УДК [004.023::004.032.26]::519.853.4

ГИБРИДНЫЙ МЕТОД ГЛОБАЛЬНОЙ ОПТИМИЗАЦИИ НА ОСНОВЕ НЕЙРОСЕТЕВЫХ АППРОКСИМАЦИЙ ИНВЕРСНЫХ ЗАВИСИМОСТЕЙ И МЕТОДА ХУКА-ДЖИВСА

Пушкарев К. В.,
научный руководитель д-р физ.-мат. наук Кошур В. Д.
Институт космических и информационных технологий
Сибирского федерального университета

Аннотация

Представлен гибридный эвристический метод глобальной оптимизации непрерывных многоэкстремальных функций многих переменных в области, имеющей форму многомерного параллелепипеда. Метод сочетает использование нейросетевых аппроксимаций инверсных зависимостей (НАИЗ) и локального поиска по методу Хука-Дживса. Приведены результаты вычислительного эксперимента по минимизации функции Растригина 50 переменных методом НАИЗ без локального поиска, НАИЗ с локальным поиском по методу Хука-Дживса и методом Хука-Дживса.

Постановка задачи

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

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

$$\Phi(\vec{\mathbf{x}}): \Omega \to R, \tag{2}$$

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

Необходимо найти точку, в которой достигается её минимум:

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

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

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:

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

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:

$$\Psi(\varphi) = \vec{\mathbf{x}}_{\min} \text{ при } \varphi \le \Phi_{\min}. \tag{6}$$

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

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

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

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.

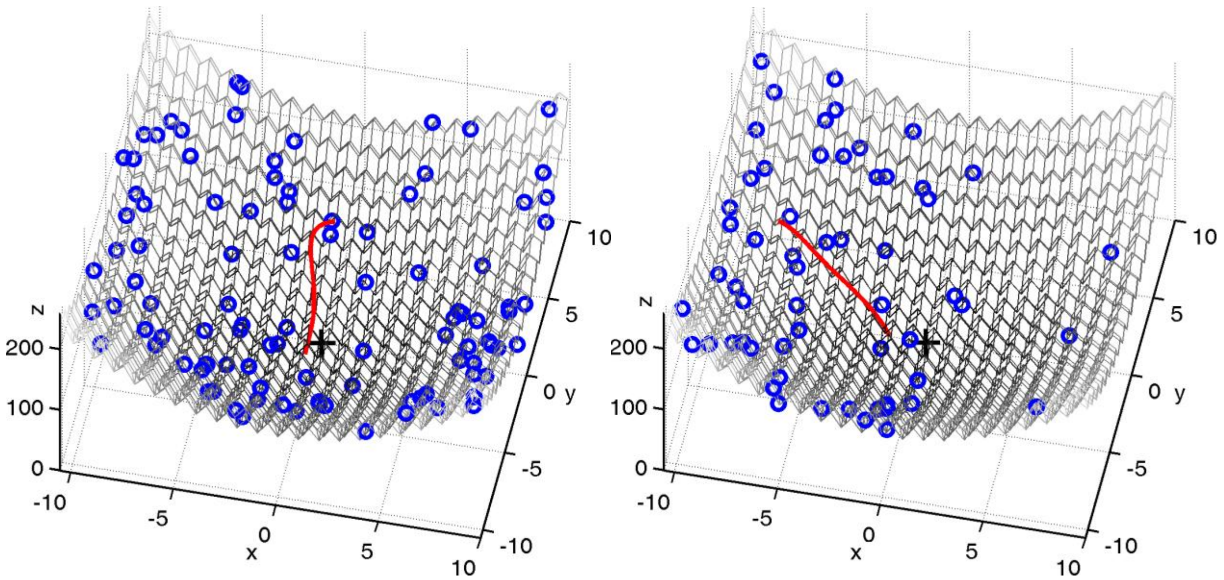
Обобщённо–регрессионные нейронные сети (GRNN) являются разновидностью радиальных базисных сетей. Их важным преимуществом перед другими сетями является высокая скорость обучения. Обычно они используются для решения задач обобщённой регрессии, анализа временных рядов и аппроксимации функций. Обученная на парах $\langle \vec{\mathbf{x}}, f(\vec{\mathbf{x}}) \rangle$ 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. Более того, ситуацию, когда глобальный минимум находится вне выпуклой оболочки, натянутой на пробные точки, невозможно разрешить с использованием только аппроксимации на основе GRNN.

Ход линии, аппроксимирующей инверсную зависимость для функции Растригина (9) двух переменных (глобальный минимум в точке (0; 0)), при равномерном и неравномерном покрытии показан соответственно на рис. 1, а и б. В первом случае линия предсказывает положение глобального минимума (–0,63; –0,63) с абсолютной погрешностью 0,89, во втором (–1,19; –0,26) с погрешностью1,22. Видно, что во втором случае линия тяготеет к области более высокой плотности расположения пробных точек.

$$f(\vec{\mathbf{x}}) = \sum_{i=1}^{n} \left[x_i^2 - 10\cos(2\pi x_i) + 10 \right]. \tag{9}$$



а б

Рис. 1. Линия аппроксимации инверсной зависимости для функции Растригина: а) при равномерном, б) при неравномерном расположении пробных точек (круглые маркеры) вокруг глобального минимума. Глобальный минимум обозначен крестиком.

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

Простейшей, необходимой, но имеющей низкую эффективность мерой является неинформированный поиск (пробные точки создаются без учёта поведения ЦФ). Для этого 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 N_g самостоятельных групп. В каждой группе выполняется отдельный экземпляр (гибридного) алгоритма НАИЗ. Когда один из алгоритмов останавливается или с заданной периодичностью, каждая группа сдвигается к той группе, в направлении которой ЦФ убывает быстрее. Если такого направления для данной группы нет, она сдвигается *от* той группы, в направлении которой ЦФ *возрастает* быстрее. Направление и скорость убывания/возрастания вычисляются между лучшими точками групп. Величина сдвига выбирается случайным образом в интервале от 0,5 до 1,5 расстояний до целевой группы.

Для преодоления ситуации, когда пробные точки оказались на «склоне» минимума, и процесс поиска с использованием аппроксимации инверсной зависимости зашёл в тупик, предлагается использовать 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».

Метод Хука-Дживса состоит из двух чередующихся стадий: исследующего поиска и поиска по образцу. В ходе исследующего поиска из базисной точки $\vec{\mathbf{b}}_i$ по *j*-й ($j = \overline{1,n}$) координате делаются шаги длиной $\pm h_j$, если значение ЦФ в новой точке меньше, то дальнейшие шаги (по координате $j+1$ и т. д.) делаются из неё. Если после просмотра всех координат значение ЦФ не уменьшилось, шаг уменьшается, и исследование повторяется. В противном случае выполняется поиск по образцу. Точка с пониженным значением ЦФ становится новой базисной точкой $\vec{\mathbf{b}}_{i+1}$, а исследование выполняется вокруг точки образца: $\vec{\mathbf{P}}_i = \vec{\mathbf{b}}_i + 2(\vec{\mathbf{b}}_{i+1} - \vec{\mathbf{b}}_i)$, то есть происходит движение в направлении $\vec{\mathbf{b}}_{i+1} - \vec{\mathbf{b}}_i$. Если исследование в точке образца потерпело неудачу, то выполняется исследование в базисной точке $\vec{\mathbf{b}}_{i+1}$, в результате которого могут быть получены новая базисная точка и новый образец. Процесс продолжается, пока шаг не станет меньше заданного значения.

В гибридном методе процедура Хука-Дживса запускается в конце итерации из точки с минимальным на данный момент значением ЦФ, если использование аппроксимации инверсной зависимости не дало результата. Если и метод Хука-Дживса не дал улучшения, тогда выполняется вбрасывание случайных точек.

Результаты экспериментов

В проведённом ранее эксперименте по сравнению метода НАИЗ (без локального поиска) с генетическим алгоритмом (GA) и методом оптимизации роем частиц (PSO) на наборе тестовых многоэкстремальных функций 2 и 50 переменных метод НАИЗ показал первый результат по точности и второй по количеству вычислений значений ЦФ. Однако функция Растригина 50 переменных не поддалась ни одному из методов.

Введение локального поиска методом Хука-Дживса позволило улучшить результат для этой функции: повысилась точность и уменьшилось количество вычислений значений ЦФ (рис. 2).

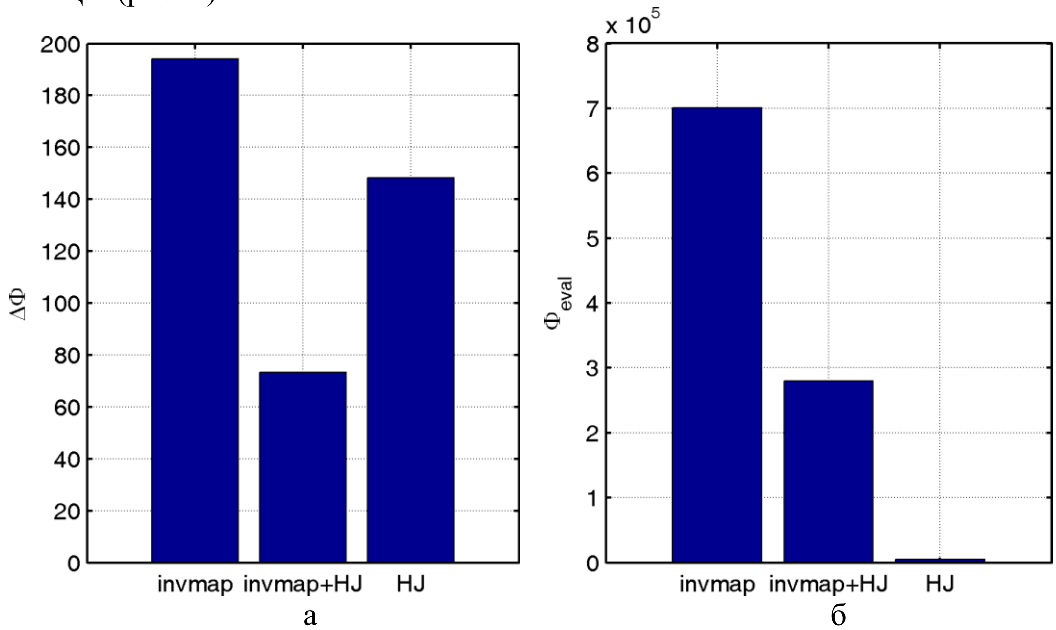


Рис. 2. Результаты вычислительного эксперимента по минимизации функции Растригина 50 переменных методом НАИЗ без локального поиска («invmap»), НАИЗ с локальным поиском по методу Хука-Дживса («invmap+HJ») и методом Хука-Дживса («HJ»): а) средние абсолютные погрешности итоговых значений ЦФ; б) средние количества вычислений значений ЦФ. Каждый метод испытывался 10 раз со случайной инициализацией.

Выводы

Совместное применение нейросетевых аппроксимаций инверсных зависимостей и локального поиска методом Хука-Дживса позволяет в некоторых случаях уменьшить погрешность нахождения глобального минимума, по сравнению с использованием каждого из этих подходов в отдельности. При этом среднее количество вычислений значений ЦФ у гибридного метода меньше, чем у чистого метода НАИЗ, но больше, чем у метода Хука-Дживса.

Можно выделить следующие направления дальнейших исследований:

1) оптимизация количества вычислений ЦФ;

2) уменьшение погрешности нахождения глобального минимума;

3) проверка эффективности гибридного метода для минимизации других многоэкстремальных функций многих переменных.