

Раздел 3. Нелинейное программирование. Безусловная многомерная оптимизация

3.1. Метод Гаусса-Зейделя

Рассматривается следующая многомерная задача безусловной оптимизации (точнее говоря, задача многомерной локальной безусловной оптимизации): найти минимум критерия оптимальности $\Phi(X)$, определенного в n -мерном евклидовом пространстве R^n ,

$$\min_{X \in R^n} \Phi(X) = \Phi(X^*) = \Phi^*.$$

При решении задачи методом Гаусса-Зейделя (методом покоординатного спуска, методом циклического покоординатного спуска) используются следующие итерационные формулы

$$\begin{aligned} X_1^{r+1} &= X^r + \lambda_1^r L_1, \\ X_2^{r+1} &= X_1^{r+1} + \lambda_2^r L_2, \\ &\dots\dots\dots, \\ X_n^{r+1} &= X_{n-1}^{r+1} + \lambda_n^r L_n = X^{r+1}, \end{aligned} \quad (3.1.1)$$

где вектор L_i определяет направление вдоль i -й координатной оси и

представляет собой n -мерный вектор с компонентами $l_{ij} = \begin{cases} 1, & \text{если } j = i \\ 0, & \text{если } j \neq i \end{cases}$,

$j = 1, 2, \dots, n$, а величины $\lambda_1^r, \lambda_2^r, \dots, \lambda_n^r$ – определяются из условий

$$\begin{aligned} \min_{\lambda \in (-\infty, +\infty)} \Phi(X^r + \lambda L_1) &= \Phi(X^r + \lambda_1^r L_1) = \Phi(X_1^{r+1}), \\ \min_{\lambda \in (-\infty, +\infty)} \Phi(X_1^{r+1} + \lambda L_2) &= \Phi(X_1^{r+1} + \lambda_2^r L_2) = \Phi(X_2^{r+1}), \\ &\dots\dots\dots, \\ \min_{\lambda \in (-\infty, +\infty)} \Phi(X_{n-1}^{r+1} + \lambda L_n) &= \Phi(X_{n-1}^{r+1} + \lambda_n^r L_n) = \Phi(X_n^{r+1}). \end{aligned} \quad (3.1.2)$$

Другими словами, величина λ_i^r , $i = 1, 2, \dots, n$ представляет собой длину шага, минимизирующего функцию $\Phi(X)$ в направлении L_i на итерации номер r , исходя из точки, полученной на предыдущем шаге.

Если положить $X_0^{r+1} = X^r$, $X_n^{r+1} = X^{r+1}$, то формулы (3.1.1), (3.1.2)

можно записать в виде

$$X_i^{r+1} = X_{i-1}^{r+1} + \lambda_i^r L_i, i=1,2,\dots,n, \quad (3.1.3)$$

$$\min_{\lambda \in (-\infty, +\infty)} \Phi(X_{i-1}^{r+1} + \lambda L_i) = \Phi(X_{i-1}^{r+1} + \lambda_i^r L_i) = \Phi(X_i^{r+1}), \quad i=1,2,\dots,n. \quad (3.1.4)$$

Таким образом, каждая итерация по методу Гаусса-Зейделя включает в себя n шагов. Каждая последующая итерация начинается из точки, полученной на последнем шаге предыдущей итерации. Поиск заканчивается при выполнении одного из стандартных условий окончания итераций:

$$\|X^{r+1} - X^r\| \leq \varepsilon_x, \quad (3.1.5)$$

$$|\Phi(X^{r+1}) - \Phi(X^r)| \leq \varepsilon_\Phi. \quad (3.1.6)$$

Заметим, что задачи (3.1.4) даже в случае одноэкстремальной функции $\Phi(X)$ могут быть задачами многоэкстремальной оптимизации и могут быть решены рассмотренными в разделе 2 методами решения задач одномерной оптимизации.

Схема метода Гаусса-Зейделя.

1. Задаем начальную точку X^0 и полагаем $r=0$, $i=1$.
2. Последовательно для $i=1,2,\dots,n$ решаем задачи (3.1.4), т.е. исходя из предыдущей точки, отыскиваем минимум функции $\Phi(X)$ вдоль i -го координатного направления.
3. Если условие окончания поиска (3.1.5) или (3.1.6) выполнено, то полагаем $X^* \approx X^{r+1}$ и заканчиваем вычисления. Иначе – полагаем $r=r+1$ и переходим к п. 2.

Метод Гаусса-Зейделя иллюстрирует рис. 3.1.1, на котором показан фрагмент линий уровня функции Химмельблау $\Phi(x_1, x_2) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$. На рис. 3.1.1 точка X_1^1 представляет

собой локальный минимум функции $\Phi(X)$ вдоль оси X_1 при исходной точке X^0 . Точка X_2^1 представляет собой локальный минимум функции $\Phi(X)$ вдоль оси X_2 при исходной точке X_1^1 . Отыскание точки X_2^1 завершает первую итерацию. Следующая итерация начинается из точки $X^1 = X_2^1$. И т.д.

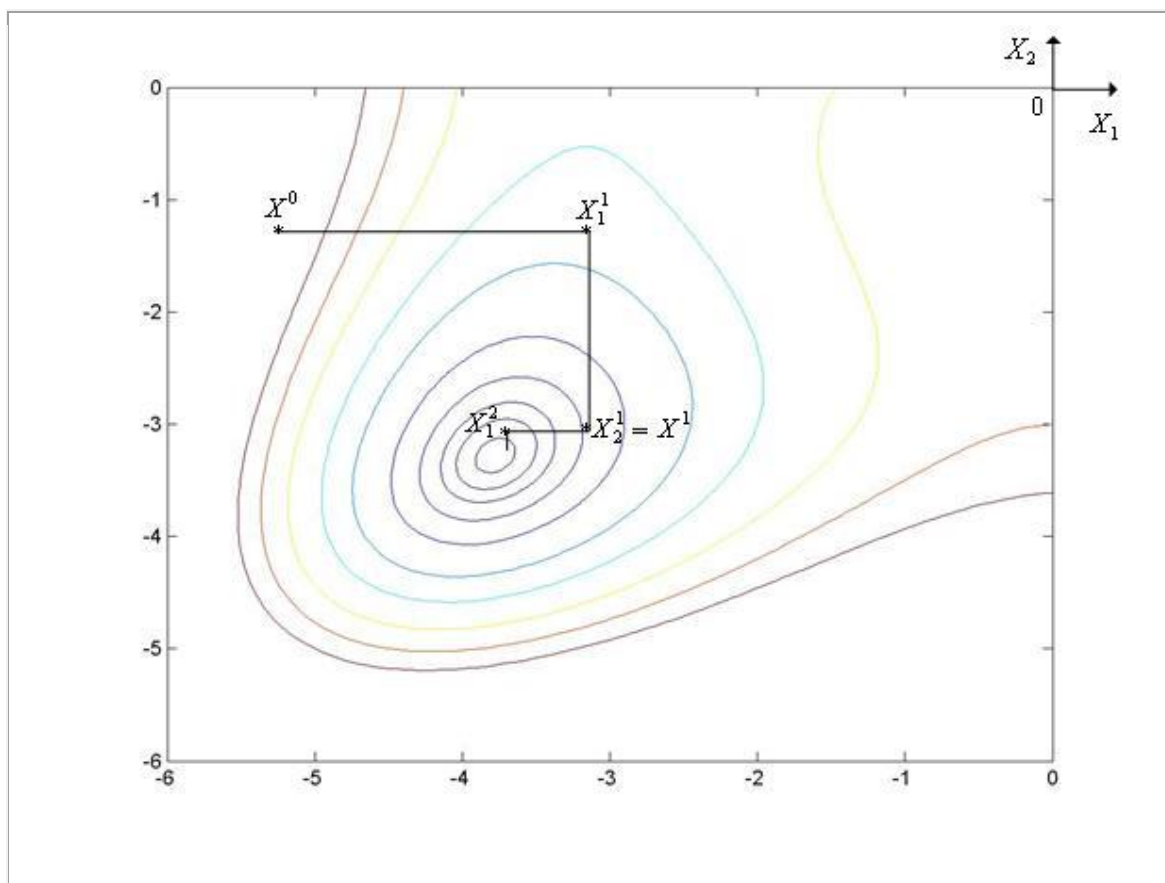


Рис. 3.1.1. Траектория поиска минимума не овражной функции Химмельблау методом Гаусса-Зейделя

Метод Гаусса-Зейделя медленно сходится на овражных функциях, в которых овраг не ориентирован в направлении какой-либо из координатных осей (см. рис. 3.1.2). На рисунке показаны линии уровня функции Розенброка ($n = 2$).

На рисунке 3.1.2. показывается, что текущая точка быстро (в данном случае – за один шаг) «скатывается» на дно оврага и очень медленно движется по дну оврага к минимуму функции $\Phi(X)$.

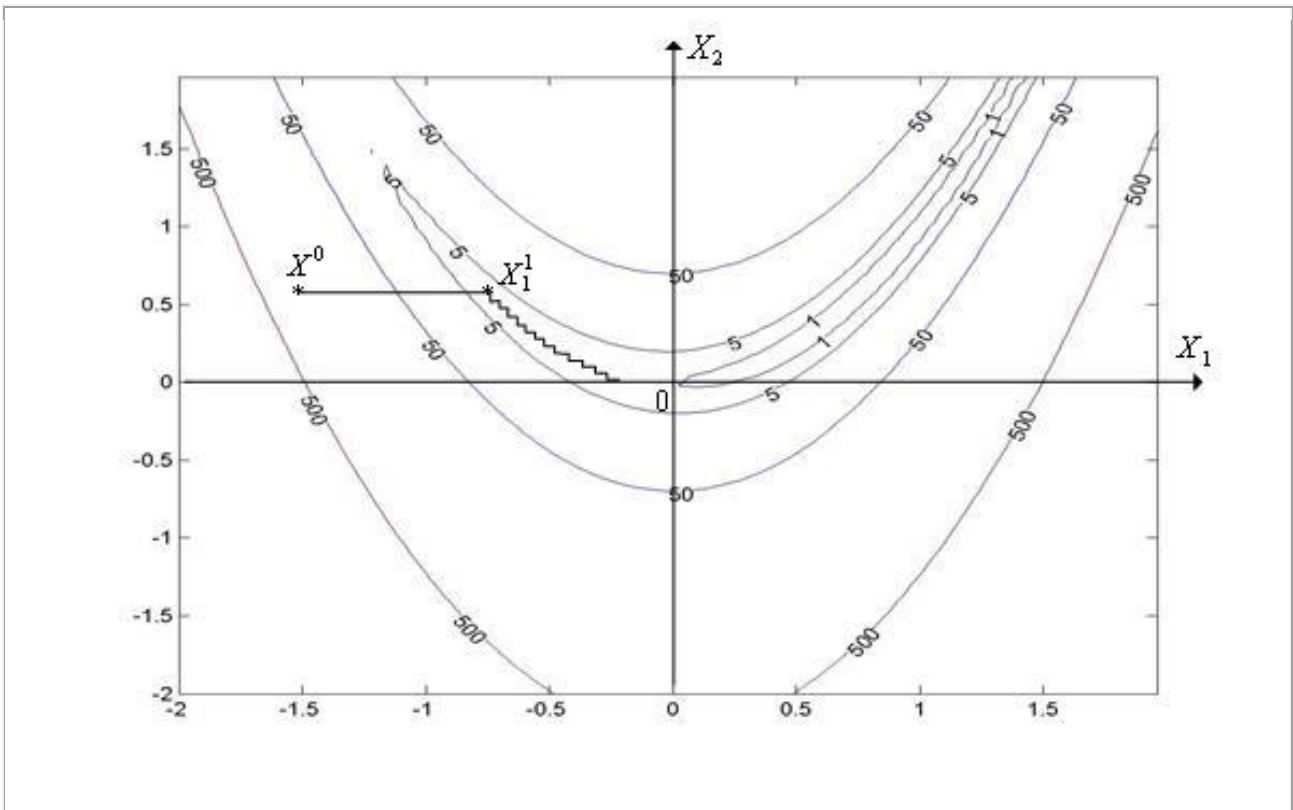


Рис. 3.1.2. Траектория поиска минимума овражной функции Розенброка методом Гаусса-Зейделя. Текущая точка быстро (в данном случае – за один шаг) «скатывается» на дно оврага и очень медленно движется по дну оврага к минимуму функции $\Phi(X)$

Пример 3.1.1. Выполните несколько итераций решения двумерной локальной задачи безусловной оптимизации $\min_{X \in R^2} \Phi(X) = \Phi(X^*) = \Phi^*$, где

$\Phi(X) = \Phi(x, y) = x^2 + y^2 + 3(x + y)^2 = 4x^2 + 4y^2 + 6xy$ методом Гаусса-Зейделя, исходя из точки $X^0 = (x^0, y^0) = (-1.5, 1.5)$.

Решение. Каждая итерация метода Гаусса-Зейделя состоит из двух шагов и имеет вид

$$X_1^{r+1} = \begin{pmatrix} x^{r+1} = x^r + \lambda_1^r \\ y^r \end{pmatrix}, \quad (3.1.7)$$

$$X_2^{r+1} = X^{r+1} = \begin{pmatrix} x^{r+1} \\ y^{r+1} = y^r + \lambda_2^r \end{pmatrix}, \quad (3.1.8)$$

где величины λ_1^r, λ_2^r - определяются из условий

$$\min_{\lambda \in (-\infty, +\infty)} \Phi(x^r + \lambda, y^r) = \Phi(x^r + \lambda_1^r, y^r) = \Phi(x^{r+1}, y^r) = \Phi(X_1^{r+1}), \quad (3.1.9)$$

$$\min_{\lambda \in (-\infty, +\infty)} \Phi(x^{r+1}, y^r + \lambda) = \Phi(x^{r+1}, y^r + \lambda_2^r) = \Phi(x^{r+1}, y^{r+1}) = \Phi(X_2^{r+1}) = \Phi(X^{r+1}). \quad (3.1.10)$$

Найдем явное решение задачи (3.1.9). Из (3.1.9) имеем

$$\begin{aligned} \Phi(x^r + \lambda, y^r) &= 4(x^r + \lambda)^2 + 4(y^r)^2 + 6(x^r + \lambda)y^r = \\ &= 4(x^r)^2 + 8x^r\lambda + 4\lambda^2 + 4(y^r)^2 + 6x^ry^r + 6\lambda y^r = 4\lambda^2 + (8x^r + 6y^r)\lambda + \dots \end{aligned} \quad (3.1.11)$$

Функция (3.1.11) относительно λ является квадратичной функцией с положительным коэффициентом при λ^2 и достигает минимума в точке, удовлетворяющей условию

$$\Phi'_\lambda(x^r + \lambda, y^r) = 8\lambda + 8x^r + 6y^r = 0.$$

из которого имеем

$$\lambda_1^r = -x^r - \frac{3}{4}y^r. \quad (3.1.12)$$

Аналогично явное решение задачи (3.1.10) равно

$$\lambda_2^r = -y^r - \frac{3}{4}x^{r+1}. \quad (3.1.13)$$

Таким образом, из (3.1.7), (3.1.8), (3.1.10), (3.1.11) имеем искомую итерационную формулу метода Гаусса-Зейделя для нашей задачи

$$X_1^{r+1} = \begin{pmatrix} x^{r+1} = x^r + \lambda_1^r \\ y^r \end{pmatrix}, \quad \lambda_1^r = -x^r - \frac{3}{4}y^r; \quad (3.1.14)$$

$$X_2^{r+1} = X^{r+1} = \begin{pmatrix} x^{r+1} \\ y^{r+1} = y^r + \lambda_2^r \end{pmatrix}, \quad \lambda_2^r = -y^r - \frac{3}{4}x^{r+1}. \quad (3.1.15)$$

Первая итерация ($r=1$).

Из формул (3.1.14) имеем $\lambda_1^1 = -x^0 - \frac{3}{4}y^0 = 1.5 - \frac{3}{4} \cdot 1.5 = \frac{3}{8} \approx 0.375$ и $x^1 = x^0 + \lambda_1^1 \approx -1.5 + 0.38 = -1.12$. Аналогично из формул (3.1.15) имеем $\lambda_2^1 = -y^0 - \frac{3}{4}x^1 \approx -1.5 - \frac{3}{4} \cdot 1.12 \approx -0.66$, $y^1 = y^0 + \lambda_2^1 \approx 1.5 - 0.66 = 0.84$.

Таким образом, $X^1 = \begin{pmatrix} x^1 \\ y^1 \end{pmatrix} \approx \begin{pmatrix} -1.12 \\ 0.84 \end{pmatrix}$ (см. рис. 3.1.3).

Вторая итерация ($r=2$).

Аналогично первой итерации, имеем

$$\lambda_1^2 = -x^1 - \frac{3}{4}y^1 = 1.12 - \frac{3}{4} \cdot 0.84 \approx 0.49, \quad x^2 = x^1 + \lambda_1^2 \approx -1.12 + 0.49 = -0.63,$$

$$\lambda_2^2 = -y^1 - \frac{3}{4}x^2 \approx -0.84 + \frac{3}{4} \cdot 0.63 \approx -0.37, \quad y^2 = y^1 + \lambda_2^2 \approx 0.84 - 0.37 = 0.47.$$

Таким образом, $X^1 = \begin{pmatrix} x^2 \\ y^2 \end{pmatrix} \approx \begin{pmatrix} -0.63 \\ 0.47 \end{pmatrix}$ (см. рис. 3.1.3).

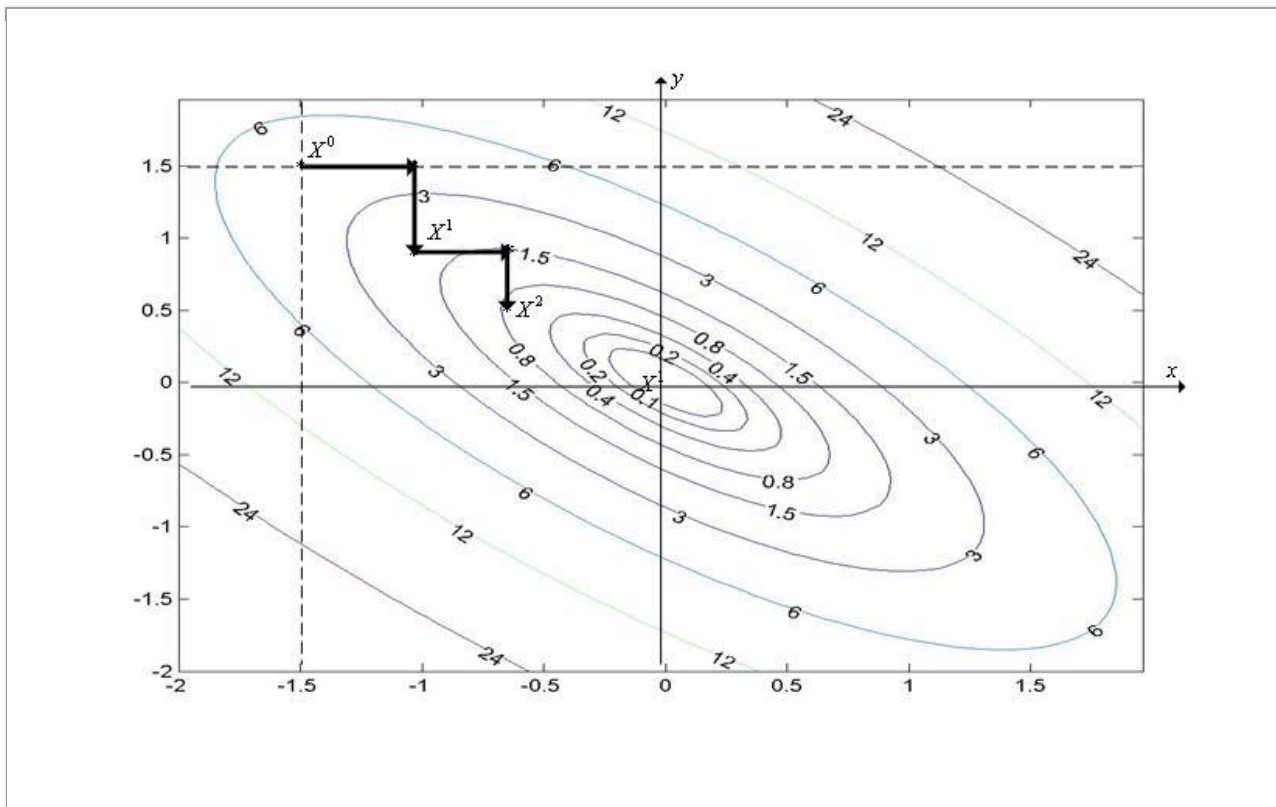


Рис. 3.1.3. Фрагмент (две итерации) траектории поиска минимума функции $\Phi(X) = \Phi(x, y) = x^2 + y^2 + 3(x+y)^2 = 4x^2 + 4y^2 + 6xy$ методом Гаусса-Зейделя из точки $X^0 = (x^0, y^0) = (-1.5, 1.5)$

3.2. Метод Хука-Дживса

Рассмотрим следующую многомерную задачу локальной безусловной оптимизации: найти минимум критерия оптимальности $\Phi(X)$, определенного в n -мерном евклидовом пространстве R^n , $\min_{X \in R^n} \Phi(X) = \Phi(X^*) = \Phi^*$.

При решении задачи методом Хука-Дживса (методом конфигураций, методом пробных шагов) используются итерационные формулы, аналогичные формулам, используемым в методе Гаусса-Зейделя

$$\tilde{X}_i^{r+1} = \tilde{X}_{i-1}^{r+1} + \lambda_i^r L_i, \quad i = 1, 2, \dots, n, \quad (3.2.1)$$

где принято $\tilde{X}_0^{r+1} = X^r$, $\tilde{X}_n^{r+1} = \tilde{X}^{r+1}$, вектор L_i определяет направление вдоль i -й координатной оси и представляет собой n -мерный вектор с компонентами

$$l_{ij} = \begin{cases} 1, & \text{если } j = i \\ 0, & \text{если } j \neq i, \quad j = 1, 2, \dots, n, \end{cases}$$

величины $\lambda_1^r, \lambda_2^r, \dots, \lambda_n^r$ - определяются из условий:

$$\lambda_n^r = \begin{cases} \Delta_i^r, & \text{если } \Phi(X_{i-1}^{r+1} + \Delta_i^r L_i) < \Phi(X_{i-1}^{r+1}), \\ -\Delta_i^r, & \text{если } \Phi(X_{i-1}^{r+1} - \Delta_i^r L_i) < \Phi(X_{i-1}^{r+1}) < \Phi(X_{i-1}^{r+1} + \Delta_i^r L_i), \\ 0, & \text{если } \min[\Phi(X_{i-1}^{r+1} + \Delta_i^r L_i), \Phi(X_{i-1}^{r+1} - \Delta_i^r L_i)] > \Phi(X_{i-1}^{r+1}). \end{cases} \quad (3.2.2)$$

После завершения n шагов выполняется спуск в направлении вектора $(\tilde{X}^{r+1} - X^r)$ по формуле

$$X^{r+1} = X^r + \alpha^r (\tilde{X}^{r+1} - X^r), \quad (3.2.3)$$

где α^r - ускоряющий множитель. В различных модификациях метода Хука-Дживса множитель α^r может

- приниматься постоянным (обычно, равным 2),
- выбираться из условия $\Phi(X^{r+1}) < \Phi(X^r)$,

- находиться из условия локального минимума функции $\Phi(X)$ при движении из точки X^r в направлении вектора $(\tilde{X}^{r+1} - X^r)$:

$$\min_{\alpha \in (-\infty, +\infty)} \Phi(X^r + \alpha(\tilde{X}^{r+1} - X^r)) = \Phi(X^r + \alpha^r(\tilde{X}^{r+1} - X^r)) = \Phi(X^{r+1}). \quad (3.2.4)$$

Заметим, что задачи (3.2.4) даже в случае одноэкстремальной функции $\Phi(X)$ могут быть многоэкстремальными задачами оптимизации и могут быть решены рассмотренными в разделе 2 методами решения задач одномерной оптимизации.

Итерации заканчиваются при выполнении одного из стандартных условий окончания итераций:

$$\|X^{r+1} - X^r\| \leq \varepsilon_x, \quad (3.2.5)$$

$$|\Phi(X^{r+1}) - \Phi(X^r)| \leq \varepsilon_\Phi. \quad (3.2.6)$$

Вектор $\Delta^0 = (\Delta_1^0, \Delta_2^0, \dots, \Delta_n^0)$ является вектором свободных параметров метода - вектором «пробных шагов» по всем n координатным осям.

Известна модификация метода Хука-Дживса, в которой точка \tilde{X}^{r+1} определяется не процедурами (3.2.1), (3.2.2), а методом Гаусса-Зейделя.

Схема метода Хука-Дживса.

1. Задаем начальную точку X^0 , вектор «пробных» шагов $\Delta^0 = (\Delta_1^0, \Delta_2^0, \dots, \Delta_n^0)$ и полагаем $r=0$.
2. Последовательно для $i=1, 2, \dots, n$ по формулам (3.2.1), (3.2.2) находим точки $\tilde{X}_1^{r+1}, \tilde{X}_2^{r+1}, \dots, \tilde{X}_n^{r+1} = \tilde{X}^{r+1}$.
3. Если $\tilde{X}^{r+1} \neq \tilde{X}^r$, то переходим к п. 4). Иначе уменьшаем длины «пробных» шагов $\Delta_1^0, \Delta_2^0, \dots, \Delta_n^0$, например, вдвое и переходим к п.2).
4. Если условие окончания поиска (3.2.5) или (3.2.6) выполнено, то полагаем $X^* \approx X^{r+1}$ и заканчиваем вычисления. Иначе выполняем спуск в

направлении вектора $(\tilde{X}^{r+1} - X^r)$ по формуле (3.2.3), в которой ускоряющий множитель находится, например, из условия (3.2.4). Полагаем $r = r+1$ и переходим к п. 2.

Метод Хука-Дживса иллюстрирует рис. 3.2.1, на котором показаны линии уровня функции Химмельблау ($n=2$). Ускоряющий множитель α^r находится из условия локального минимума функции $\Phi(X)$ при движении из точки X^r в направлении вектора $(\tilde{X}^{r+1} - X^r)$.

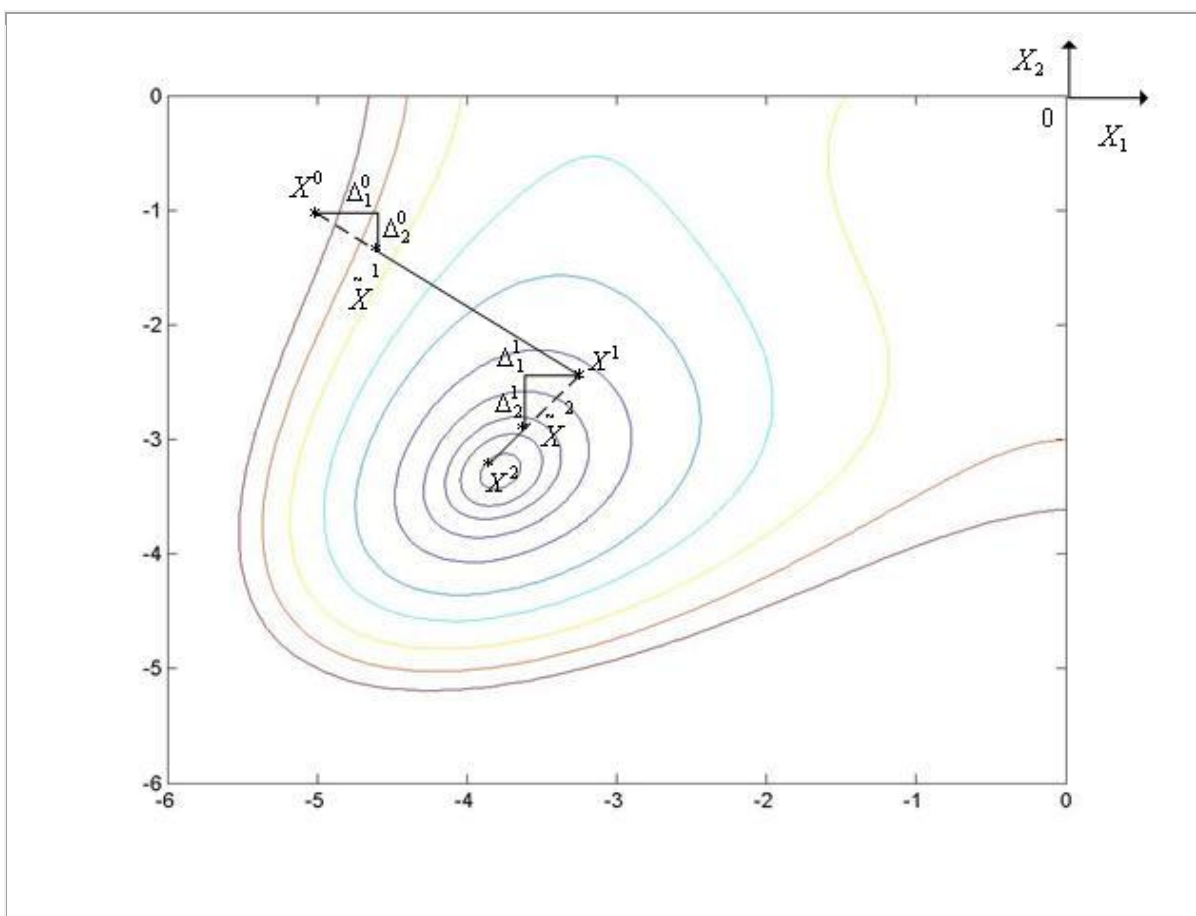


Рис. 3.2.1. Траектория поиска минимума не овражной функции Химмельблау методом Хука-Дживса

Метод Хука-Дживса имеет высокую эффективность в случае, если функция $\Phi(X)$ имеет прямолинейный овраг (не обязательно ориентированный вдоль одного из координатных направлений, как в методе Гаусса-Зейделя). При минимизации "овражных" функций, имеющих не прямолинейный овраг, процесс поиска может сильно замедлиться и

закончиться далеко от точки истинного минимума (см. рис. 3.2.2). На рисунке показаны линии уровня функции Розенброка ($n = 2$).

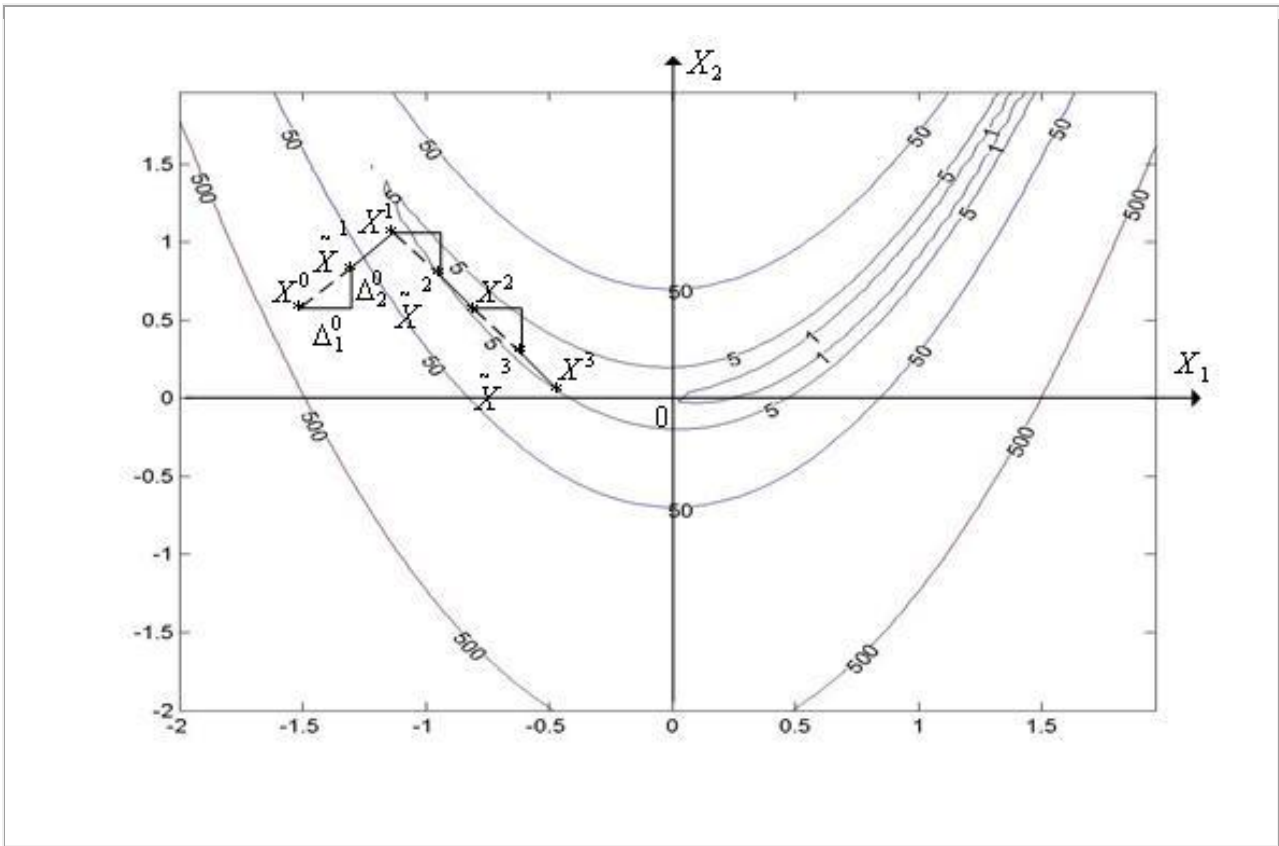


Рис. 3.2.2. Траектория поиска минимума овражной функции Химмельблау методом Хука-Дживса. Ускоряющий множитель $\alpha^r = 2$

3.3. Метод Розенброка

Рассмотрим следующую задачу локальной безусловной оптимизации: найти минимум критерия оптимальности $\Phi(X)$, определенного в n -мерном евклидовом пространстве R^n , $\min_{X \in R^n} \Phi(X) = \Phi(X^*) = \Phi^*$.

При решении задачи методом Розенброка (методом вращающихся координат) используется преобразование на каждой итерации системы координат таким образом, чтобы в новой системе координат одна из осей совпадала с направлением предыдущего шага. Остальные оси новой системы координат обычно находят с помощью процедуры ортогонализации Грамма-Шмидта.

Ортогонализация Грамма-Шмидта.

Рассмотрим произвольный набор векторов p_1, p_2, \dots, p_n пространства R^n .

Поставим задачу построить на основе этих векторов ортонормированный набор векторов e_1, e_2, \dots, e_n того же пространства R^n .

Напомним, что набор векторов e_1, e_2, \dots, e_n называется ортонормированным, если для любых двух векторов из этого набора выполняется условие

$$(e_i, e_j) = \begin{cases} 0, & i \neq j \\ 1, & i = j. \end{cases} \quad (3.3.1)$$

Или, другими словами, набор векторов e_1, e_2, \dots, e_n ортонормирован, если эти векторы линейно независимы и скалярное произведение любых двух из них равно единице.

Для построения векторов e_1, e_2, \dots, e_n применим индуктивный подход.

Положим, что

$$e_1 = \frac{p_1}{\rho_{11}}, \quad \rho_{11} = \|p_1\|, \quad (3.3.2)$$

где $\|*\|$ - символ евклидовой нормы. Полагая векторы e_1, e_2, \dots, e_{k-1} уже построенными, будем искать вектор e_k в виде

$$\rho_{kk} e_k = p_k - \sum_{i=1}^{k-1} \rho_{ik} e_i \quad (3.3.3)$$

Для отыскания неизвестных множителей ρ_{ik} умножим (3.3.3) скалярно на вектор e_j , $j = 1, 2, \dots, k-1$:

$$\rho_{kk} (e_k, e_1) = 0 = (p_k, e_1) - \sum_{i=1}^{k-1} \rho_{ik} (e_i, e_1) = (p_k, e_1) - \rho_{1k} (e_1, e_1),$$

$$\rho_{kk} (e_k, e_2) = 0 = (p_k, e_2) - \sum_{i=1}^{k-1} \rho_{ik} (e_i, e_2) = (p_k, e_2) - \rho_{2k} (e_2, e_2),$$

.....

$$\rho_{kk} (e_k, e_{k-1}) = 0 = (p_k, e_{k-1}) - \sum_{i=1}^{k-1} \rho_{ik} (e_i, e_{k-1}) = (p_k, e_{k-1}) - \rho_{k-1,k} (e_{k-1}, e_{k-1}).$$

Поскольку $(e_i, e_j) = 1$, имеем

$$\rho_{jk} (e_k, e_j) = (p_k, e_j), \quad j \in [1, \dots, k-1]. \quad (3.3.4)$$

Множитель ρ_{kk} найдем из условия $\|e_k\| = 1$:

$$\rho_{kk} = \left\| p_k - \sum_{i=1}^{k-1} \rho_{ik} e_i \right\|. \quad (3.3.5)$$

Определение 3.3.1. Процесс перехода от векторов p_1, p_2, \dots, p_n к векторам q_1, q_2, \dots, q_n согласно формулам (3.3.2)–(3.3.5) называется **ортогонализацией Грамма-Шмидта**.

Каждая итерация метода Розенброка состоит из двух этапов. В зависимости от модификации метода первый этап может выполняться с использованием различных методов. Рассмотрим применение на первом этапе итерационной формулы метода Гаусса-Зейделя. Приведем формулировку этой формулы, несколько отличную от формулировки, рассмотренной в параграфе 3.1.

Положим $X_0^{r+1} = X^r$, $X_n^{r+1} = X^{r+1}$ и пусть $e_1^r, e_2^r, \dots, e_n^r$ - орты системы координат, используемой на r -ой итерации. Тогда итерационную формулу метода Гаусса-Зейделя можно записать в виде

$$X_i^{r+1} = X_{i-1}^{r+1} + \lambda_i^r e_i^r = X_{i-1}^{r+1} + q_i^r, i=1,2,\dots,n, \quad (3.3.6)$$

где коэффициенты λ_i^r находятся из условий

$$\min_{\lambda \in (-\infty, +\infty)} \Phi(X_{i-1}^{r+1} + \lambda e_i^r) = \Phi(X_{i-1}^{r+1} + \lambda_i^r e_i^r) = \Phi(X_i^{r+1}), \quad i=1,2,\dots,n. \quad (3.3.7)$$

На втором этапе каждой из итераций система векторов $e_1^r, e_2^r, \dots, e_n^r$ с использованием ортогонализации Грамма-Шмидта заменяется новой системой линейно независимых векторов $e_1^{r+1}, e_2^{r+1}, \dots, e_n^{r+1}$.

Схема метода Розенброка.

1. Задаем начальную точку X^0 , полагаем $r=0$, $i=1$, и орты исходной системы координат обозначаем $e_1^0, e_2^0, \dots, e_n^0$.
2. Исходя из точки X^r по формулам (3.3.6), (3.3.7) выполняем одну итерацию по методу Гаусса-Зейделя – получаем точку X^{r+1} и совокупность векторов $q_1^r, q_2^r, \dots, q_n^r$.
3. Если одно из стандартных условий окончания итераций

$$\|X^{r+1} - X^r\| \leq \varepsilon_x, \quad (3.3.8)$$

$$|\Phi(X^{r+1}) - \Phi(X^r)| \leq \varepsilon_\Phi \quad (3.3.9)$$

выполнено, то полагаем $X^* \approx X^{r+1}$, и заканчиваем вычисления. Иначе переходим к п.4.

4. На основе векторов $q_1^r, q_2^r, \dots, q_n^r$ находим векторы $p_1^r, p_2^r, \dots, p_n^r$:

$$p_i^r = \sum_{j=1}^n q_j^r, \quad i=1,2,\dots,n. \quad (3.3.10)$$

5. С помощью процедуры ортогонализации Грамма-Шмидта (3.3.2) –(3.3.5) выполняем переход от системы векторов $p_1^r, p_2^r, \dots, p_n^r$ к системе векторов

$e_1^{r+1}, e_2^{r+1}, \dots, e_n^{r+1}$, полагаем $r=r+1$ и переходим к п. 2. Заметим, что из формулы (3.3.10) следует равенство $p_n^r = q_n^r$.

По сравнению с методом Гаусса-Зейделя и методом Хука-Дживса метод Розенброка имеет, как правило, более высокую эффективность на овражных функциях с не прямолинейным оврагом.

Метод Розенброка иллюстрирует рис. 3.3.1, на котором показаны линии уровня функции Химмельблау ($n=2$),

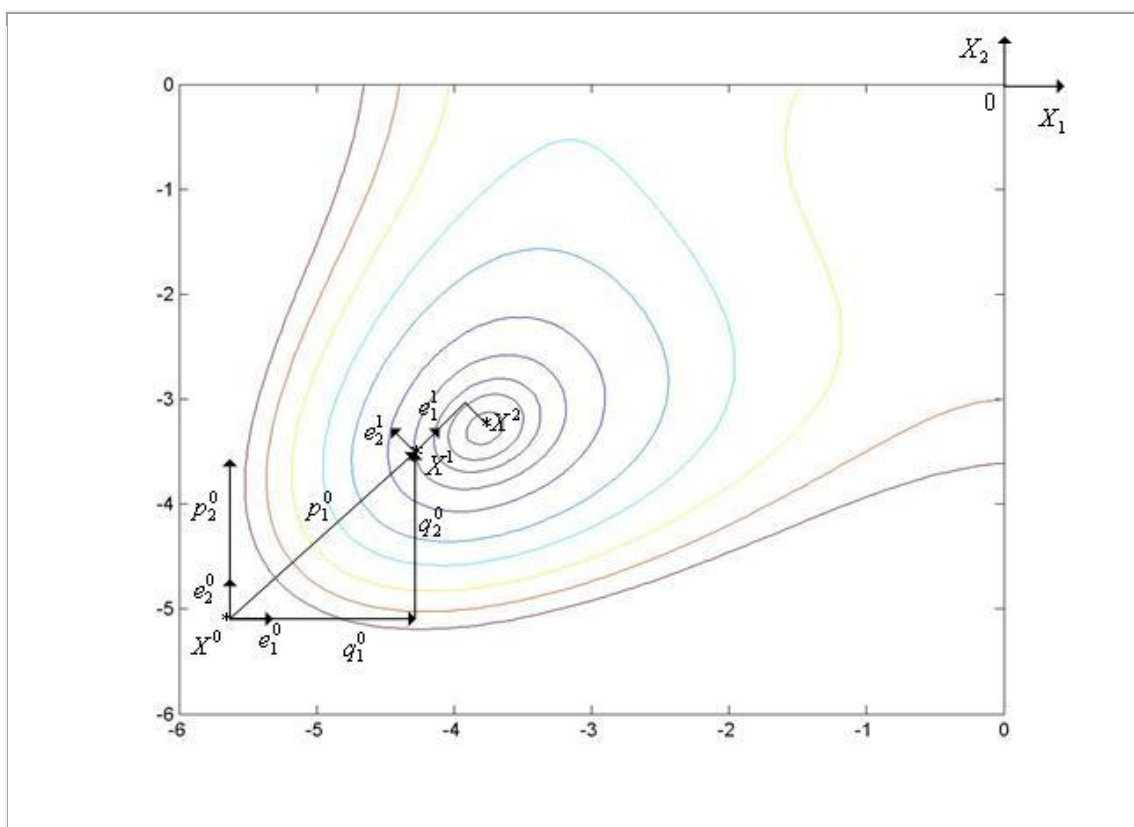


Рис. 3.3.1. Траектория поиска минимума функции Химмельблау методом Розенброка

3.4. Метод сопряженных направлений

Рассмотрим следующую задачу локальной безусловной оптимизации: найти минимум критерия оптимальности $\Phi(X)$, определенного в n -мерном евклидовом пространстве R^n , $\min_{X \in R^n} \Phi(X) = \Phi(X^*) = \Phi^*$.

Введем прежде следующие понятия: векторы p_1, p_2, \dots, p_k , принадлежащие пространству R^n , называются векторами сопряженными относительно матрицы $A(n \times n)$, если $(Ap_i p_j) = 0$ для всех $i \neq j, i, j \in [1, \dots, k]$.

В методе сопряженных направлений применяется итерационная формула метода Гаусса-Зейделя в виде, близком к использованному в параграфе 3.3.

Положим $X_0^{r+1} = X^r$ и пусть e_1, e_2, \dots, e_n - орты используемой системы координат. Тогда итерационную формулу метода Гаусса-Зейделя можно записать в виде

$$X_i^{r+1} = X_{i-1}^{r+1} + \lambda_i^r e_i = X_{i-1}^{r+1} + p_i^r, i = 1, 2, \dots, n, \quad (3.4.1)$$

где коэффициенты λ_i^r находятся из условий

$$\min_{\lambda \in (-\infty, +\infty)} \Phi(X_{i-1}^{r+1} + \lambda e_i) = \Phi(X_{i-1}^{r+1} + \lambda_i^r e_i) = \Phi(X_i^{r+1}), i = 1, 2, \dots, n. \quad (3.4.2)$$

Схема метода сопряженных направлений.

1. Задаем начальную точку X^0 и полагаем $r = 0, i = 1$.
2. Последовательно для $i = 1, 2, \dots, n$ по формулам (3.4.1), (3.4.2) находим точки $X_1^{r+1}, X_2^{r+1}, \dots, X_n^{r+1}$.
3. Исходя из точки X_n^{r+1} , еще раз находим минимум функции $\Phi(X)$ вдоль первого координатного направления - вычисляем координаты точки

$$X_{n+1}^{r+1} = X_n^{r+1} + \tilde{\lambda}_1^r e_1, \quad (3.4.3)$$

где коэффициент $\tilde{\lambda}_1^r$ находится из условия

$$\min_{\lambda \in (-\infty, +\infty)} \Phi(X_{n+1}^{r+1} + \lambda e_1) = \Phi(X_n^{r+1} + \tilde{\lambda}_1^r e_1) = \Phi(X_{n+1}^{r+1}). \quad (3.4.4)$$

4. Исходя из точки X_1^{r+1} , находим минимум функции $\Phi(X)$ вдоль вектора $p_{n+1}^r = X_{n+1}^{r+1} - X_1^{r+1}$, вычисляем

$$X^{r+1} = X_1^r + \lambda_{n+1}^r p_{n+1}^r, \quad (3.4.5)$$

где коэффициент λ_{n+1}^r находится из условия

$$\min_{\lambda \in (-\infty, +\infty)} \Phi(X_1^{r+1} + \lambda p_{n+1}^r) = \Phi(X^r + \lambda_{n+1}^r p_{n+1}^r) = \Phi(X^{r+1}). \quad (3.4.6)$$

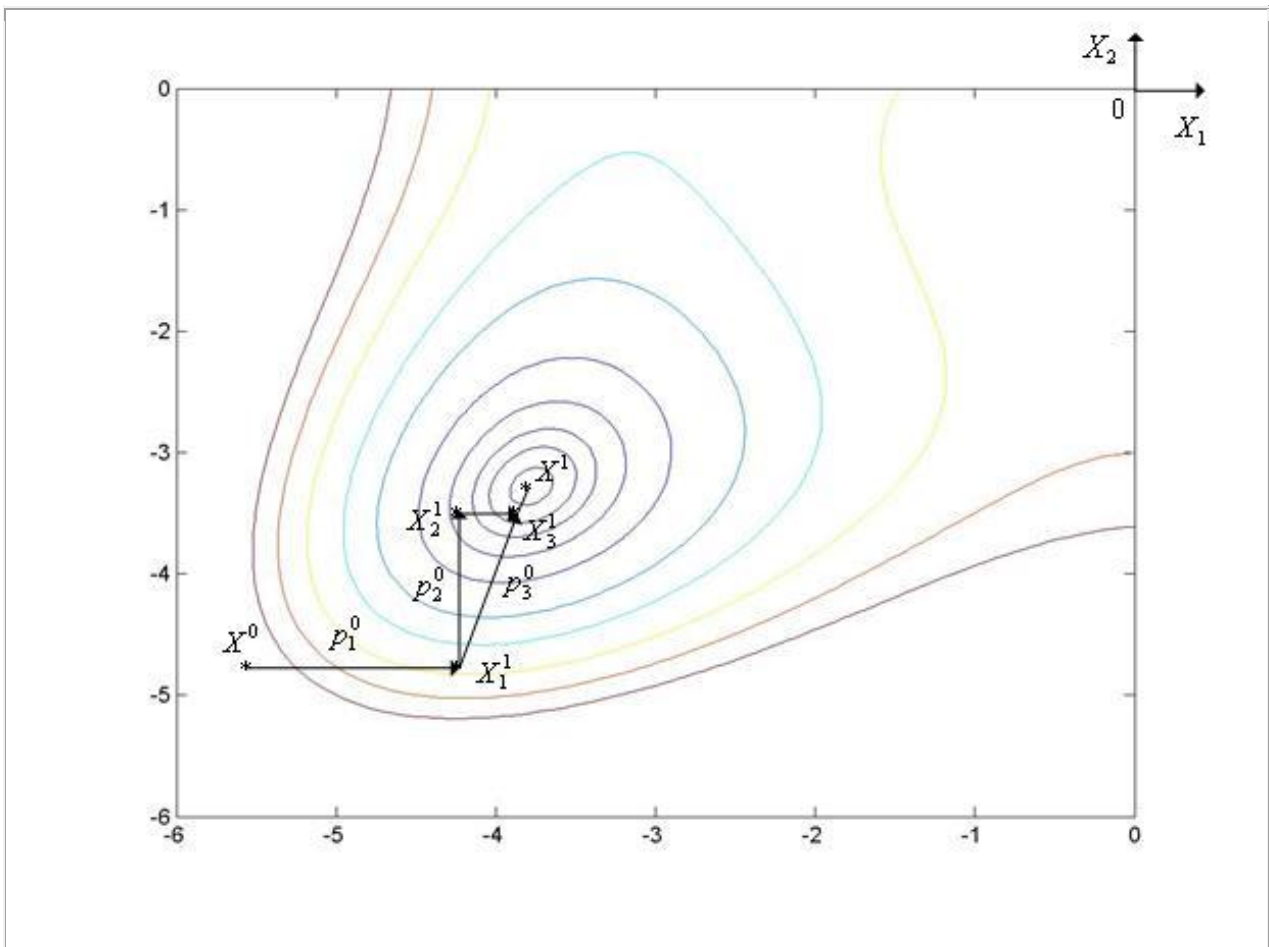


Рис. 3.4.1. Траектория поиска минимума функции Химмельблау методом сопряженных направлений

5. Если одно из стандартных условий окончания итераций

$$\|X^{r+1} - X^r\| \leq \varepsilon_x, \quad (3.4.7)$$

$$|\Phi(X^{r+1}) - \Phi(X^r)| \leq \varepsilon_\Phi \quad (3.4.8)$$

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

Метод сопряженных направлений иллюстрирует рис. 3.4.1, на котором показаны линии уровня функции Химмельблау ($n = 2$).

Рассмотрим двумерную квадратичную функцию

$$\Phi(x_1, x_2) = x_1^2 + x_2^2 + 3(x_1 + x_2)^2 = 4x_1^2 + 4x_2^2 + 6x_1x_2. \quad (3.4.9)$$

На рис. 3.4.2 показаны линии уровня двумерной квадратичной функции и траектория поиска минимума квадратичной функции (3.4.9) методом сопряженных направлений.

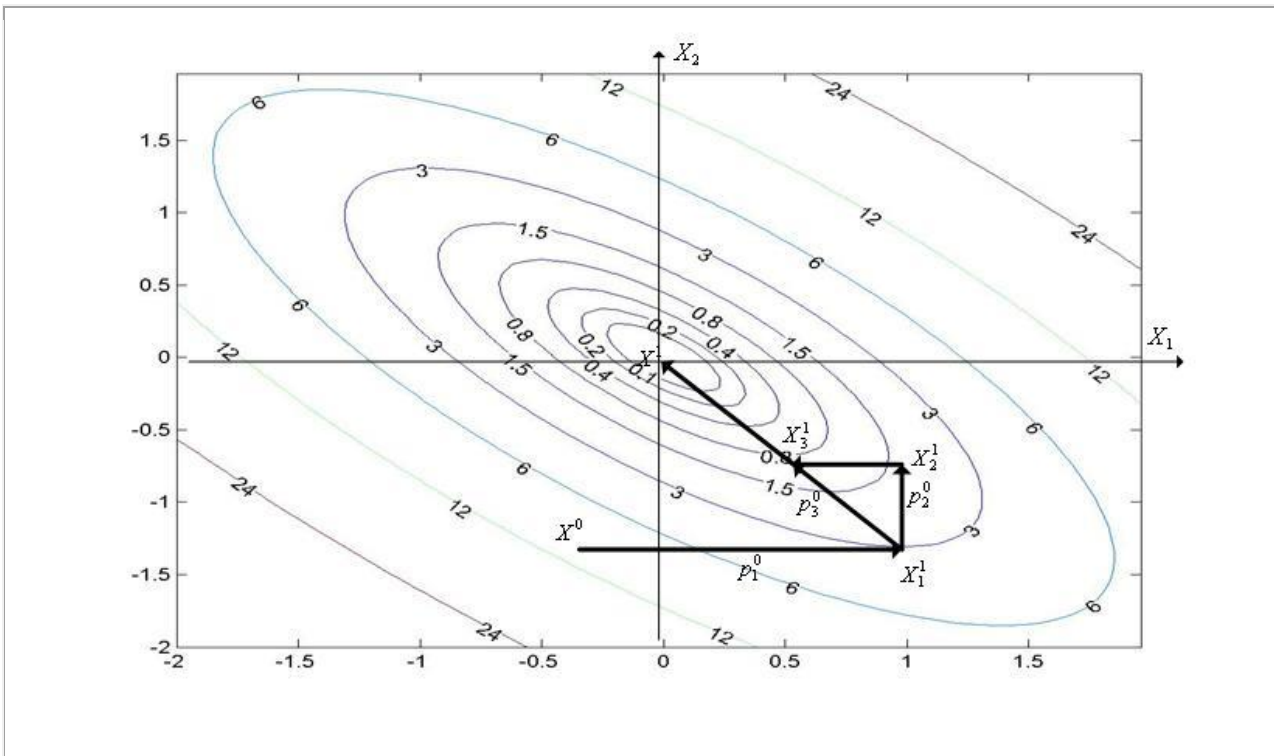


Рис. 3.4.2. Траектория поиска минимума квадратичной функции (3.4.9) методом сопряженных направлений

Произвольную n -мерную квадратичную функцию можно записать в виде

$$\Phi(X) = \frac{1}{2}(AX, X^T) + (b, X) + c, \quad (3.4.10)$$

где A – квадратная $n \times n$ матрица, b – $n \times 1$ столбец, c – скалярная константа.

Например, если положить

$$A = \begin{pmatrix} 8 & 6 \\ 6 & 8 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad c = 0,$$

то имеем функцию (3.4.9):

$$\begin{aligned} \Phi(X) &= \frac{1}{2}(AX, X^T) + (b, X) + c = \frac{1}{2}((8x_1 + 6x_2 \quad 6x_1 + 8x_2), (x_1, x_2)) = \\ &= \frac{1}{2}(8x_1^2 + 6x_1x_2 + 6x_1x_2 + 8x_2^2) = 4x_1^2 + 4x_2^2 + 6x_1x_2. \end{aligned}$$

Утверждение 3.4.1. В случае минимизации двумерной квадратичной функции (3.4.10) методом сопряженных направлений, направления p_1^r, p_3^r являются A -ортогональными.

Доказательство (см. рис. 3.4.2). По определению A -ортогональности для доказательства утверждения достаточно показать, что скалярное произведение

$$(Ap_3^r, p_1^r) = (A(X_3^r - X_1^r), p_1^r) = 0. \quad (3.4.11)$$

Легко видеть, что производная функции (3.4.10) равна

$$\Phi'(X) = AX + b.$$

Поэтому

$$A(X_3^r - X_1^r) = AX_3^r - AX_1^r = \Phi'(X_3^r) - \Phi'(X_1^r).$$

Подставляя этот результат в выражение (3.4.11), получим

$$(Ap_3^r, p_1^r) = \left((\Phi'(X_3^r) - \Phi'(X_1^r)), p_1^r \right) = (\Phi'(X_3^r), p_1^r) - (\Phi'(X_1^r), p_1^r) = 0.$$

Последнее равенство следует из ортогональности пар векторов $(\Phi'(X_3^r), p_1^r)$, $(\Phi'(X_1^r), p_1^r)$.

Утверждение 3.4.1 объясняет название рассмотренного метода.

Заметим, что при минимизации квадратичной функции методом сопряженных направлений минимум достигается за одну итерацию.

3.5. Симплекс-метод

Рассмотрим следующую задачу локальной безусловной оптимизации: найти минимум критерия оптимальности $\Phi(X)$, определенного в n -мерном евклидовом пространстве R^n , $\min_{X \in R^n} \Phi(X) = \Phi(X^*) = \Phi^*$.

Регулярным симплексом в пространстве R^n называется правильный многогранник, образованный $(n+1)$ -ой равноотстоящими друг от друга вершинами. Для случая $n=2$ – это равносторонний треугольник, для случая $n=3$ – тетраэдр.

Если в пространстве R^n необходимо построить регулярный симплекс, одна из вершин которого находится в точке $X^0 = (x_1^0, x_2^0, \dots, x_n^0)$, то координаты вершин такого симплекса удобно задавать с помощью $n \times (n+1)$

матрицы
$$R_{n \times (n+1)} = \begin{bmatrix} x_1^0 & x_1^0 + r_1 & x_1^0 + r_2 & \dots & x_1^0 + r_n \\ x_2^0 & x_2^0 + r_1 & x_2^0 + r_2 & \dots & x_2^0 + r_n \\ x_3^0 & x_3^0 + r_1 & x_3^0 + r_2 & \dots & x_3^0 + r_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n^0 & x_n^0 + r_1 & x_n^0 + r_2 & \dots & x_n^0 + r_n \end{bmatrix}. \quad (3.5.1)$$

Здесь i -й столбец представляет собой координаты i -й вершины симплекса, $i \in [1, \dots, n+1]$;

$$r_1 = l \frac{\sqrt{n+1} + n - 1}{n\sqrt{2}}, \quad r_2 = l \frac{\sqrt{n+1} - 1}{n\sqrt{2}}, \quad (3.5.2)$$

l -длина ребра симплекса.

Например, регулярный симплекс в двумерном пространстве R^2 с одной из вершин в начале координат (когда $x_1^0 = 0, x_2^0 = 0$) определяется

(2×3) матрицей $R_{2 \times 3} = \begin{pmatrix} 0 & r_1 & r_2 \\ 0 & r_2 & r_1 \end{pmatrix}$ и имеет вид, представленный на

рис. 3.5.1.

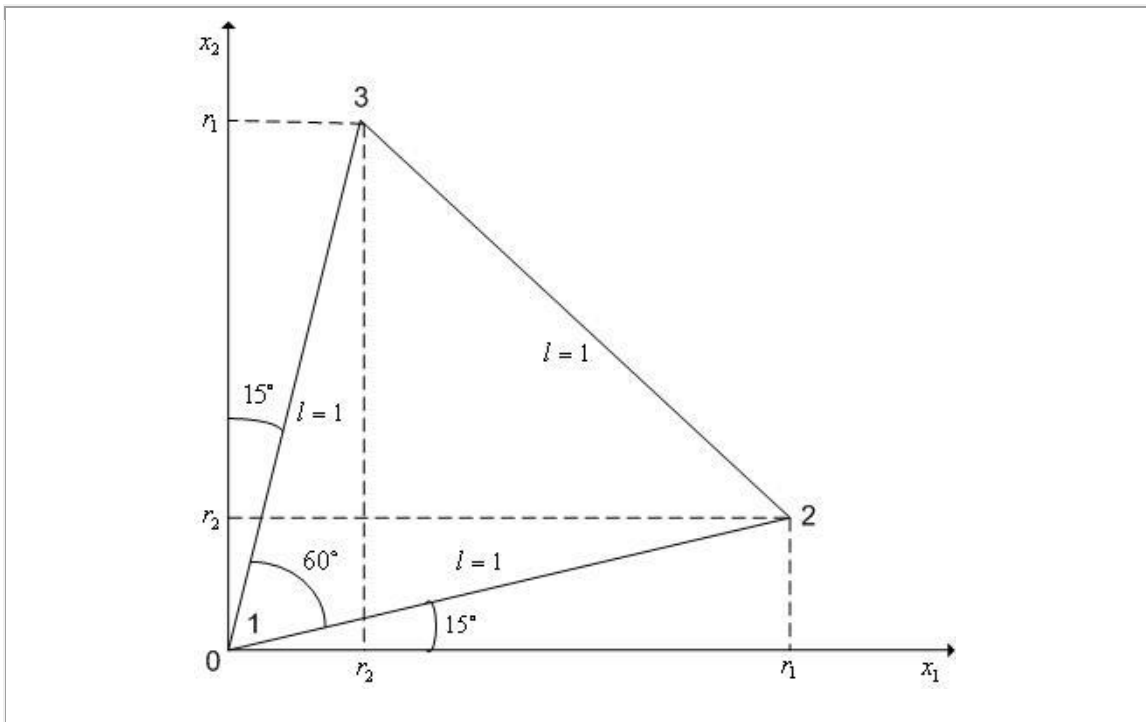


Рис. 3.5.1. Регулярный симплекс в пространстве R^2 с одной из вершин в начале координат

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 (см. рис. 3.5.2).

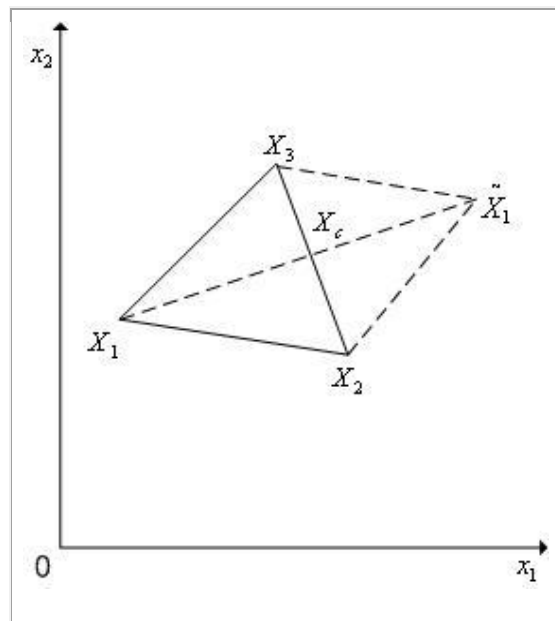


Рис. 3.5.2. Отражение вершины X_1 регулярного симплекса в пространстве R^2 относительно центра тяжести X_c остальных вершин

Будем называть эту процедуру **отражением вершины симплекса** относительно центра тяжести остальных вершин. На рис. 3.5.2 показано отражение вершины X_1 регулярного симплекса в пространстве R^2 относительно центра тяжести X_c остальных вершин.

Пусть $X_i^r, i \in [1, \dots, n+1]$ - векторы координат вершин регулярного симплекса. Тогда при выполнении операции отражения k -й вершины симплекса имеет место следующая связь координат этой вершины и новой вершины:

$$\frac{X_k^{r+1} + X_k^r}{2} = X_c^r. \quad (3.5.3)$$

Здесь

$$X_c^r = \frac{1}{n} \sum_{i=1, i \neq k}^{n+1} X_i^r \quad (3.5.4)$$

- вектор координат центра тяжести остальных вершин симплекса (за исключением отраженной вершины k).

Таким образом, после отражения k -ой вершины симплекса с координатами вершин $X_i^r, i \in [1, \dots, n+1]$, получаем новый симплекс с координатами вершин

$$X_i^{r+1} = X_i^r, i \in [1, \dots, n+1], i \neq k, X_k^{r+1} = 2X_c^r - X_k^r. \quad (3.5.5)$$

Кроме операции отражения вершины симплекса симплекс-метод может использовать операцию **редукции симплекса** - уменьшение длин всех ребер симплекса на одну и ту же величину. На рис 3.5.3 показана редукция вершин регулярного симплекса в пространстве R^2 к вершине X_1 .

Пусть $X_i^r, i \in [1, \dots, n+1]$ - векторы координат вершин регулярного симплекса. Тогда при выполнении операции редукции вершин этого симплекса к вершине X_k новые координаты остальных вершин симплекса определяются по формуле

$$X_i^{r+1} = X_k^r + \gamma(X_i^r - X_i^k), i \in [1, \dots, n+1], i \neq k,$$

где $\gamma \in (0,1)$ - коэффициент редукции. Рекомендуется использовать $\gamma = 0.5$.

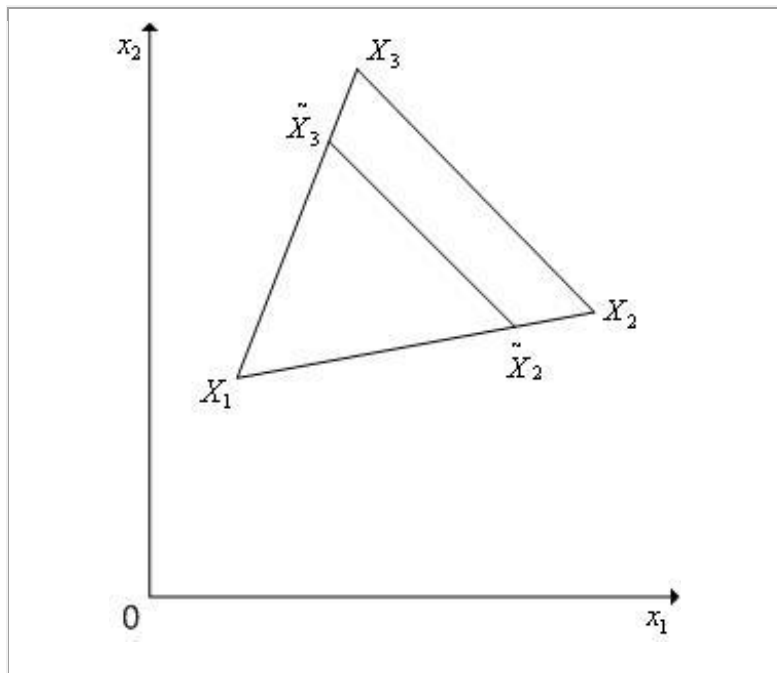


Рис. 3.5.3. редукция вершин регулярного симплекса в пространстве R^2 к вершине X_1

Таким образом, после редукции вершин симплекса $X_i^r, i \in [1, \dots, n+1]$ к вершине X_k получаем новый симплекс с координатами вершин

$$X_i^{r+1} = X_k^r + \gamma(X_i^r - X_i^k), i \in [1, \dots, n+1], i \neq k, \quad X_k^{r+1} = X_k^r. \quad (3.5.6)$$

Суть симплекс-метода раскрывает его простейший вариант.

Схема простейшего варианта симплекс-метода.

1. Задаем начальную точку X^0 , длину ребра симплекса l и полагаем $r=0$.
2. По формулам (3.5.1), (3.5.2) находим координаты $X_i^r, i \in [1, \dots, n+1]$ всех вершин симплекса.
3. Вычисляем значения $\Phi(X_i^r), i \in [1, \dots, n+1]$ минимизируемой функции во всех вершинах симплекса.

4. Находим максимальное из значений функции $\Phi(X)$ в вершинах симплекса $X_i^r, i \in [1, \dots, n+1]$

$$\max\left(\Phi(X_1^r), \Phi(X_2^r), \dots, \Phi(X_{n+1}^r)\right) = \Phi(X_k^r).$$

5. По формулам (3.5.4), (3.5.5) отражаем вершину X_k^r относительно центра тяжести остальных вершин симплекса – получаем новый симплекс с координатами вершин $X_i^{r+1}, i \in [1, \dots, n+1]$.

6. Вычисляем значение $\Phi(X_k^{r+1})$ минимизируемой функции в новой вершине симплекса.

7. Если условие окончания итераций (см. ниже) выполнено, то в качестве приближенного значения точки минимума функции $\Phi(X)$ принимаем ту вершину симплекса $X_i^{r+1} = X_i^r, i \in [1, \dots, n+1]$, в которой $\Phi(X)$ имеет минимальное значение, и заканчиваем вычисления. Иначе полагаем $r = r + 1$ и переходим к п. 4.

Поскольку размер симплекса в простейшем варианте симплекс-методе фиксирован, в качестве условия окончания итераций в данном случае можно использовать условие

$$\max_{i \in [1, \dots, n+1], i \neq k} \left| \Phi(X_i^{r+1}) - \Phi(X_k^{r+1}) \right| \leq \varepsilon_\Phi, \quad (3.5.7)$$

где ε_Φ - требуемая точность решения по Φ , $k \in [1, \dots, n+1]$ - номер произвольной вершины симплекса. Отметим, что выражение в левой части неравенства (3.5.7) есть максимальная разность значений функции $\Phi(X)$ в двух вершинах симплекса $X_i^{r+1} = X_i^r, i \in [1, \dots, n+1]$.

Простейший вариант симплекс-метода иллюстрирует рис. 3.5.4, на котором показаны линии уровня функции Химмельблау ($n = 2$).

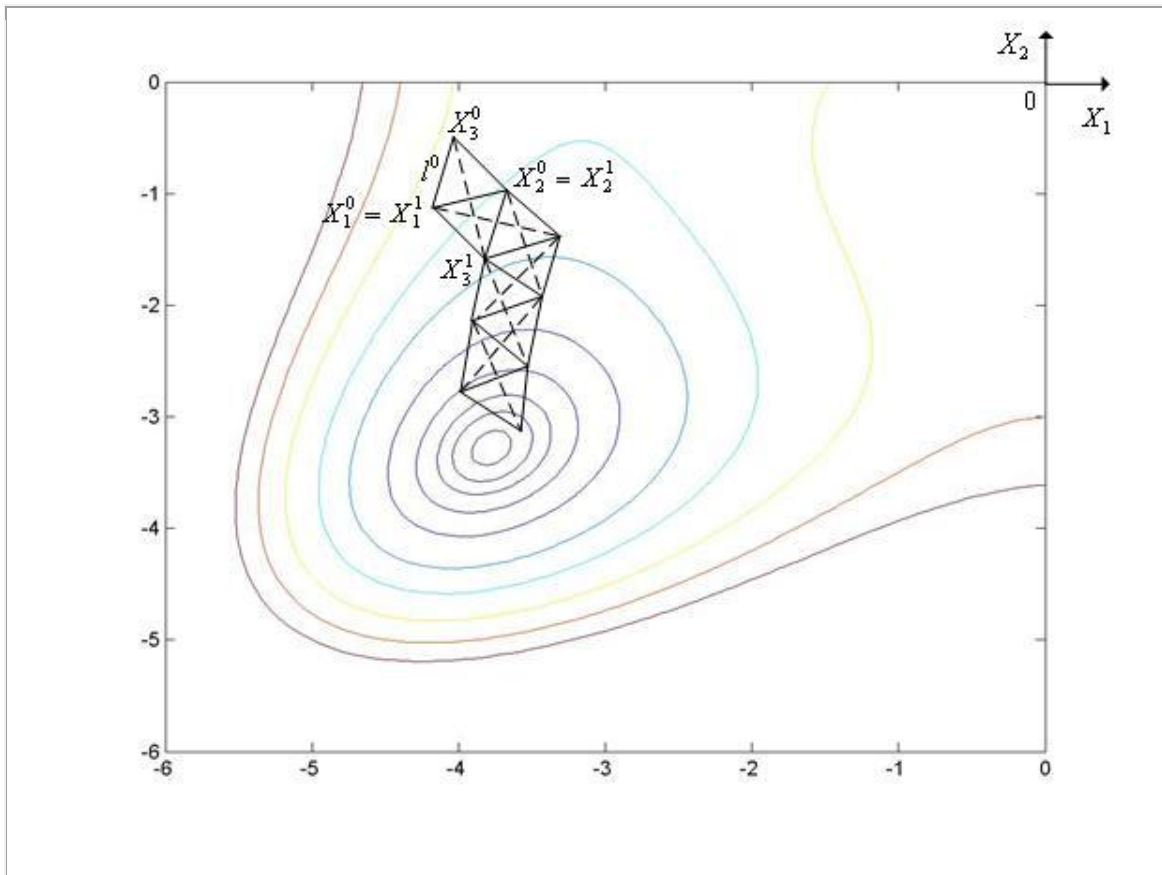


Рис. 3.5.4. Траектория поиска минимума функции Химмельблау простейшим симплекс-методом

Рассмотренный простейший симплекс-метод склонен к зацикливанию и медленно сходится, если длина ребра симплекса l выбрана малой (выбор же 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 идеей модифицированного симплекс-метода является изменение по некоторому правилу размера симплекса в процессе поиска. При этом наряду с условием (3.5.7) в качестве условия окончания итераций можно использовать условие

$$l^r \leq \varepsilon_x, \quad (3.5.8)$$

где l^r - текущая длина ребра симплекса, ε_x - требуемая точность решения по X .

Обычно размер симплекса изменяется при выполнении следующих условий:

- при «накрытии» симплексом дна оврага или точки минимума;
- при циклическом движении.

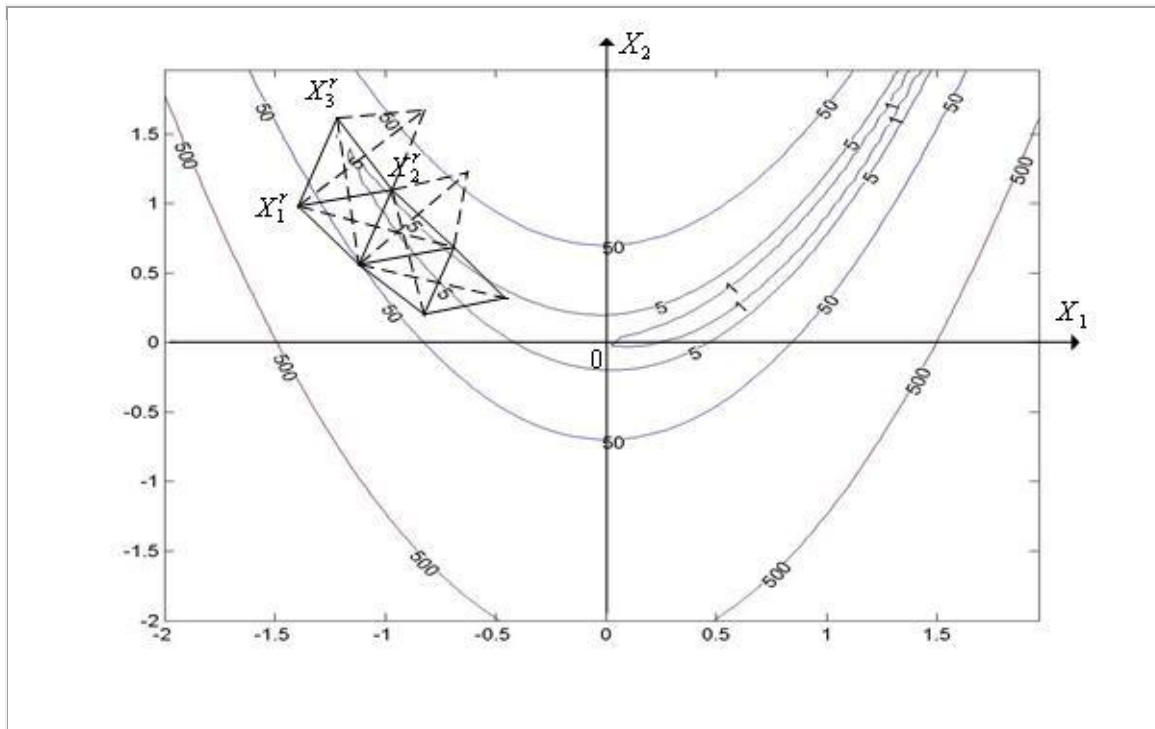


Рис. 3.5.5. Траектория поиска минимума функции Розенброка модифицированным симплекс-методом при «накрытии» дна оврага. Пунктиром показаны отвергнутые симплексы

«Накрытие» симплексом дна оврага или точки минимума. Пусть X_k^{r+1} - вершина, которая получилась на r -ой итерации в результате отражения вершины X_k^r . Так что координаты вершин нового симплекса равны $X_i^{r+1} = X_i^r, i \in [1, \dots, n+1], i \neq k, X_k^{r+1}$.

Ситуация $\Phi(X_k^{r+1}) > \Phi(X_k^r)$ интерпретируется как «накрытие» этим симплексом дна оврага или точки минимума и простейший симплекс-метод модифицируется следующим образом (см. рис. 3.5.5):

1. Полагаем $k = k + 1$ (если $k = n + 2$, то полагаем $k = 1$);
2. Выполняем отражение k -ой вершины симплекса $X_i^r, i \in [1, \dots, n + 1]$;
3. Если $\Phi(X_k^{r+1}) > \Phi(X_k^r)$ и не все вершины перебраны, то переходим к п.1.
4. Иначе - продолжаем итерации по схеме простейшего симплекс-метода.

На рис.3.5.5 показана траектория поиска минимума функции Розенброка модифицированным симплекс-методом при «накрытии» dna оврага. Пунктиром показаны отвергнутые симплексы.

Циклическое движение. Ситуация, когда некоторая вершина симплекса не исключается на протяжении m итераций, интерпретируется как «зацикливание» алгоритма.

Простейший симплекс-метод модифицируется в этом случае следующим образом:

1. Находим вершину текущего симплекса $X_i^r, i \in [1, \dots, n + 1]$, в которой функция $\Phi(X)$ принимает наименьшее значение

$$\min\left(\Phi(X_1^0), \Phi(X_2^r), \dots, \Phi(X_{n+1}^r)\right) = \Phi(X_k^r).$$

2. По формуле (3.5.6) выполняем редукцию симплекса $X_i^r, i \in [1, \dots, n + 1]$ к вершине X_k^r .
3. Продолжаем итерации по схеме простейшего симплекс-метода.

Здесь количество итераций m рекомендуется находить из условия $m = [1.65n + 0.05n^2]$, где $[*]$ - символ ближайшего целого большего.

3.6. Метод деформируемого многогранника (Нелдера-Мида)

Рассмотрим следующую задачу локальной безусловной оптимизации: найти минимум критерия оптимальности $\Phi(X)$, определенного в n -мерном евклидовом пространстве R^n , $\min_{X \in R^n} \Phi(X) = \Phi(X^*) = \Phi^*$.

В случае если функция $\Phi(X)$ является овражной функцией, эффективность симплекс-метода при решении задачи значительно снижается в силу того, что регулярный симплекс нельзя «вытянуть» вдоль оврага.

Метод Нелдера-Мида (метод деформируемого многогранника) 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;
- сжатие;
- растяжение.

Отражение вершины симплекса относительно центра тяжести остальных вершин (см. рис. 3.6.1). В результате отражения k -й вершины симплекса с координатами вершин $X_i^r, i \in [1, \dots, n+1]$, получаем новый симплекс с координатами вершин

$$X_i^{r+1} = X_i^r, i \in [1, \dots, n+1], i \neq k, X_k^{r+1} = 2X_c^r - X_k^r, \quad (3.6.1)$$

где

$$X_c^r = \frac{1}{n} \sum_{i=1, i \neq k}^{n+1} X_i^r \quad (3.6.2)$$

- вектор координат центра тяжести остальных вершин симплекса (за исключением отраженной вершины k).

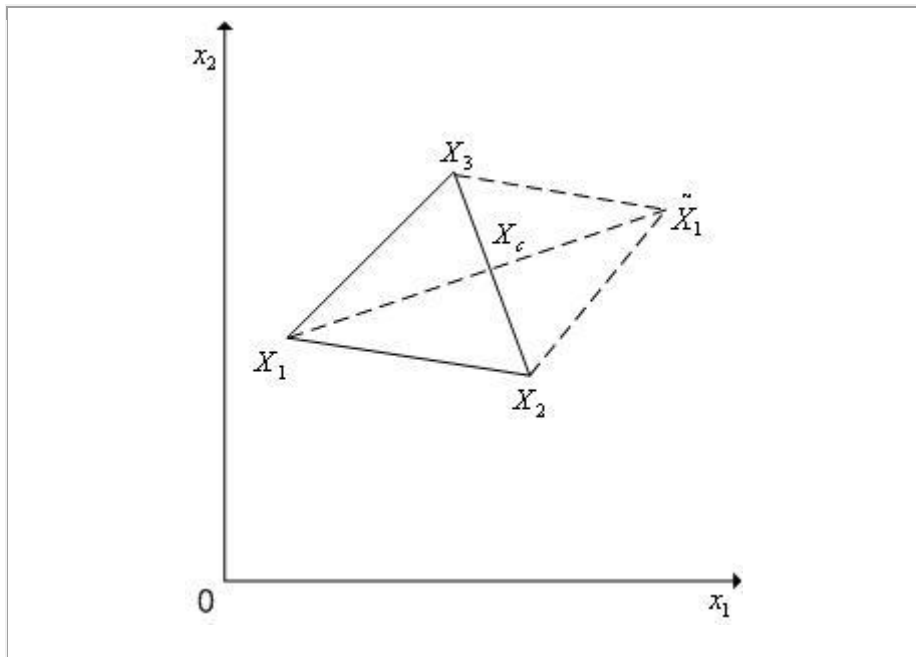


Рис. 3.6.1. Отражение вершины X_1 симплекса в пространстве R^2 относительно центра тяжести X_c остальных вершин

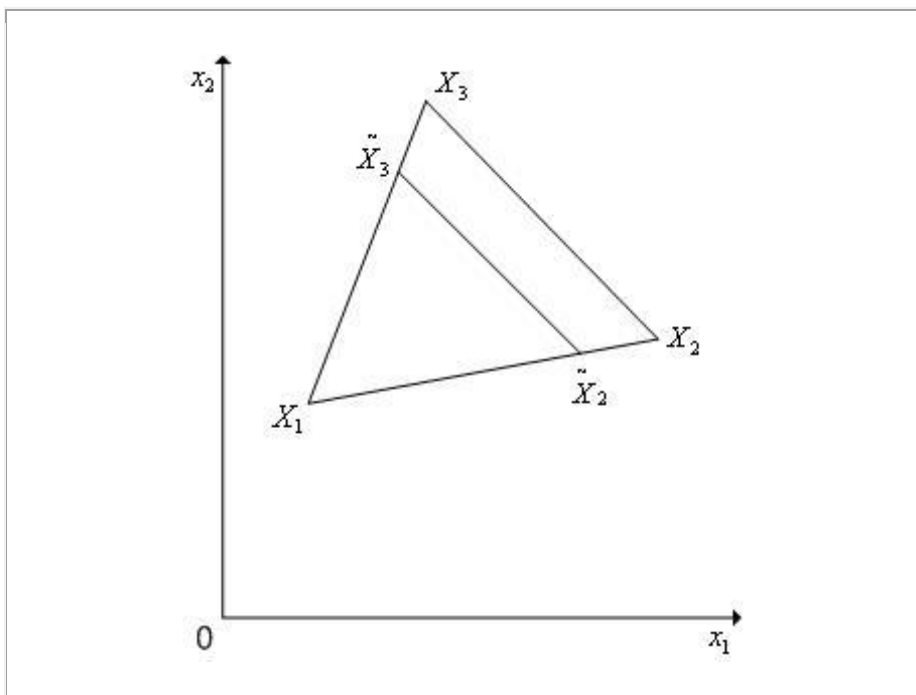


Рис. 3.6.2. Редукция вершин симплекса в пространстве R^2 к вершине X_1

Редукции симплекса (см. рис. 3.6.2) - уменьшение длин всех ребер симплекса в одно и то же количество раз. В результате выполнения редукции вершин симплекса $X_i^r, i \in [1, \dots, n+1]$ к вершине X_k получаем новый симплекс с координатами вершин

$$X_i^{r+1} = X_k^r + \gamma(X_i^r - X_i^k), i \in [1, \dots, n+1], i \neq k, \quad X_k^{r+1} = X_k^r, \quad (3.6.3)$$

где $\gamma \in (0,1)$, $\gamma \approx 0.5$ - коэффициент редукции.

Сжатие симплекса (см. рис. 3.6.3). В результате выполнения сжатия симплекса $X_i^r, i \in [1, \dots, n+1]$ в направлении $(X_k^r - X_c^r)$ получаем новый симплекс с координатами вершин

$$X_i^{r+1} = X_i^r, i \in [1, \dots, n+1], i \neq k, X_k^{r+1} = X_c^r + \beta(X_k^r - X_c^r), \quad (3.6.4)$$

где $\beta \in (0,1)$, $\beta \approx 0.4-0.6$ - коэффициент сжатия, X_c^r - вектор координат центра тяжести остальных вершин симплекса.

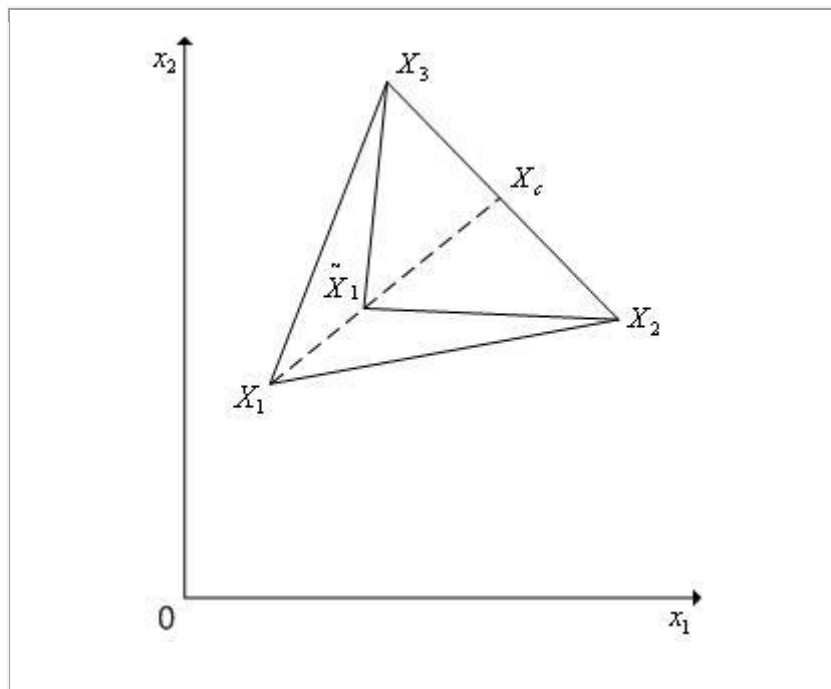


Рис. 3.6.3. Сжатие симплекса в пространстве R^2 в направлении $(X_1 - X_c)$

Растяжение симплекса (см. рис.3.6.4). В результате выполнения растяжения симплекса $X_i^r, i \in [1, \dots, n+1]$ в направлении $(X_k^r - X_c^r)$ получаем новый симплекс с координатами вершин

$$X_i^{r+1} = X_i^r, i \in [1, \dots, n+1], i \neq k, X_k^{r+1} = X_c^r + \alpha(X_k^r - X_c^r), \quad (3.6.5)$$

где $\alpha \approx 2.8-3.0$ - коэффициент растяжения, X_c^r - вектор координат центра тяжести остальных вершин симплекса.

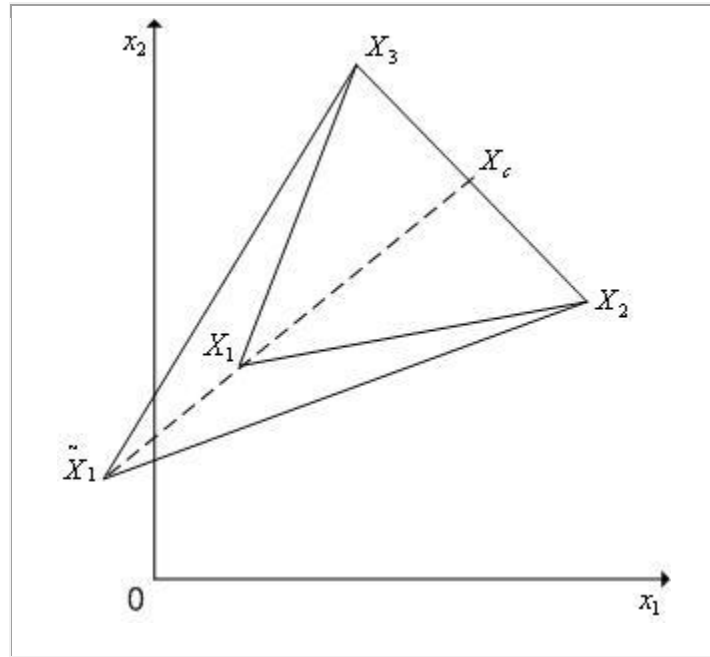


Рис. 3.6.4. Растяжение симплекса в пространстве R^2 в направлении $(X_1 - X_c)$

Схема метода Нелдера-Мида.

Симплекс с вершинами $X_i^r, i \in [1, \dots, n+1]$ обозначим S^r .

1. Задаем начальную точку X^0 , длину ребра симплекса l и полагаем $r=0$.
2. Находим координаты $X_i^0, i \in [1, \dots, n+1]$ всех вершин регулярного симплекса S^0 с длиной ребра l . Вычисляем значения $\Phi(X_i^r)$ минимизируемой функции во всех вершинах симплекса.
3. Среди вершин симплекса S^r находим вершины $X_{l=k_l}^r, X_{h=k_h}^r, X_{s=k_s}^r$, в которых функция $\Phi(X)$ принимает, соответственно, наименьшее, наибольшее и следующее за максимальным значения, а также находим значения функции $\Phi(X)$ в этих точках:

$$\Phi(X_l^r) = \min_{i \in [1, \dots, n+1]} \Phi(X_i^r),$$

$$\Phi(X_h^r) = \max_{i \in [1, \dots, n+1]} \Phi(X_i^r),$$

$$\Phi(X_s^r) = \max_{i \in [1, \dots, n+1], i \neq h} \Phi(X_i^r).$$

4. По формулам (3.6.1), (3.6.2) выполняем отражение вершину симплекса X_h^r относительно центра тяжести остальных вершин симплекса – получаем новый симплекс S^{r+1} . Вычисляем значение $\Phi(X_{k_h}^{r+1})$ минимизируемой функции в новой вершине симплекса.

5. Если условие окончания итераций (см. ниже) выполнено, то в качестве приближенного значения точки минимума функции $\Phi(X)$ принимаем ту вершину симплекса S^{r+1} , в которой $\Phi(X)$ имеет минимальное значение, и заканчиваем вычисления.

6. Если $\Phi(X_{k_h}^{r+1}) \leq \Phi(X_h^r)$ и $\Phi(X_{k_h}^{r+1}) \leq \Phi(X_l^r)$, то переходим к п.7 (растяжению симплекса) - см. рис. 3.6.5. Если $\Phi(X_{k_h}^{r+1}) \leq \Phi(X_h^r)$, но $\Phi(X_{k_h}^{r+1}) \geq \Phi(X_l^r)$, то переходим к п.3 (отражению) – см. рис. 3.6.6. Если $\Phi(X_{k_h}^{r+1}) \geq \Phi(X_h^r)$, то переходим к п.8 (сжатию симплекса) – см. рис. 3.6.7, рис. 3.6.8.

7. Ситуация $\Phi(X_{k_h}^{r+1}) \leq \Phi(X_h^r)$ и $\Phi(X_{k_h}^{r+1}) \leq \Phi(X_l^r)$. По формуле (3.6.5) выполняем растяжение симплекса S^{r+1} в направлении $(X_{k_h}^{r+1} - X_c^{r+1})$ - получаем новый симплекс S^{r+2} . Вычисляем значение минимизируемой функции в новой вершине симплекса $X_{k_h}^{r+2}$. Если $\Phi(X_{k_h}^{r+2}) \leq \Phi(X_l^r)$, то полагаем $r = r + 2$ и переходим к п.3 (отражению вершины симплекса).

Иначе полагаем $r = r + 1$ и переходим к п.3 (отражению вершины симплекса) с симплексом S^r (т.е. не используем результаты растяжения).

8. Ситуация $\Phi(X_{k_h}^{r+1}) \geq \Phi(X_h^r)$. По формуле (3.6.4) выполняем сжатие симплекса S^{r+1} в направлении $(X_{k_h}^{r+1} - X_c^{r+1})$ - получаем новый симплекс S^{r+2} . Вычисляем значение минимизируемой функции в новой вершине симплекса $X_{k_h}^{r+2}$. Если $\Phi(X_{k_h}^{r+2}) \leq \Phi(X_{k_h}^r)$, то полагаем $r = r + 2$ и переходим к п.3 (отражению вершины симплекса). Иначе по формуле (3.6.3) выполняем редукцию симплекса S^r к вершине $X_{l=k_l}^r$ - получаем новый симплекс S^{r+1} . Вычисляем значение минимизируемой функции во всех новых вершинах симплекса S^{r+1} . Полагаем $r = r + 1$ и переходим к п.3 (отражению симплекса).

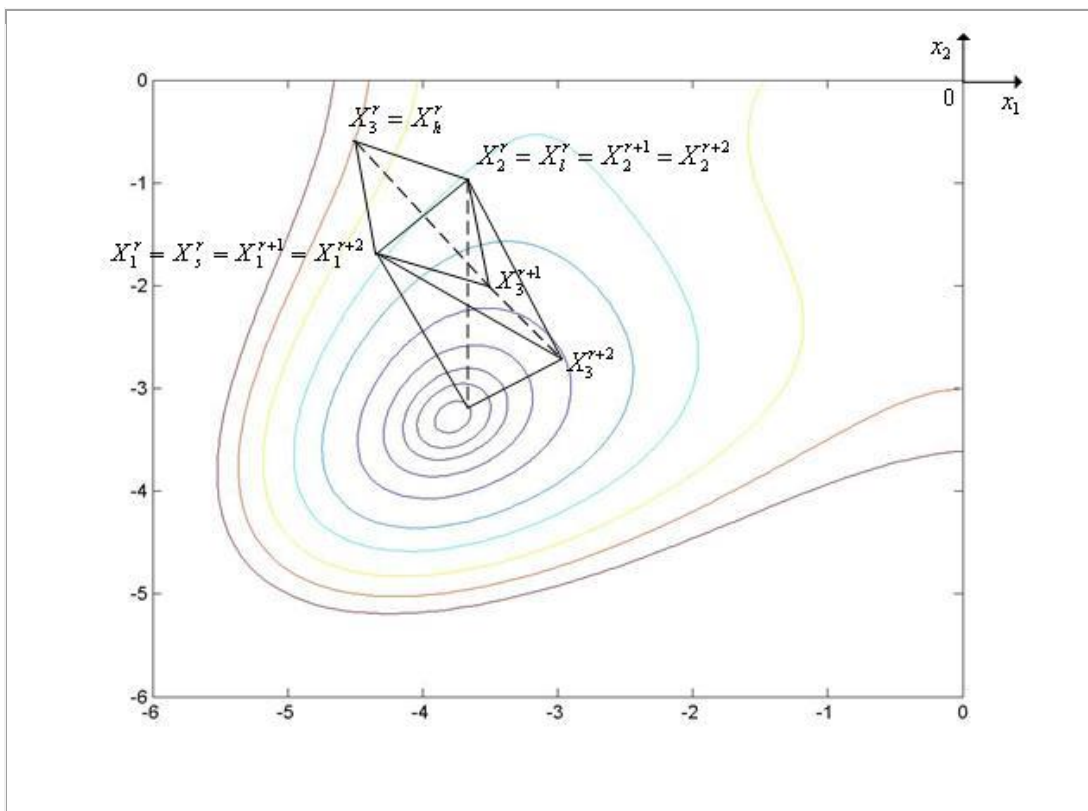


Рис. 3.6.5. Ситуация, когда метод Нелдера-Мида использует успешное растяжение симплекса

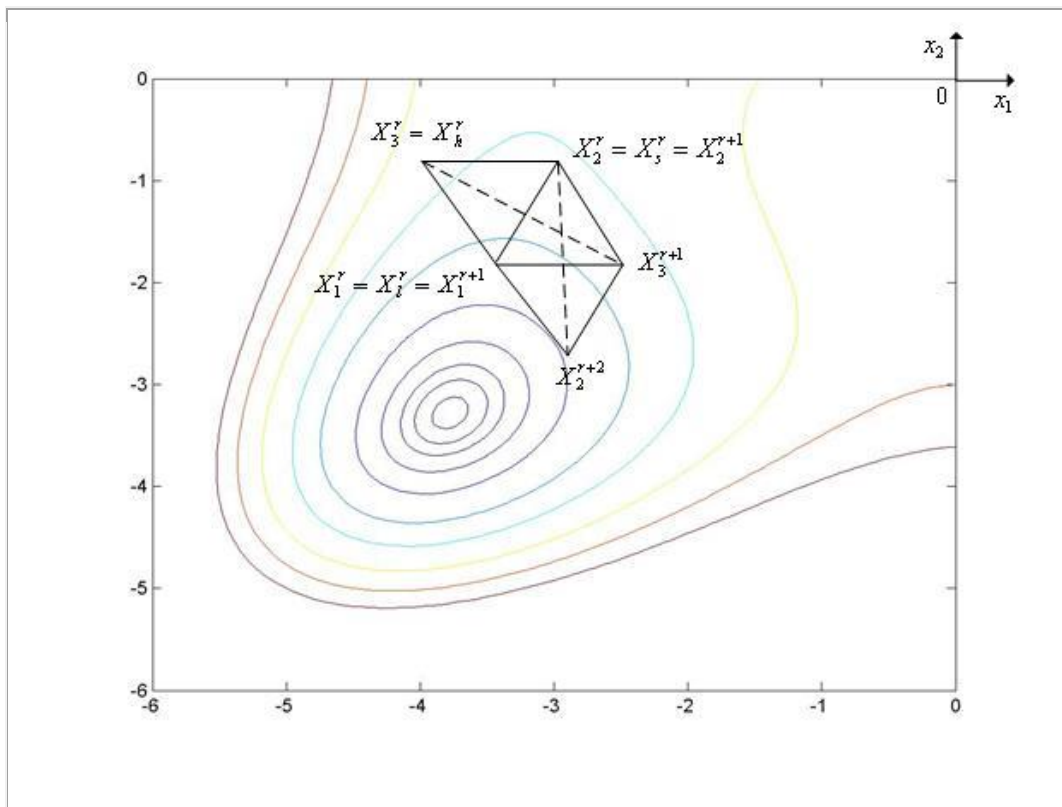


Рис. 3.6.6. Ситуация, когда метод Нелдера-Мида использует успешное отражение симплекса

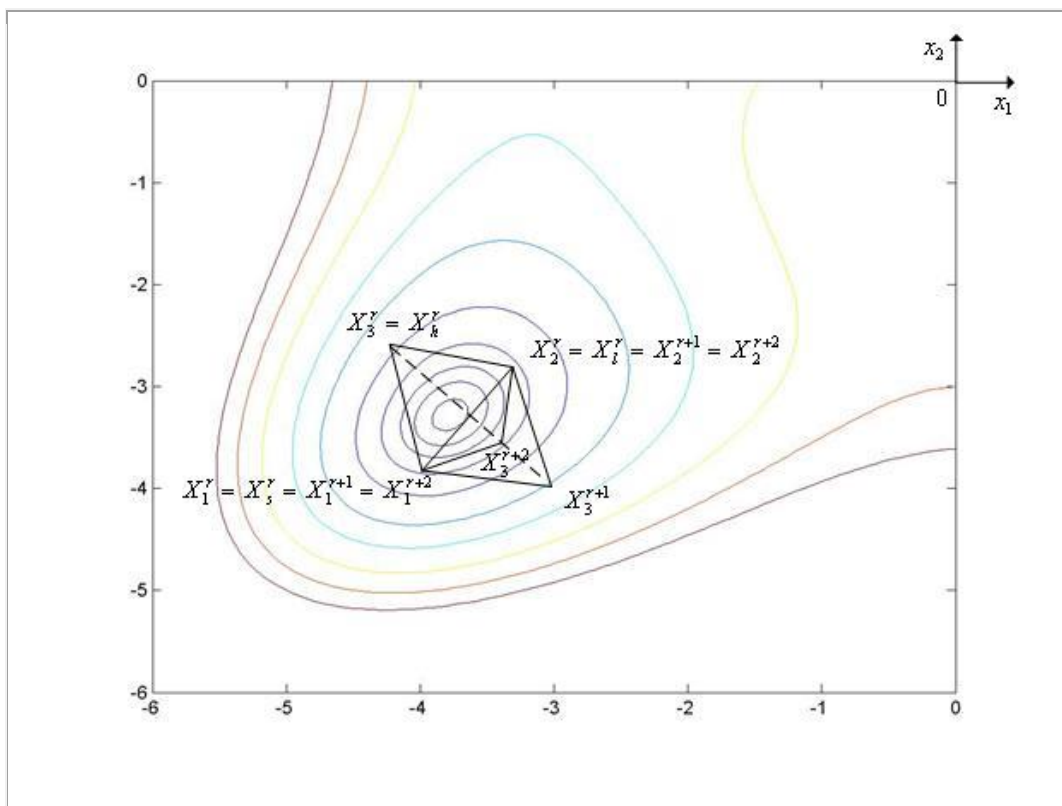


Рис. 3.6.7. Ситуация, когда метод Нелдера-Мида использует успешное сжатие симплекса

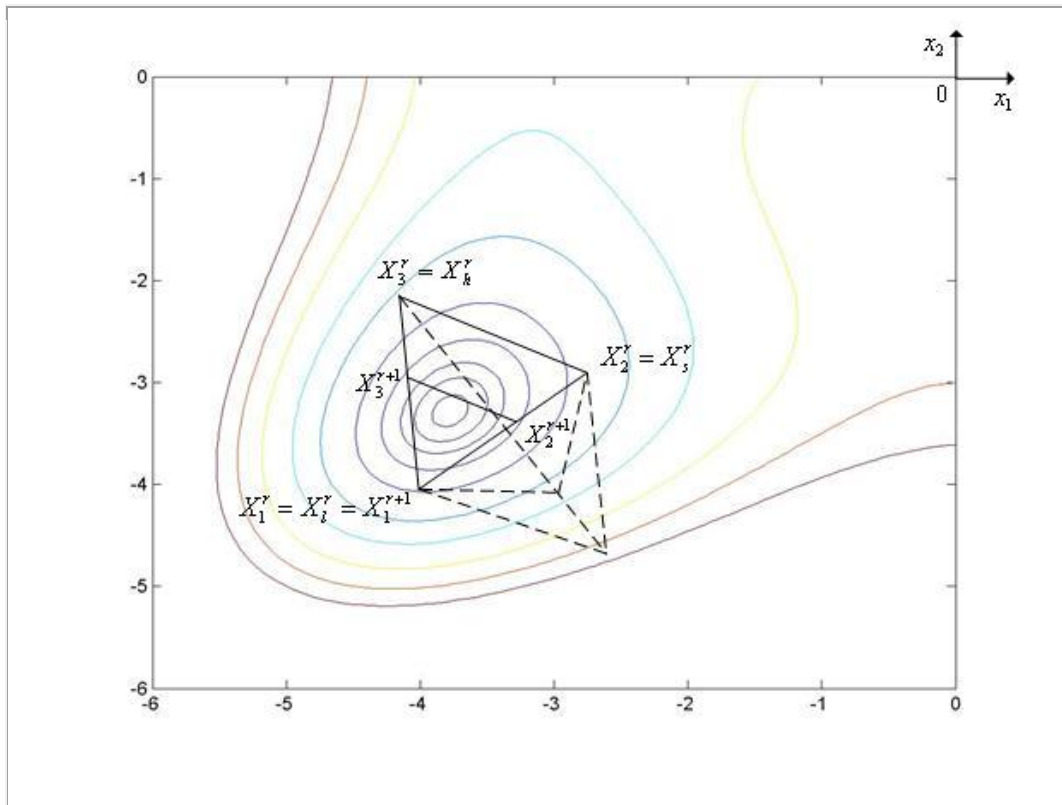


Рис. 3.6.8. Ситуация, когда метод Нелдера-Мида использует редукцию после неудачного сжатия симплекса. Пунктиром показаны отвергнутые (неудачные) итерации

На рис. 3.6.5 - рис. 3.6.8 минимизируемой функцией $\Phi(X)$ является функция Химмельблау.

В качестве условия окончания итераций в методе Нелдера-Мида можно использовать условие

$$\max_{i \in [1, \dots, n+1], i \neq k} \left| \Phi(X_i^{r+1}) - \Phi(X_k^{r+1}) \right| \leq \varepsilon_\Phi, \quad (3.6.6)$$

где ε_Φ - требуемая точность решения по Φ , $k \in [1, \dots, n+1]$ - номер произвольной вершины симплекса. Можно также завершать итерации, когда длина максимального из ребер текущего симплекса станет меньше или равна ε_X - требуемой точности решения по X .

Изложенная схема метода Нелдера-Мида имеет тот недостаток, что 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 \tilde{l} ;
- исходя из «лучшей» вершины текущего симплекса строится новый симплекс, длина ребра которого принимается равной \tilde{l} (см. параграф 3.5).

3.7. Метод наискорейшего спуска. Метод дробления шага

Рассматривается следующая многомерная задача локальной безусловной оптимизации: найти минимум критерия оптимальности $\Phi(X)$, определенного в n -мерном евклидовом пространстве R^n , $\min_{X \in R^n} \Phi(X) = \Phi(X^*) = \Phi^*$.

Положим, что функция $\Phi(X)$ всюду дифференцируема в n -мерном евклидовом пространстве R^n .

Направление спуска в градиентных методах оптимизации совпадает с направлением антиградиента минимизируемой функции $\Phi(X)$.

Итерационная формула градиентных методов оптимизации имеет вид

$$X^{r+1} = X^r + \lambda^r S^r. \quad (3.7.1)$$

Здесь λ^r - длина шага на r -ой итерации в направлении S^r , где

$$S^r = - \frac{\nabla \Phi^r}{\|\nabla \Phi^r\|} \quad (3.7.2)$$

- единичный вектор направления антиградиента функции $\Phi(X)$ в точке X^r , $\|\cdot\|$ - некоторая векторная норма, например, евклидова. Напомним, что градиент функции $\Phi(X)$ в точке X^r есть значение вектора частных производных этой функции в точке X^r :

$$\nabla \Phi^r = \nabla \Phi(X^r) = \left(\begin{array}{c} \frac{\partial \Phi(X)}{\partial x_1} \\ \frac{\partial \Phi(X)}{\partial x_2} \\ \dots \\ \frac{\partial \Phi(X)}{\partial x_n} \end{array} \right) \Bigg|_{X=X^r}.$$

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

Градиентный метод наискорейшего спуска.

Градиентный метод наискорейшего спуска в качестве длины шага λ^r использует величину, при которой достигается минимум функции $\Phi(X)$ в направлении S^r :

$$\Phi(X^{r+1}) = \Phi(X^r + \lambda^r S^r) = \min_{\lambda \in R^1} \Phi(X^r + \lambda S^r). \quad (3.7.3)$$

Задача (3.7.1) есть одномерная задача локальной безусловной оптимизации, которая может быть решена рассмотренными в разделе 2 методами.

Схема метода.

1. Задаем начальную точку X^0 и полагаем счетчик числа итераций $r=0$.
2. По формуле (3.7.2) вычисляем компоненты вектора S^r .
3. Каким-либо методом решаем одномерную задачу безусловной оптимизации (3.7.3) – определяем точку X^{r+1} .
4. Вычисляем величину $\Phi(X^{r+1})$ - значение функции $\Phi(X)$ в точке X^{r+1} .
5. Проверяем условие окончания поиска (см. ниже). Если условие окончания поиска выполнено, то полагаем $X^* \approx X^{r+1}$ и завершаем итерации. Иначе – полагаем $r=r+1$, переходим к п. 2.

В качестве критерия окончания поиска можно использоваться одно из стандартных условий окончания итераций

$$\|X^{r+1} - X^r\| = \lambda^r \leq \varepsilon_X, \quad (3.7.4)$$

$$|\Phi(X^{r+1}) - \Phi(X^r)| \leq \varepsilon_\Phi. \quad (3.7.5)$$

В качестве критерия окончания поиска можно использоваться также условие

$$\|\nabla\Phi^r\| \leq \varepsilon_{\nabla}, \quad (3.7.6)$$

где ε_{∇} - константа, определяющая требуемую точность решения по градиенту функции $\Phi(X)$.

Градиентный метод наискорейшего спуска иллюстрирует рис. 3.7.1, на котором показан фрагмент линий уровня функции Химмельблау.

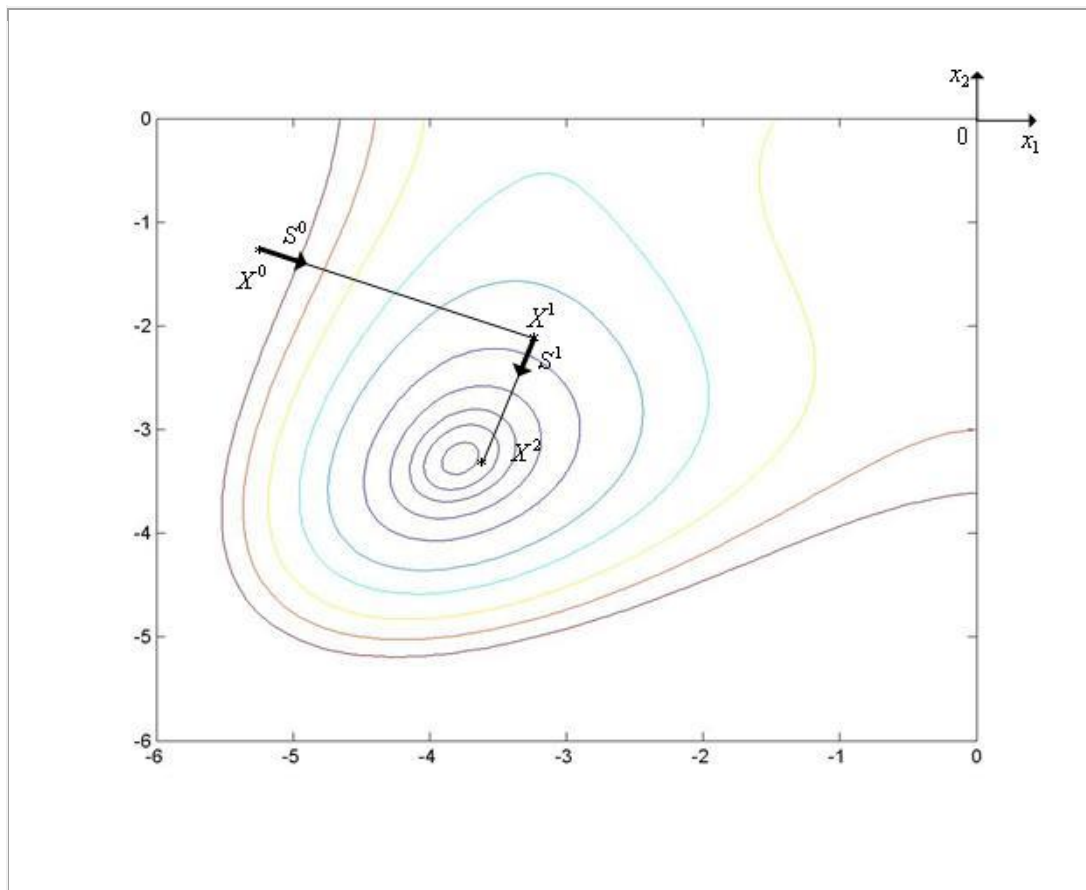


Рис. 3.7.1. Траектория поиска минимума функции Химмельблау градиентным методом наискорейшего спуска

Градиентный метод с дроблением шага.

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

$$X^{r+1} = X^r + \lambda^r S^r, \quad (3.7.7)$$

где величина шага λ^r находится из условия

$$\Phi(X^r) - \Phi(X^{r+1}) \geq 0.5\lambda^r \|\nabla\Phi^r\|. \quad (3.7.8)$$

Схема метода.

1. Задаем начальную точку X^0 , начальную величину шага λ^0 и коэффициент дробления шага $\nu \in (0,1]$. Полагаем счетчик числа итераций $r=0$.
2. По формуле (3.7.7) вычисляем компоненты вектора X^{r+1} .
3. Вычисляем величину $\Phi(X^{r+1})$ - значение функции $\Phi(X)$ в точке X^{r+1} .
4. Если условие (3.7.8) выполнено, то переходим к следующему пункту. Иначе – переходим к пункту 6.
5. Полагаем $\lambda^r = \nu\lambda^r$ и переходим к пункту 2.
6. Проверяем условие окончания поиска (см. ниже). Если условие окончания поиска выполнено, то полагаем $X^* \approx X^{r+1}$, и завершаем итерации. Иначе – полагаем $r=r+1$ переходим к п. 2.

В качестве критерия окончания поиска можно использовать условия (3.7.4) – (3.7.6).

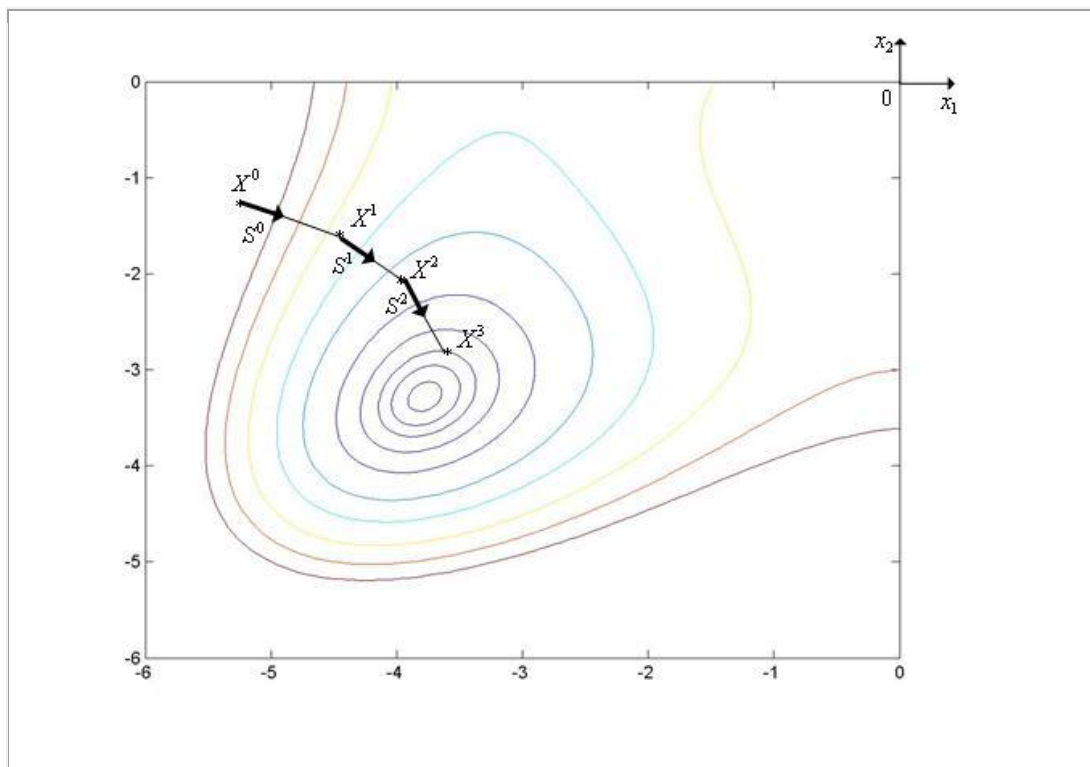


Рис. 3.7.2. Траектория поиска минимума функции Химмельблау градиентным методом дробления шага

Градиентный метод дробления шага иллюстрирует рис. 3.7.2, на котором показан фрагмент линий уровня функции Химмельблау.

Пример 3.7.1. Выполнить несколько итераций решения двумерной задачи локальной безусловной оптимизации $\min_{X \in R^2} \Phi(X) = \Phi(X^*) = \Phi^*$, где $\Phi(X) = \Phi(x, y) = x^2 + y^2 + 3(x + y)^2 = 4x^2 + 4y^2 + 6xy$, градиентным методом с дроблением шага, исходя из точки $X^0 = (x^0, y^0) = (-2.0, 1.0)$. Принять $\lambda^0 = 1.0$, $\nu = 0.7$, в качестве нормы вектора градиента использовать евклидову норму.

Решение. Итерация градиентного метода с дроблением шага для задачи имеет вид

$$X^{r+1} = \begin{pmatrix} x^{r+1} \\ y^{r+1} \end{pmatrix} = \begin{pmatrix} x^r \\ y^r \end{pmatrix} + \lambda^r \begin{pmatrix} S_x^r \\ S_y^r \end{pmatrix}, \quad (3.7.9)$$

где

$$\begin{pmatrix} S_x^r \\ S_y^r \end{pmatrix} = S^r = -\frac{\nabla \Phi^r}{\|\nabla \Phi^r\|}, \quad \nabla \Phi^r = \begin{pmatrix} \frac{\partial \Phi(x^r, y^r)}{\partial x} \\ \frac{\partial \Phi(x^r, y^r)}{\partial y} \end{pmatrix}, \quad (3.7.10)$$

а величина шага λ^r находится из условия

$$\Phi(X^r) - \Phi(X^{r+1}) \geq 0.5 \lambda^r \|\nabla \Phi^r\|. \quad (3.7.11)$$

Найдем явные выражения для частных производных нашей функции:

$$\frac{\partial \Phi(x, y)}{\partial x} = 8x + 6y; \quad \frac{\partial \Phi(x, y)}{\partial y} = 8y + 6x. \quad (3.7.12)$$

Таким образом, из (3.7.9), (3.7.10), (3.7.12) имеем искомую итерационную формулу градиентного метода с дроблением шага для нашей задачи

$$X^{r+1} = \begin{pmatrix} x^{r+1} \\ y^{r+1} \end{pmatrix} = \begin{pmatrix} x^r \\ y^r \end{pmatrix} + \lambda^r \begin{pmatrix} S_x^r \\ S_y^r \end{pmatrix}, \quad S_x^r = -\frac{8x^r + 6y^r}{\|\nabla\Phi^r\|}, \quad S_y^r = -\frac{8y^r + 6x^r}{\|\nabla\Phi^r\|},$$

$$\|\nabla\Phi(X)\| = \sqrt{(8x+6y)^2 + (8y+6x)^2}. \quad (3.7.13)$$

Первая итерация ($r=0$).

Из формул (3.7.12), (3.7.13) последовательно имеем

$$X^1 = \begin{pmatrix} x^1 \\ y^1 \end{pmatrix} = \begin{pmatrix} x^0 \\ y^0 \end{pmatrix} + \lambda^0 \begin{pmatrix} S_x^0 \\ S_y^0 \end{pmatrix},$$

$$\frac{\partial\Phi(x^0, y^0)}{\partial x} = 8x^0 + 6y^0 = 8 \cdot (-2) + 6 = -10,$$

$$\frac{\partial\Phi(x^0, y^0)}{\partial y} = 8y^0 + 6x^0 = 8 + 6 \cdot (-2) = -4,$$

$$\|\nabla\Phi^0\| = \sqrt{(-10)^2 + (-4)^2} \approx 10.77, \quad S_x^0 \approx \frac{10}{10.77} \approx 0.93, \quad S_y^0 \approx \frac{4}{10.77} \approx 0.37,$$

$$X^1 = \begin{pmatrix} x^1 \\ y^1 \end{pmatrix} = \begin{pmatrix} x^0 \\ y^0 \end{pmatrix} + \lambda^0 \begin{pmatrix} S_x^0 \\ S_y^0 \end{pmatrix} \approx \begin{pmatrix} -2 \\ 1 \end{pmatrix} + 0.5 \begin{pmatrix} 0.93 \\ 0.37 \end{pmatrix} \approx \begin{pmatrix} -1.54 \\ 1.18 \end{pmatrix}.$$

Таким образом, $X^1 = \begin{pmatrix} x^1 \\ y^1 \end{pmatrix} \approx \begin{pmatrix} -1.54 \\ 1.18 \end{pmatrix}$ (см. рис. 3.7.3).

Условие (3.7.11) на первой итерации имеет вид

$$\Phi(X^0) - \Phi(X^1) \geq 0.5\lambda^0 \|\nabla\Phi^0\|.$$

Поскольку

$$\Phi(X^0) = \Phi(x^0, y^0) = 4(x^0)^2 + 4(y^0)^2 + 6x^0y^0 = 4(-2)^2 + 4 + 6(-2) = 8,$$

$$\Phi(X^1) = \Phi(x^1, y^1) = 4(x^1)^2 + 4(y^1)^2 + 6x^1y^1 \approx 4(-1.54)^2 + 4(1.18)^2 + 6(-1.54)(1.18) \approx 4.16,$$

левая часть этого неравенства равна $8.0 - 4.16 = 3.84$. Его правая часть, легко видеть, равна $(0.5)^2 \cdot 10.77 \approx 2.69$.

Таким образом, на первой итерации условие (3.7.11) выполняется и величина шага λ должна быть изменена: $\lambda^1 = 0.7 \cdot 0.5 = 0.35$.

Вторая итерация ($r=1$).

Аналогично первой итерации последовательно имеем

$$X^2 = \begin{pmatrix} x^2 \\ y^2 \end{pmatrix} = \begin{pmatrix} x^1 \\ y^1 \end{pmatrix} + \lambda^1 \begin{pmatrix} S_x^1 \\ S_y^1 \end{pmatrix},$$

$$\frac{\partial \Phi(x^1, y^1)}{\partial x} = 8x^1 + 6y^1 = 8 \cdot (-1.54) + 6 \cdot 1.18 = -5.24,$$

$$\frac{\partial \Phi(x^1, y^1)}{\partial y} = 8y^1 + 6x^1 = 8 \cdot 1.18 + 6 \cdot (-1.54) = 0.20,$$

$$\|\nabla \Phi^1\| = \sqrt{(-5.24)^2 + (0.20)^2} \approx 5.24, \quad S_x^1 \approx \frac{5.24}{5.24} \approx 1.0, \quad S_y^1 \approx -\frac{0.2}{5.24} \approx -0.04,$$

$$X^2 = \begin{pmatrix} x^2 \\ y^2 \end{pmatrix} = \begin{pmatrix} x^1 \\ y^1 \end{pmatrix} + \lambda^1 \begin{pmatrix} S_x^1 \\ S_y^1 \end{pmatrix} \approx \begin{pmatrix} -1.54 \\ 1.18 \end{pmatrix} + 0.35 \begin{pmatrix} 1.0 \\ 0.04 \end{pmatrix} \approx \begin{pmatrix} -1.19 \\ 1.22 \end{pmatrix}.$$

Таким образом, $X^2 = \begin{pmatrix} x^2 \\ y^2 \end{pmatrix} \approx \begin{pmatrix} -1.19 \\ 1.22 \end{pmatrix}$ (см. рис. 3.7.3).

Условие (3.7.11) на второй итерации имеет вид

$$\Phi(X^1) - \Phi(X^2) \geq 0.5\lambda^1 \|\nabla \Phi^1\|.$$

Поскольку $\Phi(X^1) \approx 4.16$,

$$\begin{aligned} \Phi(X^2) = \Phi(x^2, y^2) &= 4(x^2)^2 + 4(y^2)^2 + 6x^2y^2 \approx 4(-1.19)^2 + 4(1.22)^2 + \\ &+ 6(-1.19)(1.22) \approx 2.91, \end{aligned}$$

левая часть этого неравенства равна $4.16 - 2.91 = 1.25$. Его правая часть, легко видеть, равна $0.5 \cdot 0.35 \cdot 5.24 \approx 0.92$.

Таким образом, на второй итерации условие (3.7.11.) выполняется и величина шага λ должна быть изменена: $\lambda^2 = 0.7 \cdot 0.35 \approx 0.24$.

Третья итерация ($r = 2$).

Аналогично первой итерации последовательно имеем

$$X^3 = \begin{pmatrix} x^3 \\ y^3 \end{pmatrix} = \begin{pmatrix} x^2 \\ y^2 \end{pmatrix} + \lambda^2 \begin{pmatrix} S_x^2 \\ S_y^2 \end{pmatrix},$$

$$\frac{\partial \Phi(x^2, y^2)}{\partial x} = 8x^2 + 6y^2 = 8 \cdot (-1.19) + 6 \cdot 1.22 = -2.20,$$

$$\frac{\partial \Phi(x^2, y^2)}{\partial y} = 8y^2 + 6x^2 = 8 \cdot 1.22 + 6 \cdot (-1.19) = 2.62,$$

$$\|\nabla \Phi^2\| = \sqrt{(-2.20)^2 + (2.62)^2} \approx 3.42,$$

$$S_x^2 \approx \frac{2.20}{3.42} \approx 0.64,$$

$$S_y^2 \approx -\frac{2.62}{3.42} \approx -0.77,$$

$$X^3 = \begin{pmatrix} x^3 \\ y^3 \end{pmatrix} = \begin{pmatrix} x^2 \\ y^2 \end{pmatrix} + \lambda^2 \begin{pmatrix} S_x^2 \\ S_y^2 \end{pmatrix} \approx \begin{pmatrix} -1.19 \\ 1.22 \end{pmatrix} + 0.24 \begin{pmatrix} 0.64 \\ -0.77 \end{pmatrix} \approx \begin{pmatrix} -1.04 \\ 1.03 \end{pmatrix}.$$

Таким образом, $X^3 = \begin{pmatrix} x^3 \\ y^3 \end{pmatrix} \approx \begin{pmatrix} -1.04 \\ 1.03 \end{pmatrix}$ (см. рис. 3.7.3).

Условие (3.7.11) на третьей итерации имеет вид

$$\Phi(X^2) - \Phi(X^3) \geq 0.5\lambda^2 \|\nabla \Phi^2\|.$$

Поскольку $\Phi(X^2) \approx 2.91$,

$$\begin{aligned} \Phi(X^3) = \Phi(x^3, y^3) &= 4(x^3)^2 + 4(y^3)^2 + 6x^3y^3 \approx 4(-1.04)^2 + 4(1.03)^2 + \\ &+ 6(-1.04)(1.03) \approx 2.14, \end{aligned}$$

левая часть этого неравенства равна $2.91 - 2.14 = 0.77$. Его правая часть, легко видеть, равна $0.5 \cdot 0.24 \cdot 3.42 \approx 0.41$.

Таким образом, на третьей итерации условие (3.7.11) выполняется и величина шага λ должна быть изменена: $\lambda^3 = 0.7 \cdot 0.24 \approx 0.17$.

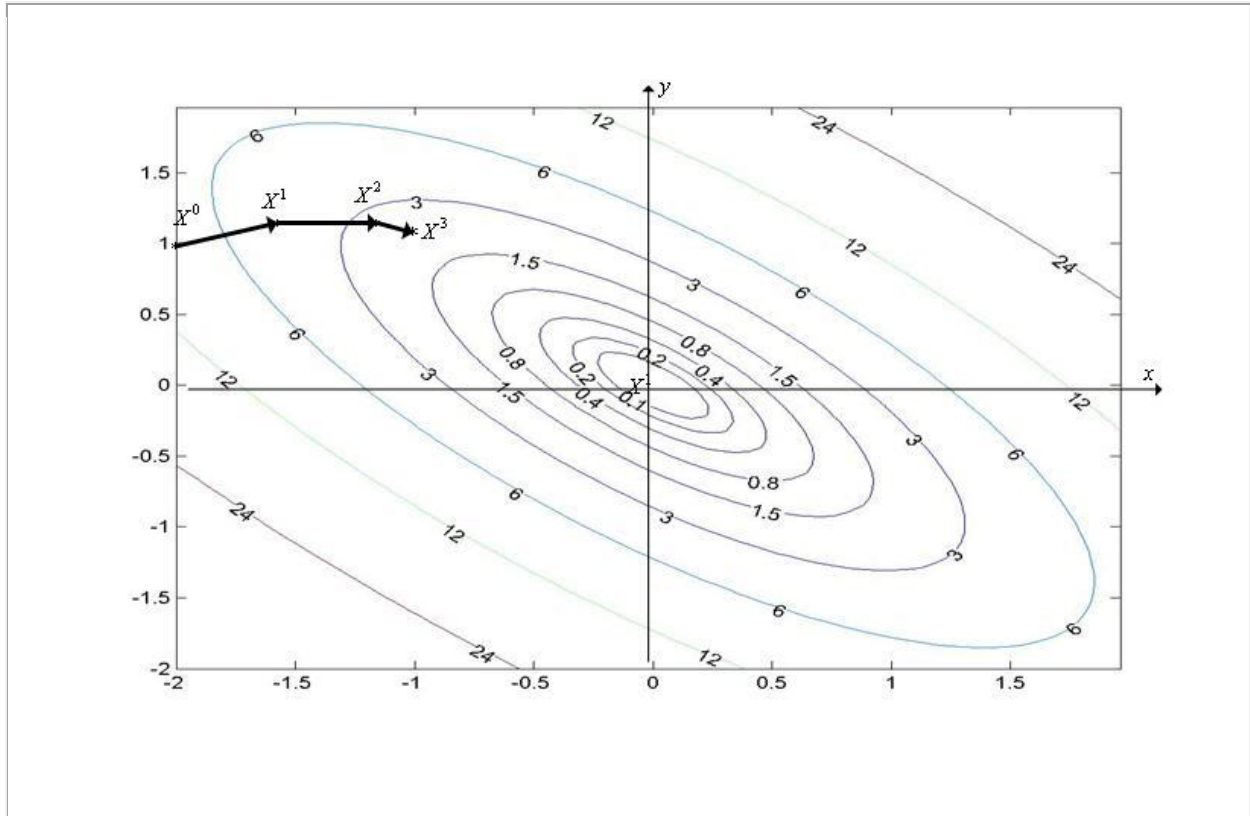


Рис. 3.7.3. Фрагмент (три итерации) траектории поиска минимума функции $\Phi(X) = \Phi(x, y) = x^2 + y^2 + 3(x + y)^2 = 4x^2 + 4y^2 + 6xy$ градиентным методом с дроблением шага, исходя из точки $X^0 = (x^0, y^0) = (-2.0, 1.0)$

3.8. Метод оптимизации Ньютона

Положим, что функция $\Phi(X)$ всюду дважды дифференцируема в n -мерном евклидовом пространстве R^n . Рассмотрим следующую многомерную задачу локальной безусловной оптимизации: найти минимум критерия оптимальности $\Phi(X)$, определенного в n -мерном евклидовом пространстве R^n , $\min_{X \in R^n} \Phi(X) = \Phi(X^*) = \Phi^*$.

Рассмотрим первые три члена разложения функции $\Phi(X)$ в ряд Тейлора в окрестности точки X^r :

$$\begin{aligned} \Phi(X) \approx \tilde{\Phi}^r(X) = & \Phi(X^r) + (\nabla \Phi(X^r), (X - X^r)) + \\ & + \frac{1}{2} (H(X^r)(X - X^r), (X - X^r)). \end{aligned} \quad (3.8.1)$$

Здесь $H(X)$ - матрица Гессе функции $\Phi(X)$. Из (3.8.1) следует, что градиент функции $\tilde{\Phi}^r(X)$ равен

$$\nabla \tilde{\Phi}^r(X) = \nabla \Phi(X^r) + H(X^r)(X - X^r). \quad (3.8.2)$$

Если матрица Гессе $H(X^r)$ положительно определена, то функция $\tilde{\Phi}^r(X)$ достигает минимума в точке, в которой градиент этой функции равен нулевому вектору.

Таким образом, в точке X^{r+1} минимума функции $\tilde{\Phi}^r(X)$ справедливо равенство

$$\nabla \Phi(X^r) + H(X^r)(X^{r+1} - X^r) = 0, \quad (3.8.3)$$

где 0 - n -мерный вектор нулей.

Отсюда получаем итерационную формулу

$$X^{r+1} = X^r - H^{-1}(X^r) \nabla \Phi(X^r) = X^r + \Delta^r \quad (3.8.4)$$

для отыскания очередного приближения к точке минимума функции $\Phi(X)$.

Здесь

$$\Delta^r = -H^{-1}(X^r)\nabla\Phi(X^r). \quad (3.8.5)$$

Выражение (3.8.4) представляет собой итерационную формулу решения системы уравнений (3.8.3) широко известным методом касательных (методом Ньютона) – см. параграф 2.6. Этим фактом объясняется название рассматриваемого метода оптимизации.

Найдем скалярное произведение градиента функции $\Phi(X)$ в точке X^r и вектора Δ^r :

$$(\nabla\Phi(X^r), \Delta^r) = (\nabla\Phi(X^r), H^{-1}(X^r)\nabla\Phi(X^r)) < 0. \quad (3.8.6)$$

Последнее неравенство справедливо в силу постулируемой положительной определенности матрицы Гессе в точке X^r . Геометрически неравенство (3.8.6) означает, что вектор Δ^r образует тупой угол с градиентом целевой функции $\Phi(X)$ в точке X^r (см. рис. 3.8.1).

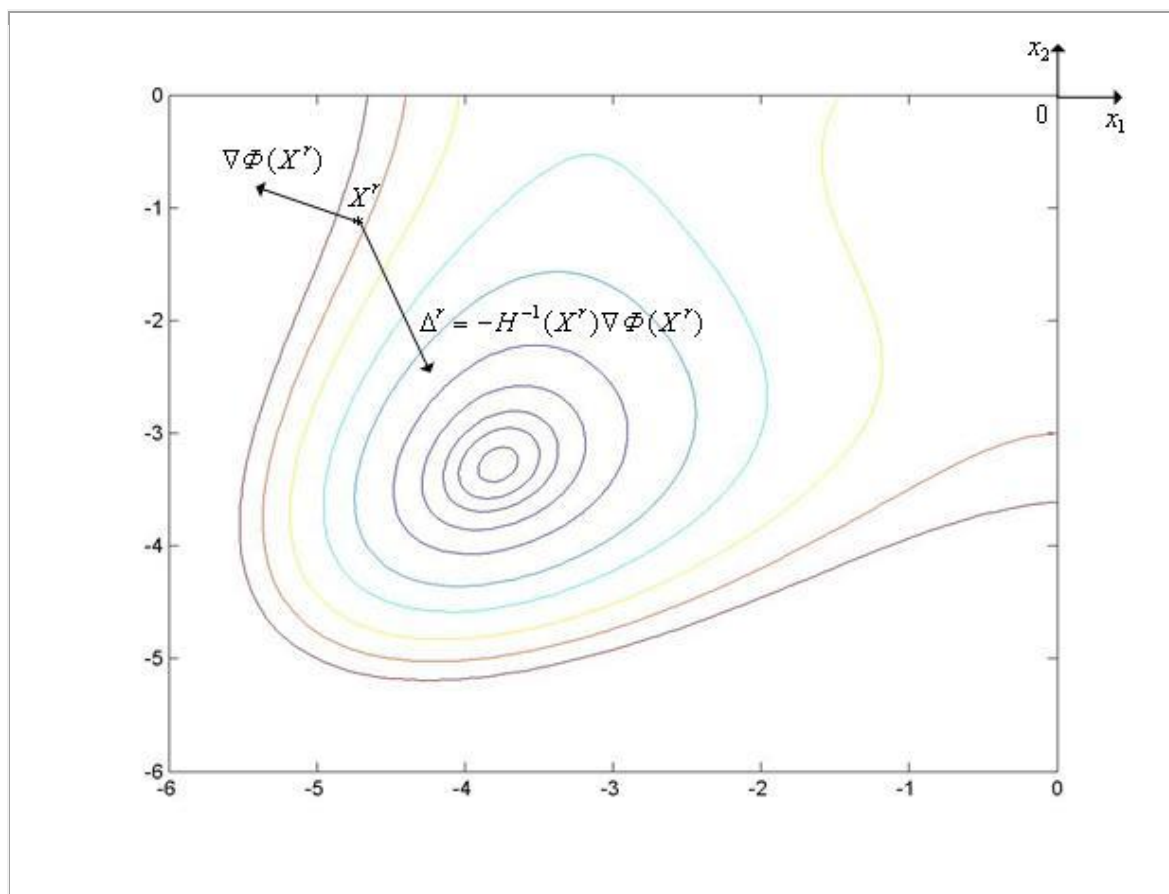


Рис. 3.8.1. К обоснованию многомерной оптимизации Ньютона

Таким образом, при минимизации овражных функций вектор Δ^r может составлять с осью оврага меньший угол, чем вектор антиградиента. Эта особенность делает метод оптимизации Ньютона, вообще говоря, более эффективным, чем градиентный метод наискорейшего спуска.

Отметим трудности, которые могут возникать при использовании итерационной формулы (3.8.4):

- Если размерность пространства R^n велика, то обращение на каждой итерации матрицы Гессе $H(X^r)$ может потребовать значительных вычислительных ресурсов;
- Значение минимизируемой функции $\Phi(X)$ в точке $X^{r+1} = X^r + \Delta^r$ может превышать значение функции в предыдущей точке X^r вследствие того, что направление Δ^r ведет к уменьшению $\Phi(X)$, но величина шага слишком велика;
- Направление спуска, определяемое вектором $\Delta^r = -H^{-1}(X^r)\nabla\Phi(X^r)$, ведет к убыванию целевой функции только при положительной определенности матрицы Гессе $H(X^r)$. Это приводит к тому, что на каждой итерации необходимы вычислительные затраты на проверку обусловленности этой матрицы. Указанная матрица может быть плохо обусловленной. Более того, указанная матрица может быть вырожденной и, поэтому, не иметь обратной матрицы.

Вследствие этих трудностей итерационная формула (3.8.4) в «чистом» виде не используется в вычислительной практике.

Для того чтобы избежать обращения матрицы Гессе, на практике вектор Δ^r находят обычно из следующей системы линейных алгебраических уравнений (СЛАУ), вытекающей из равенства (3.8.5):

$$H(X^r)\Delta^r = -\nabla\Phi(X^r). \quad (3.8.7)$$

СЛАУ (3.8.7) может быть решена различными численными методами (например, прямыми методами, итерационными методами).

Величина шага в направлении Δ^r , которая приводит к убыванию функции $\Phi(X)$, может быть обеспечена путем добавления в итерационную формулу (3.8.4) коэффициента $\lambda^r \in (0,1]$, т.е. путем использования вместо формулы (3.8.4) итерационной формулы

$$X^{r+1} = X^r - \lambda^r H^{-1}(X^r) \nabla \Phi(X^r) = X^r + \lambda^r \Delta^r, \quad (3.8.8)$$

где коэффициент λ^r выбирают тем или иным способом так, чтобы обеспечить условие $\Phi(X^{r+1}) < \Phi(X^r)$.

Для того, чтобы направление спуска независимо от определенности матрицы Гессе $H(X^r)$ вело к убыванию функции $\Phi(X)$, в качестве вектора Δ^r можно использовать вектор

$$\Delta^r = -\left(\mu^r I + H(X^r)\right)^{-1} \nabla \Phi(X^r), \quad (3.8.9)$$

где I - $(n \times n)$ -единичная матрица, а $\mu^r > 0$ - параметр, выбираемый так, чтобы матрица $\mu^r I + H(X^r)$ являлась положительно определенной.

Схема метода.

Рассмотрим схему одной из модификаций метода оптимизации Ньютона, в которой используется итерационная формула (3.8.8) и вектор Δ^r находят путем решения на каждой итерации СЛАУ (3.8.7).

1. Задаем начальную точку X^0 , начальную величину шага $\lambda^0 = 1$ и коэффициент дробления шага $\nu \in (0,1]$. Полагаем счетчик числа итераций $r = 0$.

2. Вычисляем в точке X^r вектор градиента $\nabla \Phi(X^r)$ и матрицу Гессе $H(X^r)$.

3. Решаем СЛАУ (3.8.7) и находим вектор Δ^r .
4. По формуле (3.8.8) вычисляем компоненты вектора X^{r+1} .
5. Вычисляем величину $\Phi(X^{r+1})$ - значение функции $\Phi(X)$ в точке X^{r+1} .
6. Проверяем условие окончания поиска (см. ниже). Если условие окончания поиска выполнено, то полагаем $X^* \approx X^{r+1}$, и завершаем итерации. Иначе – переходим к следующему пункту.
7. Если $\Phi(X^{r+1}) < \Phi(X^r)$, то полагаем $r = r + 1$ и переходим к п.2. Иначе – фиксированное число раз полагаем $\lambda^r = \nu \lambda^r$ и переходим к пункту 4.

В качестве условия окончания поиска можно использоваться одно из стандартных условий окончания итераций:

$$\|X^{r+1} - X^r\| = \lambda^r \leq \varepsilon_X, \quad (3.8.10)$$

$$|\Phi(X^{r+1}) - \Phi(X^r)| \leq \varepsilon_\Phi \quad (3.8.11)$$

или условие

$$\|\nabla \Phi^r\| \leq \varepsilon_\nabla, \quad (3.8.12)$$

где ε_∇ - константа, определяющая требуемую точность решения по градиенту функции $\Phi(X)$.

3.9. Метод с возвратом при неудачном шаге. Метод наилучшей пробы

Рассматривается следующая многомерная задача локальной безусловной оптимизации: найти минимум критерия оптимальности $\Phi(X)$, определенного в n -мерном евклидовом пространстве R^n ,
$$\min_{X \in R^n} \Phi(X) = \Phi(X^*) = \Phi^*.$$

При решении задачи методом с возвратом при неудачном шаге (одношаговый метод оптимизации) используется итерационная формула

$$X^{r+1} = X^r + \lambda^r \frac{\psi^r}{\|\psi^r\|}, \quad (3.9.1)$$

где λ^r - величина шага на r -ой итерации, $\psi^r = (\psi_1^r, \psi_2^r, \dots, \psi_n^r)$ - реализация n -мерного случайного вектора, $\|\cdot\|$ - некоторая векторная норма.

Обычно в качестве координат вектора ψ^r используют независимые случайные величины, равномерно распределенные в интервале $[-1, 1]$.

Схема метода с возвратом при неудачном шаге.

1. Задаем начальную точку X^0 , начальную длину шага λ^0 и полагаем счетчик числа итераций $r=0$.
2. Задаем начальное значение счетчика числа неудачных попыток $k=1$.
3. Получаем реализацию случайных чисел $\psi_1^r, \psi_2^r, \dots, \psi_n^r$ - компонент вектора ψ^r и по формуле (3.9.1) находим пробную точку X^{r+1} .
4. Вычисляем значение $\Phi(X^{r+1})$ функции $\Phi(X)$ в точке X^{r+1} .
5. Если $\Phi(X^{r+1}) < \Phi(X^r)$, то полагаем $r=r+1$ и переходим к п.3. Иначе - переходим к п.6.
6. Полагаем $k=k+1$. Если $k < K$, то переходим к п.3. Иначе - переходим к п.7. Здесь K - предельное количество неудачных попыток (свободный параметр метода). Рекомендуется $K=3n$.

7. Проверяем условие окончания поиска (см. ниже). Если условие окончания поиска выполнено, то полагаем $X^* \approx X^{r+1}$ и завершаем итерации. Иначе – полагаем $r = r + 1$, $\lambda^r = \alpha \lambda^r$ и переходим к п.2. Здесь $\alpha \in (0, 1)$ – коэффициент уменьшения шага (свободный параметр метода).

В качестве условия окончания поиска можно использоваться одно из стандартных условий окончания итераций:

$$\|X^{r+1} - X^r\| = \lambda^r \leq \varepsilon_X, \quad (3.9.2)$$

где ε_X – константа, определяющая требуемую точность решения по X ;

$$|\Phi(X^{r+1}) - \Phi(X^r)| \leq \varepsilon_\Phi, \quad (3.9.3)$$

где ε_Φ – константа, определяющая требуемую точность решения по Φ .

Метод с возвратом при неудачном шаге иллюстрирует рис. 3.9.1, на котором показан фрагмент линий уровня функции Химмельблау.

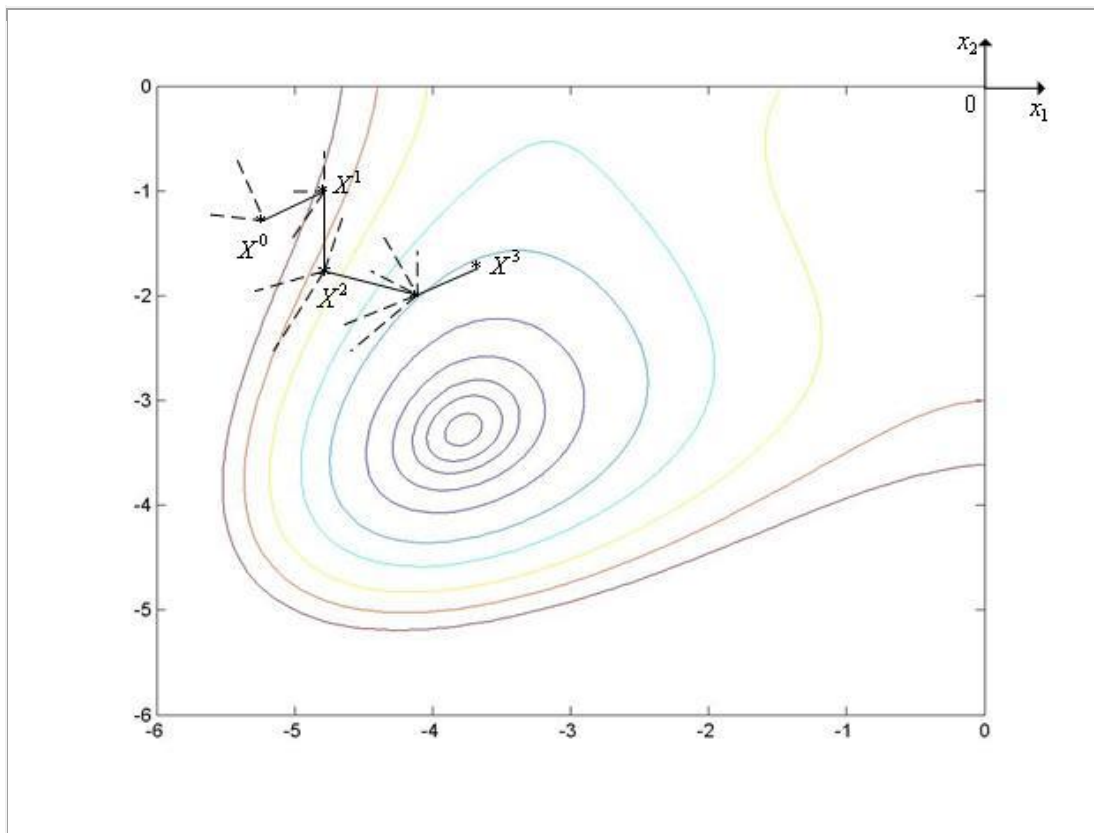


Рис. 3.9.1. Траектория поиска минимума функции Химмельблау методом с возвратом при неудачном шаге. Пунктиром на рисунке показаны неудачные шаги

Модификацией метода с возвратом при неудачном шаге является метод наилучшей пробы (также одношаговый метод оптимизации).

Схема метода наилучшей пробы.

1. Задаем начальную точку X^0 , начальную длину шага λ^0 и полагаем счетчик числа итераций $r=0$.
2. Генерируем M случайных векторов $\psi_i^r, i \in [1, \dots, M]$ и по формуле (3.9.1) находим пробные точки $X_i^{r+1}, i \in [1, \dots, M]$.
3. Вычисляем значения $\Phi(X_i^{r+1})$ функции $\Phi(X)$ в пробных точках $X_i^{r+1}, i \in [1, \dots, M]$ и находим минимальное из этих значений $\Phi(X^{r+1}) = \Phi(X_k^{r+1}) = \min_{i \in [1, \dots, M]} \Phi(X_i^{r+1})$.
4. Если $\Phi(X^{r+1}) < \Phi(X^r)$, то полагаем $r = r+1$ и переходим к п.2. Иначе – переходим к п.5.
5. Проверяем условие окончания поиска (см. (3.9.2), (3.9.3)). Если условие окончания поиска выполнено, то полагаем $X^* \approx X^{r+1}$ и завершаем итерации. Иначе – полагаем $r = r+1, \lambda^r = \alpha \lambda^r$ и переходим к п.2. Здесь $\alpha \in (0, 1)$ - коэффициент уменьшения шага (свободный параметр метода).

Метод наилучшей пробы иллюстрирует рис. 3.9.2, на котором показан фрагмент линий уровня функции Химмельблау.

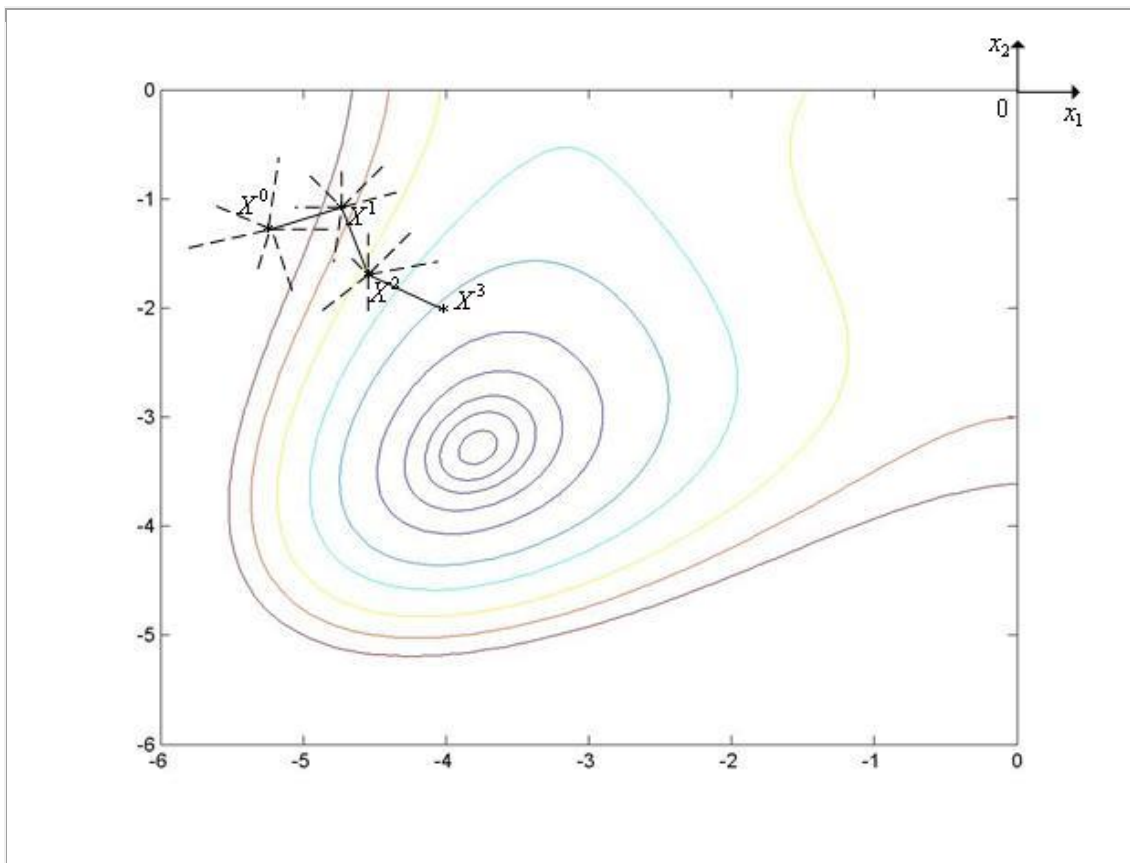


Рис. 3.9.2. Траектория поиска минимума функции Химмельблау методом наилучшей пробы. Пунктиром на рисунке показаны неудачные пробы

3.10. Метод комплексов

Рассматривается следующая многомерная задача локальной безусловной оптимизации: найти минимум критерия оптимальности $\Phi(X)$, определенного в n -мерном евклидовом пространстве R^n ,
$$\min_{X \in R^n} \Phi(X) = \Phi(X^*) = \Phi^*.$$

Комплексом называется многогранник с $N > n + 1$ вершинами (не обязательно выпуклый). Рекомендуется $N = 2n$. Вообще говоря, комплексом в комбинаторной топологии называется геометрическая фигура, которая может быть разбита на более элементарные фигуры. В нашем случае такими элементарными фигурами являются симплексы. Поэтому, говоря более строго, в данном параграфе рассматриваются **симплициальные комплексы**.

При решении **методом комплексов** используются следующие операции:

- генерация случайного комплекса;
- отражение вершины комплекса с растяжением;
- сжатие комплекса.

Генерация случайного комплекса.

В пространстве R^n координаты вершин случайного комплекса с N вершинами могут быть найдены по формуле

$$X_i = X_0 + l \frac{\Psi_i}{\|\Psi_i\|}, \quad i \in [0, \dots, N], \quad (3.10.1)$$

где X_0 - произвольная начальная точка, i - номер вершины комплекса, l - скаляр, определяющий размеры комплекса, Ψ_i - реализация n -мерного случайного вектора, $\|\cdot\|$ - некоторая векторная норма. Обычно в качестве координат вектора Ψ_i используют независимые случайные величины, равномерно распределенные в интервале $[-1, 1]$.

Отражение вершины комплекса с растяжением.

Положим, что в пространстве R^n тем или иным способом задан комплекс C^r с N вершинами X_i^r , $i \in [1, \dots, N]$, и его вершину X_k^r необходимо отразить через центр тяжести комплекса с растяжением. В новом комплексе C^{r+1} все вершины, кроме k -ой, совпадают с соответствующими вершинами исходного комплекса C^r , а k -я вершина находится на прямой, проходящей через центр тяжести этого комплекса и его вершину X_k^r (см. рис. 3.10.1).

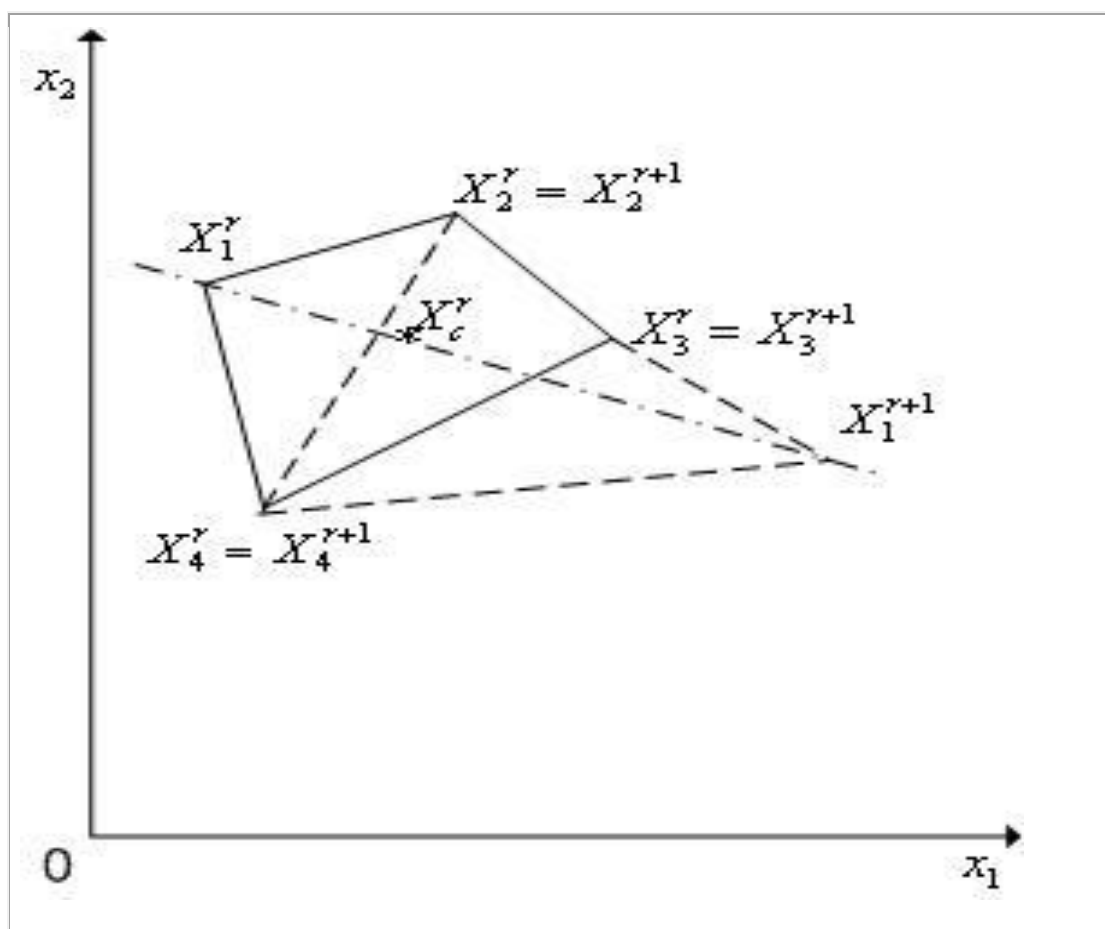


Рис. 3.10.1. Отражение вершины X_1^r комплекса C^r через центр его тяжести с растяжением. Пунктиром показан новый комплекс C^{r+1}

Обозначим координаты вершин нового комплекса X_i^{r+1} , $i \in [1, \dots, N]$.

Тогда имеем

$$X_i^{r+1} = X_i^r, \quad i \in [1, \dots, N], \quad i \neq k; \quad X_k^{r+1} = X_c^r + \alpha(X_c^r - X_k^r), \quad (3.10.2)$$

где α - коэффициент растяжения комплекса (рекомендуемое значение - 1.3),
 X_c^r - вектор координат центра тяжести комплекса C^r :

$$X_c^r = \frac{1}{N} \sum_{i=1}^N X_i^r. \quad (3.10.3)$$

Сжатие комплекса.

Положим, что в пространстве R^n тем или иным способом задан комплекс C^r с N вершинами X_i^r , $i \in [1, \dots, N]$, и его вершину X_k^r необходимо переместить ближе к центру тяжести комплекса C^r - выполнить сжатие комплекса. В новом комплексе C^{r+1} все вершины, кроме k -ой, совпадают с соответствующими вершинами исходного комплекса C^r , а k -я вершина находится на прямой, проходящей через центр тяжести этого комплекса и его вершину X_k^r (см. рис. 3.10.2).

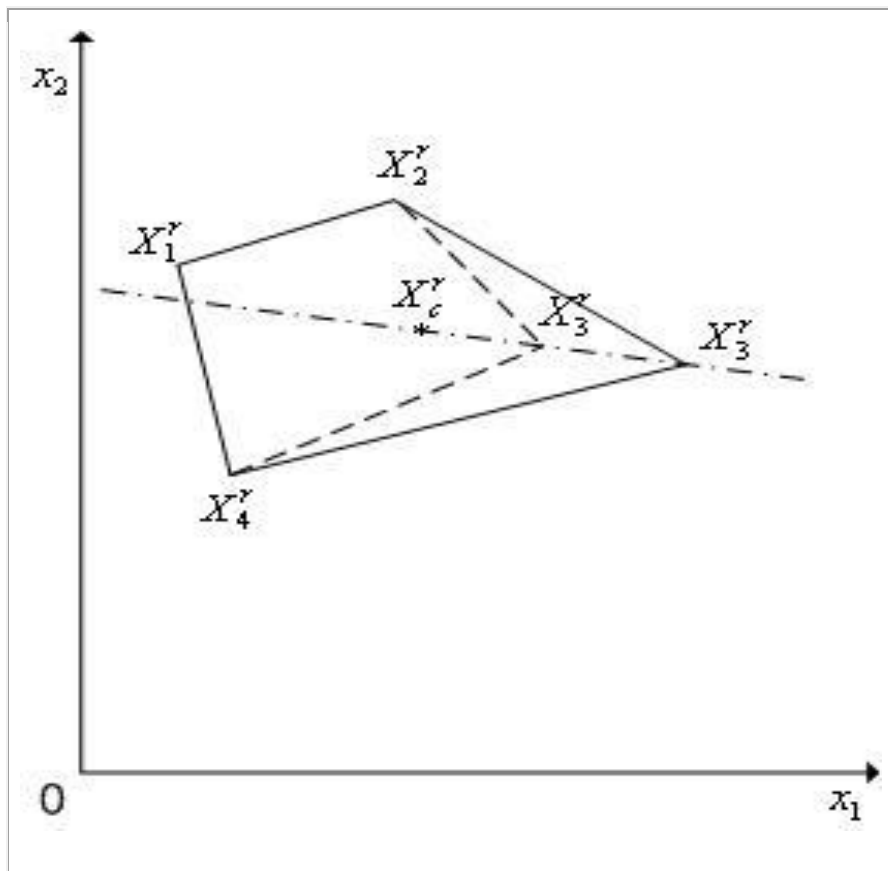


Рис. 3.10.2. Сжатие комплекса C^r . Пунктиром показан новый комплекс C^r

Обозначим координаты вершин нового комплекса X_i^{r+1} , $i \in [1, \dots, N]$.

Тогда имеем

$$X_i^{r+1} = X_i^r, \quad i \in [1, \dots, N], \quad i \neq k; \quad X_k^{r+1} = X_c^r + \beta(X_k^r - X_c^r), \quad (3.10.4)$$

где β - коэффициент сжатия комплекса (рекомендуемое значение - 2), X_c^r - вектор координат центра тяжести комплекса C^r (см. (3.10.3)).

Схема простейшего варианта метода комплексов.

1. Задаем начальную точку X_0^0 , исходя из которой должен быть построен комплекс C^0 , начальное значение величины $l = l^0$ и полагаем счетчик числа итераций $r = 0$.
2. Генерируем N случайных векторов ψ_i^r , $i \in [1, \dots, N]$ и по формуле (3.10.1) находим координаты X_i^r , $i \in [1, \dots, N]$ вершин комплекса C^r .
3. Вычисляем значения $\Phi(X_i^r)$ функции $\Phi(X)$ во всех вершинах комплекса C^r .
4. Находим максимальное из значений $\Phi(X)$ в вершинах комплекса C^r

$$\Phi(X_k^r) = \max_{i \in [1, \dots, N]} \Phi(X_i^r).$$

5. По формулам (3.10.2), (3.10.3) для комплекса C^r выполняем отражение вершины комплекса с растяжением X_k^r - получаем вершину X_k^{r+1} и новый комплекс C^{r+1} .

6. Вычисляем значение $\Phi(X_k^{r+1})$ функции $\Phi(X)$ в вершине X_k^{r+1} .

7. Если $\Phi\left(X_k^{r+1}\right) < \Phi\left(X_k^r\right)$, то полагаем $r = r + 1$ и переходим к п.8. Иначе

– по формулам (3.10.4), (3.10.3) выполняем сжатие симплекса C^{r+1} в направлении $X_k^{r+1} - X_c^{r+1}$, полагаем $r = r + 2$ и переходим к п.6.

8. Проверяем условие окончания поиска (см. ниже). Если условие окончания поиска выполнено, то в качестве точки X^* полагаем вершину комплекса C^r , к которой функция $\Phi(X)$ имеет наименьшее значение и завершаем итерации. Иначе – переходим к п. 4.

В качестве критерия окончания поиска может использоваться следующее условие: максимальная длина ребра комплекса C^r не превышает ε_X - требуемую точность решения по X .

Может использоваться также следующее аналогичное условие: максимальная разность значений функции $\Phi(X)$ в двух вершинах комплекса C^r не превышает ε_Φ - требуемую точность решения по Φ .

Могут использоваться также более сложные условия окончания поиска

$$\frac{1}{N} \sqrt{\sum_{i=1}^N \|X_i^r - X_c^r\|^2} \leq \varepsilon_X, \quad (3.10.5)$$

$$\frac{1}{N} \sqrt{\sum_{i=1}^N \|\Phi(X_i^r) - \Phi_c^r\|^2} \leq \varepsilon_\Phi. \quad (3.10.6)$$

В формуле (3.10.5) векторная норма означает расстояние вершины X_i^r до центра тяжести комплекса C^r , а сама формула (3.10.5) определяет среднее расстояние вершин комплекса C^r до его центра тяжести.

В формуле (3.10.6) $\Phi_c^r = \frac{1}{N} \sum_{i=1}^N \Phi(X_i^r)$ есть среднее значение функции $\Phi(X)$ в вершинах комплекса C^r , а сама формула (3.10.6) определяет

среднее отклонение значений функции $\Phi(X)$ в вершинах комплекса C^r от этого среднего значения.

Метод комплексов иллюстрирует рис. 3.10.3, на котором показан фрагмент линий уровня функции Химмельблау. На рисунке исходный комплекс C^r имеет вершины $X_i^r, i \in [1, \dots, 4]$. После отражения с растяжением вершины X_1^r этого комплекса, в которой функция имеет максимальное значение, получаем комплекс C^{r+1} с вершинами $X_i^{r+1}, i \in [1, \dots, 4]$. После отражения с растяжением вершины X_2^r комплекса C^{r+1} , в которой функция имеет максимальное значение, получаем комплекс C^{r+2} с вершинами $X_i^{r+2}, i \in [1, \dots, 4]$.

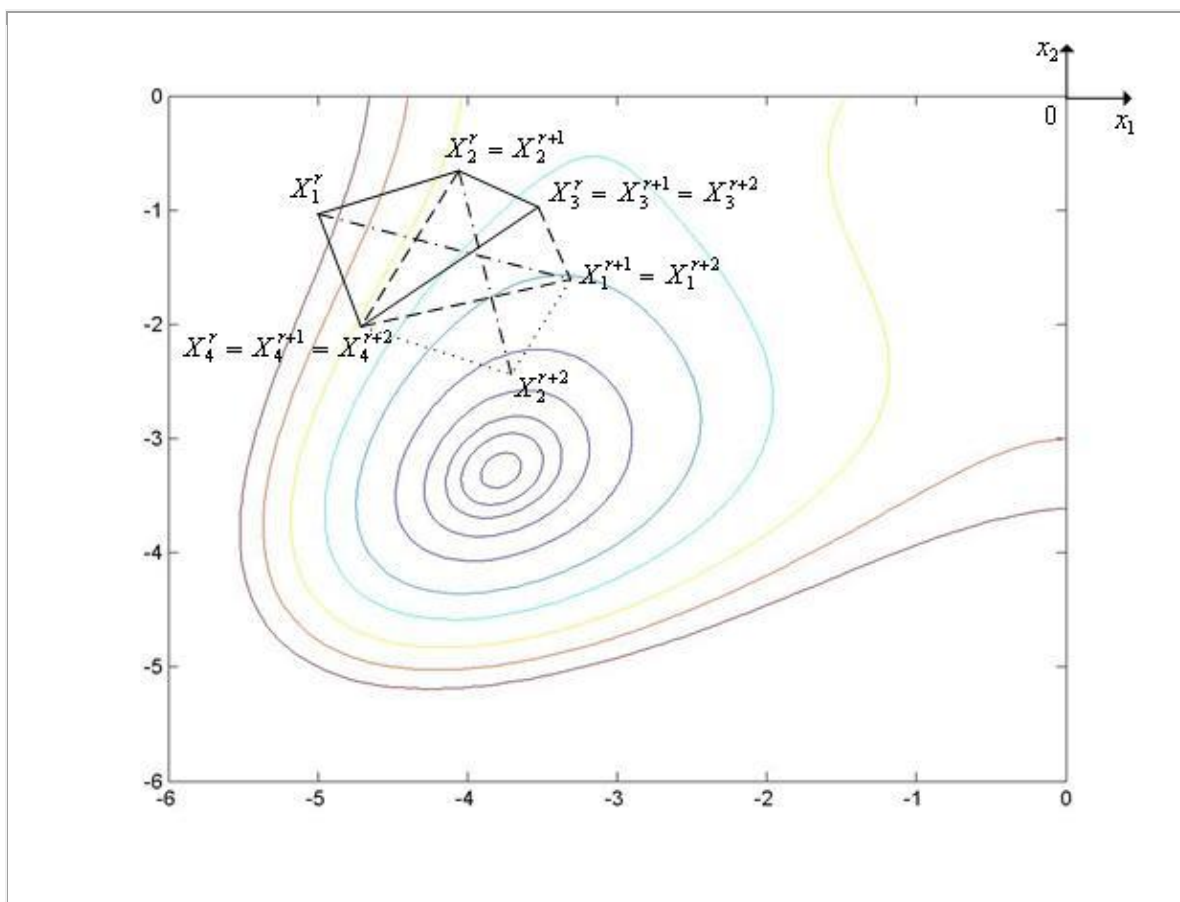


Рис. 3.10.3. Траектория поиска минимума функции Химмельблау методом комплексов

Известно множество модификаций рассмотренного метода комплексов, направленных, в частности, на преодоление «уплощения» комплекса в процессе поиска. С этой целью через фиксированное количество итераций находятся максимальная и минимальная диагонали комплекса и, если их отношение превышает заданное, то по рассмотренной схеме производится построение нового комплекса.

3.11. Метод повторяющегося случайного поиска

Рассматривается следующая многомерная задача локальной безусловной оптимизации: найти минимум критерия оптимальности $\Phi(X)$, определенного в n -мерном евклидовом пространстве R^n , $\min_{X \in R^n} \Phi(X) = \Phi(X^*) = \Phi^*$.

В методе повторяющегося случайного поиска (3-шаговый метод) используется итерационная схема (см. рис. 3.11.1)

$$X^{r+1} = X^r + \lambda^r \Delta^r, \quad (3.11.1)$$

где λ^r - величина шага (скаляр) на r -ой итерации, Δ^r - $(n \times 1)$ -вектор, определяющий направление шага на r -ой итерации:

$$\Delta^r = \left[\beta \frac{S^r}{\|S^r\|} + (1-\beta) \frac{P^r}{\|P^r\|} \right]. \quad (3.11.2)$$

Здесь $S^r = \gamma S^{r-1} + (1-\gamma) S^{r-2}$ - вектор «предыстории», определяющий среднее направление поиска на двух предыдущих шагах; $\|*\|$ - некоторая векторная норма; P^r - n -мерный вектор псевдослучайных чисел, равномерно распределенных в $[0,1]$; скаляр $\beta \in [0,1]$ - коэффициент, задающий относительные веса детерминированной и случайной компонент в векторе Δ^r (свободный параметр метода); скаляр $\gamma \in [0,1]$ - коэффициент, задающий относительные веса векторов S^{r-1}, S^{r-2} в векторе S^r (свободный параметр метода).

Заметим, что отношение $\frac{S^r}{\|S^r\|}$ представляет собой единичный вектор направления S^r , а отношение $\frac{P^r}{\|P^r\|}$ - единичный вектор направления P^r .

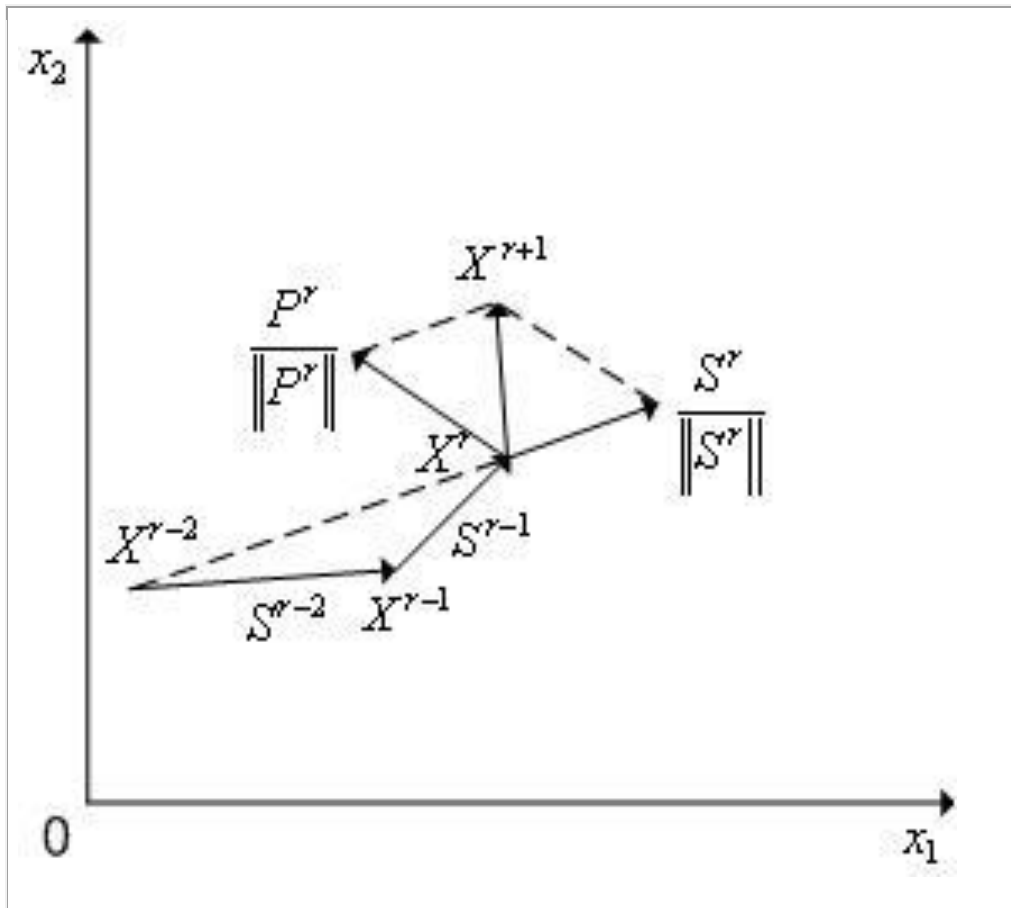


Рис. 3.11.1. К итерационной схеме метода повторяющегося случайного поиска

Принято $\gamma=0.5$, $\beta=0.5$, $\lambda^r=2$, так что $S^r=0.5(S^{r-1}+S^{r-2})$ и

$$X^{r+1} = X^r + \left[\frac{S^r}{\|S^r\|} + \frac{P^r}{\|P^r\|} \right].$$

Упрощенная схема метода повторяющегося случайного поиска.

1. Задаем начальную точку X^0 , начальный шаг λ^2 , значения коэффициентов β, γ и полагаем счетчик числа итераций $r=2$.
2. Тем или иным способом, например, с помощью одношагового метода наилучшей пробы определяем точки X^1, X^2 - этап «разгона» метода.

3. Генерируем n -мерный случайный вектор P^r и по формулам (3.11.1), (3.11.2) вычисляем координаты точки X^{r+1} и значение $\Phi(X^{r+1})$ функции $\Phi(X)$ в этой точке.

4. Если $\Phi(X^{r+1}) < \Phi(X^r)$, то проверяем условие окончания итераций (см. ниже). Если условие окончания выполнено, то полагаем $X^* \approx X^{r+1}$ и завершаем итерации. Если условие окончания итераций не выполнено, то некоторому правилу увеличиваем длину шага λ^r , например, полагая $\lambda^{r+1} = 2\lambda^r$, принимаем $r = r + 1$ и переходим к п.3. Если $\Phi(X^{r+1}) \geq \Phi(X^r)$, то переходим к п. 5.

5. Некоторое фиксированное количество раз делаем попытку, исходя из той же точки X^r , не меняя длины шага λ^r , добиться уменьшения значения функции $\Phi(X)$ путем только изменения вектора P^r , т.е., не меняя X^r и λ^r , переходим на п. 3. Если это фиксированное количество попыток не привело к успеху, то, исходя из той же точки X^r , по некоторому правилу уменьшаем длину шага λ^r , например, полагая $\lambda^r = \frac{\lambda^r}{2}$, и переходим к п.3.

В качестве условия окончания поиска можно использоваться одно из стандартных условий окончания итераций

$$\|X^{r+1} - X^r\| \leq \varepsilon_X, \quad (3.11.3)$$

где ε_X - константа, определяющая требуемую точность решения по X ;

$$|\Phi(X^{r+1}) - \Phi(X^r)| \leq \varepsilon_\Phi, \quad (3.11.4)$$

где ε_Φ - константа, определяющая требуемую точность решения по Φ .

Известно множество модификаций рассмотренной простейшей схемы метода повторяющегося случайного поиска. Например, в процессе поиска могут изменяться по некоторым правилам не только длина шага λ^r , но и коэффициенты β, γ .

Метод повторяющегося случайного поиска иллюстрирует рис. 3.11.2, на котором показан фрагмент линий уровня функции Химмельблау.

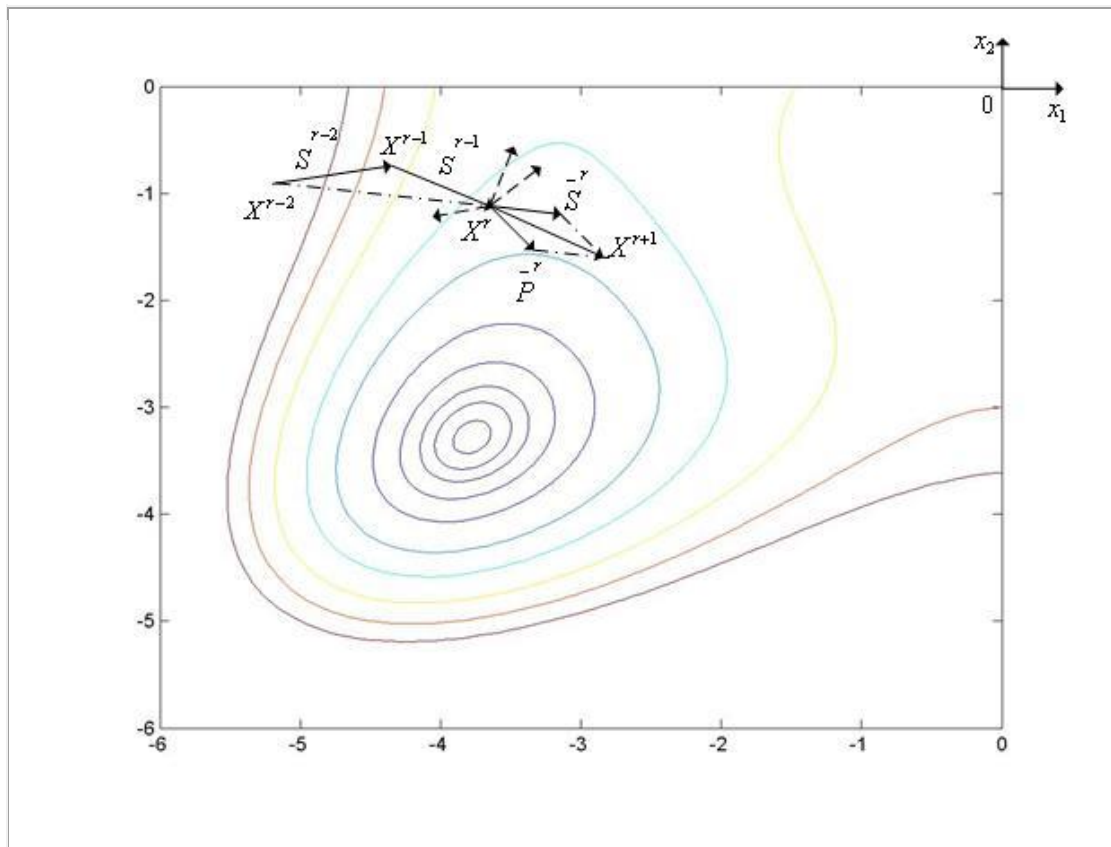


Рис. 3.11.2. Траектория поиска минимума функции Химмельблау методом повторяющегося случайного поиска

На рисунке принято $\tilde{S}^r = \frac{S^r}{\|S^r\|}$, $\tilde{P}^r = \frac{P^r}{\|P^r\|}$, $\gamma = 0.5$, $\beta = 0.5$, $\lambda^r = 2$, так что $S^r = 0.5(S^{r-1} + S^{r-2})$ и $X^{r+1} = X^r + [\tilde{S}^r + \tilde{P}^r]$. Пунктиром показаны отвергнутые векторы.

3.12. Метод случайного поиска с постоянным радиусом поиска и случайными направлениями

Рассматривается следующая многомерная задача локальной безусловной оптимизации: найти минимум критерия оптимальности $\Phi(X)$, определенного в n -мерном евклидовом пространстве R^n ,
$$\min_{X \in R^n} \Phi(X) = \Phi(X^*) = \Phi^*.$$

Метод случайного поиска с постоянным радиусом поиска и случайными направлениями использует процедуру генерации случайных точек, равномерно распределенных по поверхности гиперсферы в пространстве R^n . Пусть $X^c = (x_1^c, x_2^c, \dots, x_n^c)$ - вектор координат центра гиперсферы, ρ - радиус гиперсферы, R - вектор с началом в точке X^c и концом в искомой точке на поверхности гиперсферы, $\alpha_i, i \in [1, \dots, n]$ - углы между вектором R и ортами координатных осей $Ox_i, i \in [1, \dots, n]$.

Во введенных обозначениях схема алгоритма генерации случайных точек, равномерно распределенных по поверхности гиперсферы радиуса ρ , может быть представлена в следующем виде:

- генерируем n случайных чисел, равномерно распределенных в интервале $[0, 2\pi)$;
- вычисляем направляющие косинусы $\cos \alpha_i, i \in [1, \dots, n]$ вектора R ;
- находим координаты искомой точки $x_i = x_i^c + \rho \cos \alpha_i, i \in [1, \dots, n]$.

Упрощенная схема метода случайного поиска с постоянным радиусом поиска и случайными направлениями.

1. Задаем начальную точку X^0 , начальный радиус гиперсферы ρ^0 , и полагаем счетчик числа итераций $r = 0$.

2. Генерируем случайные точки $X_i^r, i \in [1, \dots, N]$ равномерно распределенные по поверхности гиперсферы радиуса ρ^r с центром в точке X^r . Здесь N – количество точек – свободный параметр метода.
3. Вычисляем значения минимизируемой функции $\Phi(X_i^r)$ в полученных точках и находим точку, в которой достигается минимальное значение функции $\Phi(X)$: $\Phi(X_k^r) = \min_{i \in [1, \dots, N]} \Phi(X_i^r)$.
4. Каким-либо из рассмотренных в главе 2 одномерных методов оптимизации (например, методом Пауллы) находим минимум функции $\Phi(X)$ в направлении $(X_k^r - X^r)$: $\Phi(X^{r+1}) = \min_{\lambda \in \mathbb{R}^1} \Phi(X^r + \lambda(X_k^r - X^r))$.
5. Проверяем условие окончания итераций (см. ниже). Если условие окончания выполнено, то полагаем $X^* \approx X^{r+1}$ и завершаем итерации. Иначе - принимаем $r = r + 1$ и переходим к п.2.

В качестве условия окончания поиска можно использоваться одно из стандартных условий окончания итераций:

$$\|X^{r+1} - X^r\| \leq \varepsilon_X, \quad (3.12.1)$$

где ε_X - константа, определяющая требуемую точность решения по X ;

$$|\Phi(X^{r+1}) - \Phi(X^r)| \leq \varepsilon_\Phi, \quad (3.12.2)$$

где ε_Φ - константа, определяющая требуемую точность решения по Φ .

Могут быть использованы также другие критерии окончания поиска, например, условие не превышения текущим радиусом гиперсферы величины ε_X :

$$\rho^r \leq \varepsilon_X. \quad (3.12.3)$$

В процессе поиска радиус гиперсферы может меняться, увеличиваясь при удачных шагах (вдали от точки X^*) и уменьшаясь при неудачных шагах (вблизи от точки X^*).

Поиск может быть ускорен, если точки на гиперсфере выбирать (случайным образом) в некотором секторе по отношению к предыдущему направлению. Угол раскрытия этого сектора может меняться в процессе поиска.

Метод случайного поиска с постоянным радиусом поиска и случайными направлениями иллюстрирует рис. 3.12.1, на котором показан фрагмент линий уровня функции Химмельблау.

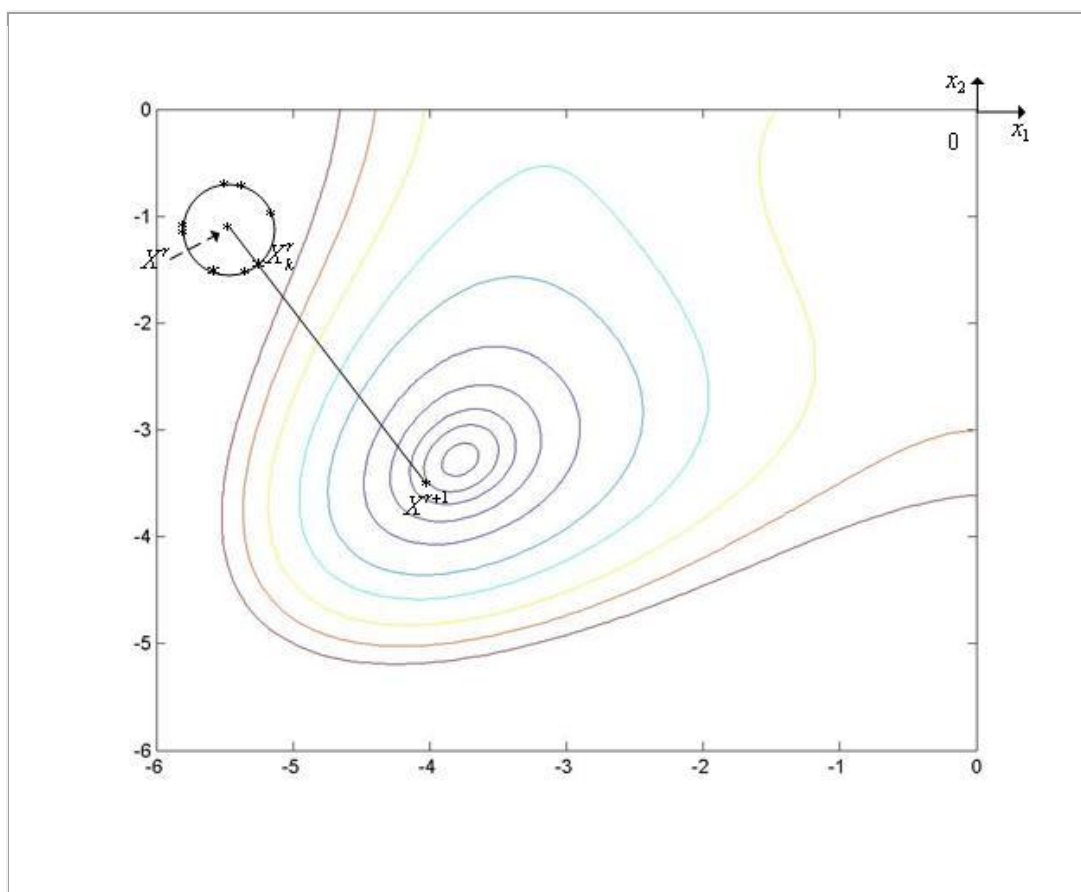


Рис. 3.12.1. Траектория поиска минимума функции Химмельблау методом случайного поиска с постоянным радиусом поиска и случайными направлениями ($n = 2$)

На рисунке точки, лежащие на окружности с центром в точке X^r , соответствуют случайным точкам X_i^r , $i \in [1, \dots, N]$.

Примечание 3.12.1. Одна итерация по методу случайного поиска с постоянным радиусом поиска и случайными направлениями может привести к уменьшению минимизируемой функции в большей степени, чем один шаг поиска в направлении антиградиента этой функции.

Данное утверждение иллюстрирует рис. 3.12.2, на котором показаны линии уровня двумерной квадратичной функции (см. параграф 3.4).

$$\Phi(x) = \Phi(x_1, x_2) = x_1^2 + x_2^2 + 3(x_1 + x_2)^2 = 4x_1^2 + 4x_2^2 + 6x_1x_2.$$

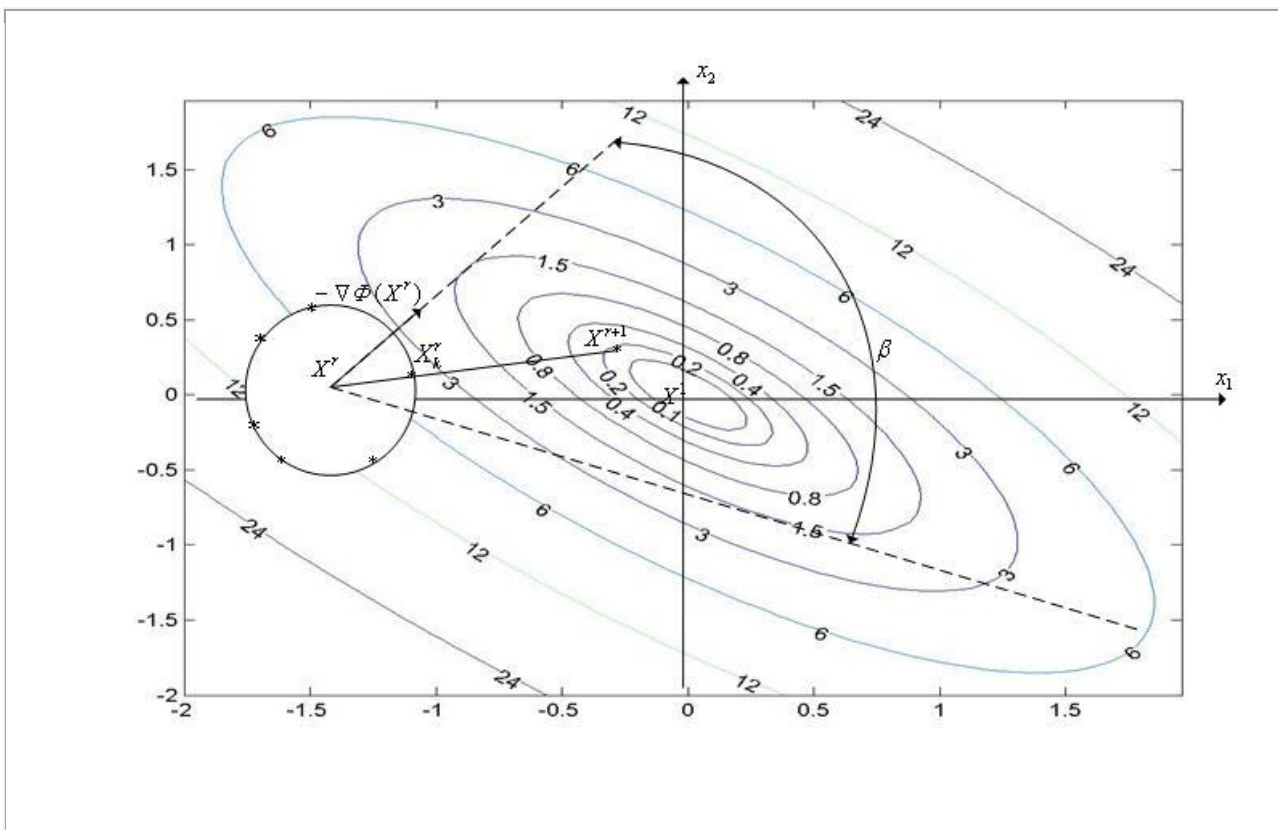


Рис. 3.12.2. Один шаг поиска в направлении антиградиента минимизируемой функции ($\Phi(X^r)$) приводит на линию уровня (1.5). В то же время одна итерация по методу случайного поиска с постоянным радиусом поиска и случайными направлениями – на линию уровня ~ 0.2 . Точки на окружности с центром в точке X^r соответствуют случайным точкам. Любое направление поиска в секторе β лучше, чем направление антиградиента