4 Метод сопряженных направлений (Пауэлла)

Определение: векторы $p_1, p_2 \ldots p_n$ называются векторами, сопряженными относительно матрицы A, если

$$p_i^T A p_j = 0$$

В методе сопряженных направлений применяется итерационная формула метода Гаусса-Зейделя в следующем виде:

Положим $\bar{X}_0^{r+1} = \bar{X}^r$ и пусть $e_1, e_2 \ldots e_n$ -- орты используемой системы координат.

Тогда итерационную формулу метода Гаусса-Зейделя можно записать в виде

$$\bar{X}_i^{r+1} = \bar{X}_{i-1}^{r+1} + \lambda_i^r e_i = \bar{X}_{i-1}^{r+1} + \lambda_i^r p_i^r, \text{ i=1,2,…n}$$

где коэффициенты λ_i^r находятся из условий:

, i=1,2,…n

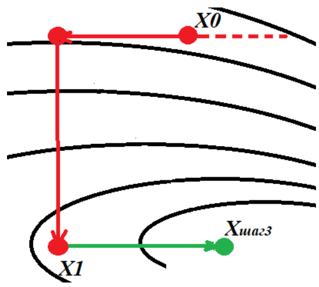
$$\min_{\lambda \in (-\infty;+\infty)} \Phi\left(\bar{X}_{i-1}^{r+1} + \lambda e_i\right) = \Phi\left(\bar{X}_{i-1}^{r+1} + \lambda_i^r e_i\right) = \Phi\left(\bar{X}_1^{r+1}\right)$$

Схема метода сопряженных направлений

1. Задаем начальную точку $\bar{X}^0$ и полагаем r =0, i =1.
2. Последовательно для i =1,2,..., находим точки $\bar{X}_1^{r+1}, \bar{X}_2^{r+1} \ldots \bar{X}_n^{r+1}$
3. Исходя из точки $\bar{X}_n^{r+1}$ еще раз находим минимум функции Ф вдоль первого координатного направления – вычисляем координаты точки $\bar{X}_{n+1}^{r+1} = \bar{X}_n^{r+1} + \lambda_1^r e_1$

где коэффициент λ_1^r находится из условия:

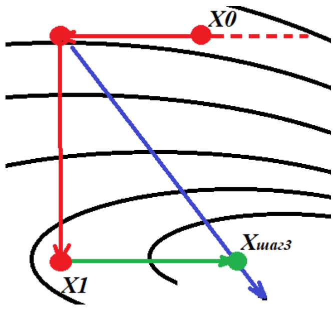
$$\min_{\lambda \in (-\infty;+\infty)} \Phi\left(\bar{X}_{n+1}^{r+1} + \lambda e_1\right) = \Phi\left(\bar{X}_n^{r+1} + \lambda_1^r e_1\right) = \Phi\left(\bar{X}_{n+1}^{r+1}\right)$$



4. Исходя из точки $\bar{X}_1^{r+1}$, находим минимум функции Ф вдоль вектора $p_{n+1}^r = \bar{X}_{n+1}^{r+1} + \bar{X}_1^{r+1}$ вычисляем

$\bar{X}^{r+1} = \bar{X}_1^1 + \lambda_{n+1}^{r+1} p_{n+1}^r$, где коэффициент λ находится из условия:

$$\min_{\lambda \in (-\infty;+\infty)} \Phi\left(\bar{X}_1^{r+1} + \lambda p_{n+1}^r\right) = \Phi\left(\bar{X}^r + \lambda_{n+1}^r p_{n+1}^r\right) = \Phi\left(\bar{X}^{r+1}\right)$$

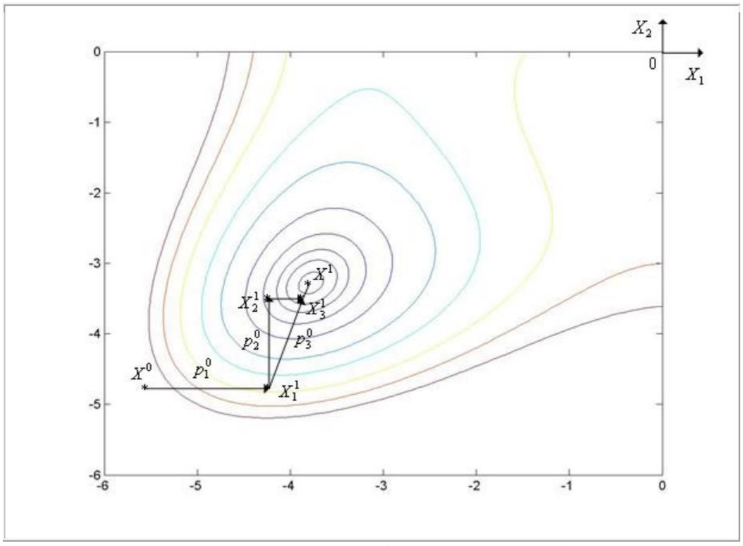


5. Если одно из стандартных условий окончания итераций

$$\left\|\bar{X}^{r+1} - \bar{X}^r\right\| \le epsx$$

$$\left|\Phi\left(\bar{X}^{r+1}\right) - \Phi\left(\bar{X}^r\right)\right| \le eps\Phi$$

выполнено, то полагаем $\bar{X}^* \approx \bar{X}^{r+1}$ и заканчиваем вычисления. Иначе – полагаем r=r+1 и переходим к п.2.



Траектория поиска минимума функции Химмельблау методом сопряженных направлений

Метод ориентирован, прежде всего, на минимизацию квадратических функций и существенным образом использует специфику последних. Минимизация квадратических функций занимает важное место в теории и практике оптимизационных задач. Во-первых, это простейшая задача нелинейной безусловной оптимизации. Во-вторых, в окрестности экстремума любые нелинейные функции хорошо аппроксимируются квадратической функцией.
Заметим, что при **минимизации квадратичной функции методом сопряженных направлений минимум достигается за одну итерацию**.

Пример
Метод сопряженных направлений для квадратичной функции
Найти минимум Ф(X) методом сопряженных направлений

$$\Phi(x, y) = 4x^2 + 4y^2 + 6xy$$

Начальное приближение решения (-1.0; -1.0)

Точность по X epsx=0.1

Точность по Ф epsФ=0.1

В матричной форме

$$\Phi(x, y) = \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} x & y \end{bmatrix}$$

Каноническая форма записи квадратичной функции

$$\Phi(\overline{X}) = \frac{1}{2}\left(A\overline{X}, \overline{X}^T\right) + \left(\overline{b}^T, \overline{X}\right) + c$$

$$\Phi(x,y)=\frac{1}{2}\begin{bmatrix}8 & 6\\ 6 & 8\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}\begin{bmatrix}x & y\end{bmatrix}+\begin{bmatrix}0 & 0\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}+0$$

Метод сопряженных направлений:

1. Задаем начальную точку: $\overline{X}^0$= (-1, -1)^T, номер итерации r=0
2. Выполняем последовательный поиск из вдоль координатных осей, шаг выбирается из решения одномерной задачи минимизации Ф в направлении поиска.

   По х

   $$\min_{\lambda}\Phi(-1+\lambda,-1)=4(-1+\lambda)^2+4(-1)^2+6(-1+\lambda)(-1)=4\lambda^2-14\lambda+14$$

   $$\lambda=1{,}75$$

   Делаем шаг вдоль оси х:

   $\overline{X}_1^1$= (0,75 , -1)^T

   $\Phi(\overline{X}_1^1)=1{,}75$

   Из новой точки $\overline{X}_1^1$ шаг по у:

   $$\min_{\lambda}\Phi(0{,}75,-1+\lambda)=4(0{,}75)^2+4(-1+\lambda)^2-6*0{,}75*(-1+\lambda)=\ldots$$

   $$\lambda=0{,}4375$$

   Делаем шаг вдоль оси у

   $\overline{X}_2^1$ = (0,75 , -0,5625)^T

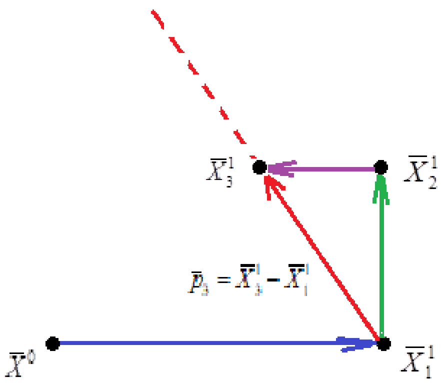
   $\Phi(\overline{X}_2^1)=0{,}9844$

   Из последней точки еще раз вдоль оси х

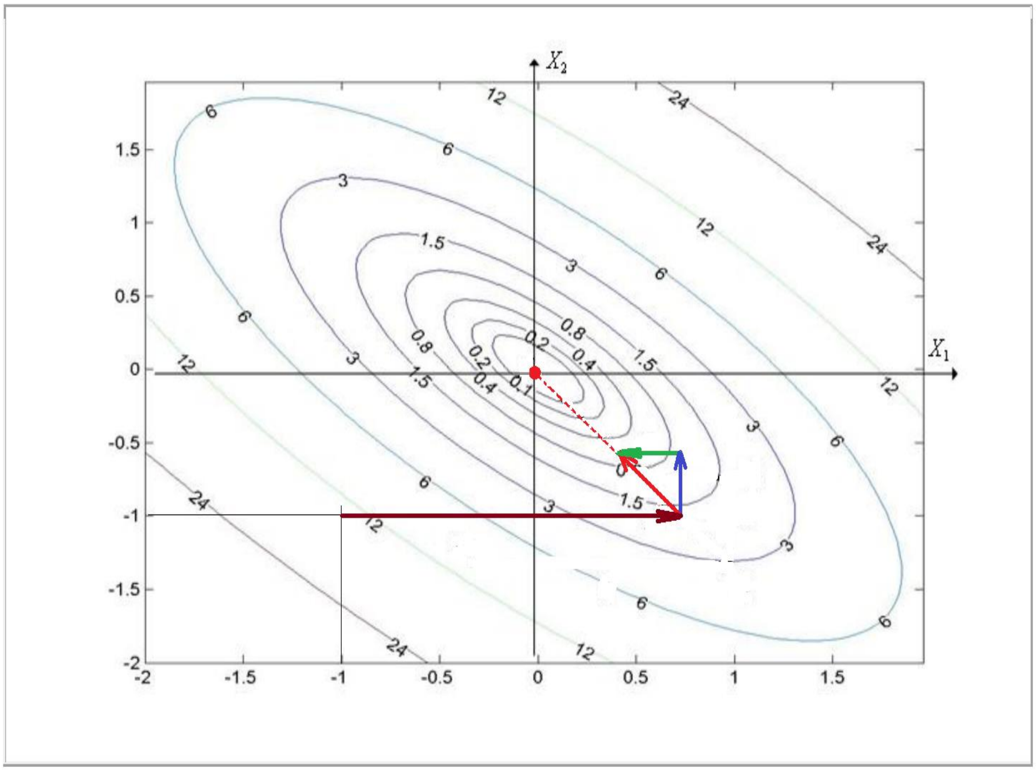
   $$\min_{\lambda}\Phi(0{,}75+\lambda,-0{,}5625)=\ldots$$

   $$\lambda=-0{,}32813$$

   $\overline{X}_3^1$= (0,42187 , -0,5625)^T



Шаг вдоль $\overline{p}_3=\overline{X}_3^1-\overline{X}_1^1=\begin{bmatrix}0{,}42187\\ -0{,}5625\end{bmatrix}-\begin{bmatrix}0{,}75\\ -1\end{bmatrix}=\begin{bmatrix}-0{,}32813\\ 0{,}4375\end{bmatrix}$



Ф – квадратичная, минимум достигается за одну итерацию.

Замечание: вектора p1 и p3 являются сопряженным относительно матрицы А.

Проверка сопряженности

$$\overline{p}_1 A \overline{p}_3^{\,T} = \begin{bmatrix} 1{,}75 & 0 \end{bmatrix} \begin{bmatrix} 8 & 6 \\ 6 & 8 \end{bmatrix} \begin{bmatrix} -0{,}32813 \\ 0{,}4375 \end{bmatrix} \approx 0$$

5 Симплекс-метод

Регулярным симплексом в n-мерном пространстве R называется правильный многогранник, образованный (n+1) равноотстоящими друг от друга вершинами.

n = 2– это равносторонний треугольник

n =3 тетраэдр

Если в пространстве R необходимо построить регулярный симплекс, одна из вершин которого находится в точке $\overline{X}^0 = \left(x_1^0, x_2^0, \ldots, x_n^0\right)$, то координаты вершин такого симплекса удобно задавать с помощью матрицы n×(n+1) матрицы вида

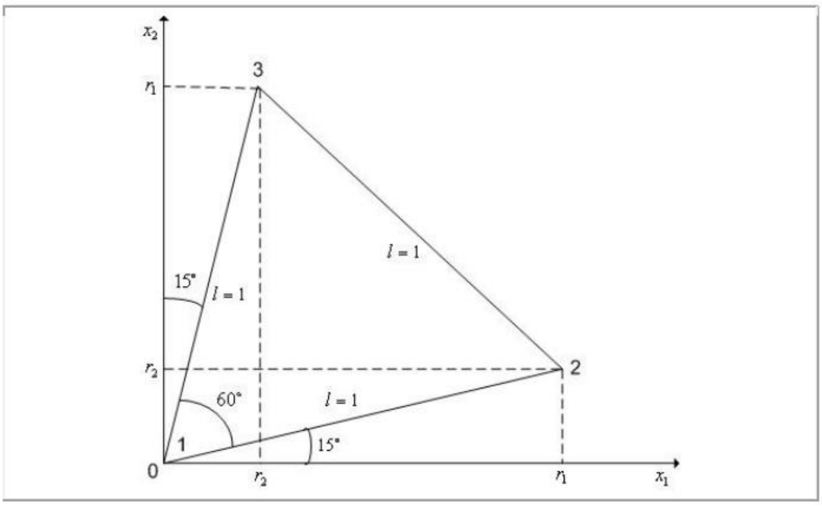
$$R_{n\times(n+1)} = \begin{bmatrix} x_1^0 & x_1^0 + r_1 & x_1^0 + r_2 & \ldots & x_1^0 + r_2 \\ x_2^0 & x_2^0 + r_2 & x_2^0 + r_1 & \ldots & x_2^0 + r_2 \\ x_3^0 & x_3^0 + r_2 & x_3^0 + r_2 & \ldots & x_3^0 + r_2 \\ \ldots & \ldots & \ldots & \ldots & \\ x_n^0 & x_n^0 + r_2 & x_n^0 + r_2 & & x_n^0 + r_1 \end{bmatrix}$$

Здесь i -й столбец представляет собой координаты i -й вершины симплекса

$$r_1 = l\,\frac{\sqrt{n+1}+n-1}{n\sqrt{2}}; \quad r_2 = l\,\frac{\sqrt{n+1}-1}{n\sqrt{2}}, \quad l - длина\ ребра\ симплекса$$

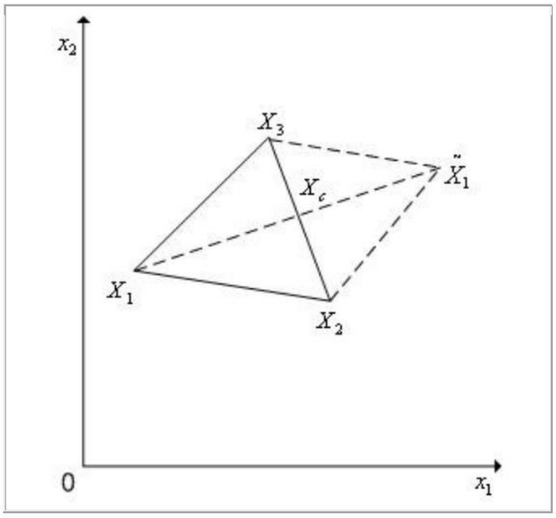
Например, регулярный симплекс в двумерном пространстве с одной из вершин в начале координат определяется матрицей

$R_{2\times3} = \begin{bmatrix} 0 & r_1 & r_2 \\ 0 & r_2 & r_1 \end{bmatrix}$ и имеет вид



Преобразования симплекса

В алгоритме симплекс-метода используется следующее важное свойство регулярного симплекса: если одну из вершин регулярного симплекса перенести на надлежащее расстояние вдоль прямой, соединяющей данную вершину и центр тяжести оставшихся вершин, то вновь получится регулярный симплекс



Будем называть эту процедуру **отражением вершины симплекса** относительно центра тяжести остальных вершин.

Пусть X_i^r – векторы координат вершин регулярного симплекса (i=1,…,n+1).
Тогда при выполнении операции отражения k -й вершины симплекса имеет место следующая связь координат этой вершины и новой вершины:

$$\frac{\overline{X}_k^r + \overline{X}_k^{r+1}}{2} = \overline{X}_c^r,$$ где

координаты центра тяжести остальных вершин симплекса (за исключением отраженной вершины k)

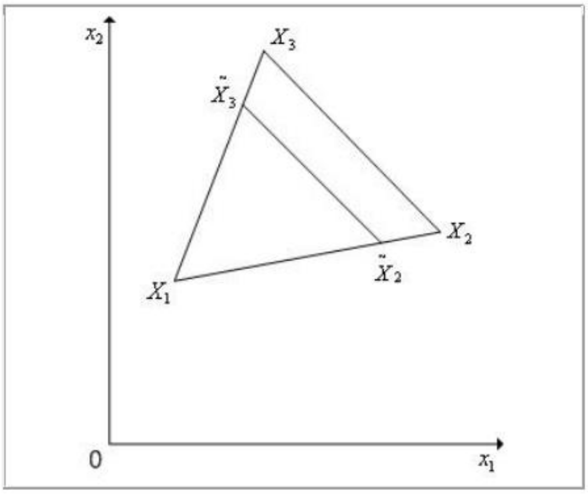
$$\overline{X}_c^r = \frac{1}{n} \sum_{i=1, i \neq k}^{n+1} \overline{X}_i^r$$

Таким образом, после отражения k -ой вершины симплекса получаем новый симплекс с координатами вершин

$$\overline{X}_i^{r+1} = \overline{X}_i^r, \quad i \neq k$$

$$\overline{X}_k^{r+1} = 2\overline{X}_c^r - \overline{X}_k^r$$

Кроме операции отражения вершины симплекса, симплекс-метод может использовать операцию **редукции симплекса** – уменьшение длин всех ребер симплекса на одну и ту же величину.
Пример – редукция вершин регулярного симплекса к вершине X1 .



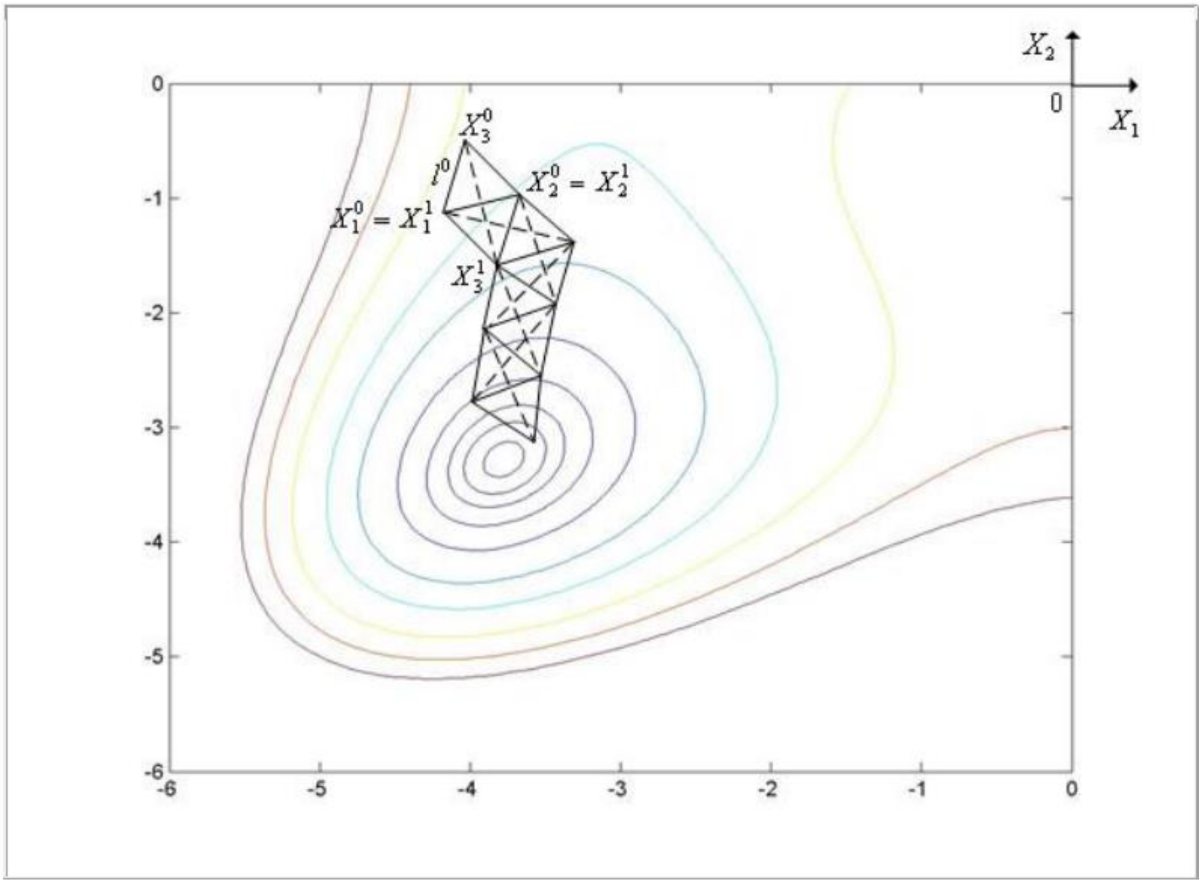
Таким образом, после редукции вершин симплекса к вершине Xk получаем новый симплекс с координатами вершин

$$\overline{X}_i^{r+1} = \overline{X}_k^r + \gamma\left(\overline{X}_i^1 - \overline{X}_i^r\right), \quad i \neq k$$

$$\overline{X}_k^{r+1} = \overline{X}_k^r$$

Схема простейшего варианта симплекс-метода

1. Задаем начальное приближение экстремума, длину ребра симплекса ***l***
2. Вычисляем координаты всех вершин симплекса
3. Вычисляем Ф во всех вершинах симплекса
4. Находим максимальное из вычисленных значений Ф – пусть максимальное значение в вершине Xk
5. Отражаем вершину Xk относительно центра тяжести остальных вершин симплекса – получаем новый симплекс
6. Вычисляем Ф в новой вершине
7. Если условие окончания итераций выполнено, то заканчиваем. Иначе – п.4



Траектория поиска минимума функции Химмельблау простейшим симплекс-методом

Рассмотренный простейший симплекс-метод склонен к зацикливанию и медленно сходится, если длина ребра симплекса ***l*** выбрана малой (выбор большой длины ребра симплекса обеспечивает высокую скорость сходимости, но дает малую точность решения).

Поэтому в вычислительной практике используются различные модификации простейшего метода, направленные на преодоление его указанных недостатков.

Модифицированный симплекс-метод

Основной идей модифицированного симплекс-метода является ***изменение по некоторому правилу размера симплекса*** в процессе поиска.

При этом наряду с условием достижения точности по Ф в качестве условия окончания итераций можно использовать условие $l \le \text{epsx}$, где l – текущая длина ребра симплекса, epsx - требуемая точность решения по X .

Обычно размер симплекса изменяется при выполнении следующих условий:

– при «накрытии» симплексом дна оврага или точки минимума;

– при циклическом движении (ситуация, когда некоторая вершина симплекса не исключается на протяжении m итераций, интерпретируется как «зацикливание» алгоритма).

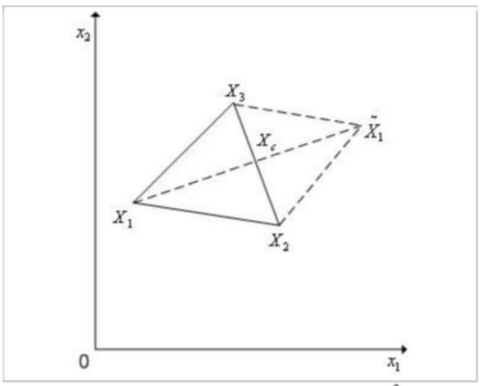
6 Метод деформируемого многогранника (Нелдера-Мида) – модификация симплекс-метода

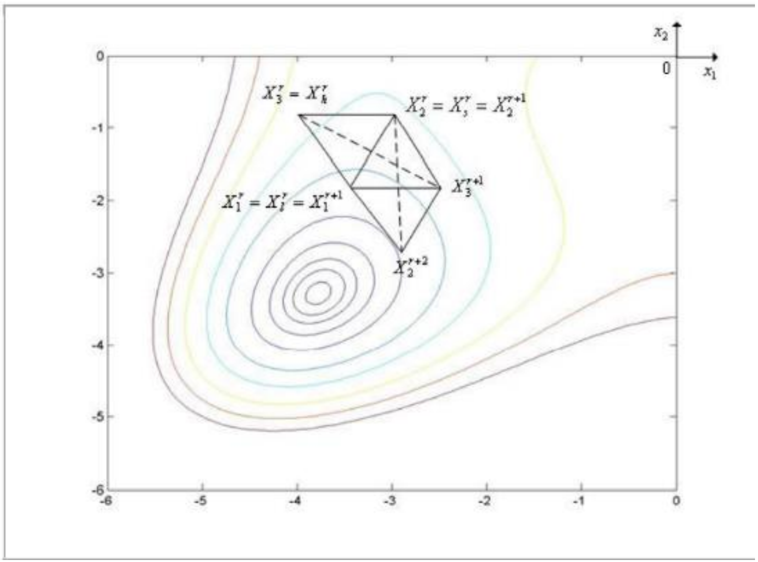
Метод является развитием симплекс-метода и использует в процессе поиска деформацию (изменение размеров и формы) текущего симплекса (не обязательно регулярного).

Метод использует следующие операции над симплексами:

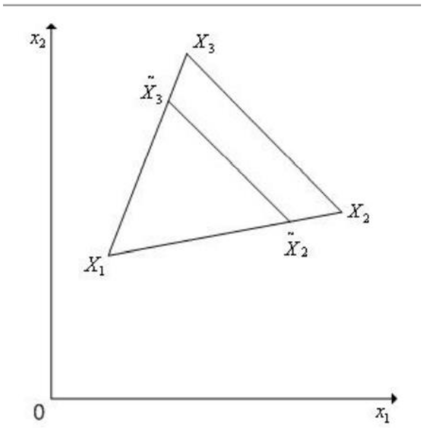
- отражение;
- редукция (уменьшение длин всех ребер в одно и то же количество раз);
- сжатие;
- растяжение.

Отражение

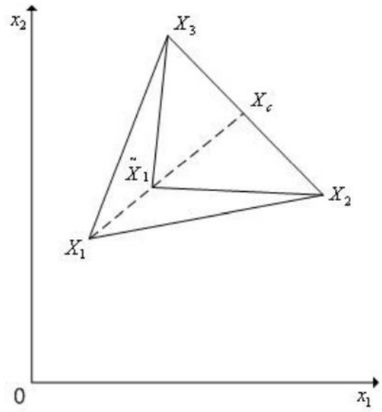


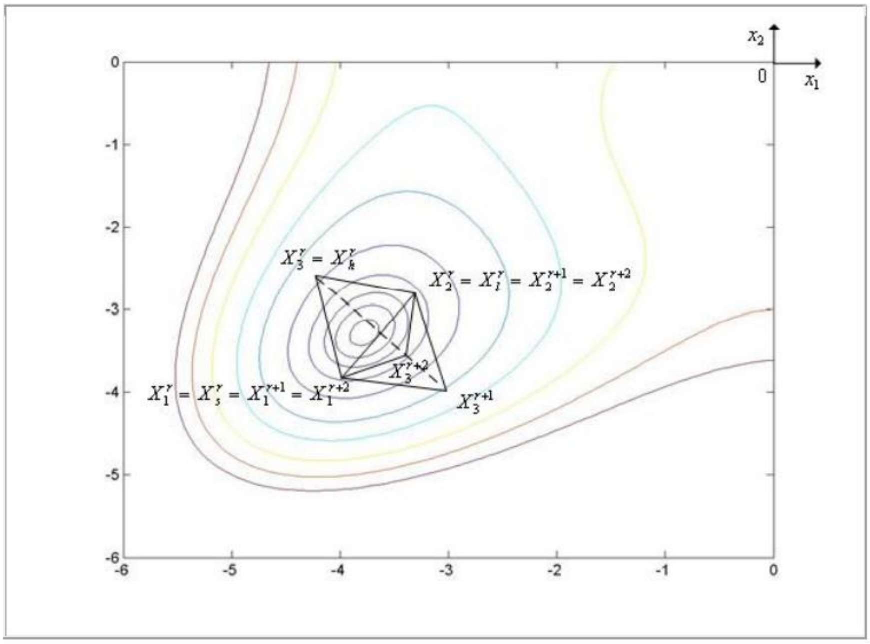


Редукция к X1

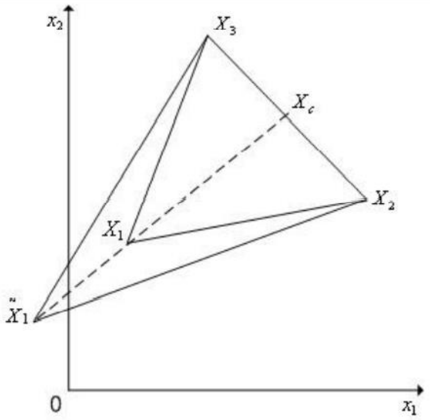


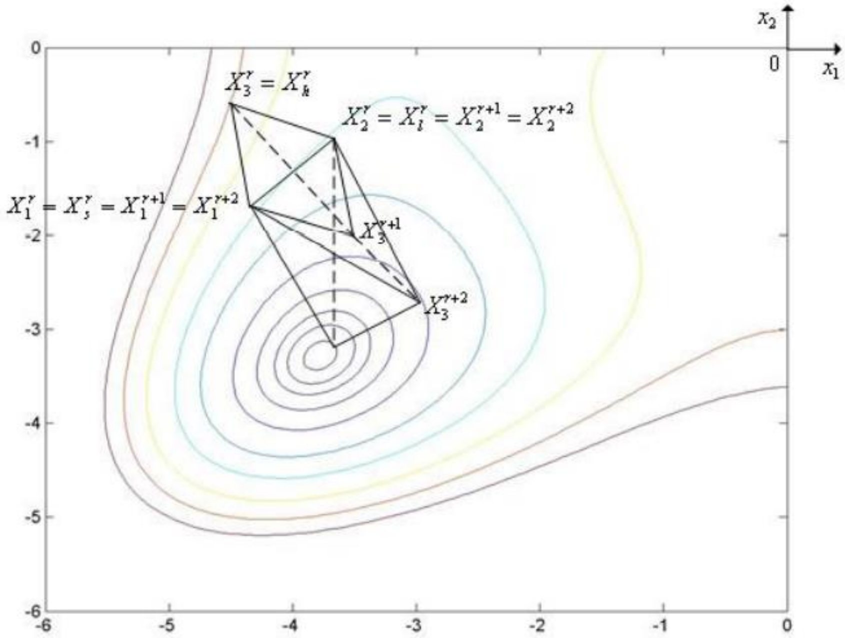
Сжатие в направлении X1 – Xc:



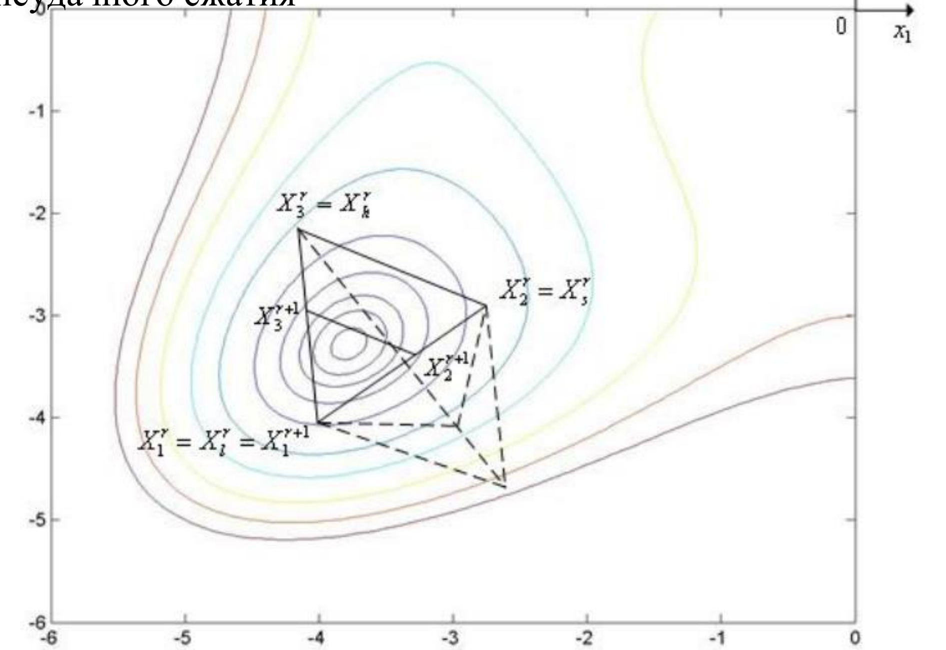


Растяжение в направлении X1 – Xc:





Пример редукции после неудачного сжатия



Недостатки метода Нелдера-Мида: для сильно овражных функций может происходить вырождение («сплющивание») симплекса. Поэтому к рассмотренной схеме метода Нелдера-Мида добавляется этап периодического (через N итераций) восстановления симплекса, который заключается в следующем:

- в текущем симплексе выбираются две «лучшие» вершины и определяется расстояние между ними l;
- исходя из «лучшей» вершины текущего симплекса строится новый симплекс, длина ребра которого принимается равной l.

Градиентные методы

Найти минимум Φ

$$\min_{\overline{X} \in R^n} \Phi\left(\overline{X}\right) = \Phi\left(\overline{X}*\right) = \Phi*$$

Пусть Φ всюду дифференцируема в n-мерном евклидовом пространстве R^n. Направление спуска в градиентных методах оптимизации совпадает с направлением антиградиента минимизируемой функции Φ.

Итерационная формула градиентных методов:

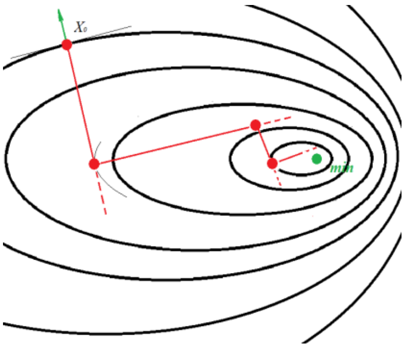
$\overline{X}^0$ - начальное приближение решения

$$\overline{X}^{r+1} = \overline{X}^r + \lambda^r \overline{S}^r,$$

где $\overline{S}^r$ - единичный вектор направления антиградиента функции Φ в точке $\overline{X}^r$:

$$\overline{S}^r = -\frac{\nabla \Phi^r}{\left\|\nabla \Phi^r\right\|}$$

Различные градиентные методы оптимизации отличаются между собой правилами выбора длины шага λ^r.



7 Градиентный метод наискорейшего спуска

Градиентный метод наискорейшего спуска в качестве длины шага λr использует величину, при которой достигается минимум функции Ф в направлении Sr:

$$\Phi\left(\bar{X}^{r+1}\right)=\Phi\left(\bar{X}^{r}+\lambda^{r}\bar{S}^{r}\right)=\min_{\lambda}\Phi\left(\bar{X}^{r}+\lambda\bar{S}^{r}\right)$$

Схема метода

1. Задаем начальную точку $\bar{X}^{0}$ и счетчик количества итераций r=0.
2. Вычисляем антиградиент $\bar{S}^{r}=-\dfrac{\nabla\Phi^{r}}{\left\|\nabla\Phi^{r}\right\|}$
3. Решаем одномерную задачу безусловной минимизации – находим точку $\bar{X}^{r+1}$:
   $$\Phi\left(\bar{X}^{r+1}\right)=\Phi\left(\bar{X}^{r}+\lambda^{r}\bar{S}^{r}\right)=\min_{\lambda}\Phi\left(\bar{X}^{r}+\lambda\bar{S}^{r}\right)$$
4. Вычисляем значение функции в найденной точке $\Phi\left(\bar{X}^{r+1}\right)$
5. Если одно из условий окончание поиска выполнено, то $\bar{X}*\approx\bar{X}^{r+1}$, конец.
   Иначе r=r+1, переход на п.2

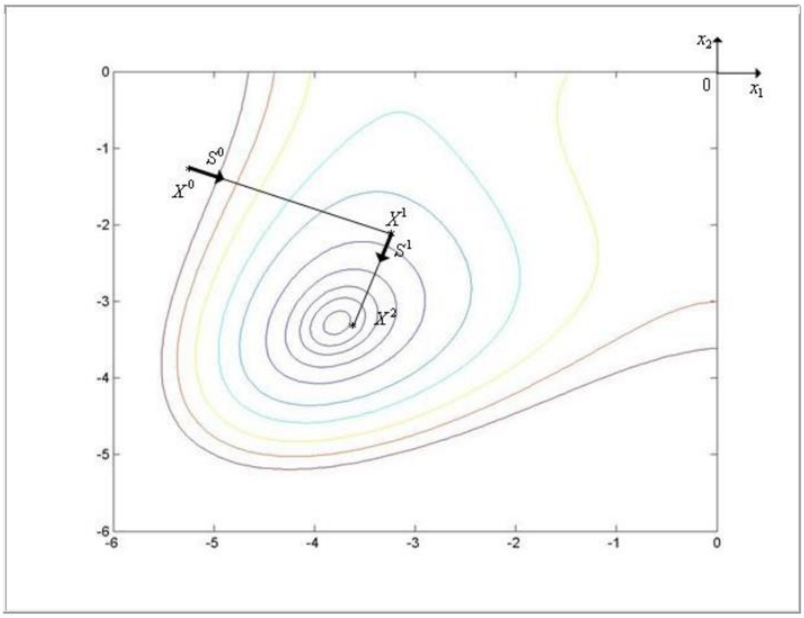
В качестве **критерия окончания поиска** можно использовать одно из стандартных условий:

а) $\left\|\bar{X}^{r+1}-\bar{X}^{r}\right\|\leq epsx$

б) $\left|\Phi\left(\bar{X}^{r+1}\right)-\Phi\left(\bar{X}^{r}\right)\right|\leq eps\Phi$

в) В качестве критерия окончания поиска можно использоваться также условие $\left\|\nabla\Phi^{r}\right\|\leq eps\_grad$,

где eps_grad -- требуемая точность решения по градиенту функции Ф.



Траектория поиска минимума функции Химмельблау градиентным методом наискорейшего спуска

Пример

Минимизировать методом наискорейшего спуска целевую функцию

$$f(x_1, x_2) = x_1^2 + 2x_2^2 + e^{x_1+x_2}, \quad x_1, x_2 \in R$$

Критерий завершения вычислений:

$$\left| \frac{\partial f}{\partial x_i} \right| \le 0.05, i = 1,2$$

Начальное приближение решения:

$$x_1 = x_2 = 0$$

$$grad\ \ f(x_1, x_2) = \begin{pmatrix} 2x_1 + e^{x_1+x_2} \\ 4x_2 + e^{x_1+x_2} \end{pmatrix}$$

Шаг 1.

$$grad\ \ f(0,0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Для определения величины шага λr решаем задачу одномерной минимизации

$$\min\left(f\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} - \lambda \cdot grad\ f(0,0) \right) \right) = \min\left(f\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} - \lambda \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \right) = \min\ f(-\lambda, -\lambda) =$$

$$= \min\left(\lambda^2 + 2\lambda^2 + e^{-2\lambda}\right)$$

$$3\lambda - e^{-2\lambda} = 0$$

$$\lambda = 0.22$$

$$\overline{X}^{r+1} = \overline{X}^r + \lambda^r \overline{S}^r$$

Таким образом, $x_1 = x_2 = -0.22;\ f(-0.22, -0.22) = 0.789$

Шаг	X1	X2	F(x1,x2)	df/dx1	df/dx2	λ	комментарий
1	0	0	1	1	1	0.22	продолжаем
2	-0.22	-0.22	0.789				

Шаг 2.

$$grad\ \ f(-0.22, -0.22) = \begin{pmatrix} 0.204 \\ -0.236 \end{pmatrix}$$

Для определения величины шага λ решаем задачу одномерной минимизации

$$\min\left(f\left(\begin{pmatrix} -0.22 \\ -0.22 \end{pmatrix} - \lambda \cdot grad\ f(-0.22, -0.22) \right) \right) ==$$

$$= \min\left((-0.22 - 0.204\lambda)^2 + (-0.22 + 0.236\lambda)^2 + eps(-0.44 + 0.032\lambda)\right)$$

...

$\lambda = 0.32$

Таким образом,

$$x_1 = -0.2853\,;\, x_2 = -0.1445\,;\; f(-0.2853, -0.1445) = 0.7738$$

Шаг	X1	X3	F(x1,x2)	df/dx1	df/dx2	λ	комментарий
1	0	0	1	1	1	0.22	продолжаем
2	-0.22	-0.22	0.789	0.204	-0.236	0.32	продолжаем
3	-0.2853	-0.1445	0.774				

Шаг 3.

$$grad\ f(-0.2853, -0.1445) = \begin{pmatrix} ? \\ ? \end{pmatrix}$$

Для определения величины шага λr решаем задачу одномерной минимизации

$$\min\left(f\left(\begin{pmatrix} -0.2853 \\ -0.1445 \end{pmatrix} - \lambda \cdot grad\ f\left(-0.2853, -0.1445\right) \right) \right) ==$$

$$= \min\left(\left(-0.2853 - 0.08007\lambda\right)^2 + \left(-0.1445 - 0.07268\lambda\right)^2 + \exp(-0.429 + 0.15275\lambda) \right)$$

. . .

$\lambda = 0.24$

Таким образом, $x_1 = -0.3045\,;\, x_2 = -0.1619\,;\; f(-0.3045, -0.1619) = 0.77240$

Шаг	X1	X3	F(x1,x2)	Df/dx1	Df/dx2	λ	комментарий
1	0	0	1	1	1	0.22	продолжаем
2	-0.22	-0.22	0.789	0.204	-0.236	0.32	продолжаем
3	-0.2853	-0.1445	0.774	0.08007	0.07268	0.24	продолжаем
4	-0.3045	-0.1619	0.772	0.01821	-0.02051	-	Точность достигнута

Ответ: minf=f(-0.305,-0.162)=0.772

8 Градиентный метод с дроблением шага

В градиентном методе с дроблением шага точка $\overline{X}^{r+1}$ определяется по формуле

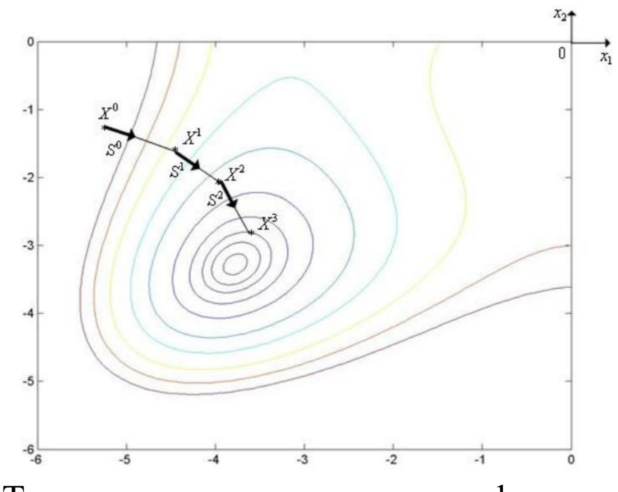
$$\overline{X}^{r+1} = \overline{X}^{r} + \lambda^{r}\overline{S}^{r}$$

где величина шага λ^{r} находится из условия

$$\Phi\left(\overline{X}^{1}\right) - \Phi\left(\overline{X}^{r+1}\right) \geq 0.5\lambda^{r}\left\|\nabla\Phi^{r}\right\|$$

Схема метода

1. Задаем начальную точку $\overline{X}^{0}$, начальную величину шага λ^{0}, коэффициент дробления шага $\nu \in (0,1]$. Счетчик количества итераций r=0.
2. Вычисляем компоненты вектора $\overline{X}^{r+1}$
   $$\overline{X}^{r+1} = \overline{X}^{r} + \lambda^{r}\overline{S}^{r}$$
3. Вычисляем значение функции в найденной точке $\Phi\left(\overline{X}^{r+1}\right)$
4. Если условие
   $$\Phi\left(\overline{X}^{r}\right) - \Phi\left(\overline{X}^{r+1}\right) \geq 0.5\lambda^{r}\left\|\nabla\Phi^{r}\right\|$$
   Выполнено – п.5, иначе – п.6.
5. Полагаем
   $\lambda^{r} = \nu\,\lambda^{r}$ и переходим на п.2
6. Проверяем условие окончания поиска можно использовать одно из стандартных условий:
   $$\left\|\overline{X}^{r+1} - \overline{X}^{r}\right\| \leq epsx$$
   $$\left|\Phi\left(\overline{X}^{r+1}\right) - \Phi\left(\overline{X}^{r}\right)\right| \leq eps\Phi$$
   $$\left\|\nabla\Phi^{r}\right\| \leq eps\_grad$$



Траектория поиска минимума функции Химмельблау градиентным методом с дроблением шага

Эффективность градиентных методов зависит от вида минимизируемой функции. В тех ситуациях, когда линии уровня минимизируемой функции представляют собой прямолинейный или, хуже того, криволинейный «овраг», эффективность метода оказывается очень низкой.

Пример

Минимизировать целевую функцию

$$f(x) = f(x_1, x_2) = x_1^2 + 2x_2^2 + e^{x_1 + x_2},\ x \in R^2,$$

методом градиентного спуска, завершив расчет при

$$\left|\frac{\partial f(x^{(k)})}{\partial x_i}\right| \le 0.05,\ i = 1, 2.$$

Решение

Выберем начальное приближение $x^{(0)} = (0, 0)$ и величину шага спуска $\alpha = 1$, построим последовательность (5.3) (с дроблением шага спуска α), записывая результаты вычислений в таблицу.

k	$x_1^{(k)}$	$x_2^{(k)}$	$f(x^{(k)})$	$\frac{\partial f(x^{(k)})}{\partial x_1}$	$\frac{\partial f(x^{(k)})}{\partial x_2}$	α	Примечание
0	0	0	1	1	1	1	
	–1	–1	3.145	–	–		Условие (5.4) нарушено. Уменьшим α в 2 раза
	0	0	1	1	1	0.5	
	–0.5	–0.5	1.118	–	–		Условие (5.4) нарушено. Уменьшим α в 2 раза
	0	0	1	1	1	0.25	
1	–0.25	–0.25	0.794	0.106	–0.393	0.25	Условие (5.4) выполнено
2	–0.277	–0.152	0.774	0.098	0.045	0.25	Условие (5.4) выполнено
3	–0.301	–0.163	0.772	0.026	0.023	–	Точность достигнута

Ответ: min f=f(-0.301,-0.163)=0.772