

1. ЛЕКЦИЯ. Безусловная оптимизация многих переменных. Методы нулевого порядка

Любое исследование основывается на применении научного метода, предоставляет научную информацию и теории для объяснения природы, и свойств окружающего мира. При этом не существует непреступных границ – и подходы, и методы могут переплетаться в едином исследовании. При этом научные подходы нацеливают на научные разработки и открытия, а методы помогают совершать их.

Когда наш знаменитый соотечественник Пётр Леонидович Капица работал в Великобритании, с ним произошла забавная история. Получив звание доктора и вступив в профсоюз научных работников, он подписал договор, по которому не имел права бесплатно консультировать промышленных клиентов. Более того, вознаграждение за консультацию не должно было опускаться ниже определенной суммы, соответствующей ученому званию. Как-то раз одна из фирм попросила Пётра Леонидовича, чей авторитет был очень высок, помочь устранить вибрацию новой модели турбины. Осмотрев ротор разобранной турбины и провернув его, Капица ударил куда-то молотком. Когда турбину вновь собрали и запустили, она уже не вибрировала. Благодарные клиенты спросили, сколько стоит оказанная услуга. Ответ был предельно четок: «1000 фунтов». Довольные клиенты не возражали, но спросили, что им лучше написать в финансовом отчете. «Напишите так — сказал знаменитый физик. — Удар молотком — 1 фунт; знание того, куда ударить, — 999 фунтов». Понимающие толк в юморе англичане согласились...

Итак, только благодаря надежным методам возникают новые плодотворные теории, а наша практическая деятельность становится более эффективной. Хороший специалист, вооруженный научной методологией, не только знает правильные ответы на сложные вчерашние вопросы и умеет самостоятельно находить грамотные ответы на сегодняшние, но и не растеряется перед непростыми завтрашними.

Мы с Вами прошли методы системного анализа и принятия решений в рамках соответствующих предметов. Из изученного материала можно сделать вывод, что данные методы базируются на комплексе математических методов, которые используются в условиях полной или неполной определенности об объекте управления.

В условиях полной определенности используются «методы математического программирования» в конечномерных или бесконечномерных

пространствах. Здесь мы с Вами прошли методы линейного программирования: графический, симплекс и двойственные задачи.

При условии неопределенности мы с вами рассматривали методы многокритериальной информации: оптимальность по Парето, «Электра» и анализа иерархий.

Методологическую основу этих методов составляет системный анализ как совокупность методологических средств, используемых для подготовки и обоснования решений по сложным проблемам различного характера.

Поэтому продолжим наши исследования оптимизационных методов и начнем с методов безусловной оптимизации, которые широко применяются в различных областях при нахождении минимумов функций и главное для нас это методы которые лежат в обучении нейронных сетей.

Задача многомерной оптимизации формулируется следующим образом: найти точку локального минимума $\bar{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$ целевой функции $f(\bar{x}) = f(x_1, x_2, \dots, x_n)$ на множестве допустимых значений $\bar{x} = (x_1, x_2, \dots, x_n) \in R^n$. Символически задачу записывают так:

$$f(\bar{x}^*) = \min f(\bar{x}), \quad \bar{x} \in R^n.$$

Все методы нахождения точки локального минимума основаны на *итерационной процедуре*, реализуемой в соответствии с формулой

$$\bar{x}^{(k+1)} = \bar{x}^{(k)} + h_k \bar{p}^{(k)}, \quad (2.1)$$

где k – номер итерации $k = 0, 1, \dots$;

$\bar{x}^{(k)}$ – текущее приближение;

h_k – величина шага;

\bar{p}^k – вектор, определяющий направление убывания функции $f(\bar{x})$.

Переходы от точек $\bar{x}^{(k)}$ к точкам $\bar{x}^{(k+1)}$ выполняются таким образом, чтобы обеспечивалось убывание значения целевой функции $f(\bar{x}^{(k+1)}) \leq f(\bar{x}^{(k)})$. Такие методы принято называть *методами спуска*. Название метода спуска определяется способом выбора вектора $\bar{p}^{(k)}$, а его варианты связываются с различными способами выбора величины шага h_k .

Методы решения задач безусловной оптимизации принято разделять на группы: *нулевого, первого и второго порядков*, в зависимости от уровня используемой в методе информации о целевой функции.

Методы нулевого порядка

Методы нулевого порядка используют информацию только о значениях целевой функции $f(\bar{x})$. Направление минимизации в данном случае полностью

определяется последовательными вычислениями значений целевой функции. Ниже рассматриваются методы, довольно часто применяемые на практике: *метод Хука–Дживса*, *симплексный метод*, *метод Нелдера–Мида* (деформируемого многогранника). Основное достоинство этих методов состоит в том, что они не требуют непрерывности целевой функции и существования производных.

Симплексный метод

Регулярным симплексом в n -мерном пространстве называется правильный многогранник с $n+1$ вершиной. При $n = 2$ симплексом является правильный треугольник, при $n = 3$ – тетраэдр и т.д. Отрезок, соединяющий 2 вершины симплекса, называется *ребром симплекса*.

Поиск симплексным методом ведется по следующей схеме. Устанавливаются координаты вершин симплексов. Определяется вершина с наибольшим значением целевой функции. Вместо нее строится новая вершина отражением старой через центр тяжести остальных вершин симплекса. На рисунке 1 представлен процесс построения нового симплекса на плоскости.

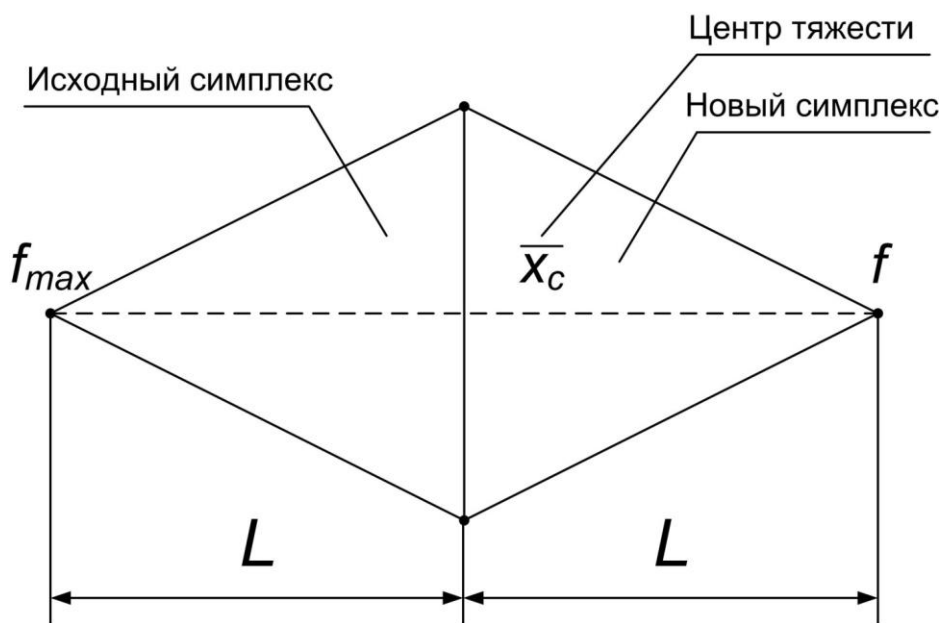


Рис 1 Построение нового симплекса

Если попытка отражения не приводит к уменьшению целевой функции, то выполняется операция редукции, в результате которой формируется новый симплекс с уменьшенными вдвое сторонами. При операции редукции в качестве базовой точки выбирается вершина старого симплекса, в которой функция принимает минимальное значение.

В результате исключения вершин симплексов с наибольшим значением целевой функции процесс поиска сходится к минимальному значению.

Поиск завершается, когда разности между значениями функции в центре тяжести симплекса и вершинах становится достаточно малым.

На рисунке 2 представлена иллюстрация построения симплекса для двумерной области.

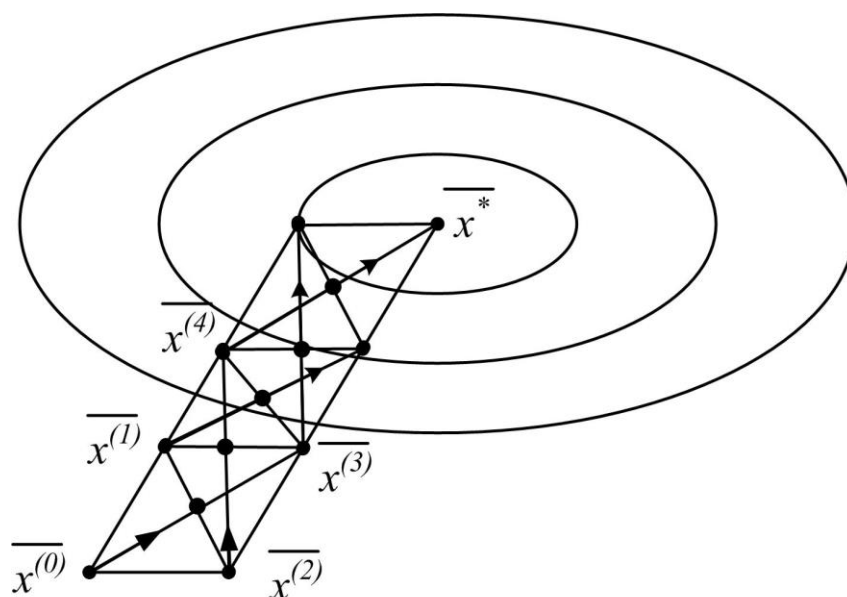


Рис. 2. Графическая иллюстрация поиска точки минимума симплексным методом

Объяснение рисунка.2

Пусть требуется решить задачу:

$$f(\bar{x}^*) = \min f(\bar{x}), \quad \bar{x} \in R^n$$

В двумерном пространстве R^2 решению такой задачи можно дать геометрическую иллюстрацию. Пусть точка $x = (x_1, x_2)$ лежит на плоскости Ox_1x_2 . Введем третью координату x_3 так, чтобы ось координат Ox_3 была перпендикулярна плоскости Ox_1x_2 (рис.3). Уравнению $x_3 = f(x_1, x_2)$ соответствует поверхность в трехмерном пространстве.

Если функция $f(x)$ достигает локального минимума в точке $x^* \in R^2$, то поверхность в некоторой окрестности точки x^* имеет форму чаши (рис.3).

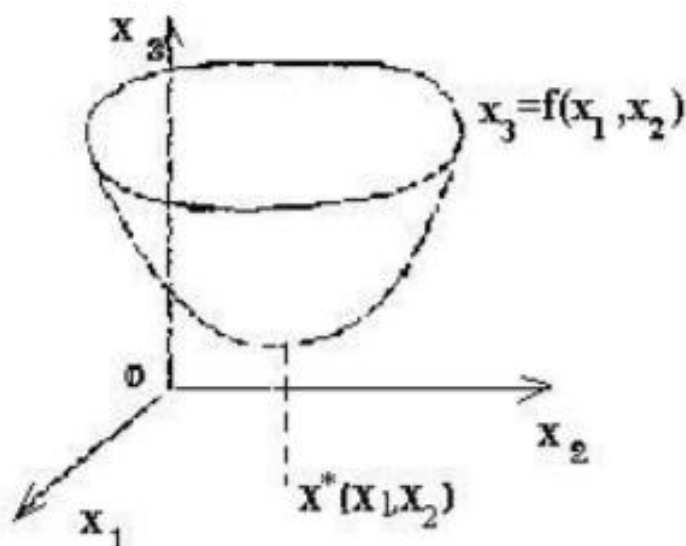


Рис. 3.

Линиями уровня функции $f(x_1, x_2)$ называют семейство линий плоскости R^2 , на которых функция принимает постоянное значение. Неявным уравнением линии уровня является уравнение $f(x_1, x_2) = C$. Если функция $f(x)$ имеет в R^2 единственную точку локального минимума $x^*(x_1^*, x_2^*)$, то такая функция называется мономодальной. Взаимное расположение ее линий уровня имеет вид, изображенный на рис. 4.

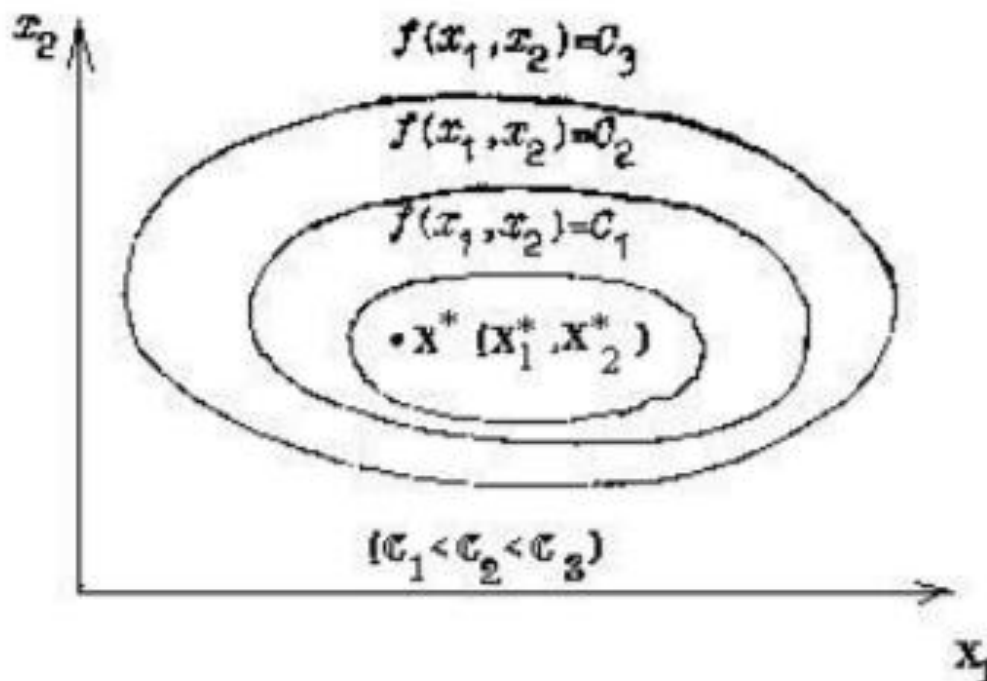


Рис. 4.

Мультимодальными называются функции, которые имеют более одного экстремума. Такова, например, функция Химмельблау

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

имеющая четыре изолированные точки минимума.

Продолжим.

Алгоритм поиска симплексным методом

1. Задать размерность задачи оптимизации n , координаты начальной точки симплекса $\overline{x^{(0)}} = \{x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}\}$, длину ребра симплекса m , точность поиска ε .

2. Вычислить координаты остальных n вершин симплекса $\overline{x^{(1)}}$, $\overline{x^{(2)}}$, ..., $\overline{x^{(n)}}$ по формулам

$$\overline{x^{(i)}} = \begin{cases} x_j^{(0)} + \delta_1, & \text{если } i = j \\ x_j^{(0)} + \delta_2, & \text{если } i \neq j \end{cases} \quad \text{для } i, j = \overline{1, n}$$

где приращения δ_1 и δ_2 определяются по формулам

$$\delta_1 = \left(\frac{\sqrt{n+1}-1}{n\sqrt{2}} \right) * m, \quad \delta_2 = \left(\frac{\sqrt{n+1}+n-1}{n\sqrt{2}} \right) * m.$$

3. Определить номер вершины k с наибольшим значением целевой функции $f_{\max} = f(\overline{x^{(k)}})$.

4. Определить центр тяжести всех вершин многогранника за исключением вершины $\overline{x^{(k)}}$:

$$\overline{x_c} = \frac{1}{n} \sum_{\substack{i=0 \\ i \neq k}}^n \overline{x^{(i)}}.$$

5. Отразить вершину $\overline{x^{(k)}}$ относительно центра тяжести $\overline{x} = 2\overline{x_c} - \overline{x^{(k)}}$. Вычислить значение целевой функции в отраженной точке $f(\overline{x})$ и перейти к пункту 6.

6. Проверить условие. Если $f(\overline{x}) < f(\overline{x^{(k)}})$, то операция закончилась успешно. Положить $\overline{x^{(k)}} = \overline{x}$, $f(\overline{x^{(k)}}) = f(\overline{x})$ и перейти к пункту 8. В противном случае перейти к пункту 7.

7. Выполнить операцию редукции. Для этого отделить номер вершины r с минимальным значением целевой функции $f_{\min} = f(\overline{x^{(r)}})$. Используя соотношение $\overline{x^{(i)}} = \overline{x^{(r)}} + 0,5(\overline{x^{(i)}} - \overline{x^{(r)}})$, $i = \overline{0, n}$, $i \neq r$, сформировать новый многогранник с уменьшенными вдвое сторонами. Перейти к шагу 8.

8. Определить центр тяжести симплекса $\overline{x_c} = \frac{1}{n+1} \sum_{i=0}^n \overline{x^{(i)}}$ значение функции в этой точке $f(\overline{x_c})$.

9. Проверить условие окончания процесса вычислений. Если $|f(\overline{x^{(i)}}) -$

$f(\bar{x}_c) \mid < \varepsilon, i = \overline{1, n}$, то процесс вычислений завершен. В качестве приближенного решения принять вершину с минимальным значением целевой функции. В противном случае перейти к пункту 3.

Информация о симплекс формируется и хранится в двумерном массиве размером $(n+1) \times (n+1)$. При этом каждая i -я строка массива ($i = \overline{0, n}$) содержит информацию о вершине симплекса $\bar{x}^{(i)} (x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)})$ и значении целевой функции в этой вершине $f(\bar{x}^{(i)})$ (таблица 1).

При успешном отражении информация в массиве о старой вершине симплекса аннулируется, и на ее место записывается информация о новой вершине. Таким образом, размер массива в процессе итераций остается неизменным и характеризует текущий симплекс.

Таблица 1

		Координаты вершины				Значение функции
		1	2	...	n	
Номер вершины	0	$x_1^{(0)}$	$x_2^{(0)}$...	$x_n^{(0)}$	$f(\bar{x}^{(0)})$
	1	$x_1^{(1)}$	$x_2^{(1)}$...	$x_n^{(1)}$	$f(\bar{x}^{(1)})$

	n	$x_1^{(2)}$	$x_2^{(2)}$...	$x_n^{(2)}$	$f(\bar{x}^{(n)})$

Пример: Найти минимум целевой функции.

$$f(\bar{x}) = x_1^2 - x_1 x_2 + 3x_2^2 - x_1$$

Симплексным методом с точностью $\varepsilon = 0,1$.

Размерность задачи $n = 2$.

Решение. Зададим начальную точку симплекса $\bar{x}^{(0)} = (x_1^{(0)}, x_2^{(0)})^T = (0, 0)^T$ и длину ребра симплекса $m = 0,25$.

Можем сразу для понимания отобразить таблицу.

		Координаты вершины		Значение функции
		1	2	
Номер вершины	0	$x_1^{(0)} = 0$	$x_2^{(0)} = 0$	$f(\bar{x}^{(0)}) = 0$

Вычислим приращения

$$\delta_1 = \left(\frac{\sqrt{n+1}-1}{n\sqrt{2}} \right) * m = \left(\frac{\sqrt{3}-1}{2\sqrt{2}} \right) * 0,25 = 0,065,$$

$$\delta_2 = \left(\frac{\sqrt{n+1}+n-1}{n\sqrt{2}} \right) * m = \left(\frac{\sqrt{3}+2-1}{2\sqrt{2}} \right) * 0,25 = 0,241.$$

Используя δ_1 и δ_2 , вычислим координаты двух остальных вершин симплекса

$$\overline{x^{(1)}} = (x_1^{(0)} + \delta_1, x_2^{(0)} + \delta_2)^T = (0 + 0,065; 0 + 0,241)^T = (0,065; 0,241)^T$$

$$\overline{x^{(2)}} = (x_1^{(0)} + \delta_2, x_2^{(0)} + \delta_1)^T = (0 + 0,241; 0 + 0,065)^T = (0,241; 0,065)^T$$

Итерация $k = 0$. Вычислим значение целевой функции $f(\bar{x})$ в вершинах $\overline{x^{(0)}}$, $\overline{x^{(1)}}$, $\overline{x^{(2)}}$:

$$f(\overline{x^{(0)}}) = 0, f(\overline{x^{(1)}}) = 0,099, f(\overline{x^{(2)}}) = -0,186.$$

Отобразим эти значения в таблице

		Координаты вершины		Значение функции
		1	2	
Номер вершины	0	$x_1^{(0)} = 0$	$x_2^{(0)} = 0$	$f(\overline{x^{(0)}}) = 0$
	1	$x_1^{(1)} = 0,065$	$x_2^{(1)} = 0,241$	$f(\overline{x^{(1)}}) = 0,099$
	2	$x_1^{(2)} = 0,241$	$x_2^{(2)} = 0,065$	$f(\overline{x^{(2)}}) = -0,186$

Наибольшее значение целевой функции соответствует вершине $\overline{x^{(1)}}$, поэтому необходимо отразить ее относительно центра тяжести остальных вершин $\overline{x^{(0)}}$ и $\overline{x^{(2)}}$. центр тяжести расположен в точке

$$\overline{x_{c1}} = \frac{1}{2}(x_1^{(0)} + x_1^{(2)}, x_2^{(0)} + x_2^{(2)})^T = (0,121; 0 + 0,032)^T.$$

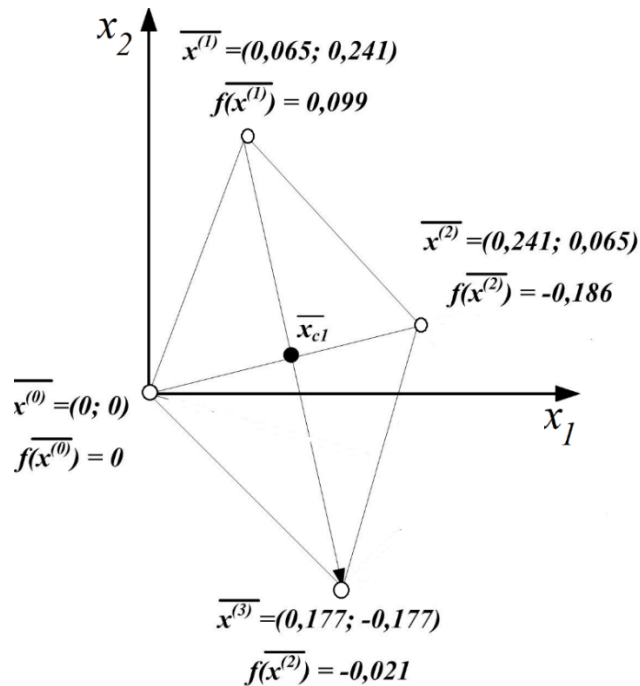
Используя свойство регулярности, найдем координаты отраженной вершины

$$\overline{x^{(3)}} = 2\overline{x_{c1}} - \overline{x^{(1)}} = 2(0,121; 0 + 0,032)^T - (0,065; 0,241)^T = (0,177; -0,177)^T.$$

В полученной вершине значение целевой функции $f(\overline{x^{(3)}}) = -0,021$.

Таким образом, таблица дополняется

		Координаты вершины		Значение функции
		1	2	n+1
Номер вершины	0	$x_1^{(0)} = 0$	$x_2^{(0)} = 0$	$f(\overline{x^{(0)}}) = 0$
	1	$x_1^{(1)} = 0,065$	$x_2^{(1)} = 0,241$	$f(\overline{x^{(1)}}) = 0,099$
	2	$x_1^{(2)} = 0,241$	$x_2^{(2)} = 0,065$	$f(\overline{x^{(2)}}) = -0,186$
	3	$x_1^{(3)} = 0,177$	$x_2^{(3)} = -0,177$	$f(\overline{x^{(3)}}) = -0,021$



Следовательно, наблюдается уменьшение целевой функции $f(\bar{x}^{(3)}) < f(\bar{x}^{(1)})$. Новый симплекс образован вершинами $\bar{x}^{(0)}$, $\bar{x}^{(2)}$, $\bar{x}^{(3)}$, которым соответствует значение целевой функции $f(\bar{x}^{(0)}) = 0$, $f(\bar{x}^{(2)}) = -0,186$, $f(\bar{x}^{(3)}) = -0,021$.

Проверим условие окончания поиска $|f(\bar{x}^{(i)}) - f(\bar{x}_c)| < \varepsilon$, $i = 0,2,3$.
Определим центр тяжести симплекса

$$\bar{x}_c = \frac{1}{3}(\bar{x}^{(0)} + \bar{x}^{(2)} + \bar{x}^{(3)})^T = \frac{1}{3}(x_1^{(0)} + x_1^{(2)} + x_1^{(3)}; x_2^{(0)} + x_2^{(2)} + x_2^{(3)})^T = (0,139; -0,037)^T.$$

В полученной точке $f(\bar{x}_c) = -0,111$. Вычислим $|f(\bar{x}^{(0)}) - f(\bar{x}_c)| = 0,111 > \varepsilon$, $|f(\bar{x}^{(2)}) - f(\bar{x}_c)| = 0,075 < \varepsilon$, $|f(\bar{x}^{(3)}) - f(\bar{x}_c)| = 0,090 < \varepsilon$.

Так как все условия окончания поиска не выполняются, то процесс итераций должен быть продолжен.

Итерация $k = 1$. Вершину $\bar{x}^{(0)}$, которой соответствует наибольшее значение целевой функции в новом симплексе ($\bar{x}^{(0)}$, $\bar{x}^{(2)}$ и $\bar{x}^{(3)}$), необходимо отразить относительно центра тяжести вершин $\bar{x}^{(2)}$ и $\bar{x}^{(3)}$. Центр тяжести расположен в точке

$$\bar{x}_{c2} = \frac{1}{2}(x_1^{(2)} + x_1^{(3)}; x_2^{(2)} + x_2^{(3)})^T = (0,209; -0,056)^T.$$

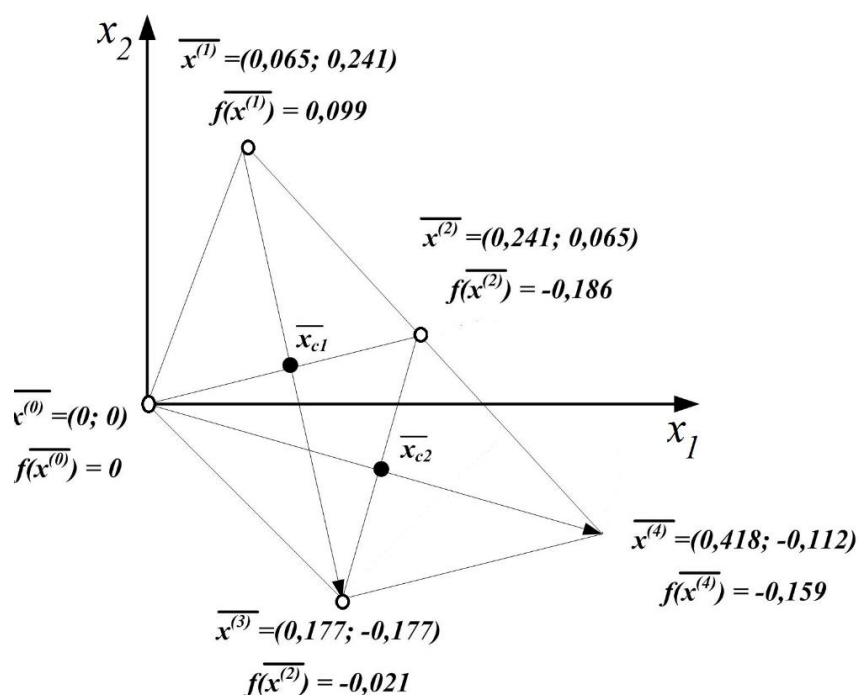
Используя свойство регулярности, найдем координаты отраженной вершины

$$\bar{x}^{(4)} = 2\bar{x}_{c2} - \bar{x}^{(0)} = 2(0,209; -0,056)^T - (0; 0)^T = (0,418; -0,112)^T.$$

В полученной вершине значение целевой функции $f(\bar{x}^{(4)}) = -0,159$.

		Координаты вершины		Значение функции
		1	2	n+1
Номер вершины	0	$x_1^{(0)} = 0$	$x_2^{(0)} = 0$	$f(\bar{x}^{(0)}) = 0$
	1	$x_1^{(1)} = 0,065$	$x_2^{(1)} = 0,241$	$f(\bar{x}^{(1)}) = 0,099$
	2	$x_1^{(2)} = 0,241$	$x_2^{(2)} = 0,065$	$f(\bar{x}^{(2)}) = -0,186$
	3	$x_1^{(3)} = 0,177$	$x_2^{(3)} = -0,177$	$f(\bar{x}^{(3)}) = -0,021$
	4	$x_1^{(4)} = 0,418$	$x_2^{(4)} = -0,112$	$f(\bar{x}^{(4)}) = -0,159$

Следовательно, наблюдается уменьшение целевой функции $f(\bar{x}^{(4)}) < f(\bar{x}^{(0)})$. Новый симплекс образован вершинами $\bar{x}^{(2)}$, $\bar{x}^{(3)}$, $\bar{x}^{(4)}$, которым соответствует значение целевой функции $f(\bar{x}^{(2)}) = -0,186$, $f(\bar{x}^{(3)}) = -0,021$, $f(\bar{x}^{(4)}) = -0,159$.



Проверим условие окончания поиска $|f(\bar{x}^{(i)}) - f(\bar{x}_c)| < \varepsilon$, $i = 2,3,4$.

Определим центр тяжести симплекса

$$\bar{x}_c = \frac{1}{3}(\bar{x}^{(2)} + \bar{x}^{(3)} + \bar{x}^{(4)})^T = \frac{1}{3}(x_1^{(2)} + x_1^{(3)} + x_1^{(4)}; x_2^{(2)} + x_2^{(3)} + x_2^{(4)})^T = (0,279; -0,075)^T.$$

В полученной точке $f(\bar{x}_c) = -0,164$. Вычислим $|f(\bar{x}^{(2)}) - f(\bar{x}_c)| = 0,022 < \varepsilon$, $|f(\bar{x}^{(3)}) - f(\bar{x}_c)| = 0,143 > \varepsilon$, $|f(\bar{x}^{(4)}) - f(\bar{x}_c)| = 0,005 < \varepsilon$.

Так как все условия окончания поиска не выполняются, то процесс итераций должен быть продолжен.

Итерация $k = 2$. Вершину $\overline{x^{(3)}}$, которой соответствует наибольшее значение целевой функции, необходимо отразить относительно центра тяжести вершин $\overline{x^{(2)}}$ и $\overline{x^{(4)}}$. Центр тяжести расположен в точке

$$\overline{x_{c3}} = \frac{1}{2}(\overline{x_1^{(2)}} + \overline{x_1^{(4)}}, \overline{x_2^{(2)}} + \overline{x_2^{(4)}})^T = (0,330; -0,025)^T.$$

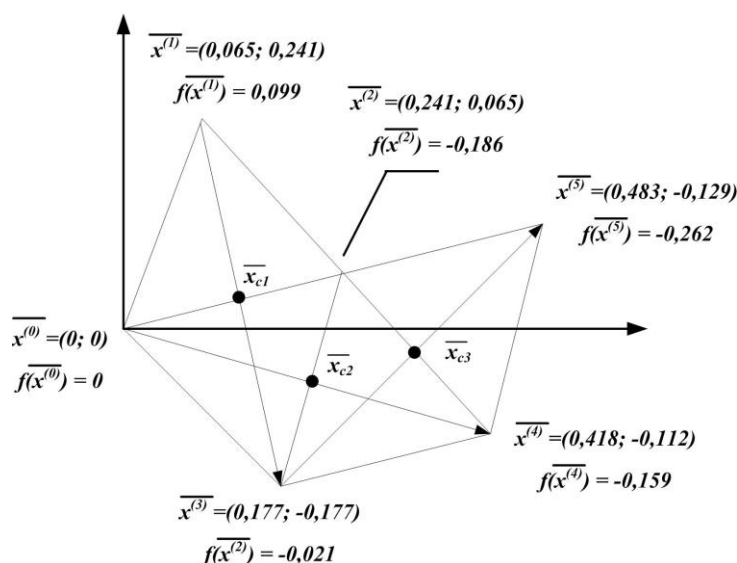
Используя свойство регулярности, найдем координаты отраженной вершины

$$\overline{x^{(5)}} = 2\overline{x_{c3}} - \overline{x^{(3)}} = 2(0,330; -0,025)^T - (0,177; -0,177)^T = (0,483; -0,129)^T.$$

В полученной вершине значение целевой функции $f(\overline{x^{(5)}}) = -0,262$.

		Координаты вершины		Значение функции
		1	2	n+1
Номер вершины	0	$\overline{x_1^{(0)}} = 0$	$\overline{x_2^{(0)}} = 0$	$f(\overline{x^{(0)}}) = 0$
	1	$\overline{x_1^{(1)}} = 0,065$	$\overline{x_2^{(1)}} = 0,241$	$f(\overline{x^{(1)}}) = 0,099$
	2	$\overline{x_1^{(2)}} = 0,241$	$\overline{x_2^{(2)}} = 0,065$	$f(\overline{x^{(2)}}) = -0,186$
	3	$\overline{x_1^{(3)}} = 0,177$	$\overline{x_2^{(3)}} = -0,177$	$f(\overline{x^{(3)}}) = -0,021$
	4	$\overline{x_1^{(4)}} = 0,418$	$\overline{x_2^{(4)}} = -0,112$	$f(\overline{x^{(4)}}) = -0,159$
	5	$\overline{x_1^{(5)}} = 0,483$	$\overline{x_2^{(5)}} = -0,129$	$f(\overline{x^{(5)}}) = -0,262$

Следовательно, наблюдается уменьшение целевой функции $f(\overline{x^{(5)}}) < f(\overline{x^{(3)}})$. Новый симплекс образован вершинами $\overline{x^{(2)}}$, $\overline{x^{(4)}}$, $\overline{x^{(5)}}$, которым соответствует значение целевой функции $f(\overline{x^{(2)}}) = -0,186$, $f(\overline{x^{(4)}}) = -0,159$, $f(\overline{x^{(5)}}) = -0,262$.



Проверим условие окончания поиска $|f(\overline{x^{(i)}}) - f(\overline{x_c})| < \varepsilon$, $i = 2, 4, 5$.

Определим центр тяжести симплекса

$$\overline{x_c} = \frac{1}{3}(\overline{x^{(2)}} + \overline{x^{(4)}} + \overline{x^{(5)}})^T = \frac{1}{3}(x_1^{(2)} + x_1^{(4)} + x_1^{(5)}; x_2^{(2)} + x_2^{(4)} + x_2^{(5)})^T = (0,381; -0,027)^T.$$

В полученной точке $f(\overline{x_c}) = -0,244$. Вычислим $|f(\overline{x^{(2)}}) - f(\overline{x_c})| = 0,085 < \varepsilon$,
 $|f(\overline{x^{(4)}}) - f(\overline{x_c})| = 0,018 < \varepsilon$, $|f(\overline{x^{(5)}}) - f(\overline{x_c})| = 0,058 < \varepsilon$.

Так как все условия окончания поиска выполняются, то процесс итерации завершен.

В качестве приближенного решения $\overline{x^*}$ выбирается $\overline{x^{(5)}} = (0,483; 0,129)^T$, которой соответствует наименьшее значение целевой функции $f(\overline{x^{(5)}}) = -0,262$.