

**ГИБРИДНЫЙ МЕТОД ГЛОБАЛЬНОЙ ОПТИМИЗАЦИИ НА ОСНОВЕ НЕЙРОСЕТЕВЫХ АППРОКСИМАЦИЙ ИНВЕРСНЫХ ЗАВИСИМОСТЕЙ, МЕТОДА ХУКА-ДЖИВСА И ПОДВИЖНЫХ ГРУПП ПРОБНЫХ ТОЧЕК**

**Пушкарев К.В.,**

**научный руководитель д-р физ.-мат. наук Кошур В.Д.,**

*Сибирский федеральный университет*

*Институт космических и информационных технологий*

**Аннотация**

Представлен гибридный эвристический параллельный метод глобальной оптимизации непрерывных многоэкстремальных функций многих переменных в области, имеющей форму многомерного параллелепипеда. Метод сочетает использование нейросетевых аппроксимаций инверсных зависимостей (НАИЗ), локального поиска по методу Хука-Дживса и подвижных групп пробных точек. Описана программная реализация метода, приспособленная к выполнению на кластерных вычислительных системах. Приведены результаты вычислительных экспериментов по минимизации многоэкстремальной функции Растригина 500 переменных.

**Постановка задачи**

Рассматривается ограниченная функция от  $\bar{x} \in \Omega \subset R^n$

$$\varphi = \Phi(\bar{x}), \tag{1}$$

$$\Phi(\bar{x}) : \Omega \rightarrow R, \tag{2}$$

которая непрерывна в ограниченной области  $\Omega = \{ \bar{x} : \bar{x} \in [ \underline{L}; \bar{U} ] \subset R^n$ .

Необходимо найти точку, в которой достигается её минимум:

$$\bar{x}_{\min} = \arg \min_{\bar{x} \in \Omega} \Phi(\bar{x}); \tag{3}$$

$$\Phi_{\min} = \Phi(\bar{x}_{\min}). \tag{4}$$

**Инверсные зависимости**

Ключевым для метода является предположение о возможности глобальной оптимизации функций через использование инверсных зависимостей.

Инверсные зависимости являются средством перехода от значений целевой функции (ЦФ) к координатам. Они представляют собой покоординатные отображения значений ЦФ в пространство поиска:

$$\bar{x} = \Psi(\varphi) = (\Psi_1(\varphi), \Psi_2(\varphi), \dots, \Psi_n(\varphi)), \varphi \in R, \bar{x} \in \Omega. \tag{5}$$

Инверсные зависимости *не являются, вообще говоря, обратными функциями*. Инверсных зависимостей может быть много, но в данной работе рассматриваются только те из них, которые удовлетворяют требованию:

$$\Psi(\varphi) = \bar{x}_{\min} \text{ при } \varphi \leq \Phi_{\min}. \tag{6}$$

Тогда решение исходной  $n$ -мерной задачи (3) сведётся к решению  $n$  одномерных задач:

$$x_i = \Psi_i(\Phi_{\min}), i = \overline{1, n}; \tag{7}$$

$$\Phi_{\min} = \arg \min_{\varphi} \Phi(\Psi_1(\varphi), \Psi_2(\varphi), \dots, \Psi_n(\varphi)). \tag{8}$$

Из-за ограниченности информации о ЦФ возможно строить только аппроксимации  $\Psi^{[k]}(\varphi)$  подходящей инверсной зависимости, справедливые в некотором интервале значений ЦФ  $[\varphi_1; \varphi_2]$ . Подходящим инструментом для этого являются обобщённо-регрессионные нейронные сети (GRNN). Можно показать, что условие (6) в предельном смысле для них выполняется.

GRNN обладает аппроксимирующими и сглаживающими свойствами. Степень сглаживания определяется параметром сети spread.

### Описание метода

Суть метода состоит в итеративном понижении значения целевой функции с отображением его в координаты с помощью инверсной зависимости. В благоприятном случае истинное значение ЦФ в точке с этими координатами окажется не больше, чем предполагаемое. Для реализации такого случая необходимо, чтобы «склоны» минимума были покрыты пробными точками равномерно (плотность при этом не имеет решающего значения), что вытекает из свойств GRNN.

Обеспечить равномерное покрытие области поиска в многомерных задачах крайне сложно, поэтому были предложены альтернативные меры по преодолению подобных затруднительных ситуаций.

Простейшей и необходимой мерой является неинформированный поиск (пробные точки создаются без учёта поведения ЦФ). В данном случае в некоторой области вокруг наилучшей пробной точки случайным образом создаются новые пробные точки, которые могут заменять собой точки с более высокими значениями ЦФ.

Перед добавлением пробных точек в рамках неинформированного поиска отбрасываются  $K_d$  точек с наибольшими значениями ЦФ из текущего множества пробных. Благодаря этому,  $K_d$  худших пробных точек безусловно замещаются новыми на каждом шаге неинформированного поиска. Это позволяет внести возмущение в процесс поиска, улучшая исследование области и снижая вероятность попадания в тупик, когда все новые пробные точки отвергаются, так как не дают улучшения значения ЦФ.

Множество пробных точек разбито на группы, каждая из которых исследует определённую часть области поиска с помощью (гибридного) алгоритма НАИЗ. Группы объединены в кластеры (рис. 1). Будем называть ведущей точкой группы (кластера) пробную точку с наименьшим значением ЦФ в данной группе (кластере), глобальной ведущей точкой — ведущую точку кластера с наименьшим значением ЦФ.

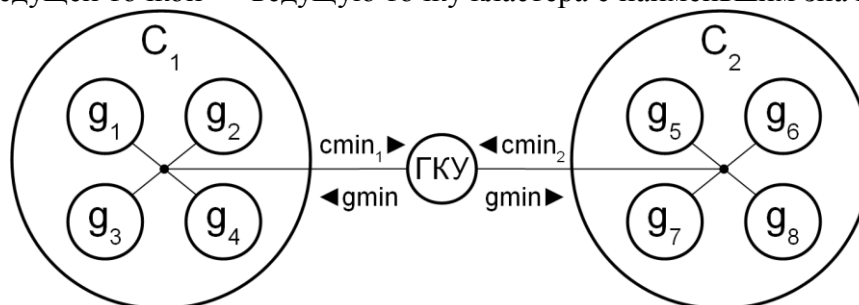


Рис. 1. Информационная топология соседства групп пробных точек.  $C_1, C_2$  — кластеры групп  $g_i$  — связаны через глобальный коммуникационный узел (ГКУ);  $cmin_1, cmin_2$  — ведущие точки кластеров;  $gmin$  — глобальная ведущая точка.

Группа имеет доступ к локальным ориентирам — ведущим точкам других групп своего кластера, а также к глобальному ориентиру — глобальной ведущей точке. С заданной периодичностью, точки группы копируются, и копии сдвигаются к тому

ориентир, в направлении которого ЦФ убывает быстрее. Если такого направления для данной группы нет, точки сдвигаются *от* того ориентира, в направлении которого ЦФ *возрастает* быстрее. Направление и скорость убывания/возрастания, а также вектор сдвига вычисляются между ведущей точкой группы и ориентиром. Величина сдвига выбирается случайным образом в интервале от 0,25 до 0,75 расстояния до ориентира.

Количество точек в группе ограничено, поэтому сдвинутые образы точек группы конкурируют с прообразами за место в группе. Точки с наибольшими значениями ЦФ отбрасываются.

Преимущества вышеописанного подхода:

- группа может сконцентрироваться на изучении определённого перспективного района области поиска;
- если область вокруг глобальной ведущей точки является перспективной, группы будут тяготеть к ней, обеспечивая более тщательное её исследование;
- он позволяет легко повысить точность или адаптироваться к увеличению размера или размерности области поиска через увеличение числа групп или кластеров;
- он позволяет легко использовать дополнительные вычислительные ресурсы для увеличения точности;
- он естественным образом отображается на кластерные вычислительные системы.

Для преодоления ситуации, когда пробные точки оказались на «склоне» минимума, и процесс поиска с использованием аппроксимации инверсной зависимости зашёл в тупик, используется метод Хука-Дживса, который является локальным методом прямого поиска. Он подходит для решения многомерных задач оптимизации и не требует определения производной ЦФ, а значит не ограничивает применимость гибридного метода при совмещении с методом нейросетевой аппроксимации инверсных зависимостей. Он учитывает поведение ЦФ, что эффективнее поиска «вслепую». В гибридном методе процедура Хука-Дживса запускается регулярно из ведущей точки группы.

### **Программная реализация**

Программная реализация метода написана на языке C++ и является кроссплатформенной. Используется гибридная модель параллельных вычислений, т. е. сочетание обмена сообщениями с непосредственным доступом к общей памяти. Для обмена сообщениями используется интерфейс MPI (Message Passing Interface) версии 2.2.

Кластеры групп пробных точек отображаются в процессы (рис. 2), а группы — в потоки внутри процессов (оптимизационные агенты, ОА). Также существуют служебные потоки, отвечающие за передачу данных: локальные коммуникационные агенты (ЛКА; по одному в каждом процессе, держат таблицу ведущих точек кластера и обмениваются данными с глобальным коммуникационным агентом), глобальный коммуникационный агент (ГКА; один на все процессы, собирает сведения о ведущих точках кластеров, определяет и распространяет сведения о глобальной ведущей точке).

Главный поток каждого процесса становится потоком управления (ПУ), отвечающим за запуск и завершение других потоков и передачу итоговых результатов и статистики главному процессу.

При использовании кластерной вычислительной системы, на вычислительном узле помещается один или несколько оптимизационных процессов, обменивающихся сообщениями через MPI. Оптимизационные агенты одного процесса обмениваются информацией через общую память, что позволяет минимизировать задержки.

Количество оптимизационных процессов, помещаемых на одном узле, определяется количеством ОА в процессе и вычислительных ядер на узле. При определении

количества ОА в процессе следует учитывать, что повышение этого количества увеличивает вычислительные затраты в самих ОА (за счёт увеличения числа локальных ориентиров), а также нагрузку на ЛКА.

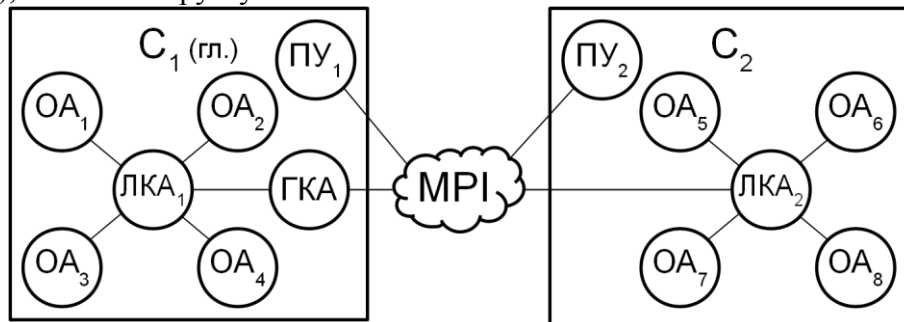


Рис. 2. Процессы и потоки в программной реализации метода, показанные для 2 кластеров по 4 группы. Прямоугольниками обозначены процессы, кругами — потоки. Главным процессом является  $C_1$ .  $OA_i$  — оптимизационный агент, соответствующий  $i$ -й группе пробных точек; ЛКА — локальный коммуникационный агент; ГКА — глобальный коммуникационный агент; MPI — интерфейс MPI (Message Passing Interface); ПУ — поток управления.

### Результаты экспериментов

Эксперименты проводились с двумя кластерами по 4 группы. В качестве тестовой функции использовалась многоэкстремальная функция Растргина 500 переменных:

$$f(\vec{x}) = \sum_{i=1}^{500} x_i^2 - 10 \cos(2\pi x_i) + 10. \quad (9)$$

Функция Растргина имеет огромное количество локальных минимумов и потому является стандартной тестовой функцией для методов глобальной оптимизации.

Область поиска:  $\Omega = [-50; 50]^{500}$ . Было проведено 50 испытаний метода со случайной инициализацией. Аппаратное обеспечение: 4-ядерный процессор Intel Core i5-2400 (3,1 ГГц), 2 ГБ ОЗУ. Результаты представлены в табл. 1.

Таблица 1 — Результаты экспериментов

	Мин.	Макс.	Среднее	Медиана	Среднекв. отклонение
$\Delta\Phi^1$	$6 \cdot 10^{-5}$	64,800	9,371	0,003	16,245
$\Phi_{eval}^2$	2 962 702	6 002 105	4 268 869,96	4 203 539,5	804 712,6
Кол-во итераций	3150	6075	4385,5	4337,5	798,281
Время, с	14	172	91,16	88,5	47,733

Примечания: 1) абсолютное отклонение итогового значения ЦФ от истинного глобального минимума; 2) количество вычислений значений ЦФ.

### Выводы

Совместное применение нейросетевых аппроксимаций инверсных зависимостей, локального поиска методом Хука-Дживса, а также подвижных групп пробных точек позволило с высокой точностью найти глобальный минимум многоэкстремальной функции Растргина 500 переменных.

Можно выделить следующие направления дальнейших исследований:

- 1) оптимизация количества вычислений ЦФ;
- 2) уменьшение средней погрешности нахождения глобального минимума;
- 3) проверка эффективности гибридного метода для минимизации других многоэкстремальных функций многих переменных.