

2 Методы последовательной безусловной оптимизации

$$\min_{\bar{X} \in D} \Phi(\bar{X}) = \Phi(\bar{X}^*) = \Phi^*$$

где $\Phi(X)$ – произвольная функция,

$$D = \{\bar{X} \mid h_i(\bar{X}) = 0, g_j(\bar{X}) \geq 0, i \in [1, \dots, m], j \in [1, \dots, l]\}$$

Основная идея – сведение задачи условной минимизации к задаче (к последовательности задач) безусловной минимизации

$$\min_{\bar{X} \in R^n} Q^\alpha(\bar{X}) = \min_{\bar{X} \in R^n} (\Phi(\bar{X}) + P^\alpha(\bar{X})) = Q^\alpha(\bar{X}^*)$$

где $P^\alpha(\bar{X})$ – функции, которые возрастают вблизи границ области D тем быстрее, чем больше значение α .

В качестве приближенного решения исходной задачи принимается решение \bar{X}^* вспомогательной задачи при достаточно большом α

Пример

Найти минимум функции

$$\Phi(\bar{X}) = (x_1 - 3)^2 + (x_2 - 2)^2$$

При ограничении

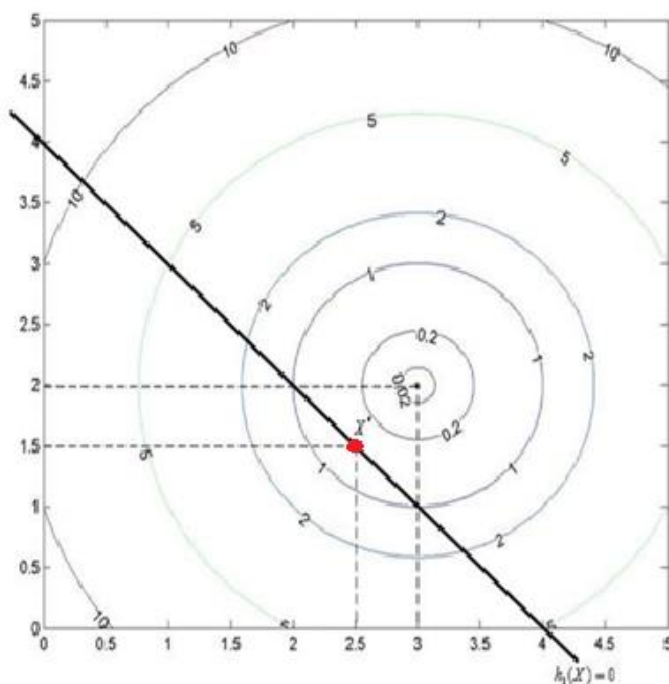
$$h(\bar{X}) = x_1 + x_2 - 4 = 0$$

Решение - точка X^* с координатами (2.5, 1.5).

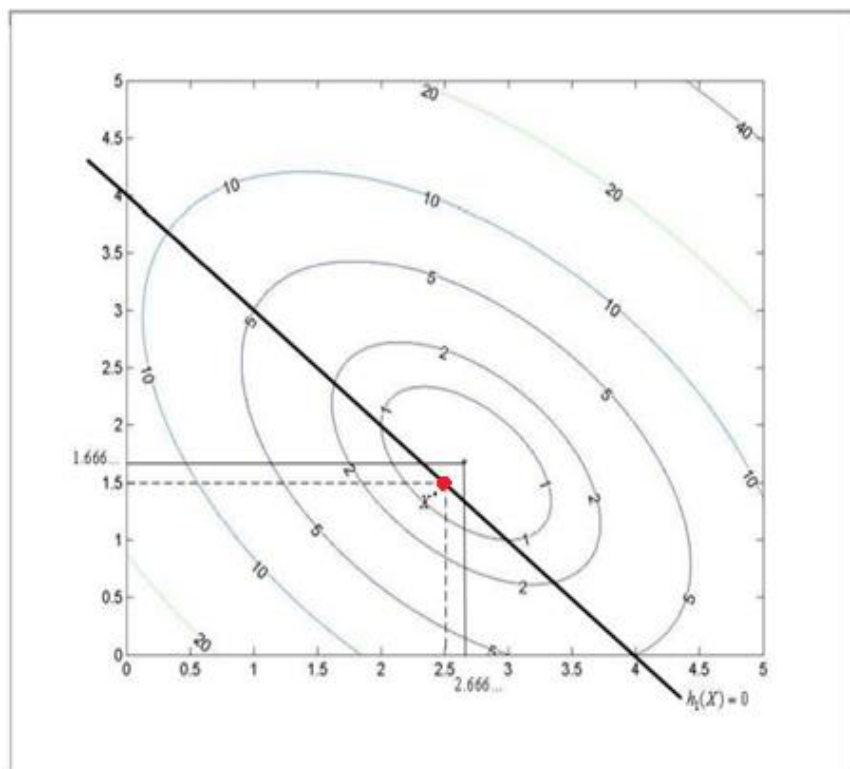
Пусть $P^\alpha(\bar{X}) = \alpha(h(\bar{X}))^2 = \alpha(x_1 + x_2 - 4)^2$, где $\alpha \geq 0$ -- константа

Линии уровня $Q^\alpha(\bar{X}) = \Phi(\bar{X}) + P^\alpha(\bar{X})$ при разных $\alpha \geq 0$:

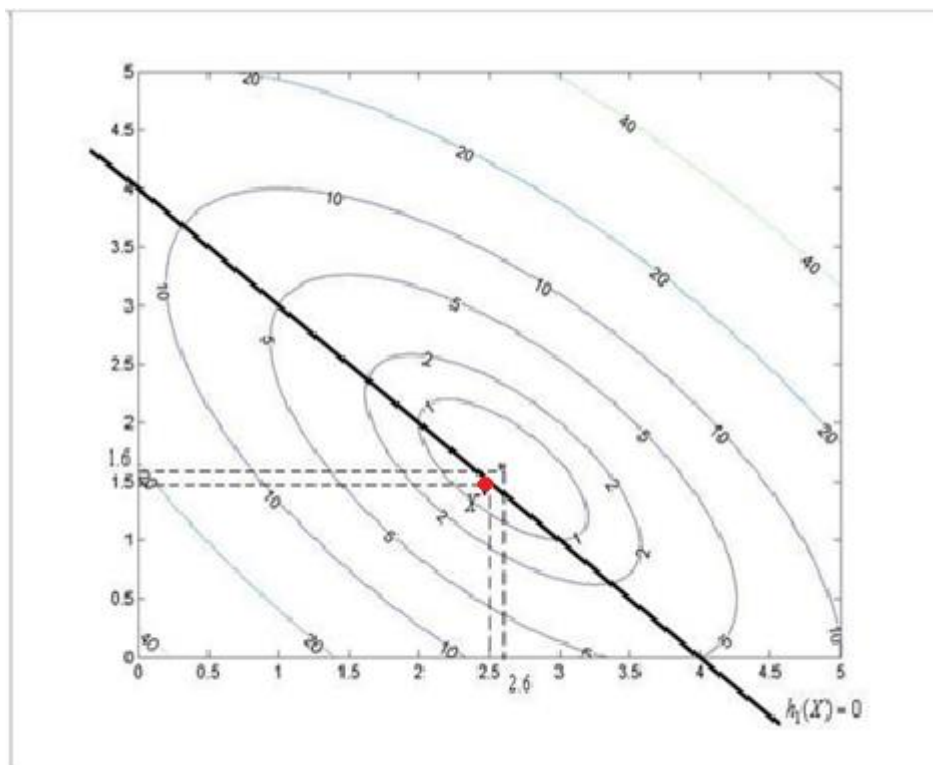
А) Точка минимума функции $Q^\alpha(\bar{X})$ при $\alpha = 0$ имеет координаты (3, 2).



Б) Точка минимума функции $Q^\alpha(\bar{X})$ при $\alpha = 1$ имеет координаты $(2.66..., 1.66...)$



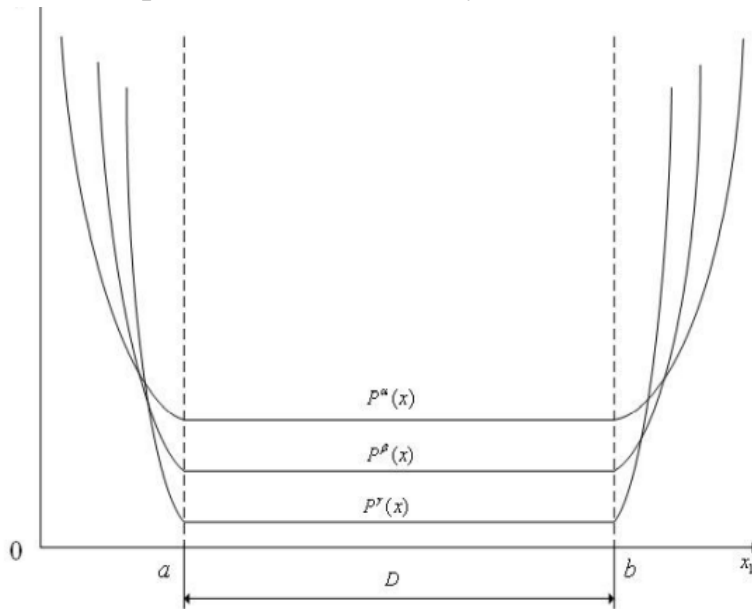
В) Точка минимума функции $Q^\alpha(\bar{X})$ при $\alpha = 2$ имеет координаты $(2.6, 1.6)$



Таким образом, при увеличении α минимум $Q^\alpha(\bar{X})$ приближается к решению исходной задачи.

Среди методов последовательной безусловной оптимизации выделяют 2 подхода - **метод штрафных функций** и метод **барьерных функций**.

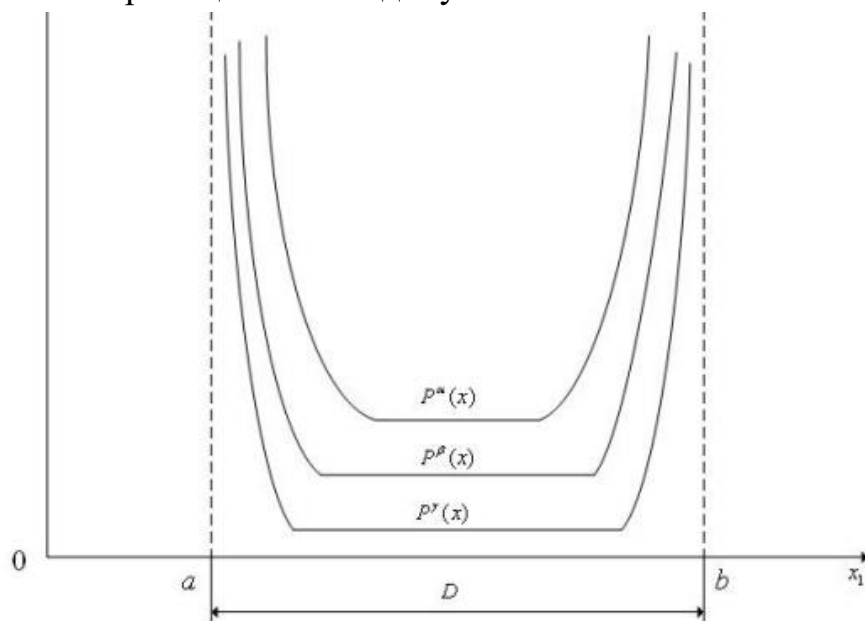
2а В методе штрафных функций функцию P , которая в этом случае называется штрафной функцией, подбирают таким образом, чтобы при больших α функция Q мало отличалась от функции Φ при $X \in D$ и быстро возрастала при удалении точки $X \notin D$ от границы области допустимых значений D .



$D=[a,b]$ – область допустимых значений, $\alpha < \beta < \gamma$

В методе штрафных функций точка X в процессе поиска может выходить за границы области D (см. рис.), т.е. метод штрафных функций относится к классу **методов внешней точки**.

26 В методе барьерных функций функцию P , которая в этом случае называется барьерной функцией, подбирают таким образом, чтобы при больших α функция Q мало отличалась от функции Φ при $X \in D$ и быстро возрастала при приближении точки $X \in D$ к границе области допустимых значений D .



$D=[a,b]$ – область допустимых значений, $\alpha < \beta < \gamma$

В методе барьерных функций точка X в процессе поиска не может выходить за границы области D (см. рис.), т.е. метод барьерных функций относится к классу **методов внутренней точки**.

В вычислительной практике преимущественно используется метод штрафных функций.

МЕТОД ШТРАФНЫХ ФУНКЦИЙ

Штрафная функция в общем случае имеет вид:

$$P(\bar{X}, \alpha_h^r, \alpha_g^r) = \alpha_h^r \sum_{i=1}^m \mu_i H(h_i(\bar{X})) + \alpha_g^r \sum_{j=1}^l \lambda_j G(g_j(\bar{X}))$$

где $\alpha^r = (\alpha_h^r, \alpha_g^r)$ – двумерный вектор параметров штрафной функции

μ_i, λ_j – весовые коэффициенты, которые могут изменяться в процессе итераций
 H и G – функционалы.

Функционалы H и G должны удовлетворять очевидным требованиям:

$$H(h_i(\bar{X})) \rightarrow 0 \quad \text{при} \quad h_i(\bar{X}) \rightarrow 0; \quad \lim_{r \rightarrow \infty} \left(\alpha_h^r \sum_{i=1}^m \mu_i H(h_i(\bar{X})) \right) = \begin{cases} 0, & \bar{X} \in D \\ \infty, & \bar{X} \notin D \end{cases}$$

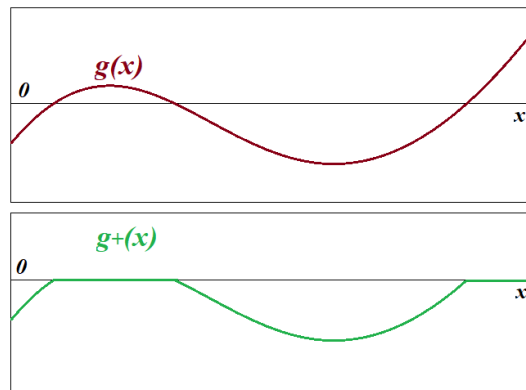
$$G(g_j(\bar{X})) \rightarrow 0 \quad \text{при} \quad g_j(\bar{X}) \rightarrow 0; \quad \lim_{r \rightarrow \infty} \left(\alpha_g^r \sum_{j=1}^l \lambda_j G(g_j(\bar{X})) \right) = \begin{cases} 0, & \bar{X} \in D \\ \infty, & \bar{X} \notin D \end{cases}$$

В качестве функционалов H и G можно взять расстояния в какой-либо метрике от точки X до соответствующей границы множества D . Однако, вычисление этих расстояний, а значит и значений штрафной функции, может быть затруднительным. Поэтому обычно применяют штрафные функции более удобного вида, например

$$H(h_i(\bar{X})) = (h_i(\bar{X}))^2$$

$$G(g_j(\bar{X})) = (g_j^+(\bar{X}))^2$$

$$\text{где } g_j^+(\bar{X}) = \begin{cases} 0, & g_j(\bar{X}) \geq 0 \\ g_j(\bar{X}), & g_j(\bar{X}) < 0 \end{cases}$$



В качестве **критерия окончания итераций** в методе последовательной безусловной оптимизации можно использовать неравенство

$$\|\bar{X}^r - \bar{X}^{r/2}\| \leq \epsilon p s x$$

где r – четное число итераций, $\epsilon p s x$ – требуемая точность решения по X .

Схема метода штрафных функций

1. Задаем начальную точку X^0 и полагаем счетчик числа итераций $r = 0$.
2. Исходя из точки X^r , одним из методов локальной безусловной оптимизации решаем задачу – находим точку X^{r+1} .
3. Проверяем условие окончания поиска

$$\|\bar{X}^r - \bar{X}^{r/2}\| \leq \epsilon p s x$$

Если условие окончания поиска выполнено, то полагаем $X^* = X^{r+1}$ и завершаем итерации.

Иначе – по некоторому правилу увеличиваем значения параметров α , полагаем $r = r + 1$ и переходим к п.3.

Примечание

В зависимости от метода локальной безусловной оптимизации, который используется для решения задач, метод последовательной безусловной оптимизации может быть детерминированным и случайным, нулевого, первого или второго порядка

Недостатком метода последовательной безусловной оптимизации является значительное усложнение структуры минимизируемой функции.

3 Метод скользящего допуска

Найти минимум критерия оптимальности $\Phi(X)$, определенного на множестве D евклидова пространства R^n

$$\min_{\bar{X} \in D} \Phi(\bar{X}) = \Phi(\bar{X}^*) = \Phi^*$$

где область определения

$$D = \{\bar{X} \mid g_j(\bar{X}) \geq 0, j \in [1, \dots, J]\}$$

Пусть $T(X)$ – неотрицательно определенный функционал над множеством всех ограничивающих функций g_j такой, что $T(X)=0$ при $X \in D$ и значение $T(X)$ возрастает по мере удаления точки X от границы области допустимых значений D .

Рассмотрим множество

$$D^r = \{\bar{X} \mid T(\bar{X}) \leq \Delta^r\}, \text{ где } \Delta^r \geq 0 - \text{критерий скользящего допуска (скаляр)}.$$

Критерий скользящего допуска Δ^r определяет требуемую точность выполнения ограничений, которые формируют область допустимых значений D , и конструируется таким образом, чтобы обеспечить его уменьшение с ростом количества итераций r .

Точка X называется

- допустимой точкой, если $T(X) = 0$
- почти допустимой точкой – если $0 < T(X) \leq \Delta^r$
- недопустимой точкой – если $T(X) > \Delta^r$

Поскольку величина Δ^r с ростом номера итерации уменьшается, отклонение от границы области D , при котором точка считается допустимой, сужается, так что в пределе рассматриваются только допустимые точки.

Метод скользящего допуска может быть скомбинирован со многими из многомерных методов локальной безусловной оптимизации.

Метод, с которым комбинируется метод скользящего допуска – **базовый метод**.

Одна итерация метода скользящего допуска состоит из одного или двух этапов:

1. С помощью базового метода, исходя из точки X^r , выполняем итерацию по решению задачи локальной безусловной оптимизации:

$$\min_{\bar{X} \in R^n} \Phi(\bar{X}) = \Phi(\bar{X}^{r+1*}), \text{ находим точку } X^{r+1*}$$

Если $T(\bar{X}^{r+1*}) \leq \Delta^{r+1}$ (точка X^{r+1*} является допустимой точкой или почти допустимой точкой), то полагаем $X^{r+1} = X^{r+1*}$ и закидываем данную итерацию.

2. Если $T(\bar{X}^{r+1*}) > \Delta^{r+1}$ (точка X^{r+1*} является недопустимой), то отыскиваем точку X^{r+1} , лежащую ближе к границе области D .

Для этого с помощью того же базового метода, исходя из точки X^{r+1*} , решаем задачу локальной безусловной оптимизации

$$\min_{\bar{X} \in R^n} T(\bar{X}) = T(\bar{X}^{r+1})$$

с условием окончания итераций $T(\bar{X}^{r+1}) \leq \Delta^{r+1}$ и заканчиваем данную итерацию.

Достоинством метода скользящего допуска является то, что степень нарушения ограничений по мере приближения к минимуму минимизируемой функции постепенно уменьшается. Т.е. на первых итерациях ограничения могут удовлетворяться приближенно, а высокая точность удовлетворения ограничений необходима лишь в окрестности решения. Это обстоятельство позволяет сократить полный объем вычислений по сравнению с другими методами.

Одна из сложностей применения метода скользящего допуска – возможные осцилляция решения относительно границы области D (рассмотрим позже).

Пример. Комбинация метода скользящего допуска с методом Нелдера-Мида

При комбинации метода скользящего допуска с методом Нелдера-Мида можно предложить разные виды критерия скользящего допуска.

Например, в качестве этого критерия можно использовать следующую функцию координат вершин деформируемого многогранника $(\bar{X}_1^r, \bar{X}_2^r, \dots, \bar{X}_{n+1}^r)$:

$$\begin{cases} \Delta^0 = \frac{1}{n+1} \sum_{i=1}^{n+1} \|\bar{X}_c^0 - \bar{X}_i^0\| \\ \Delta^r = \min \left\{ \Delta^{r-1}, \frac{1}{n+1} \sum_{i=1}^{n+1} \|\bar{X}_c^r - \bar{X}_i^r\| \right\} \end{cases}$$

Где \bar{X}_c^r -- вектор координат центра тяжести многогранника $(\bar{X}_1^r, \bar{X}_2^r, \dots, \bar{X}_{n+1}^r)$

Таким образом построенный критерий Δ^r является положительно определенной функцией координат многогранника.

Т.к. размер многогранника уменьшается при приближении к точке минимума X^* (в пределе до 0), то

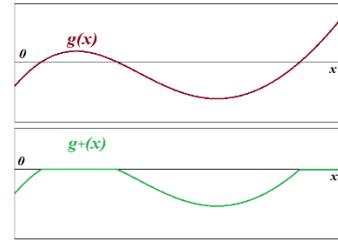
$$\lim_{r \rightarrow \infty} \Delta^r = 0$$

Каким образом строится $T(X)$?

В качестве функционала $T(X)$ обычно используется

$$T(\bar{X}) = \sum_{j=1}^l (g_j^+(\bar{X}))^2$$

$$\text{где } g_j^+(\bar{X}) = \begin{cases} 0, & g_j(\bar{X}) \geq 0 \\ g_j(\bar{X}), & g_j(\bar{X}) < 0 \end{cases}$$



Функционал $T(X)$ обладает следующим свойством:

$T(X)=0$ при $X \in D$, $T(X)>0$ иначе

Если $T(X)$ мало, то X находится недалеко от границы области D .

Упрощенная схема комбинации метода скользящего допуска и метода Нелдера-Мида

Симплекс с вершинами \bar{X}_i^r $i=1, \dots, n+1$ обозначим S^r

1. Задаем начальный симплекс S^0 и полагаем счетчик числа итераций $r=0$.

2. С помощью метода Нелдера-Мида, исходя из симплекса S^r , выполняем одну итерацию по решению задачи локальной безусловной оптимизации – находим

симплекс \tilde{S}^{r+1} с вершинами $\tilde{\bar{X}}_i^r$

3. Вычисляем значения функционала $T(X)$ во всех вершинах симплекса \tilde{S}^{r+1}

и значение критерия скользящего допуска Δr . Находим вершину симплекса \tilde{S}^{r+1} , в которой значение функционала $T(X)$ максимально, т.е. вершину, которая расположена дальше всех от границы области D . Обозначим эту вершину $\tilde{\bar{X}}_T^{r+1}$

4. Если $T(\tilde{\bar{X}}_T^{r+1}) \leq \Delta^{r+1}$ (точка $\tilde{\bar{X}}_T^{r+1}$ является **допустимой точкой или почти допустимой точкой**), то проверяем условие окончания поиска (см. схему метода Нелдера-Мида). Если это условие выполнено, то завершаем итерации. Если условие окончания поиска не выполнено, то формируем симплекс S^{r+1} с вершинами

$\bar{X}_i^{r+1} = \tilde{\bar{X}}_i^{r+1}$, полагаем $r=r+1$ и переходим к п.2.

5. Если $T(\tilde{\bar{X}}_T^{r+1}) > \Delta^{r+1}$ (точка $\tilde{\bar{X}}_T^{r+1}$ является **недопустимой точкой**), то с помощью метода Нелдера-Мида, исходя из точки $\tilde{\bar{X}}_T^{r+1}$, решаем задачу локальной безусловной оптимизации

$$\min_{\bar{X} \in R^n} T(\bar{X}) = T(\bar{X}^{r+1})$$

с условием окончания итераций $T(\bar{X}^{r+1}) \leq \Delta^{r+1}$ и находим точку \bar{X}_T^{r+1} .

Формируем новый симплекс S_{r+1} с вершинами $\bar{X}_i^{r+1} = \tilde{\bar{X}}_i^{r+1}$, $i \neq T$, \bar{X}_T^{r+1} полагаем $r=r+1$ и переходим к п.2.

Эффект осцилляций решения

Как отмечалось выше, одной из сложностей применения метода скользящего допуска являются возможные осцилляции решения относительно границы области D.

Пример

Рассмотрим двумерную задачу условной оптимизации, когда критерий оптимальности равен

$$\Phi(\bar{X}) = -x_1 - x_2$$

и множество допустимых значений D определяется ограничениями:

$$\begin{cases} g_1(\bar{X}) = -x_1^2 - x_2^2 + 9 \geq 0 \\ g_2(\bar{X}) = -x_1 \geq 0 \\ g_3(\bar{X}) = -x_2 \geq 0 \end{cases}$$

Пусть на r -й итерации координаты вершин текущего симплекса S^r равны $\bar{X}_1^r, \bar{X}_2^r, \bar{X}_3^r$. Тогда после одной итерации по решению задачи локальной безусловной оптимизации методом Нелдера-Мида получим симплекс S^{r+1} с вершинами $\tilde{\bar{X}}_1^r, \tilde{\bar{X}}_2^r, \tilde{\bar{X}}_3^r$ (см.рис)

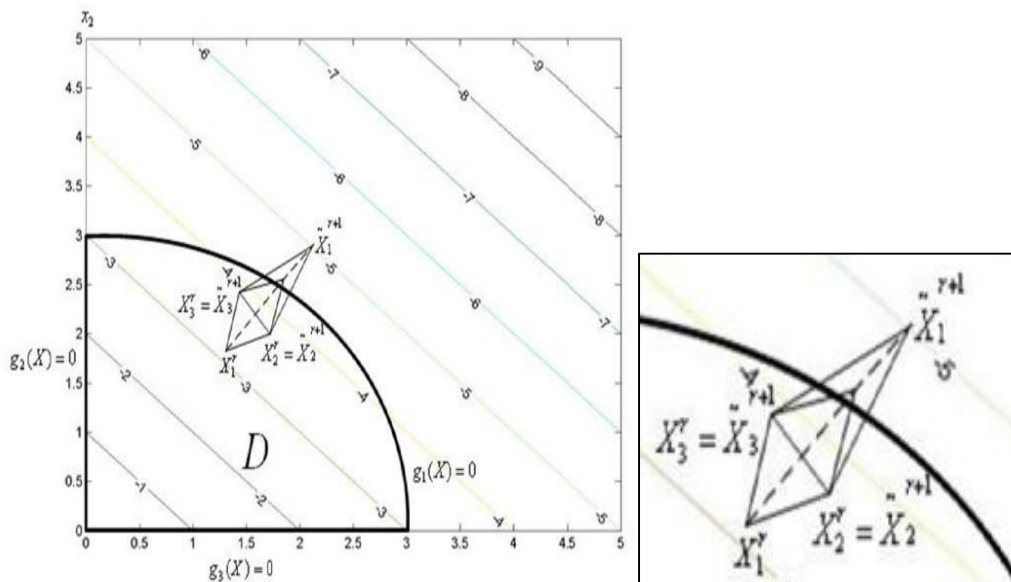


Рис. После успешного отражения вершины \bar{X}_1^r выполнено успешное растяжение симплекса

Пусть при этом $\tilde{\bar{X}}_1^{r+1}$, расположенная далее всех от границы области D , является недопустимой точкой, т.е. $T(\tilde{\bar{X}}_1^{r+1}) > \Delta^{r+1}$

Тогда при решении помощью метода Нелдера-Мида задачи локальной безусловной оптимизации возможна ситуация, приведенная на рис.

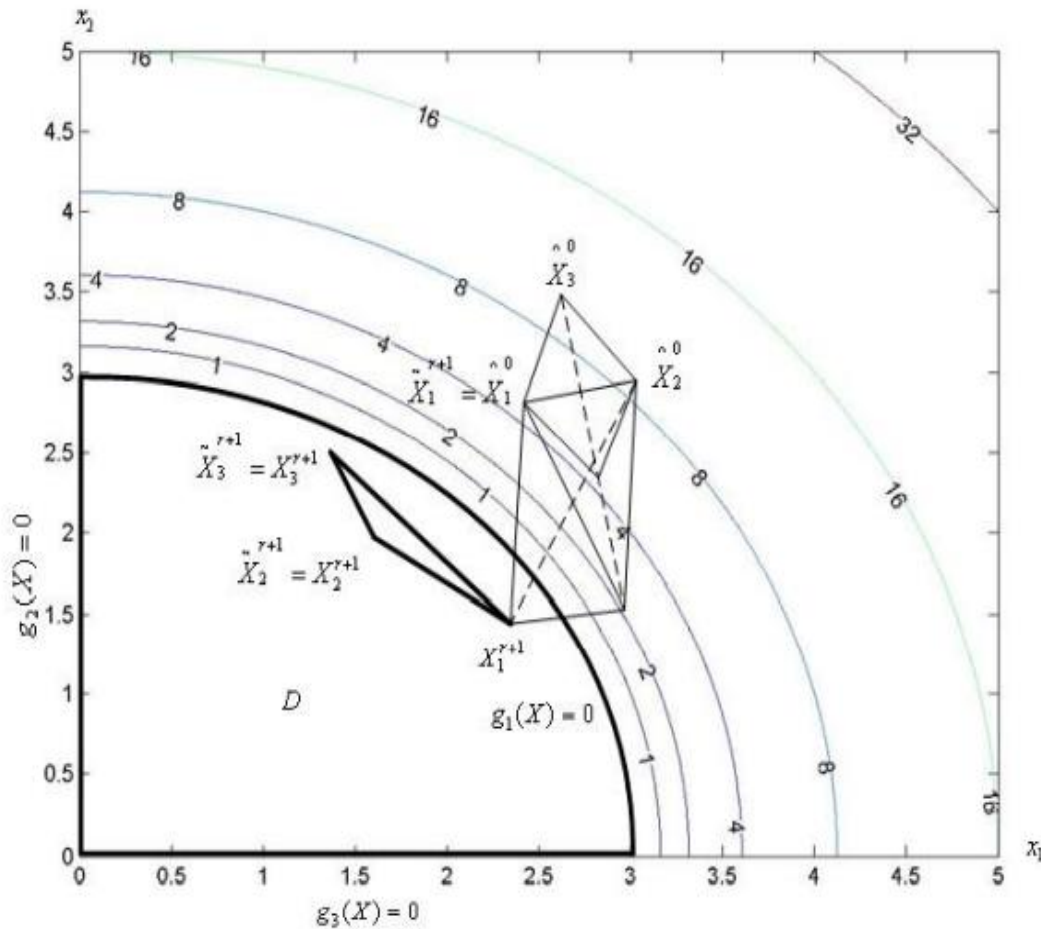


Рис. После успешного отражения вершины \bar{X}_3^0 выполнено растяжение симплекса и отражение вершины \bar{X}_2^0

Если $X \notin D$ и точка X лежит в первой четверти системы координат (см рис), то $T(\bar{X}) = x_1^2 + x_2^2 - 9$ ($T(X)$ показана линиями уровня).

Т.к. вершина X_1^r симплекса S^r расположена далеко от границы области D , то после операций отражения и растяжения точка X^{r+1} может оказаться глубоко в недопустимой области. В результате в процессе минимизации функционала $T(X)$ может получиться точка X^{r+1} , которая снова оказывается далеко от границы области D , и т.д. — этот эффект называется **осцилляцией решения относительно границы области D** .

Для ослабления влияния осцилляций:

а) в простейшем случае можно вместо точки $\tilde{\tilde{X}}_T^{r+1}$ использовать точку A^r – середину отрезка $\left[\bar{X}_T^r, \tilde{\tilde{X}}_T^{r+1} \right]$

б) чаще с этой целью используют квадратичную интерполяцию функции $T(X)$ на отрезке $\left[\bar{X}_T^r, \tilde{\tilde{X}}_T^{r+1} \right]$ по трем точкам $\left(\bar{X}_T^r, \Phi(\bar{X}_T^r) \right), \left(A^r, \Phi(A^r) \right), \left(\tilde{\tilde{X}}_T^{r+1}, \Phi(\tilde{\tilde{X}}_T^{r+1}) \right)$, где A^r – также середина отрезка $\left[\bar{X}_T^r, \tilde{\tilde{X}}_T^{r+1} \right]$.

Обозначим эту интерполирующую функцию $y(x) = \alpha x^2 + \beta x + \gamma$.

Вместо точки $\tilde{\tilde{X}}_T^{r+1}$ в этом случае можно использовать один из нулей функции $y(x)$ либо его приближенное значение, найденное, например, методом касательных.

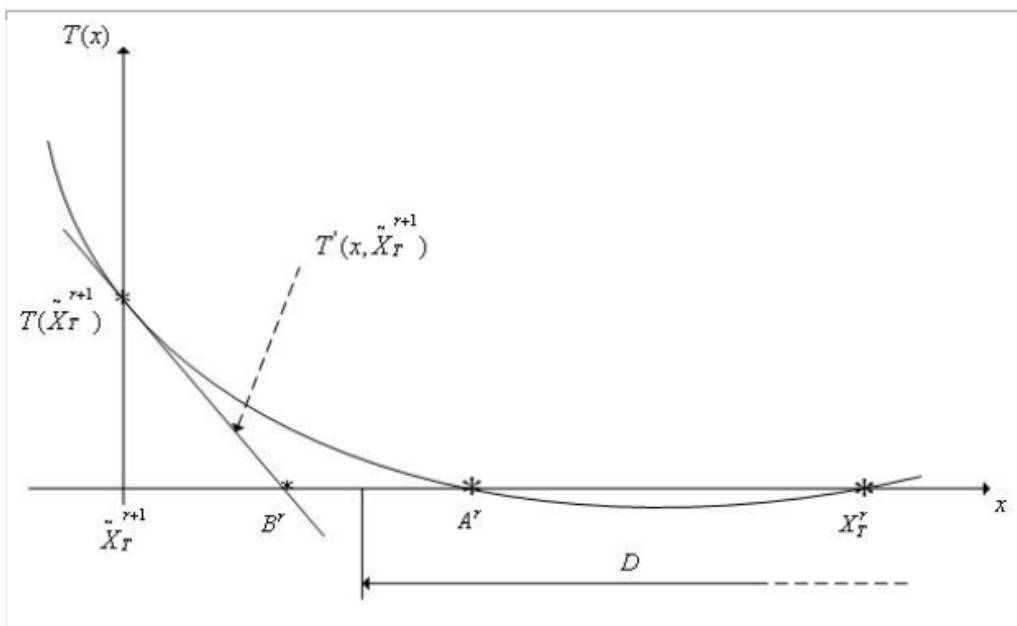


Иллюстрация. Использование квадратичной интерполяции функции $T(X)$ на отрезке по трем точкам для ослабления осцилляций. Случай, когда точка A^r принадлежит области допустимых значений D .

4 Модифицированный метод комплексов

Рассмотрим многомерную задачу локальной условной оптимизации:

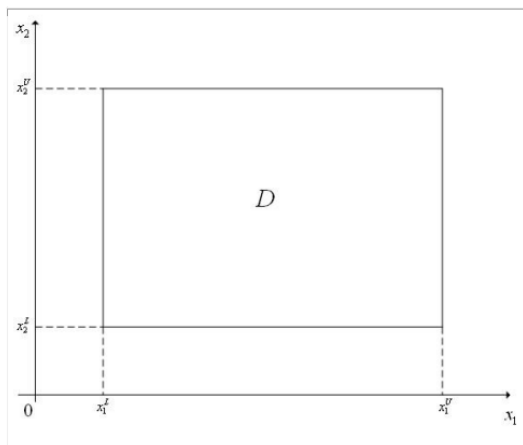
$$\min_{\bar{X} \in D \subset R^n} \Phi(\bar{X}) = \Phi(\bar{X}^*) = \Phi^*$$

где множество допустимых значений определяется только *ограничениями типа неравенств* и представляет собой *гиперпараллелепипед*, т.е.

$$D = \{ \bar{X} / x_i^U \geq x_i \geq x_i^L \}, \quad i \in [1, \dots, n]$$

с условием окончания итераций $T(\bar{X}^{r+1}) \leq \Delta^{r+1}$ и

Здесь x_i^U, x_i^L -- верхняя и нижняя границы области допустимых значений D по i-ому измерению (см. рис).



Область допустимых значений D в виде гиперпараллелепипеда, $n = 2$

Метод комплексов в многомерной задаче безусловной оптимизации – см лекция.

Рассмотрим модификацию этого метода для решения многомерной задачи *условной* оптимизации – т.н. *модифицированный метод комплексов*.

Основные операции метода комплексов

Напомним, что комплексом называется многогранник с $N > n+1$ вершинами (не обязательно выпуклый). Рекомендуется использовать комплекс с $N=2n$ вершинами.

Так же, как при решении задачи безусловной оптимизации, при решении задачи условной минимизации методом комплексов используются:

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

Упрощенная схема модифицированного метода комплексов

1. Задаем начальную точку $X_0^0 \in D$, исходя из которой должен быть построен комплекс C_0 , величину l^0 , α и β - коэффициенты сжатия/растяжения; полагаем счетчик числа итераций $r=0$.

2. Строим начальный комплекс C_0 :

– поочередно для $i=1,2,\dots,N$ находим координаты вершин X_i^r комплекса C_r ;

– если вершина X_p^r является **НЕ**допустимой (выходит за границы области D), то выполняем сжатие уже построенного комплекса с p вершинами, вдоль направления $(X_p^r, (X_c^r)_p)$, где $(X_c^r)_p$ — центр тяжести уже найденных $(p-1)$ вершин комплекса;

– если после сжатия комплекса вершина X_p^r по-прежнему является недопустимой, повторяем процедуру сжатия;

– вычисляем значения функции $\Phi(X)$ во всех вершинах построенного комплекса C_r .

3. Находим максимальное из значений функции $\Phi(X)$ в вершинах комплекса C_r :

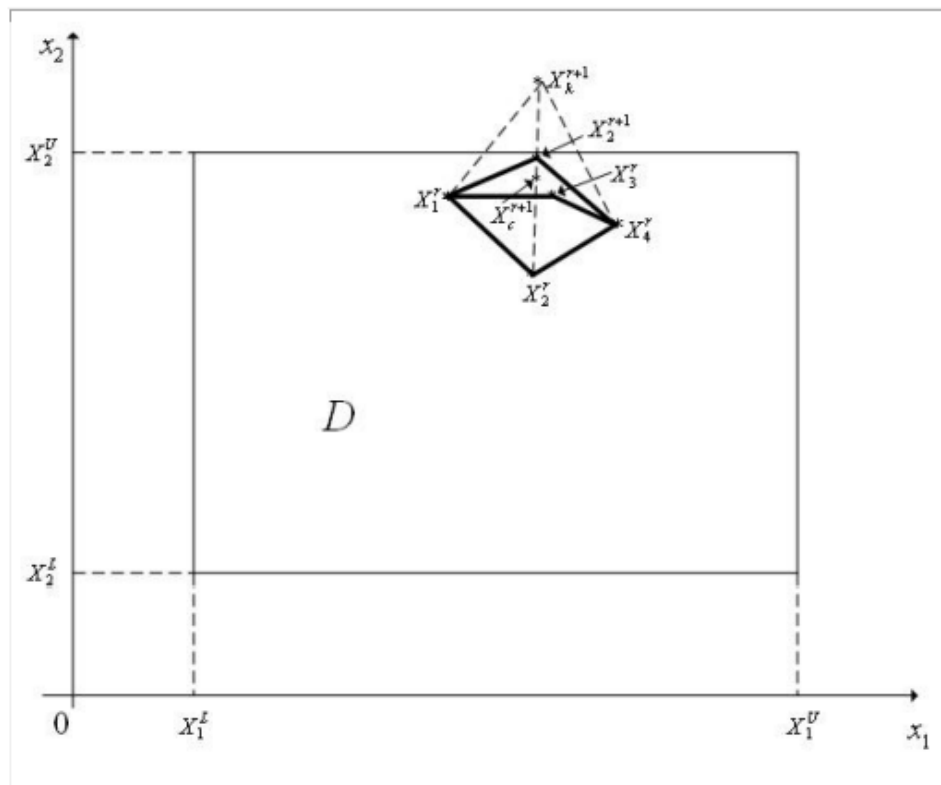
$$\Phi(X_k^r) = \max_{i=1,\dots,N} \Phi(X_i^r)$$

4. Отражаем с растяжением вершину X_k^r комплекса C_r — получаем вершину X_k^{r+1} и новый комплекс C_{r+1} :

– если точка X_k^{r+1} является **НЕ**допустимой (выходит за границы области D) и

$\Phi(X_i^{r+1}) > \Phi(X_i^r)$, то выполняем сжатие комплекса C_{r+1} вдоль направления

(X_k^{r+1}, X_c^{r+1}) , где X_c^{r+1} — центр тяжести комплекса C_{r+1} , до тех пор, пока точка X_k^{r+1} не станет допустимой. Переходим к п.5;



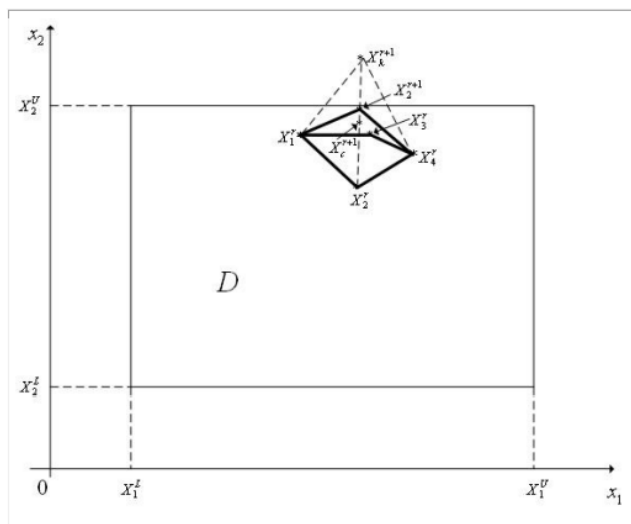
– если точка X_k^{r+1} является допустимой (не выходит за границы области D) и $\Phi(X_i^{r+1}) < \Phi(X_i^r)$, то переходим к шагу 5;

– если точка X_k^{r+1} является **НЕ**допустимой, но $\Phi(X_i^{r+1}) < \Phi(X_i^r)$, то переходим к п. 6.

5. Проверяем условие окончания поиска (ниже). Если условие окончания поиска выполнено, то в качестве точки X^* принимаем вершину комплекса C_{r+1} , к которой функция $\Phi(X)$ имеет наименьшее значение, вычисляем соответствующие значения $\Phi(X)$ и завершаем итерации. Иначе – переходим к п. 3.

6. Если $(X_k^{r+1})_i < X_i^L$, то полагаем $(X_k^{r+1})_i = X_i^L$;

если $(X_k^{r+1})_i > X_i^U$ то полагаем $(X_k^{r+1})_i = X_i^U$ (см. рис.). Переходим к п.3.



В качестве критерия окончания поиска могут использоваться следующие условия:

1. максимальная длина ребра комплекса C_r не превышает eps_x — требуемую точность решения по X .
2. максимальная разность значений функции $\Phi(X)$ в двух вершинах комплекса C_r не превышает eps_Φ — требуемую точность решения по Φ .
3. более сложные условия окончания поиска, учитывающие текущий размер комплекса или среднее значение функции $\Phi(X)$ в его вершинах.

Изложенная схема метода комплексов приводит к "уплощению" комплекса вблизи границы области допустимых значений D , что может значительно уменьшить эффективность метода. С целью преодоления этого недостатка через фиксированное количество итераций находятся максимальная и минимальная диагонали комплекса и, если их отношение превышает заданное, то по рассмотренной схеме производится построение нового комплекса.