

«Оптимизация и математические методы принятия решений»

ст. преп. каф. СС и ПД
Владимиров Сергей Александрович

Лекция 5

Численные методы оптимизации.

СОДЕРЖАНИЕ

Введение

Учебные вопросы:

1. Структура и постановка задач оптимизации.
2. Условия оптимальности и типы вычислительных процедур оптимизации.
3. Методы одномерной оптимизации
4. Поиска экстремума функций многих переменных. Метод Гаусса-Зайделя.
5. Методы наискорейшего спуска.

Заключение

Литература:

1. Щекотихин В.М., Терентьев В.М. Прикладная математика.- Орел: Академия ФАПСИ, 2002.
2. Терентьев В.М. Методика обоснования требований к показателям качества АСМКРС. -Л.: ВАС, 1991.
3. Терентьев В.М., Паращук И.Б. Теоретические основы управления сетями многоканальной радиосвязи. - С-Петербург: ВАС, 1995.
4. Таха Х. Введение в исследование операций . В двух книгах .Пер. с англ. М.: Мир 1985.

Введение

Оптимизация в широком смысле слова - это поиск лучшего из возможных вариантов.

Методы оптимизации – это численные методы решения задач оптимизации (задач, имеющих множество допустимых решений, из которых необходимо выбрать одно, лучшее в каком-либо смысле). Численные методы оптимизации, как методы численного (приближенного) программирования основаны на поисковых регулярных либо случайных процедурах выбора одного из множества возможных путей достижения экстремумов критериев.

Критерий оптимальности — мера целесообразности принимаемого на его основе решения.

Структура и постановка задач оптимизации

Формулировка задачи оптимизации включает три этапа:

1. словесное представление о параметрах задачи, множестве ее решений и поставленной цели;
2. запись критерия оптимальности (целевой функции) как функции параметров задачи;
3. запись условий, определяющих область допустимых значений параметров.

Параметры задачи x_i , ($i=1,2,\dots,n$) - это переменные, значения которых необходимо определить в результате ее решения

Критерий оптимальности может быть представлен в виде функции параметров - целевой функции - $f(X)$, где $X=(x_1, x_2, \dots, x_n)$ – одно из допустимых решений задачи, или функционала $I(X)$, который при фиксированных значениях параметров задачи представляет собой не число, а функцию времени или пространственной координаты, например,

$$f(X) = \sum_{i=1}^n (x_i - x_i^{(0)})^2, \quad I(X) = \int_0^t \varphi(X, t) dt.$$

Область допустимых значений параметров x_i , ($i=1,2,\dots,n$) может быть определена с помощью ограничений - $g_j(X) \leq 0$; ($j=1,2,\dots,m$) - условия типа неравенств и связей - $h_k(X) = 0$; ($k=1,2,\dots,p$) - условия типа равенств.

Таким образом, обобщенная постановка оптимизационной задачи имеет вид: найти решение $X^*=(x_1^*, x_2^*, \dots, x_n^*)$, которому соответствует экстремум функции $f(X)$ - минимум или максимум, при выполнении условий:

$$g_j(X) \leq 0; (j=1,2,\dots,m) \quad (1)$$

$$h_k(X) = 0; (k=1,2,\dots,p) \quad (2)$$

Задачи оптимизации классифицируют по следующим признакам:

1. наличие или отсутствие ограничений и связей - задачи на
 - *безусловный экстремум* - условия (1), (2) отсутствуют и на
 - *условный экстремум* - имеется условие (1), условие (2) или оба;
2. вид критерия оптимальности:
 - *вариационные задачи* (критерий — функционал) и
 - *задачи математического программирования* (критерий - функция);
3. характер функций f, g, h
 - *задачи линейного программирования* (все функции - линейные)
 - *задачи нелинейного программирования* (хотя бы одна из функций - нелинейная);
4. характер параметров
 - *непрерывные значения*
 - если параметры задачи могут принимать только строго определенные значения то это задача *дискретного (целочисленного) программирования*,
 - а если число этих значений конечно – *комбинаторная задача*.

Условия оптимальности и типы вычислительных процедур оптимизации

Теоретической основой методов решения задач оптимизации являются условия оптимальности решения различных типов задач, т.е. условия, при которых критерий оптимальности той или иной задачи достигает минимального или максимального значения. Условия оптимальности подразделяются на необходимые и достаточные.

Примеры необходимых условий оптимальности - условие А необходимо для выполнения В, если при выполнении В всегда выполняется А):

- если в окрестности некоторой точки x^* ($x \in [(x^* - \delta), (x^* + \delta)]$, $\delta > 0$) $f(x) > f(x^*)$, то в точке $x = x^*$ функция $f(x)$ имеет локальный минимум (в окрестности другой точки x значения $f(x)$ могут быть еще меньше);
- если дифференцируемая функция $f(x)$ имеет в точке $x = x^*$ экстремум, то $f'(x^*) \equiv 0$ для функции многих переменных ($df(x_1^*, x_2^*, \dots, x_n^*)/dx_i = 0$, $i = 1, 2, \dots, n$).

Примеры достаточных условий оптимальности (условие А достаточно

для выполнения В если при выполнении А всегда выполняется В):

- если в точке $x=x^*$ функция $f(x)$ имеет минимум, то в некоторой окрестности этой точки x^* ($x \in [(x^*-\delta), (x^*+\delta)]$, $\delta > 0$) выполняется условие $f(x) > f(x^*)$;

- если $f'(x^*)=0$ и $f''(x^*) \neq 0$, то в точке x^* $f(x)$ имеет экстремум (при $f''(x^*) > 0$ – минимум, при $f''(x^*) < 0$ – максимум);

- для того, чтобы дифференцируемая функция $f(x_1, x_2, \dots, x_n)$ имела в точке $(x_1^*, x_2^*, \dots, x_n^*)$ минимум достаточно, чтобы $(df(x_1^*, x_2^*, \dots, x_n^*)/dx_i = 0, i=1, 2, \dots, n)$ и все миноры матрицы Гессе $\{ d^2f(x_1^*, x_2^*, \dots, x_n^*)/dx_1^2 \dots d^2f(x_1^*, x_2^*, \dots, x_n^*)/(dx_1 dx_n) ; \dots ; d^2f(x_1^*, x_2^*, \dots, x_n^*)/(dx_n dx_1) \dots d^2f(x_1^*, x_2^*, \dots, x_n^*)/dx_n^2 \}$ были положительны, а если миноры нечетного порядка отрицательны, а четного - положительны, то в точке $(x_1^*, x_2^*, \dots, x_n^*)$ функция $f(x_1, x_2, \dots, x_n)$ имеет максимум.

После того как сформулированы условия оптимальности критерия конкретной задачи, необходимо организовать вычислительную процедуру поиска оптимального решения. Имеется **пять основных типов вычислительных процедур решения задач оптимизации**, на основе которых разрабатываются методы оптимизации:

Сравнение значений целевой функции на сетке значений аргументов. Сетка образуется в результате разбиения областей допустимых значений аргументов на равные интервалы (рис.1). Оптимальному решению соответствует минимальное или максимальное значение целевой функции в “узлах” сетки. Процедура обычно применяется многократно: вначале шаг сетки “крупный”, а затем вокруг лучшей точки строится более “мелкая” сетка. Аналогом этой процедуры для задач дискретного программирования и комбинаторных является поиск лучшего допустимого решения путем полного или локального перебора. Эта процедура применяется всегда, когда это не связано со слишком большим объемом вычислений.

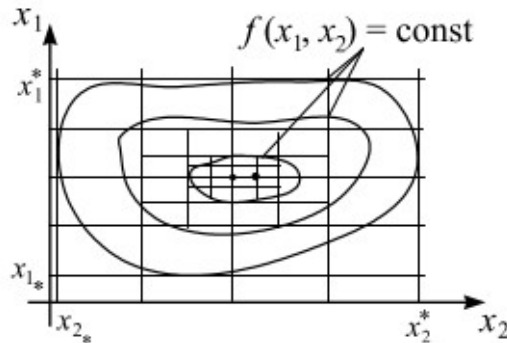


Рисунок 1. Сетка значений аргументов функции $f(x_1, x_2)$

Использование необходимых условий экстремума целевой функции, т.е. формирование и решение систем уравнений вида $(df(x_1^*, x_2^*, \dots, x_n^*)/dx_i = 0, i=1, 2, \dots, n)$. Применение этой процедуры ограничено, т.к. аналитические выражения производных целевой функции существуют далеко не всегда.

Использование достаточных условий оптимальности: образуется вспомогательная задача, множество решений которой шире допустимого, а критерий оптимальности на допустимом множестве совпадает с критерием исходной задачи - например задача условной оптимизации заменяется задачей безусловной оптимизации, в критерий которой вводится "штраф" за выход из допустимой области, т.е. за невыполнение условий (1), (2).

Отсечение множеств заведомо неоптимальных решений (рис. 2) на основе правил, различных для каждой конкретной задачи (метод "золотого сечения", ветвей и границ для экстремальных комбинаторных задач).

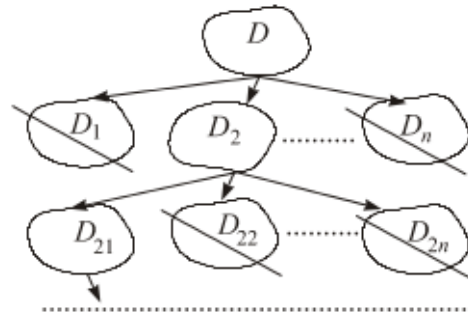


Рисунок 2. Отсечение множеств неоптимальных решений.

Построение оптимизирующей последовательности допустимых решений

задачи: выбирается одно из допустимых решений $X(0)$, называемое начальным приближением, и на его базе строится последовательность допустимых решений $X(0), X(1), X(2), \dots, X(n)$, где $X(k+1) = X(k) + \Delta X(k)$, $k = 0, 1, \dots, n$, причем $\Delta X(k)$ выбирается так, чтобы при поиске $\min f(X)$ выполнялось условие $f(X(k)) > f(X(k+1))$, а при поиске $\max f(X)$ – условие $f(X(k)) < f(X(k+1))$. Приращение $\Delta X(k)$ чаще всего является функцией одного или нескольких предыдущих членов последовательности: $\Delta X(k) = \varphi(X(k))$ или $\Delta X(k) = \varphi(X(k), X(k-1))$. Построение последовательности заканчивается в момент выполнения необходимых условий $\text{ext } f(X)$. Большинство методов оптимизации основаны именно на этой процедуре.

Методы одномерной оптимизации

К числу наиболее популярных численных методов поиска экстремума функции одного переменного одномерной оптимизации относятся: метод «золотого сечения» и пошаговый метод. Первый из них ориентирован на поиск $\text{ext } f(x)$ внутри фиксированного интервала (a, b) оси x , последний - на поиск $\text{ext } f(x)$ в окрестности заданной точки x_0 .

Метод «золотого сечения»

Определение: «золотым сечением» отрезка называется его деление на две части таким образом, что отношение длины отрезка к его большей части равно отношению большей части к меньшей. Следовательно, для отрезка единичной длины: $1/t=t/(1-t)$

$$\Rightarrow t^2+t-1=0 \Rightarrow t=-1/2 \pm \sqrt{(1/4+1)};$$

$$|t|<1 \Rightarrow t=(\sqrt{5}-1)/2=0.618; \quad (1-t)=0.382$$

Алгоритм метода «золотого сечения» при поиске минимума функции $f(x)$ включает операции, см. рис. 3:

- 1) интервал (a, b) делится точками x_1, x_2 в отношении «золотого сечения»:
 $x_1 = a + (b - a) \times 0.382$, $x_2 = b - (b - a) \times 0.382$;
- 2) вычисляются значения $f(x_1)$ и $f(x_2)$;
- 3) если $f(x_1) < f(x_2)$, то от интервала (a, b) отсекается его правая часть: $b = x_2$, в противном случае – левая: $a = x_1$;
- 4) в случае $b = x_2$ точка x_1 осуществляет «золотое сечение» нового интервала (a, b) и играет в нем роль точки x_2 , а точка x_1 нового интервала определяется аналогично п.1); при $a = x_1$ наоборот – $x_1 = x_2$, а точка x_2 нового интервала (a, b) определяется как в п.1).

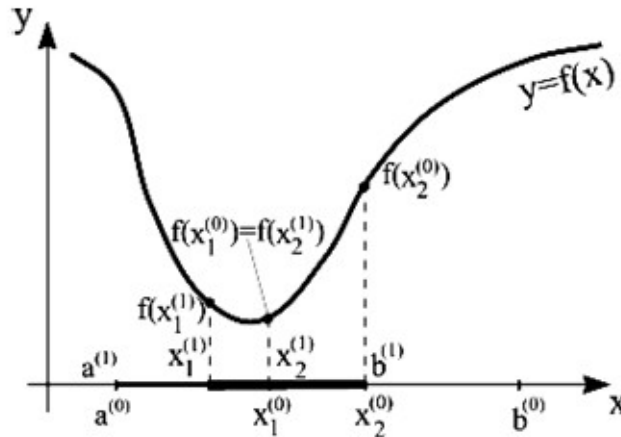


Рисунок 3. Метод «золотого сечения».

Для нового интервала (a, b) вновь выполняются действия п.п.2)-4), причем в п.2) значение функции $f(x)$ вычисляется один раз: только для вновь определяемой точки x_1 или x_2 .

Процесс деления интервала продолжается до тех пор, пока его длина не станет меньше заданной точности: $b-a < \epsilon$. При завершении процесса поиска за точку минимума принимается значение $x^* = (a+b)/2$.

Достаточные условия сходимости алгоритма метода «золотого сечения»:

- функция $f(x)$ непрерывна внутри интервала (a, b) ;
- $f(x)$ унимодальна на интервале (a, b) , т.е. имеет внутри него единственный экстремум;
- в некоторой окрестности искомой точки x^* $f(x)$ является монотонной (с одной стороны возрастает, с другой — убывает).

Пошаговый метод (простой итерации)

Этот метод применяется в тех случаях, когда интервал (a, b) оси x , содержащий точку экстремума функции $f(x)$ неизвестен, но известно, что экстремум находится в окрестности экспериментально найденной точки x_0 . Этот метод применяется на практике значительно чаще метода «золотого сечения», так как условие сходимости его алгоритма намного проще: для этого достаточно, чтобы функция $f(x)$ была непрерывна в окрестности точки x_0 . Остановка вычислений — точка x_{opt} , при $h < \varepsilon$.

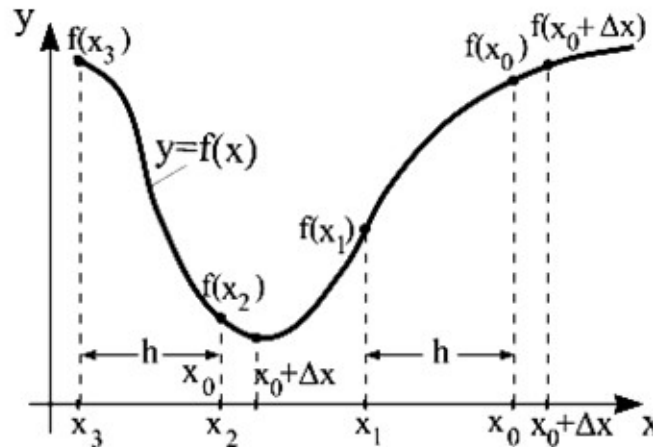


Рисунок 4. Пошаговый метод.

При поиске минимума функции **пошаговый метод** содержит следующее, см. рис. 4:

- 1) выполняется пробный шаг от точки x_0 с целью выбора направления поиска: $x = x_0 + \Delta x$ ($\Delta x \sim 0.5 \times \varepsilon$) и вычисляются значения $f(x_0), f(x)$;
- 2) если $f(x) < f(x_0)$, то величина основного шага, с которым осуществляется движение в направлении убывания функции, положительна ($h > 0$), в противном случае – отрицательна ($h < 0$);
- 3) движение в выбранном направлении с шагом h : $x_{k+1} = x_k + h, \quad k = 0, 1, 2, \dots$ осуществляется до тех пор, пока $f(x_{k+1}) < f(x_k)$;
- 4) если $f(x_{k+1}) \geq f(x_k)$, то при выполнении условия $h < \varepsilon$ процесс поиска заканчивается, а если $h \geq \varepsilon$, то шаг дробится: $h = |h|/p, \quad p > 1$ и осуществляется возврат к п.1) с начальной точкой $x_0 = x_k$.

В качестве коэффициента дробления шага p используют 2, 3, 5, но чаще всего $p = e = 2.71828$. По завершении процесса поиска за точку экстремума принимается значение $x^* = (x_{k+1} + x_k)/2$.

Поиска экстремума функций многих переменных

Несмотря на большой выбор приближенных методов вычислений далее рассмотрим три основополагающих из них: метод Гаусса-Зайделя; метод наискорейшего спуска и методы случайного поиска.

Метод Гаусса-Зайделя.

Данный метод использует информацию об унимодальности функционала качества, что позволяет существенно сократить число просматриваемых точек по сравнению с полным перебором. Суть метода заключается в эквивалентной замене общей многопараметрической задачи поиска экстремума последовательностью однопараметрических задач поиска частных экстремумов. Частная производная оптимизируемого функционала имеет вид:

$$dI(\vec{x})/dx_i = dI(x_1, \dots, x_i, \dots, x_n)_{x_{l \neq i} = \text{const}} / dx_i, i, l = 1, \dots, I. \quad (3)$$

при этом оптимальное значение x_i^{onm} может быть найдено из условия:

$$dI(\vec{x})/dx_i=0, x_i=x_i^{onm}. \quad (4)$$

Как видно из выражения (3) поиск оптимальных значений параметров x_i^{onm} , удовлетворяющих экстремуму $\partial I(\vec{x})/\partial \vec{x}=0$, может быть осуществлен на основе итеративной последовательной процедуры оптимизации по каждому i -му параметру при фиксированных значениях остальных l -х параметров. Сходимость такой процедуры к оптимальному решению по всем оптимизируемым переменным гарантируется при наличии унимодальности и дифференцируемости целевой функции.

Эквивалентная система - $x_i=Cx_{j \neq i}+d$, где $d_i=\frac{b_i}{a_{ii}}$, $c_{ij}=\begin{cases} \frac{-a_{ij}}{a_{ii}}, & i \neq j, \\ 0, & i=j. \end{cases}$, итерации

$$x_i^k=\sum_{j=1}^{i-1} c_{ij}x_j^k+\sum_{j=i+1}^n c_{ij}x_j^{k-1}+d_i, \quad \text{где } x_i^0 \text{ выбирается произвольно, } i=1,2,\dots,n.$$

Алгоритм вычисления экстремальных координат критерия включает следующие шаги:

- выбор начальной оптимизируемой переменной x_1 и отыскание частного экстремума (min, max) $\partial I(\vec{x})/\partial x_1=0$ при фиксированных значениях остальных переменных;
- фиксация значения $x_1=x_1^{opt}$ и осуществление поиска экстремума по переменной x_2 , до обращения в нуль частной производной $\partial I(\vec{x})/\partial x_2=0$;
- фиксация значений переменных x_1 и x_2 на уровне частных экстремумов и поиск оптимальных значений оставшихся переменных;
- переход к повторному циклу поиска частных экстремумов (шагу 1) до тех пор, пока найденная точка экстремума окажется общей для всех переменных (момент выполнения условий по точности решения задачи оптимизации для всех переменных).

Пример 1. Пусть критерий оптимальности имеет следующий вид:

$$\min_{x_1, x_2} I(x_1, x_2) = x_1^2 + x_2^2 + 1,5x_1x_2.$$

Шаг_1. Поиск начинается из точки $x_1=3$, $x_2=3$. Начальной оптимизируемой переменной выбирается x_1 . Тогда экстремальное ее значение определяется из условия

$$\partial I(x_1, 3) / \partial x_1 = 2x_1 + 4,5 = 0,$$

т.е. $x_1^{opt} = -2,25$.

Шаг_2. Находим минимум по x_2 .

$$\partial I[(-2,25), x_2] / \partial x_2 = (5 + x_2^2 - 3,4x_2)' = 2x_2 - 3,4 = 0,$$

откуда $x_2^{opt} = 1,7$.

Шаг_3. Повторяем цикл вычислений для координаты x_1 при фиксации $x_2 = x_2^{opt} = 1,7$. Получаем $x_1^{opt}(2) = -1,27$.

Далее цикл вычислений повторяется для переменной $x_2(2)$ и так далее до достижения заданной точности ε приближения оптимальных решений к точке общего экстремума ($x_1^{opt}(\infty) = 0$, $x_2^{opt}(\infty) = 0$).

Методы наискорейшего спуска

Градиентный метод

Для нахождения оптимального вектора \vec{x}' используют следующую рекуррентную схему:

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} + \alpha^{(k)} \vec{L}^{(k)}; k=0, 1, 2, \dots \quad (5)$$

В выражении (5) $\vec{x}^{(k+1)}$ и $\vec{x}^{(k)}$ – значения вектора \vec{x} на k -м и $(k+1)$ -м шаге итерационной процедуры; $\alpha^{(k)}$ – множитель, определяющий длину шага; $\vec{L}^{(k)}$ – вектор, определяющий направление движения от точки $\vec{x}^{(k)}$ к точке $\vec{x}^{(k+1)}$. В градиентном методе в качестве $L^{(k)}$ выбран градиент целевой функции $\nabla f(\vec{x})$ в случае максимизации $f(\vec{x})$ и антиградиент $\vec{L}^{(k)} = -\nabla f(\vec{x})$ в случае минимизации $f(\vec{x})$. В качестве начала итераций выбирается точка $\vec{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$. Далее будем предполагать, что решается задача минимизации целевой функции.

Решение задачи градиентным методом заключается в следующем. Находится вектор $\vec{L}^{(0)} = -\nabla f(\vec{x}^{(0)})$, тогда луч $\vec{x}^{(1)} = \vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)}$ (где $\alpha \geq 0$) определяет направление, вдоль которого скорость уменьшения $f(\vec{x})$ будет наибольшей. При этом шаг $\alpha^{(0)} = \alpha$ предполагается постоянным. При этом его выбор производится таким образом, чтобы выполнялось условие

$$f(\vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)}) < f(\vec{x}^{(0)}) . \quad (6)$$

Если это условие не выполняется, то производится коррекция длины шага, например $\alpha^{(0)} = \frac{\alpha}{2}$, и опять проверяется выполнение неравенства. Получив точку $\vec{x}^{(1)}$, в которой значение функции будет меньше, чем в точке $\vec{x}^{(0)}$, начинаем следующую итерацию с начальной точкой $\vec{x}^{(1)}$ и длиной шага $\alpha^{(1)} = \frac{\alpha}{2}$. Процесс завершается в точке, где $\nabla f(\vec{x}) = 0$ либо приращение значений переменных на соседних шагах оптимизации не превышает ранее заданной величины погрешности вычислений.

Рассмотрим **пример** решения задачи оптимизации градиентным методом. Пусть есть целевая функция $\min_{x_1, x_2} I(x_1, x_2) = x_1^2 + x_2^2 + 1,5x_1x_2$, начальные условия $\vec{x}^{(0)} = (3, 3)$, шаг итераций $\alpha = \alpha^0 = 1$.

1. Найдем градиент функции в точке $\vec{x}^{(0)} = (3, 3)$

$$\left. \frac{\partial f(x)}{\partial x_1} \right|_{\vec{x}^{(0)}} = (2x_1 + 1,5x_2) \Big|_{\vec{x}^{(0)}} = 10,5;$$

$$\left. \frac{\partial f(\vec{x})}{\partial x_2} \right|_{\vec{x}^{(0)}} = (2x_2 + 1,5x_1) \Big|_{\vec{x}^{(0)}} = 10,5.$$

$\nabla f(\vec{x}^{(0)}) = (10,5; 10,5)$, так как $f(\vec{x}) \rightarrow \min$, следовательно, $\vec{L}^{(0)}$

$$= -\nabla f(\vec{x}^{(0)}) = (-10,5; -10,5).$$

2. Уравнение луча имеет вид: $\vec{x}^{(1)} = \vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)}$

$$=(3;3) + 1 \cdot (-10,5; -10,5) = (-7,5; -7,5).$$

3. Проверим выполнение неравенства $f(\vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)}) < f(\vec{x}^{(0)})$, для чего произведем вычисление $f(\vec{x}^{(0)})$ и $f(\vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)})$.

$$f\left(\vec{x}^{(0)}\right) = 31,5, \quad f\left(\vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)}\right) \approx 196,9.$$

Учитывая, что неравенство не выполняется, произведем изменение длины шага $\alpha = \alpha^{(0)} = \frac{1}{2}$.

В этом случае $f\left(\vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)}\right) = f(-2,25; -2,25) \approx 17,7 < f\left(\vec{x}^{(0)}\right)$, следовательно,

$$\vec{x}^{(1)} = (-2,25; -2,25).$$

Повторяем рассуждения, взяв за начальную точку $\vec{x}^{(1)} = (-2,25; -2,25)$ и длину шага $\alpha^{(1)} = \frac{1}{2}$.

$$\vec{L}^{(1)} = -\nabla f\left(\vec{x}^{(1)}\right) = (7,9; 7,9), \quad \alpha = \frac{1}{2}. \quad \text{В этом случае } f(\vec{x}^{(1)} + \alpha^{(1)}\vec{L}^{(1)}) = 10,1 <$$

$$f\left(\vec{x}^{(1)}\right), \quad \text{следовательно } \vec{x}^{(2)} = (1,7; 1,7)$$

Продолжим вычисления для точки $\vec{x}^{(2)} = (1,7; 1,7)$

$$\vec{L}^{(2)} = -\nabla f\left(\vec{x}^{(2)}\right) = (-5,44; -5,44), \quad \alpha = \frac{1}{2}. \quad \text{В этом случае}$$

$$f\left(\vec{x}^{(2)} + \alpha^{(2)}\vec{L}^{(2)}\right) = 3,5 < f\left(\vec{x}^{(2)}\right), \quad \text{следовательно } \vec{x}^{(3)} = (-1; -1),$$

Эти операции повторяются до тех пор, пока $\nabla f(\vec{x}) = 0$.

Для ускорения процесса поиска оптимального вектора \vec{x}' может быть использован *метод наискорейшего спуска*. Основное отличие этого метода от градиентного состоит в способе выбора шага $\alpha^{(k)}$, а поиск оптимального вектора осуществляется с помощью следующей итерационной процедуры:

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} - \alpha^{(k)} \frac{\nabla f(\vec{x}^{(k)})}{\|\nabla f(\vec{x}^{(k)})\|}, \quad (7)$$

где $\|\nabla f(\vec{x}^{(k)})\| = \sqrt{\sum_{i=1}^n \left(\frac{\partial f(\vec{x}^{(k)})}{\partial x_i}\right)^2}$ – норма градиента целевой функции в точке $\vec{x}^{(k)}$, а все остальные элементы выражения (7) соответствуют введенным ранее.

То есть, в отличие от градиентного метода, в методе наискорейшего спуска шаг спуска $\alpha^{(k)}$ определяется из решения уравнения

$$\frac{\partial f(\vec{x}^{(k)} + \alpha^{(k)} \vec{L}^{(k)})}{\partial \alpha^{(k)}} = 0. \quad (8)$$

Метод хорошо "работает" при минимизации гладких функций и если начальное приближение выбрано достаточно далеко от оптимума. Если же очередная точка $\vec{x}^{(k)}$

окажется в окрестности оптимума, то уменьшение целевой функции будет очень медленным. Это происходит из-за того, что для получения оптимума с высокой точностью необходимо выполнить большое число мелких шагов [2].

Как видно из существа метода, условием его применения является непрерывная дифференцируемость $f(\vec{x})$.

Пример 2. Найти координаты минимума целевой функции следующего вида:

$$f(x_1, x_2) = x_1^2 + x_2^2 + 1,5 x_1 x_2.$$

Для начальной точки $(x_1=2, x_2=3)$ находим градиент (частные производные):

$$\nabla f(\vec{x}) = \left[\frac{\partial f(\vec{x})}{\partial x_1}; \frac{\partial f(\vec{x})}{\partial x_2} \right]^m = [2x_1 + 1,5x_2; 2x_2 + 1,5x_1]^m.$$

Определим норму градиента:

$$\|\nabla f(\vec{x})\| = \sqrt{(2x_1 + 1,5x_2)^2 + (2x_2 + 1,5x_1)^2} = \sqrt{6,25x_1^2 + 12x_1x_2 + 6,25x_2^2}.$$

Далее определяется единичный вектор, имеющий направление антиградиента:

$$\vec{L} = -\frac{\nabla f(\vec{x})}{\|\nabla f(\vec{x})\|} = \left[-\frac{2x_1 + 1,5x_2}{\sqrt{6,25x_1^2 + 12x_1x_2 + 6,25x_2^2}}; -\frac{2x_2 + 1,5x_1}{\sqrt{6,25x_1^2 + 12x_1x_2 + 6,25x_2^2}} \right].$$

Выражение, определяющее величину оптимального шага, находим из уравнения (8):

$$\frac{\partial f(\vec{x} + \alpha \vec{L})}{\partial \alpha} = \frac{\partial \left((x_1 + \alpha L(x_1))^2 + (x_2 + \alpha L(x_2))^2 + 1,5(x_1 + \alpha L(x_1))(x_2 + \alpha L(x_2)) \right)}{\partial \alpha} = 0 .$$

Решая данное уравнение получаем:

$$\alpha = - \frac{x_1 L(x_1) + x_2 L(x_2) + 0,75 x_1 L(x_2) + 0,75 x_2 L(x_1)}{L^2(x_1) + 1,5 L(x_1) L(x_2) + L^2(x_2)} ,$$

где $L(x_i)$ - проекция вектора \vec{L} на ось x_i .

Первый шаг: $x_1^{(0)} = 2, x_2^{(0)} = 3$;

$$\nabla f(\vec{x}^{(0)}) = [2x_1^{(0)} + 1,5x_2^{(0)}; 2x_2^{(0)} + 1,5x_1^{(0)}]^m = [8,5; 9]^m ;$$

$$\|\nabla f(\vec{x}^{(0)})\| = \sqrt{6,25(x_1^{(0)})^2 + 12x_1^{(0)}x_2^{(0)} + 6,25(x_2^{(0)})^2} \approx 12,38 ;$$

$$\vec{L}^{(0)} = \left[-\frac{8,5}{12,38}; -\frac{9}{12,38} \right]^m = [-0,69; -0,73]^m$$

$$\alpha^{(0)} \approx 3,5 ;$$

$$x_1^{(1)} = x_1^{(0)} + \alpha^{(0)} L(x_1^{(0)}) = 2 - 3,5 \cdot 0,69 = -0,415 ;$$

$$x_2^{(1)} = x_2^{(0)} + \alpha^{(0)} L(x_2^{(0)}) = 3 - 3,5 \cdot 0,73 = 0,445 .$$

Второй шаг: $x_1^{(1)} = -0,415, x_2^{(1)} = 0,445$;

$\nabla f(\vec{x}^{(1)}) = [-0,16; 0,24]^m$; $\|\nabla f(\vec{x}^{(1)})\| \approx 0,32$; $\vec{L}^{(1)} = [0,5; -0,75]^m$; $\alpha^{(1)} \approx 0,56$;

$$x_1^{(2)} = x_1^{(1)} + \alpha^{(1)} L(x_1^{(1)}) = 0,415 - 0,56 \cdot 0,5 = -0,135 ;$$

$$x_2^{(2)} = x_2^{(1)} + \alpha^{(1)} L(x_2^{(1)}) = 0,445 - 0,56 \cdot 0,75 = 0,025 .$$

Сравнение результатов применения первого и второго методов показывает большую скорость сходимости к оптимальным значениям переменных при применении метода наискорейшего спуска, чем метода Гаусса –Зайделя.

Заключение.

Регулярные и случайные методы поиска оптимальных значений переменных обладают тем основным достоинством по сравнению с аналитическими методами оптимизации, что остаются эффективными в более широком классе целевых функций (нелинейные, динамические, нестационарные).