

Лекция 11

1. Основные понятия и определения
2. Аналитический метод отыскания точки минимума функции двух переменных
3. Методы спуска
4. Метод градиентного спуска с дроблением шага (ГДШ)

Основные понятия и определения

Задача, требующая нахождения оптимального значения функции n переменных $f(\mathbf{X}) = f(x_1, x_2, \dots, x_n)$, называется задачей *многомерной оптимизации*. Так же, как и для случая одномерной оптимизации, задача нахождения максимума функции сводится к задаче нахождения минимума путем замены целевой функции f на $-f$.

Пусть функция $f(\mathbf{X}) = f(x_1, x_2, \dots, x_n)$ определена на некотором множестве \mathbf{X} . Если ограничения на переменные x_1, x_2, \dots, x_n отсутствуют, принято говорить о задаче *многомерной безусловной минимизации*. В противном случае говорят о задаче *многомерной условной минимизации*. Мы ограничимся решением задачи безусловной минимизации.

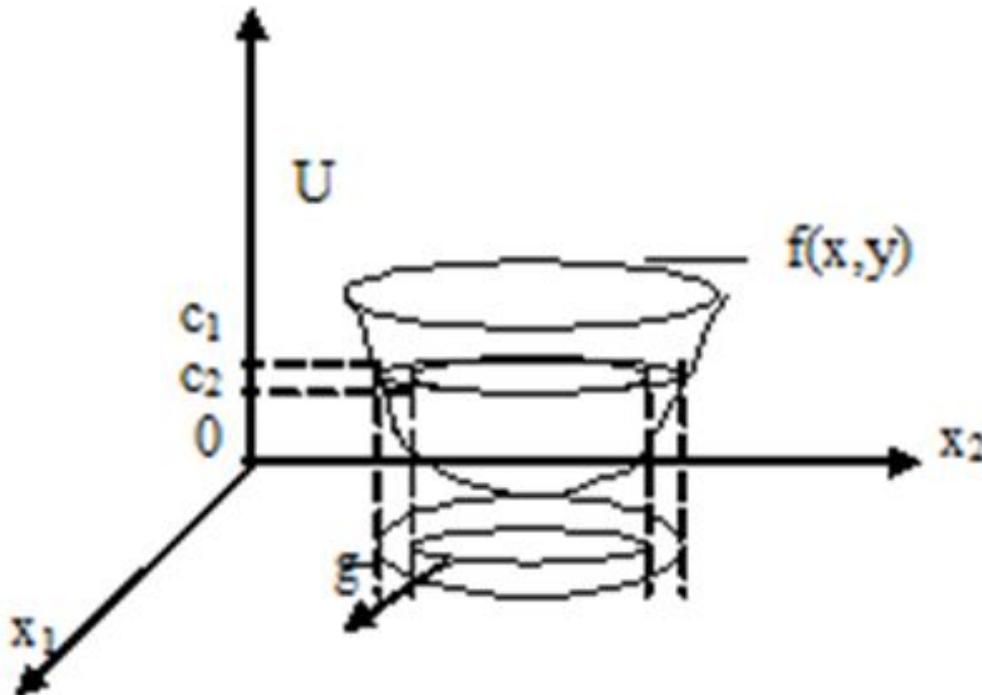
Как и в случае одномерной оптимизации, различают *глобальный* и *локальные* минимумы функции $f(\mathbf{X})$. Точка \mathbf{X}^* называется точкой глобального минимума функции f на множестве \mathbf{X} , если для всех \mathbf{X} выполняется неравенство $f(\mathbf{X}^*) \leq f(\mathbf{X})$.

Точка \mathbf{X}^* называется точкой локального минимума функции f , если существует такая δ - окрестность ($\delta > 0$), что для всех $\mathbf{X} \in \mathbf{X}_\delta$ выполняется неравенство $f(\mathbf{X}^*) \leq f(\mathbf{X})$.

Поверхности и линии уровня

Множество точек, для которых целевая функция принимает постоянное значение $f(x_1, x_2, \dots, x_n) = c$, называется *поверхностью уровня*. Для функции двух переменных ($n = 2$) это множество называется *линией уровня*.

Функция $f(x_1, x_2)$ задает в трехмерном пространстве некоторую поверхность $U=f(x_1, x_2)$ (рис. 1.8.1-1), низшая точка которой и дает решение задачи минимизации. Для того чтобы изобразить рельеф этой поверхности, проведем несколько плоскостей ($U = \text{const}$): $U=c_1$, $U=c_2$, $U=c_3$. Проекции на плоскость Ox_1x_2 линий пересечений этих плоскостей с поверхностью и дают *линии уровня*.



Градиент и антиградиент

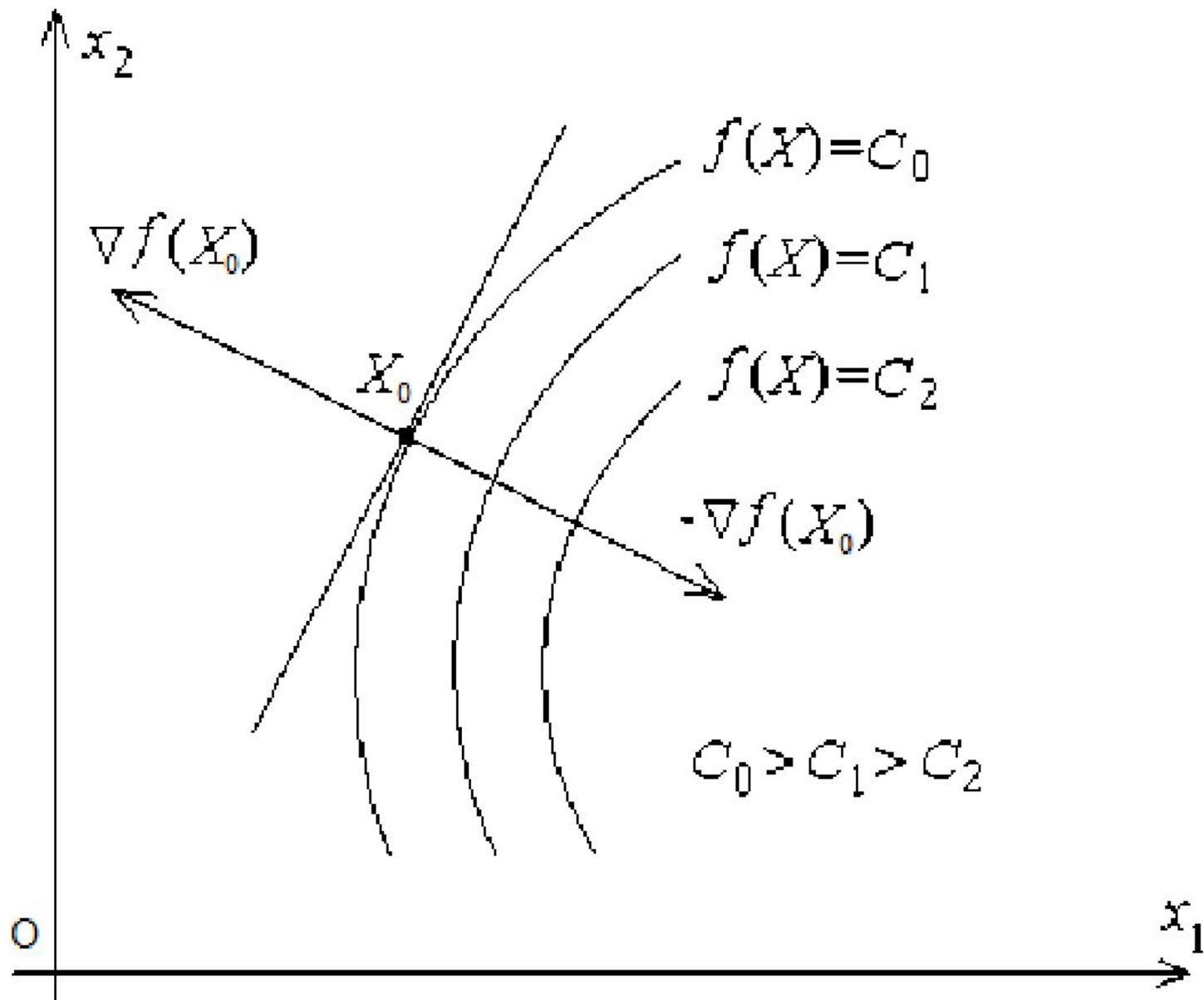
Для дифференцируемой функции $f(\mathbf{X})$ можно определить вектор из первых частных производных, который называется *градиентом*. Для функции двух переменных

$$\overline{\mathbf{grad}(f(\mathbf{X}))} = \left[\frac{\partial f(\mathbf{X})}{\partial \mathbf{x}_1}, \frac{\partial f(\mathbf{X})}{\partial \mathbf{x}_2} \right]$$

Направление вектора градиента указывает направление наибольшего возрастания функции, а его модуль (длина) равен скорости возрастания функции. Вектор градиента нормален к линии уровня в каждой ее точке (перпендикулярен касательной к линии уровня).

Вектор $-\overline{\mathbf{grad}(f(\mathbf{X}))}$ называется *антиградиентом* и показывает направление наискорейшего убывания функции.

Векторы градиента и антиградиента



Необходимое условие существования локального минимума

Равенство нулю градиента в точке \mathbf{X} является необходимым условием того, чтобы точка \mathbf{X} была точкой локального минимума дифференцируемой функции f . Точка \mathbf{X} , для которой выполняется равенство $\mathbf{f}'(\mathbf{X}) = \mathbf{0}$, называется *стационарной точкой функции*. Для функции двух переменных $Q(x, y)$ это равенство представляет собой систему из **2** нелинейных уравнений относительно x и y :

$$\begin{cases} \frac{\partial Q(x, y)}{\partial x} = 0, \\ \frac{\partial Q(x, y)}{\partial y} = 0. \end{cases}$$

Достаточное условие существования ТОЧКИ МИНИМУМА

Однако не всякая стационарная точка является точкой минимума. Для всякой непрерывно дифференцируемой функции f достаточным условием того, чтобы стационарная точка X была точкой локального минимума, является положительная определенность матрицы вторых производных (*матрицы Гессе*).

Для функции двух переменных $Q(x, y)$ матрица Гессе имеет вид:

$$G(x, y) = \begin{vmatrix} \frac{\partial^2 Q}{\partial x^2} & \frac{\partial^2 Q}{\partial x \partial y} \\ \frac{\partial^2 Q}{\partial y \partial x} & \frac{\partial^2 Q}{\partial y^2} \end{vmatrix},$$

а достаточное условие существования минимума:

$$\Delta_1 = \frac{\partial^2 Q}{\partial x^2} > 0,$$

$$\Delta_2 = \frac{\partial^2 Q}{\partial x^2} \cdot \frac{\partial^2 Q}{\partial y^2} - \left(\frac{\partial^2 Q}{\partial x \partial y} \right)^2 > 0.$$

Аналитический метод отыскания точки минимума функции двух переменных

Если достаточное условие существования минимума выполняется, то функция $Q(x,y)$ является выпуклой и имеет единственную точку минимума, глобальную и локальную одновременно.

Алгоритм решения задачи отыскания точки минимума функции двух переменных $Q(x,y)$ аналитическим методом следующий:

1. Составляется и решается система уравнений

$$\begin{cases} \frac{\partial Q(x,y)}{\partial x} = 0; \\ \frac{\partial Q(x,y)}{\partial y} = 0, \end{cases}$$

решением которой находятся (x^*, y^*) .

Аналитический метод отыскания точки минимума функции двух переменных

2. Проверяются достаточные условия существования минимума

$$\left\{ \begin{array}{l} \frac{\partial^2 Q(x, y)}{\partial x^2} > 0 \\ \frac{\partial^2 Q(x, y)}{\partial x^2} \cdot \frac{\partial^2 Q(x, y)}{\partial y^2} - \left(\frac{\partial^2 Q(x, y)}{\partial x \partial y} \right)^2 > 0. \end{array} \right.$$

Если (x^*, y^*) – единственное решение, и в этой точке выполняются достаточные условия, то это точка минимума. Если хотя бы в одном из неравенств получается знак “<”, то минимума не существует.

Пример отыскания точки минимума функции двух переменных аналитическим методом

Пример. Найдем точку минимума функции

$$f(x,y) = x^2 + y^2 - 4x + 100 - 8y$$

$$\begin{cases} \frac{\partial f}{\partial x} = 2x - 4, \\ \frac{\partial f}{\partial y} = 2y - 8, \end{cases} \quad \begin{cases} 2x - 4 = 0, \\ 2y - 8 = 0, \end{cases} \quad x = 2; y = 4.$$

$$\frac{\partial^2 f}{\partial x^2} = 2; \quad \frac{\partial^2 f}{\partial y^2} = 2; \quad \frac{\partial^2 f}{\partial x \partial y} = 0.$$

$$\begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix}, \quad \Delta_1 = 2 > 0; \quad \Delta_2 = 4 - 0 = 4 > 0.$$

Следовательно, в точке $x^* = 2$ и $y^* = 4$ функция имеет единственный минимум.

Методы спуска

Аналитический метод поиска минимума применим только для ограниченного круга задач. В основном это связано с необходимостью решения системы нелинейных уравнений, которая, как правило, может быть решена лишь численными методами. Оказывается, гораздо проще сразу решать задачу минимизации численными методами. Рассмотрим некоторые из этих методов.

Большинство итерационных методов, применяемых для решения задач безусловной минимизации функции нескольких переменных, относятся к классу *методов спуска*. Это такие методы, для которых каждая итерация (шаг) приводит к уменьшению значения целевой функции: $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$, для всех $k \geq 0$.

Типичный алгоритм метода спуска

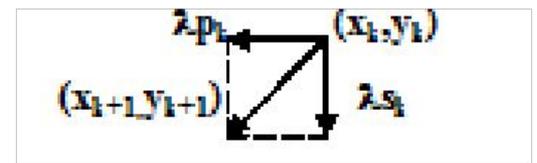
Типичный алгоритм метода спуска для функции двух переменных $Q(x,y)$ состоит в следующем:

1. Задается начальная точка (x_0, y_0) , принадлежащая области допустимых значений функции.
2. На текущей k -й итерации ($k=0,1,\dots$) определяется вектор $\bar{G}(p_k, s_k)$, задающий направление спуска, причем такой, чтобы для всех достаточно малых значений $\lambda > 0$ (где λ - коэффициент, определяющий шаг поиска) выполнялось неравенство:

$$f(x_k + \lambda p_k, y_k + \lambda s_k) < f(x_k, y_k).$$

3. Определяется шаг поиска λ_k , для которого выполняется условие п.2, и за очередное приближение к точке минимума принимается точка $(x_{k+1}; y_{k+1})$, где

$$x_{k+1} = x_k + \lambda_k p_k, \text{ а } y_{k+1} = y_k + \lambda_k s_k.$$



4. Проверяется выполнение условия окончания итераций. Если эти условия выполняются, то полагают $(x^*, y^*) \approx (x_{k+1}, y_{k+1})$. В противном случае осуществляется переход к п.2 и выполняется следующая итерация.

Последовательность точек (x_k, y_k) , получаемую методом спуска, называют *траекторией спуска*.

Разновидности методов спуска. Методы случайного спуска

Различные методы спуска различаются способом выбора вектора направления спуска (\mathbf{p}_k ; \mathbf{s}_k) и (или) способом определения шага λ_k на каждой итерации поиска минимума.

В методах *случайного спуска* направление и шаг спуска на каждой итерации выбираются случайным образом путем генерации равномерно распределенных случайных чисел. При этом на каждой итерации необходимо обеспечивать условие уменьшения целевой функции $\mathbf{f}(\mathbf{x}_{k+1}) < \mathbf{f}(\mathbf{x}_k)$.

Метод случайного спуска весьма трудоемок, так как может потребовать очень большого числа итераций. Гораздо более эффективными являются методы, в которых направление спуска выбирается не случайным образом, а целенаправленно.

Градиентные методы спуска

В *градиентных методах спуска* направление движения к точке минимума целевой функции совпадает с направлением *антиградиента*, и выбирается по формулам:

$$p_k = \frac{-\partial Q(x,y)}{\partial x} \Big|_{\substack{x = x_k \\ y = y_k}} \quad s_k = -\frac{\partial Q(x,y)}{\partial y} \Big|_{\substack{x = x_k \\ y = y_k}}.$$

Различные разновидности градиентных методов спуска отличаются правилом выбора шага λ_k на каждой итерации.

При решении вопроса о выборе шага λ_k следует учитывать, что выбор малого шага на каждой итерации приведет к малым изменениям аргумента и функции, и, следовательно, к увеличению числа итераций, необходимых для решения задачи. Выбор слишком большого шага λ_k может привести не к убыванию целевой функции $Q(x,y)$, а к ее увеличению, и, следовательно, процесс не будет сходиться к точке минимума.

Наиболее распространенными методами градиентного спуска являются *метод градиентного спуска с дроблением шага* (ГДШ) и *метод наискорейшего спуска* (НС).

Методы покоординатного спуска

В *методах покоординатного спуска* в качестве очередного направления спуска выбирают направление одной из координатных осей. Наиболее известным является *метод циклического покоординатного спуска*, в котором направления спуска циклически меняются от итерации к итерации.

Условия останова

Условием прекращения итераций в методах спуска может быть достаточно малое изменение значения целевой функции на двух последовательных итерациях:

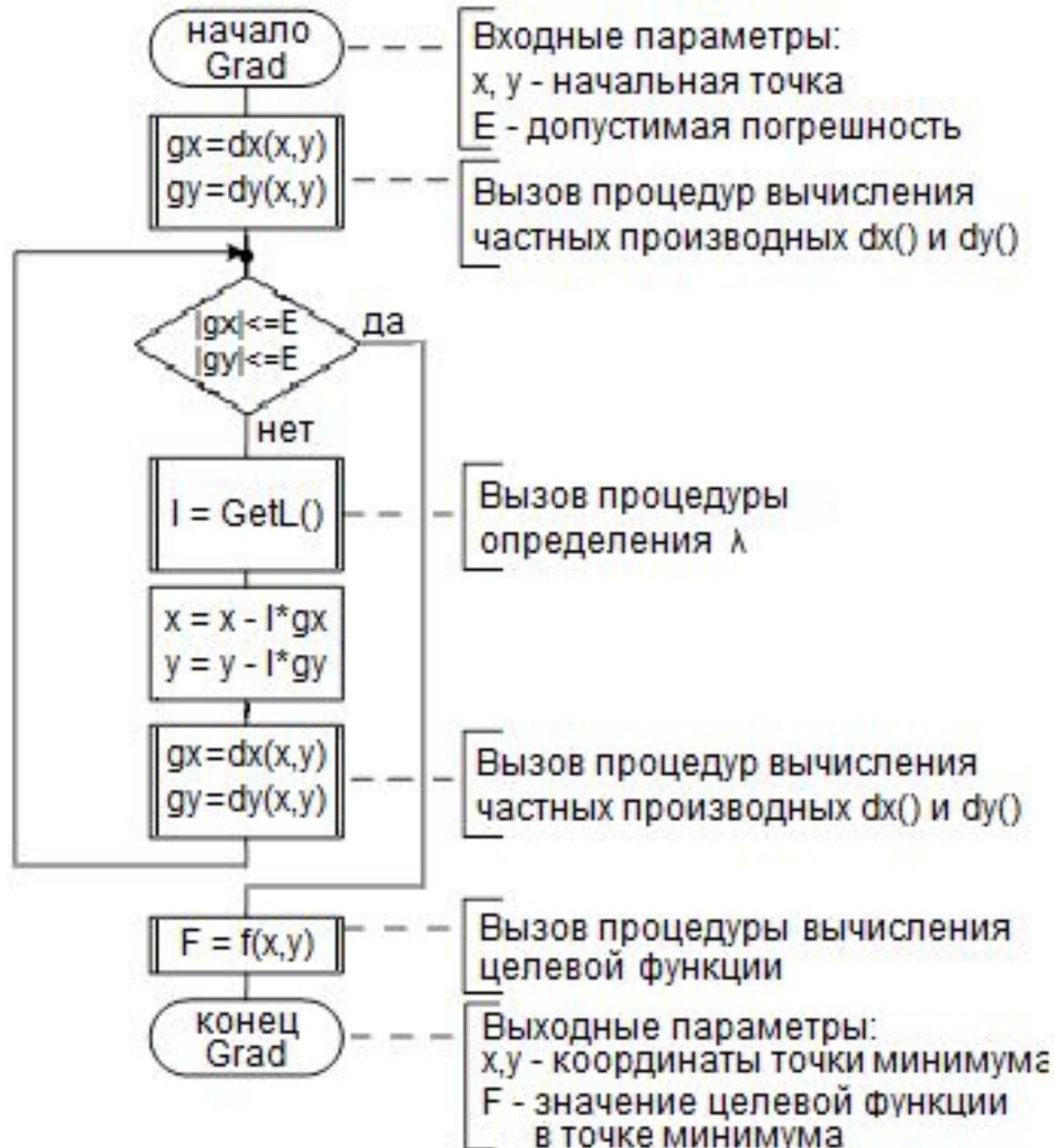
$$|Q(x_{k+1}, y_{k+1}) - Q(x_k, y_k)| \leq \varepsilon$$

где ε – допустимая погрешность.

В градиентных методах спуска условием окончания процесса итераций обычно служат достаточно малые значения частных производных целевой функции:

$$\left| \frac{\partial Q}{\partial x} \right|_{x_{k+1}, y_{k+1}} \leq \varepsilon \quad \left| \frac{\partial Q}{\partial y} \right|_{x_{k+1}, y_{k+1}} \leq \varepsilon$$

Схема алгоритма градиентных методов спуска



Метод градиентного спуска с дроблением шага (ГДШ)

В методе ГДШ на каждой итерации шаг спуска λ_k выбирается таким образом, чтобы выполнялось условие сходимости:

$$Q(x_k, y_k) - Q(x_{k-1}, y_{k-1}) \geq \frac{\lambda_k}{2} \left(\rho_k^2 + s_k^2 \right)$$

Процедура выбора шага в методе ГДШ заключается в следующем.

- 1) Задается начальное значение шага $\lambda_k = \lambda_0$ (как правило, $\lambda_0 = 0.5$).
- 2) Проверяется условие сходимости, приведенное выше.
- 3) Если условие сходимости выполняется, то шаг спуска для данной итерации выбран.
- 4) Если условие сходимости не выполняется, то принимают новый шаг $\lambda_k = \lambda_k/2$, и возвращаются к п. 2 и т.д.

На слайде 19 приведен пример возможной траектории спуска методом ГДШ.

Пример траектории спуска методом ГДШ



Схема алгоритма процедуры выбора

шага

