процессу с наименьшей нагрузкой, а само испытание считается незавершенным.

4. **Проверка условия остановки**. Завершение глобального поиска происходит, если в правиле 3 выбирается интервал (x_{t-1}, x_t)с максимальной характеристикой $R(t)$, длина которого оказывается меньшей заданной точности поиска ε>0, т.е. выполняется условие

$$x_t - x_{t-1} \le \varepsilon \, .$$

Приведенное правило остановки допускает два варианта применения: "слабое" условие прекращает вычисления при выполнении правила остановки для любой задачи семейства F_l; "строгий" вариант условия осуществляет проверку правила остановки только для функции самого верхнего уровня $\tilde{\varphi}_1(y_1)$.

Характеристическая схема представимости параллельных алгоритмов глобального поиска может быть детализирована для получения того или иного метода многоэкстремальной оптимизации указанием правила вычисления характеристики интервалов (7) и правила определения точки очередного испытания (8). В качестве этих правил могут быть использованы правила практически любого существующего последовательного алгоритма поиска глобального поиска и, как результат, на его основе может быть сформирован параллельный метод многоэкстремальной оптимизации, соответствующий исходному последовательному алгоритму. Так, в частности, могут быть задействованы правила метода последовательного сканирования, метода ломаных, а также правила всех информационно-статистических алгоритмов [3].

3. Распределенная схема параллельного глобального поиска для адаптивной многошаговой схемы редукции размерности

Централизованная схема организации параллельных вычислений является достаточно эффективной при сравнительно небольшом количестве используемых процессоров. В случае же многопроцессорных систем с существенным числом (сотни и тысячи) процессоров управляющий узел системы, обеспечивающий выполнение решающих правил алгоритмов, обработку поисковой информации и определение точек очередных испытаний может оказаться «узким» местом. Управляющий процессор может не успевать обеспечивать нагрузку для всех процессоров-исполнителей даже если этот управляющий процессор будет обладать повышенной производительностью. Все дальнейшее рассмотрение будет выполнено на примере для информационно-статистических многомерных методов многоэкстремальной оптимизации с адаптивной многошаговой схемы редукции размерности.

Введем множество номеров задач семейства F_l из (6):

$$L = \{ 1, 2, \dots , l \}, \tag{9}$$

и пусть для проведения вычислений имеется $p>1$ процессоров. Распределим имеющиеся задачи между процессорами – данное распределение можно зафиксировать при помощи соответствующего разделения множества L на подмножества:

$$\Pi = \{ \pi_1, \pi_2, \dots , \pi_p \}, \tag{10}$$

$$\pi_i = \{ j_s : j_s \in L,\ 1 \le s \le l_i \},\ 1 \le i \le p,$$

$$\forall\ i \in L\ \exists\ j : i \in \pi_j,\ \forall\ i,j \Rightarrow \pi_i \cap \pi_j = \{\varnothing\}$$

где π_i, $1 \le i \le p$, есть множество задач, распределенных для решения на процессоре i, $1 \le i \le p$.

Построенная схема является децентрализованной – все процессоры работают параллельно и самостоятельно генерируют точки проведения испытаний. С другой стороны, распределение задач между процессорами приводит к сложным информационным зависимостям между процессорами. Для решения данной проблемы в статье предлагается эффективная децентрализованная схема распределенных вычислений, в которой учитывается структура иерархической зависимости решаемых задач оптимизации.

Пусть номера терминальных задач образуют множество

$$L^N = \{ i_j : i_j \in L,\ 1 \le j \le l^N \}. \tag{11}$$

Терминальные задачи распределяются между процессорами; выбранное распределение фиксируется по аналогии с (10):

$$\Pi^N = \{ \pi_0, \pi_1, \pi_2, \dots , \pi_p \}, \tag{12}$$

$$\pi_0 = L \setminus L^N,\ \pi_i = \{ j_s : j_s \in L^N,\ 1 \le s \le l_i \},\ 1 \le i \le p,$$

$$\forall\ i \in L\ \exists\ j : i \in \pi_j,\ \forall\ i,j \Rightarrow \pi_i \cap \pi_j = \{\varnothing\}.$$

В (12) каждое подмножество π_i, $1 \le i \le p$, определяет список терминальных задач, распределенных для решения на процессоре i, $1 \le i \le p$. Подмножество π_0 содержит номера всех структурных задач, обработку которых может осуществлять дополнительный процессор. Подобное распределение задач обеспечивает систематический характер информационных взаимодействий между процессорами - процессоры с терминальными задачи пересылают получаемые оценки минимальных значений исходной оптимизируемой задачи только управляющему процессору, а управляющий процессор занимается обработкой только структурных задач без выполнения трудоемких вычислений значений функционалов исходной решаемой задачи оптимизации.

Приведем результаты выполненных численных экспериментов для оценки эффективности распределенной схемы параллельного глобального поиска для адаптивной многошаговой схемы редукции размерности. Для проведения вычислительных экспериментов использовался класс тестовых задач многоэкстремальной оптимизации Растригина:

$$f(y_1,...,y_N)=\sum_{i=1}^{N}(y_i^2-\cos(18*y_i^2)),\ y_i\in[-1.0;1.5]\forall i=1,..,N$$

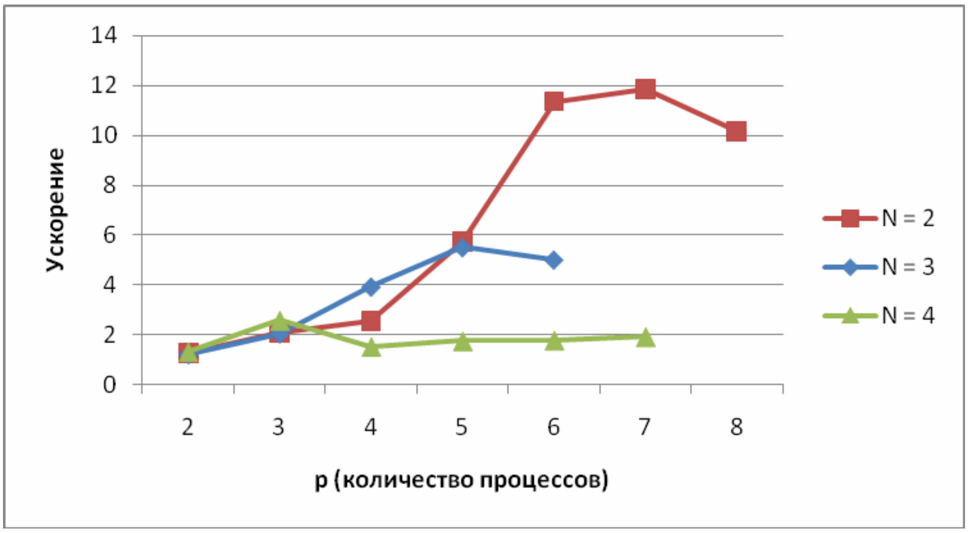

Рисунок 1. Оценка эффективности распределенной схемы параллельного глобального поиска для адаптивной многошаговой схемы редукции размерности.

Литература

1. Р.Г. Стронгин . Численные методы в многоэкстремальных задачах. М.: Наука, 1978.
2. R.G. Strongin , Y.D. Sergeyev Global Optimization with non-convex constraints: Sequential and parallel algorithms. Kluwer Academic Publishers, Dordrecht, 2000.
3. С.Ю. Городецкий , В.А. Гришагин Нелинейное программирование и многоэкстремальная оптимизация. – Н.Новгород: Изд-во ННГУ, 2007.
4. Y.D. Sergeyev , V.A. Grishagin Parallel asynchronous global search and the nested optimization scheme. // J. Comput. Anal. Appl. – 2001. – Vol. 3, № 2. – P. 123-145.
5. В.П. Гергель Теория и практика параллельных вычислений. – М.: Интернет-Университет Информационных технологий; БИНОМ. Лаборатория знаний, 2007.