

МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ
ИНСТИТУТ

КАФЕДРА ДИСКРЕТНОЙ МАТЕМАТИКИ

Метод Нелдера — Мида

Работа студентки 2 курса, гр. 295
Ичаловой Дианы

Москва
2014 г.

1 Введение

Метод Нелдера—Мида — один из наиболее популярных методов безусловной оптимизации, применяемый в случае когда градиент функции неизвестен или отсутствует. Высокая скорость работы, простота реализации и возможность применения к негладким функциям позволяет использовать его во многих задачах, например, в оптимизации целевой функции в машинном обучении или в максимизации функции правдоподобия в математической статистике. Также некоторой вариацией метода является стандартный алгоритм `fminsearch` в MATLAB.

2 Постановка задачи

Пусть на множестве \mathbb{R}^n определена функция $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Задачей минимизации функции называется нахождение следующей величины

$$f_* = \min_{\mathbb{R}^n} f(x),$$

а также, если возможно, множества $X_* \subset \mathbb{R}^n$, на котором это значение достигается.

3 Описание

Впервые метод Нелдера—Мида был опубликован в 1965 году в [6]. Алгоритм является симплекс-методом, т. е. строит последовательность симплексов, сходящихся к точке минимума.

Перед началом работы выбираются коэффициенты, каждый из которых влияет на одну из фаз метода: ρ — коэффициент отражения, χ — коэффициент растяжения, γ — коэффициент сжатия, σ — коэффициент уменьшения.

Начальный симплекс выбирается исходя из условий задачи так, чтобы точки симплекса находились не очень далеко от предполагаемой точки минимума.

На каждой итерации алгоритма, кроме тех случаев, для которых выполняется шаг 6, точка с наибольшим значением функции в ней заменяется на новую по следующим правилам.

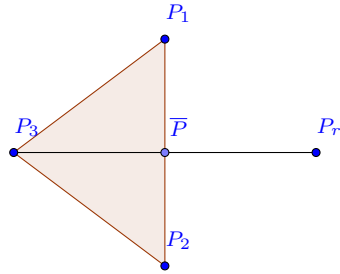
1. Упорядочим точки симплекса по возрастанию значения функции в них и обозначим P_1, P_2, \dots, P_{n+1} . Будем называть точку P_1 *наилучшей*, а точку P_{n+1} *наихудшей*.
2. Посчитаем центр масс всех точек, кроме наихудшей:

$$\bar{P} = \frac{\sum_{i=1}^n P_i}{n}$$

3. **Отражение.** Определим точку

$$P_r = \bar{P} + \rho(\bar{P} - P_{n+1}),$$

полученную отражением точки P_1 относительно гиперплоскости $P_1 P_2 \dots P_n$ вдоль прямой $P_{n+1} \bar{P}$.

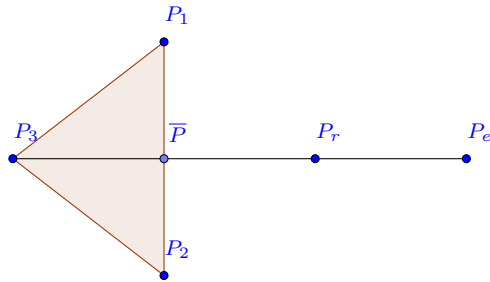


Если $f(P_1) \leq f(P_r) < f(P_n)$, то заменим худшую точку на P_r и завершим итерацию, если $f(P_r) < f(P_1)$, то перейдем к шагу 3, иначе перейдем к шагу 4.

4. **Растяжение.** Определим точку

$$P_e = \bar{P} + \chi(P_r - \bar{P}),$$

полученную растяжением отрезка $\bar{P}P_r$ вдоль прямой $P_{n+1}\bar{P}$ в χ раз.



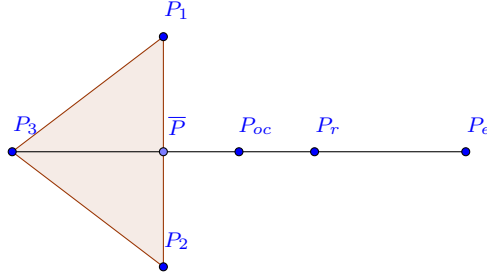
Если $f(P_e) < f(P_r)$, то заменим худшую точку на P_e , иначе — на P_r , и завершим итерацию.

5. **Сжатие.** Если $f(P_n) \leq f(P_r) < f(P_{n+1})$ то проведем сжатие наружу, иначе, если $f(P_r) \geq f(P_{n+1})$, — внутрь.

(а) **Сжатие наружу.** Определим точку

$$P_{oc} = \bar{P} + \gamma(P_r - \bar{P}),$$

полученную сжатием отрезка $\bar{P}P_r$ вдоль прямой $P_{n+1}\bar{P}$ в γ раз.

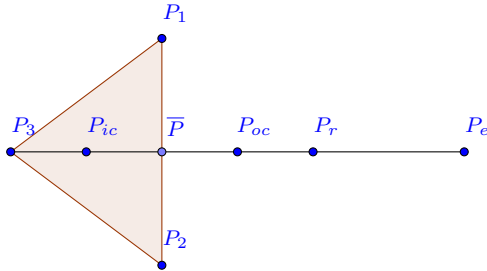


Если $f(P_{oc}) \leq f(P_r)$, то заменим худшую точку на P_{oc} и завершим итерацию, иначе перейдем к шагу 6.

(б) **Сжатие внутрь.** Определим точку

$$P_{ic} = \bar{P} - \gamma(P_r - \bar{P}),$$

полученную сжатием отрезка $P_{n+1}\bar{P}$ вдоль прямой $P_{n+1}\bar{P}$ в γ раз.



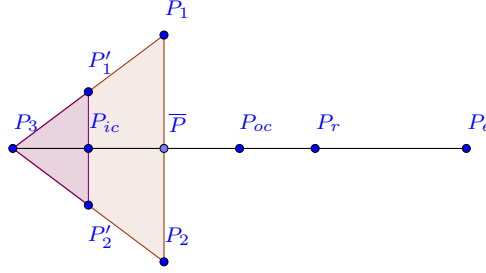
Если $f(P_{ic}) < f(P_{n+1})$, то заменим худшую точку на P_{ic} и завершим итерацию, иначе перейдем к шагу 6.

6. Уменьшение.

В этом случае определим новые точки

$$P'_i = P_1 + \sigma(P_i - P_1), i = \overline{1, n}.$$

Эти точки будут вершинами симплекса на следующей итерации.



4 Выбор коэффициентов

Из алгоритма следует, что коэффициенты должны удовлетворять следующим условиям:

$$\rho > 0, \chi > 1, \chi > \rho, 0 < \gamma < 1, 0 < \sigma < 1.$$

Стандартным выбором является следующий набор коэффициентов:

$$\rho = 1, \chi = 2, \gamma = \frac{1}{2}, \sigma = \frac{1}{2}.$$

Однако при больших размерностях данные коэффициенты неэффективны. В [1] описан выбор коэффициентов в зависимости от размерности задачи:

$$\rho = 1, \chi = 1 + \frac{2}{n}, \gamma = \frac{3}{4} - \frac{1}{2n}, \sigma = 1 - \frac{1}{n}.$$

5 Условие завершения

В качестве условия останова было выбрано следующее условие:

$$\sqrt{\frac{\sum_{i=1}^{n+1} (f(P_i) - f(\bar{P}))^2}{n+1}} < \varepsilon.$$

По умолчанию $\varepsilon = 10^{-6}$.

6 Сходимость

Несмотря на большую популярность метода, о его сходимости известно очень мало. На данный момент доказана одномерная сходимость при некоторых специальных условиях, а также известно, что при этих же условиях в \mathbb{R}^2 метод может расходиться.

6.1 Некоторые определения

Определение 1. Функция f называется *строго выпуклой*, если для любых $x_1 \neq x_2$ из области определения функции и для любого числа $t \in (0, 1)$ выполнено

$$f(tx_1 + (1-t)x_2) < tf(x_1) + (1-t)f(x_2).$$

Определение 2. Пусть определена функция $f : D \rightarrow \mathbb{R}$, $D \subset \mathbb{R}^n$. Множество

$$L_c(f) = \{x \in D \mid f(x) \leq c\}$$

называется (*нижним*) *уровневым множеством*.

Определение 3. Будем говорить, что функция $f : D \rightarrow \mathbb{R}$, $D \subset \mathbb{R}^n$ имеет *ограниченные уровневые множества*, если для любого c множество $L_c(f)$ ограничено.

6.2 Сходимость в \mathbb{R}^1

Теорема 1 ([2]). Пусть $f : \mathbb{R} \rightarrow \mathbb{R}$ — строго выпуклая функция с ограниченными уровневыми множествами. Пусть параметры метода удовлетворяют условиям: $\rho > 0$, $\chi > 1$, $\chi > \rho$, $\rho\chi \geq 1$, $0 < \gamma < 1$. Пусть также начальный симплекс Δ_0 невырожден.

Тогда точки симплексов $\{\Delta_k\}$ сходятся к точке глобального минимума.

Замечание. Из условия строгой выпуклости и ограниченности уровневых множеств следует, что точка минимума единственна.

Условие $\rho\chi \geq 1$ гарантирует, что полученная на шаге растяжения точка P_e лежит вне отрезка (симплекса).

Теорема 2 ([2]). Пусть $f : \mathbb{R} \rightarrow \mathbb{R}$ — строго выпуклая функция с ограниченными уровневыми множествами. Пусть параметры метода удовлетворяют условиям: $\rho = 1$, $\chi > 1$, $0 < \gamma < 1$. Пусть также начальный

симплекс Δ_0 невырожден.

Тогда существует число $M(\chi, \gamma)$ такое, что метод сходится линейно с шагом M (M-step linear convergence), т. е.

$$\overline{\lim}_{k \rightarrow \infty} \frac{|x_{k+M} - x_{\min}|}{|x_k - x_{\min}|} = q \in (0, 1).$$

6.3 Сходимость в \mathbb{R}^2

О двумерной сходимости известно, что с использованием стандартного набора коэффициентов при условии строгой выпуклости функции и ограниченности уровней множеств, диаметр симплексов $\{\Delta_k\}$ стремится к нулю. Однако этот факт не гарантирует сходимость. Более того, в [4] приведен общий вид семейства функций, на которых при некотором выборе начального симплекса метод не находит точку минимума. Например,

$$f(x, y) = \begin{cases} 360x^2 + y + y^2, & x \leq 0 \\ 6x^2 + y + y^2, & x > 0 \end{cases}.$$

Функция f строго выпукла и непрерывно дифференцируема, тем не менее, при $P_1^0 = (0, 0)$, $P_2^0 = (1, 1)$, $P_3^0 = (\frac{1+\sqrt{33}}{8}, \frac{1-\sqrt{33}}{8})$ метод сходится в точку P_1^0 , а не в точку минимума $(0, -\frac{1}{2})$, совершая бесконечно много сжатий внутрь.

В 2012 году [3] была доказана сходимость метода без шага растяжения при определенных условиях на функцию f .

Открытым остается вопрос, существует ли функция, на которой метод сходится всегда.

7 Достоинства метода

- Не использует градиент функции
- Высокая скорость сходимости на практике
- Простота реализации
- На каждой итерации вычисляется не более 4-х новых значений функции, поэтому метод можно применять, когда вычисление функции является ресурсозатратным

8 Недостатки метода

- Находит только локальный экстремум
- Неэффективен при больших размерностях
- Неисследованная сходимост

9 Детали реализации

Для реализации метода был выбран язык `Python` (версии 3.3) по следующим причинам:

- `Python` ближе к математическому языку, чем другие языки общего назначения
- `Python` предоставляет такие же возможности, как языки типа `MATLAB`, но позволяет просто использовать код в других задачах

Для большей точности вычислений использовался модуль `decimal`. Все графики в следующей части построены с помощью `gnuplot`.

10 Тесты

Существует много статей, в которых метод протестирован на различных функциях, где главным критерием является количество итераций и ошибка. Поэтому было решено протестировать метод на небольшом количестве функций, сделав упор на исследование зависимости поведения метода от выбора начального симплекса и выявление общих закономерностей работы метода.

Во всех запусках набор коэффициентов стандартный, $\varepsilon = 10^{-6}$.

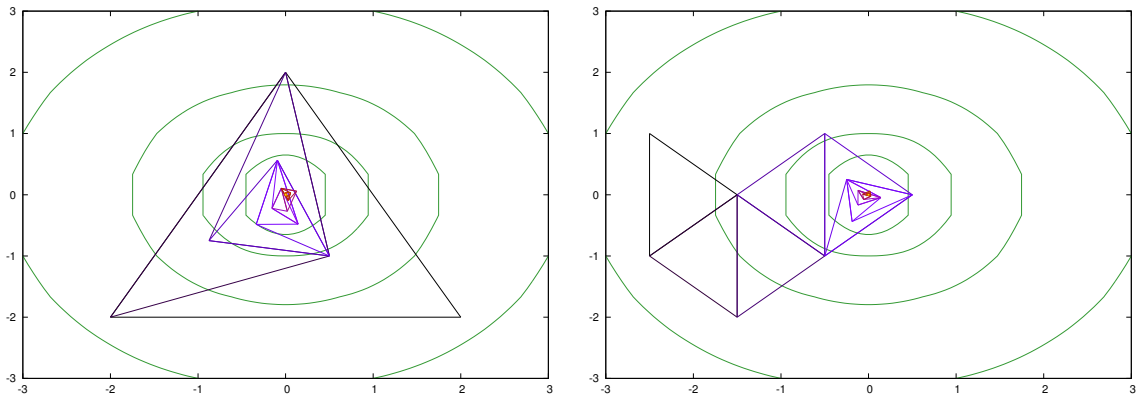
В таблице используются следующие обозначения: N_r — количество отражений, N_e — количество растяжений, N_{oc} — количество сжатий наружу, N_{ic} — количество сжатий внутрь, N_s — количество уменьшений, N — общее число итераций.

10.1 Сферическая функция

$$f(x_1, x_2) = x_1^2 + x_2^2$$

$$f_{min} = 0$$

$$x_{min} = (0, 0)$$



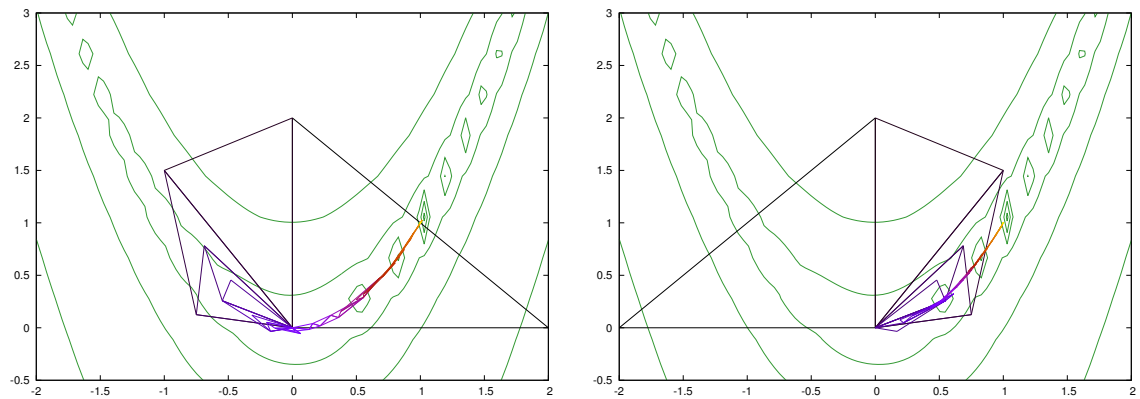
Начальный симплекс	N	N_r	N_e	N_{oc}	N_{ic}	N_s	$ x - x_{min} $
$(-2, -2), (2, -2), (0, 2)$	22	1	0	1	20	0	0.0014
$(-2,5, -1), (-2,5, 1), (-1,5, 0)$	27	8	0	1	18	0	0.0009

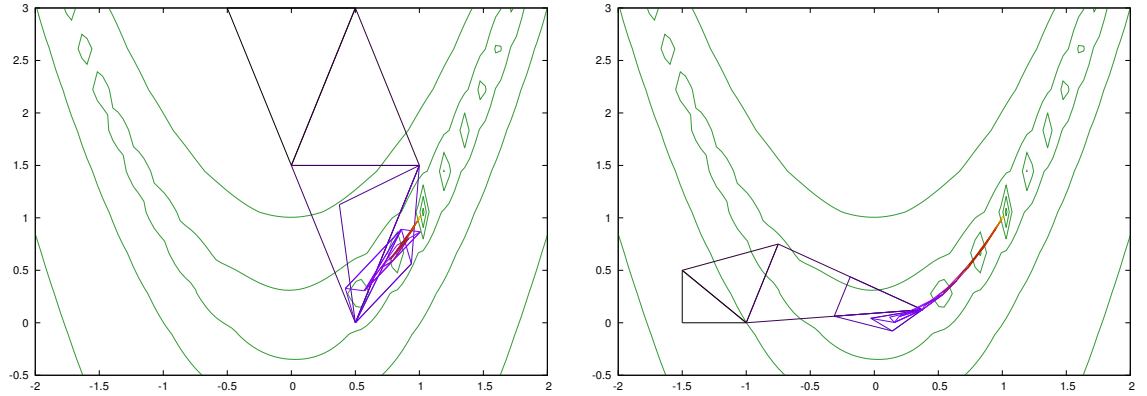
10.2 Функция Розенброка

$$f(x_1, x_2) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2$$

$$f_{min} = 0$$

$$x_{min} = (1, 1)$$





Начальный симплекс	N	N_r	N_e	N_{oc}	N_{ic}	N_s	$ x - x_{min} $
(0, 0), (0, 2), (2, 0)	49	10	8	6	25	0	0.0035
(0, 0), (0, 2), (-2, 0)	63	29	5	9	20	0	0.0016
(-0,5, 3), (0,5, 3), (0, 1,5)	31	11	0	2	17	1	0.0035
(-1.5, 0), (-1.5, 0,5), (-1, 0)	36	4	7	5	19	1	0.0033

10.3 Функция Химмельблау

$$f(x_1, x_2) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$$

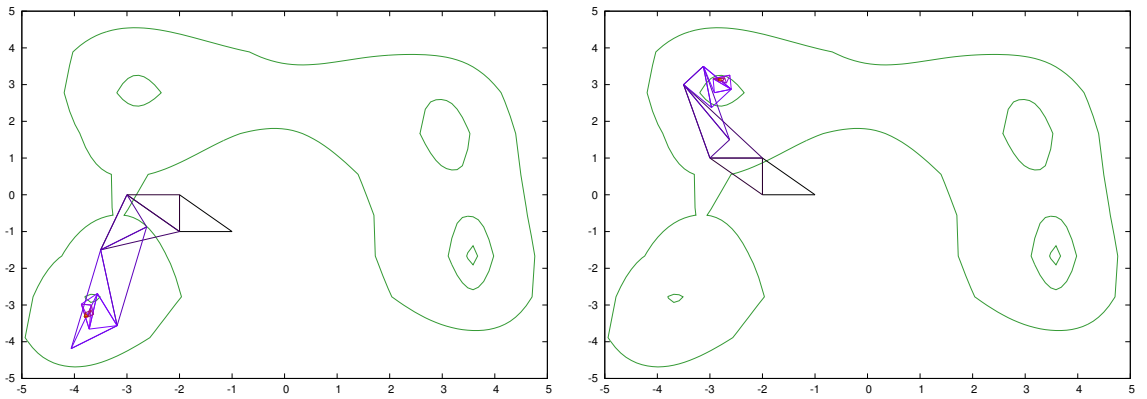
$$f_{min} = 0$$

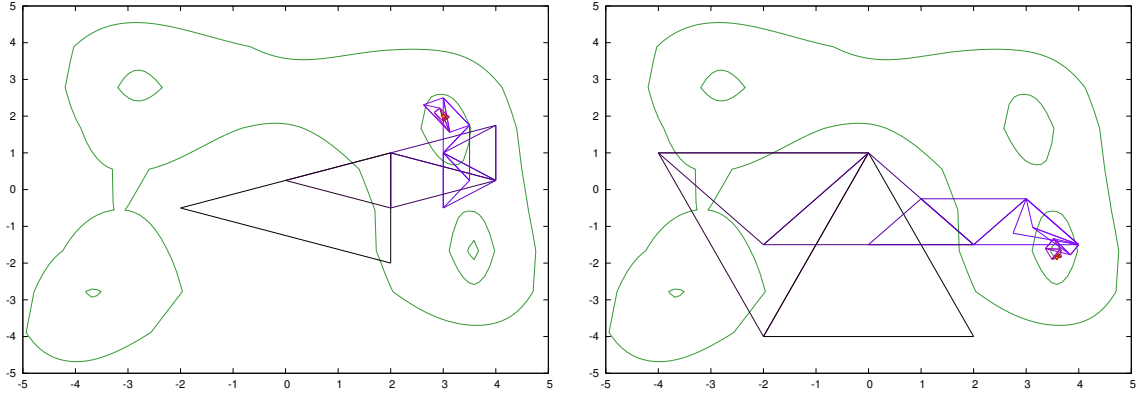
$$x_{min}^1 = (3,00, 2,00),$$

$$x_{min}^2 = (-2,81, 3,13),$$

$$x_{min}^3 = (-3,78, -3,28),$$

$$x_{min}^4 = (3,58, -1,85)$$





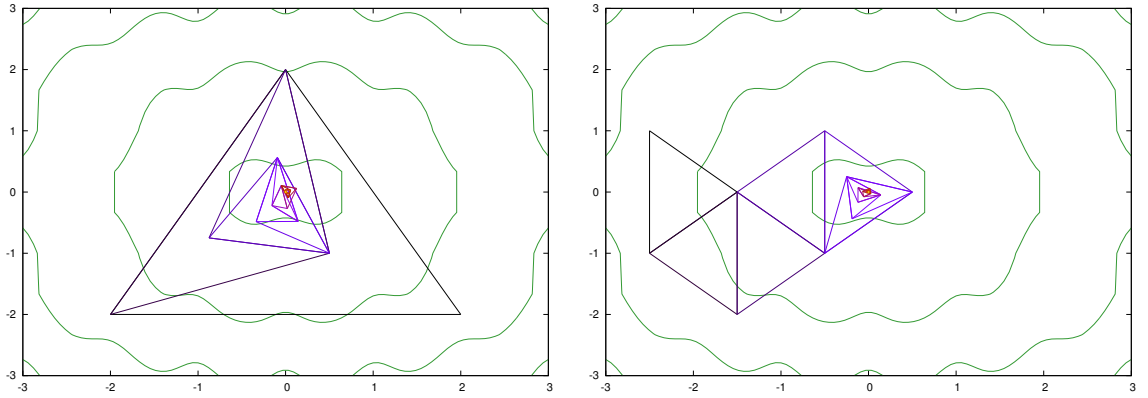
Начальный симплекс	N	N_r	N_e	N_{oc}	N_{ic}	N_s	$ x - x_{min} $
$(-2, -1), (-2, 0), (-1, -1)$	30	4	2	6	18	0	0.00035
$(-2, 0), (-2, 1), (-1, 0)$	31	6	1	3	21	0	0.00030
$(2, 1), (2, -2), (-2, -0,5)$	34	10	0	1	22	1	0.00036
$(-2, -4), (2, -4), (0, 1)$	35	10	0	4	20	1	0.00053

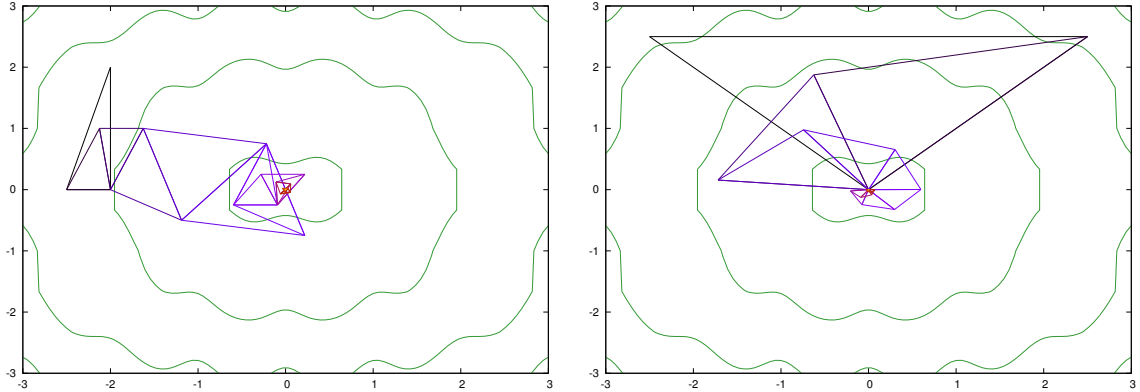
10.4 Функция Экли

$$f(x_1, x_2) = -20e^{-0,2\sqrt{0,5(x_1^2+x_2^2)}} - e^{0,5(\cos(2\pi x_1)+\cos(2\pi x_2))} + 20 + e$$

$$f_{min} = 0$$

$$x_{min} = (0, 0)$$





Начальный симплекс	N	N_r	N_e	N_{oc}	N_{ic}	N_s	$ x - x_{min} $
$(-2, -2), (2, -2), (0, 2)$	22	1	0	1	20	0	0.0014
$(-2.5, -1), (-2.5, 1), (-1.5, 0)$	27	8	0	1	18	0	0.0009
$(-2.5, 0), (-2, 0), (-2, 2)$	27	4	2	3	18	0	0.0008
$(0, 0), (-2.5, 2.5), (2.5, 2.5)$	23	0	0	18	5	0	0.0000

10.5 Несколько наблюдений

Во-первых, видно, что количество уменьшений N_s очень мало, что свидетельствует о том, что почти на каждом шаге происходит уменьшение наихудшего значения в точках симплекса. Во-вторых, отражения позволяют быстро подойти к окрестности точки минимума в случае, если начальный симплекс находится далеко от нее. В-третьих, во всех тестах было очень много сжатий внутрь. Они отвечают за уменьшение диаметра симплекса и стремление к одной точке. В целом, видно, что поведение метода довольно непредсказуемо, что затрудняет его анализ, но обеспечивает хорошую работу на практике.

Список литературы

- [1] Lixing Han and Fuchang Gao. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications*, 45:1–19, 2012.

- [2] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the Nelder—Mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9:112–147, 1998.
- [3] Jeffrey C. Lagarias, Bjorn Poonen, and Margaret H. Wright. Convergence of the restricted Nelder—Mead algorithm in two dimensions. *SIAM Journal on Optimization*, 22(2):501–532, 2012.
- [4] K. I. M. McKinnon. Convergence of the nelder—mead simplex method to a nonstationary point. *SIAM J. on Optimization*, 9(1):148–158, May 1998.
- [5] Jorge J. Moré, Burton S. Garbow, and Kenneth E. Hillstom. Testing unconstrained optimization software. *ACM Trans. Math. Softw.*, 7(1):17–41, 1981.
- [6] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, January 1965.