ЧИСЛЕННЫЕ МЕТОДЫ ОПТИМИЗАЦИИ

Глава 1. Численные методы оптимизации

1.1. Основные определения

Пусть задано множество и функция $f(x)$, определенная на этом множестве. Требуется найти точки минимума или максимума функции $f(x)$ на множестве . Задачу на минимум записывают следующим образом:

$$f(x) \to \min, x \in X \tag{1}$$

При этом функцию $f(x)$ называют целевой функцией; множество – допустимым множеством; любое $x \in X$ – допустимой точкой.

Ниже будем рассматривать так называемые конечномерные задачи оптимизации, то есть задачи допустимое множество которых является подмножеством евклидового пространства E^n.

Определение 1.1. *Точка $x^* \in X$ называется:*

- *точкой глобального минимума функции $f(x)$ на множестве , если*

$$f(x^*) \leq f(x) \; \forall x \in X \tag{2}$$

- *точкой локального минимума функции $f(x)$ на множестве , если*

$$f(x^*) \leq f(x) \; \forall x \in X \cap U_\epsilon(x^*) \tag{3}$$

Отметим, что глобальный минимум всегда одновременно является локальным, но не наоборот.

Если в (2) и (3) при $x \neq x^*$ имеем строгие неравенства, то точка x^* называется соответственно точкой строгого глобального минимума, строгого локального минимума.

Определение 1.2. *Вектор h называется направлением убывания функции $f(x)$ в точке x, если выполняется*

$$f(x + \alpha h) < f(x) \tag{4}$$

при всех достаточно малых $\alpha > 0$.

Множество всех направлений убывания функции $f(x)$ в точке x обозначим $U(x, f)$

Теорема 1. *Пусть функция $f(x)$, дифференцируемая в точке $x \in E^n$. Если вектор h является направлением убывания функции $f(x)$ в точке x, то справедливо*

$$\langle \nabla f(x), h \rangle \leq 0 \tag{5}$$

Если при некотором $h \in E^n$ выполняется

$$\langle \nabla f(x), h \rangle < 0 \tag{6}$$

тогда вектор h является направлением убывания функции $f(x)$ в точке x.

Доказательство 1. ***Необходимость.*** *Пусть вектор $h \in U(x, f)$ и при этом $\langle \nabla f(x), h \rangle > 0$. В этом случае имеем*

$$f(x + \alpha h) - f(x) = \langle \nabla f(x), \alpha h \rangle + O(\alpha) = \alpha \left(\langle \nabla f(x), h \rangle + \frac{O(\alpha)}{\alpha} \right) > 0 \tag{7}$$

Здесь неравенство следует из того, что при достаточно малых $\alpha > 0$ знак выражения определяется знаком первого слагаемого. Таким образом, получили противоречие с (4), что и доказывает справедливость (5)

Достаточность. *Пусть выполняется (6). В этом случае*

$$f(x + \alpha h) - f(x) = \alpha \left(\langle \nabla f(x), h \rangle + \frac{O(\alpha)}{\alpha} \right) < 0 \tag{8}$$

при всех достаточно малых $\alpha > 0$. Следовательно, вектор h является направлением убывания функции $f(x)$ в точке x, то есть $h \in U(x, f)$.

Определение 1.3. *Вектор $h \in E^n$ задает возможное направление относительно множества в точке $x \in X$, если $x + \alpha h \in X$ при всех достаточно малых $\alpha > 0$. Вектор h в этом случае будем называть возможным направлением в точке x относительно множества X. Множество всех таких векторов h обозначим через $V(x, X)$ (множество возможных на-*

правлений в точке $x \in X$).

Теорема 2. *(необходимое условие локальной оптимальности). Если x^* – локальное решение задачи (1), то*

$$U(x^*, f) \cap V(x^*, X) = \emptyset \tag{9}$$

Доказательство 2. *Пусть $x^* \in X$ локальное решение (1) и при этом (9) неверно, то есть существует вектор $h \in E^n$ такой, что $f(x^* + \alpha h) < f(x^*)$ и при этом $x^* + \alpha h \in X$ при всех достаточно малых $\alpha > 0$. А это означает, что в любой сколь угодно малой окрестности точки x^* существует точка $x = x^* + \alpha h \in X \cap U_\epsilon(x^*)$ такая, что справедливо $f(x) < f(x^*)$, что противоречит определению локального минимума.*

Глава 2. Методы безусловной оптимизации

Рассмотрим задачу оптимизации:

$$f(x) \to \min, x \in E^n \tag{10}$$

Теорема 3. *(необходимое условие). Пусть функция $f(x)$ является дифференцируемой в точке $x^* \in E^n$. Если x^* – локальное решение задачи (10), то*

$$\nabla f(x^*) = 0 \tag{11}$$

Определение 2.1. *Точка x^*, удовлетворяющая условию (11), называется стационарной точкой функции $f(x)$.*

Теорема 4. *(необходимое условие оптимальности 2-го порядка). Пусть функция $f(x)$ дважды дифференцируемая в точке $x^* \in E^n$. Если x^* – локальное решение задачи (10), то матрица Гессе функции $f(x)$ в точке x^* неотрицательно определена, то есть*

$$\langle \nabla^2 f(x^*)h, h \rangle \geq 0 \; \forall h \in E^n \tag{12}$$

Теорема 5. *(достаточное условие). Пусть функция $f(x)$ дважды дифференцируемая в точке $x^* \in E^n$. И пусть в этой точке выполняется условие стационарности (11) и матрица Гессе положительно определена, то есть*

$$\langle \nabla^2 f(x^*)h, h \rangle > 0 \; \forall h \in E^n \tag{13}$$

Тогда x^* – строгое локальное решение задачи (10).

2.1. Метод наискорейшего спуска

Общая схема методов спуска, в которых последовательность приближений $x^1, x^2, \dots$ к точке минимума строится по правилу:

$$x^{k+1} = x^k + \alpha_k h^k \tag{14}$$

где направление h^k принадлежит множеству направлений убывания функции $h^k \in U(x^k, f)$; $\alpha_k \geq 0$ – параметр, определяющий длину шага вдоль направления h^k.

В градиентных методах направление h^k берется равным антиградиенту функции $f(x)$ в точке x^k, то есть $h^k = -\nabla f(x^k)$. В градиентных методах используются различные методы выбора шага α_k. Если длина шага выбирается из минимизации функции вдоль направления антиградиента, то получаем вариант градиентного метода, называемый методом наискорейшего спуска.

Итак, в методе наискорейшего спуска шаг α_k вдоль направления h^k выбирается из решения оптимизационной задачи:

$$f(x^k - \alpha_k \nabla f(x^k)) = \min_{\alpha \geq 0} f(x^k - \alpha_k \nabla f(x^k)) \tag{15}$$

Теорема 6. *Пусть функция $f(x)$ дифференцируема на всем пространстве E^n, и ее градиент удовлетворяет условию Липшица:*

$$||\nabla f(x + \Delta x) - \nabla f(x)|| \leq L||\Delta x||,\ \forall x, x + \Delta x \in E^n \tag{16}$$

Тогда для для остаточного члена в разложении $\Delta f(x) = \langle \nabla f(x), \Delta x \rangle + O(||\Delta x||)$ справедлива оценка:

$$O(||\Delta x||) \leq \frac{L}{2}||\Delta x||^2 \tag{17}$$

Доказательство 3. *Для любых $x, x + \Delta x \in E^n$ справедливо соотношение*

$$\Delta f(x) = f(x + \Delta x) - f(x) = \int_0^1 \langle \nabla f(x + \alpha \Delta x), \Delta x \rangle d\alpha \tag{18}$$

где $\alpha \in [0, 1]$. Действительно, рассмотрим функцию переменной α : $g(\alpha) = f(x + \alpha \Delta x)$. Её производная по переменной α имеет следующий вид:

$$\frac{dg}{d\alpha}(\alpha) = \langle \nabla f(x + \alpha \Delta x), \Delta x \rangle \tag{19}$$

а отсюда следует справедливость

$$\int_0^1 \frac{dg}{d\alpha}(\alpha) = g(1) - g(0) = f(x + \Delta x) - f(x) \tag{20}$$

Таким образом можем представить

$$\begin{aligned} \Delta f(x) = \int_0^1 \langle \nabla f(x + \alpha \Delta x), \Delta x \rangle d\alpha = \\ = \int_0^1 \langle \nabla f(x), \Delta x \rangle d\alpha + \int_0^1 \langle \nabla f(x + \alpha \Delta x) - \nabla f(x), \Delta x \rangle d\alpha \leq \\ \leq \int_0^1 \langle \nabla f(x), \Delta x \rangle d\alpha + \int_0^1 ||f(x + \alpha \Delta x) - \nabla f(x)|| \cdot ||\Delta x|| d\alpha \end{aligned} \tag{21}$$

Так как градиент функции $f(x)$ удовлетворяет условию Липшица

$$\begin{aligned} \Delta f(x) \leq \langle \nabla f(x), \Delta x \rangle + \int_0^1 ||f(x + \alpha \Delta x) - \nabla f(x)|| \cdot ||\Delta x|| d\alpha \leq \\ \leq \langle \nabla f(x), \Delta x \rangle + L \int_0^1 ||\alpha \Delta x|| \cdot ||\Delta x|| d\alpha = \langle \nabla f(x), \Delta x \rangle + L ||\Delta x||^2 \int_0^1 \alpha d\alpha = \\ = \langle \nabla f(x), \Delta x \rangle + \frac{L}{2} ||\Delta x||^2 \end{aligned} \tag{22}$$

Таким образом имеем

$$\Delta f(x) = \langle \nabla f(x), \Delta x \rangle + O(||\Delta x||) \leq \langle \nabla f(x), \Delta x \rangle + \frac{L}{2} ||\Delta x||^2 \tag{23}$$

2.2. Метод Ньютона

Метод Ньютона решения задач безусловной минимизации относится к методам 2-го порядка, то есть к методам, используемым информацию о вторых производных целевой функции $f(x)$.

Предположим, что функция $f(x)$ дважды непрерывно дифференцируема в E^n. Пусть начальное приближение $x^0 \in E^n$ задано и с помощью метода Ньютона уже найдено k-ое приближение $x^k \in E^n$. В некоторой окрестности точки x^k функцию $f(x)$ аппроксимируем квадратичной функцией $\psi_k(x)$:

$$\psi_k(x) = f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle \nabla^2 f(x^k) \cdot [x - x^k], x - x^k \rangle \qquad (24)$$

Рассмотрим вспомогательную задачу минимизации функции $\psi_k(x)$:

$$\psi_k(x) \to \min, x \in E^n \qquad (25)$$

Предположим, что решение задачи (25) $\tilde{x}^k$ существует. Очевидно, что в этом случае в точке $\tilde{x}^k$ выполняется $\nabla \psi_k(\tilde{x}^k) = 0$. Так как $\nabla \psi_k(x) = \nabla f(x^k) + \nabla^2 f(x^k) \cdot [x - x^k]$, то имеем:

$$\tilde{x}^k = x^k - [\nabla^2 f(x^k)]^{-1} \nabla f(x^k) \qquad (26)$$

Таким образом в качестве направления спуска можно принять:

$$h^k = -[\nabla^2 f(x^k)]^{-1} \nabla f(x^k) \qquad (27)$$

Рассмотрим точки, лежащие на отрезке $[x^k, \tilde{x}^k] : x^k(\alpha) = x^k + \alpha(\tilde{x}^k - x^k)$, где $\alpha \in [0, 1]$. Здесь заметим, что в силу (26): $\tilde{x}^k - x^k = -[\nabla^2 f(x^k)]^{-1} \nabla f(x^k)$. Конкретную точку из отрезка $[x^k, \tilde{x}^k]$ выберем, найдя из условия минимума функции $f(x^k(\alpha))$. Следующее приближение определим по формуле $x^{k+1} = x^k(\alpha_k)$. С учётом решения вспомогательной задачи (25) схема метода Ньютона примет вид:

$$\begin{cases} x^{k+1} = x^k - \alpha_k [\nabla^2 f(x^k)]^{-1} \nabla f(x^k), \\ \alpha_k : f(x^k - \alpha_k [\nabla^2 f(x^k)]^{-1} \nabla f(x^k)) \to \min, \alpha \in [0, 1] \end{cases} \qquad (28)$$

2.3. Методы сопряженных направлений

Рассмотри задачу минимизации квадратичной функции:

$$f(x) = \frac{1}{2} \langle Ax, x \rangle + \langle b, x \rangle \to \min, x \in E^n \qquad (29)$$

где A - симметричная положительно определенная $n \times n$ матрица.

Идея методов сопряженных направлений основана на стремлении найти минимум квадратичной функции (29) за конечное число шагов. Согласно методу, требуется найти направления $h^0, h^1, \ldots, h^{n-1}$ такие, что последовательная одномерная минимизация функции $f(x)$ вдоль этих направлений, начиная с любой точки $x^0 \in E^n$:

$$\begin{cases} f(x^k + \alpha_k h^k) = \min_{\lambda_k} f(x^k + \lambda_k h^k), \\ x^{k+1} = x^k + \alpha_k h^k, (k = 0, 1, \ldots, n-1) \end{cases} \tag{30}$$

приводит к отысканию минимума функции (29).

Определение 2.2. *Вектора h^1 и h^2 называются сопряженными (относительно матрицы), если они отличны от нуля и скалярное произведение $\langle Ah^1, h^2 \rangle = 0$.*

Определение 2.3. *Вектора $h^0, h^1, \ldots, h^{k-1}$ называются взаимно сопряженными (относительно матрицы), если все они отличны от нуля и скалярное произведение $\langle Ah^i, h^j \rangle = 0 \; \forall i \neq j, 0 \leq i, j \leq k$.*

Теорема 7. *Пусть вектора $h^0, h^1, \ldots, h^{k-1}$ являются взаимно сопряженными, тогда они линейно не зависимы.*

Доказательство 4. *Доказательство проведем от противного. Пусть вектора $h^0, h^1, \ldots, h^{k-1}$ являются взаимно сопряженными, но при этом они являются линейно зависимыми, то есть в этом случае один из векторов можно представить в виде линейной комбинации остальных векторов, например $h^i = \sum_{j=0, j\neq i}^{k-1} \lambda_j h^j$. Тогда $\langle Ah^i, h^i \rangle = \sum_{j=0, j\neq i}^{k-1} \lambda_j \langle Ah^i, h^j \rangle$. И отсюда в силу взаимной сопряженной векторов $\langle Ah^i, h^i \rangle = 0$, что возможно, если вектор $h^i = 0$, так как матрица является по условию симметричной положительно определенной $n \times n$ матрицей. Получили противоречие с тем, что по условию взаимной сопряженности векторов $h^i \neq 0$.*

Определение 2.4. *Если в методе минимизации функции $f(x)$ (29) вектора $h^0, h^1, \ldots, h^{k-1}$ взаимно сопряжены, то метод (30) называется методом сопряженных направлений*

Теорема 8. *Если в методе минимизации (30) функции $f(x)$ вектора $h^0, h^1, \ldots, h^{k-1}$ взаимно сопряжены, то для функции $f(x)$, заданной формулой (29) справедливо $f(x^m) = \min_{x \in X_m} f(x)$, где $X_m = x^0 + lin\ h^0, h^1, \ldots, h^{m-1}$ линейное подпространство, натянутое на указанные векторы.*

Доказательство 5. *Предварительно заметим справедливость соотношения*

$$f(x^k + \lambda_k h^k) - f(x^k) = \lambda_k \langle Ax^0 + b, h^k \rangle + \frac{1}{2}\lambda_k^2 \langle Ah^k, h^k \rangle \tag{31}$$

при любом $k = 0, 1, \ldots, m-1$. Действительно, учитывая, что справедливо $x^k = x^0 + \sum_{i=0}^{k-1} \alpha_i h^i$ и $\langle Ah^i, h^k \rangle = 0$, получим

$$\begin{aligned} f(x^k + \lambda_k h^k) &= \frac{1}{2}\langle A(x^k + \lambda_k h^k), (x^k + \lambda_k h^k) \rangle + \langle b, (x^k + \lambda_k h^k) \rangle = \\ &= \frac{1}{2}\langle Ax^k, x^k \rangle + \langle b, x^k \rangle + \lambda_k \langle Ax^k + b, h^k \rangle + \frac{1}{2}\lambda_k^2 \langle Ah^k, h^k \rangle = \\ &= f(x^k) + \lambda_k \langle Ax^0 + \sum_{i=0}^{k-1} \alpha_i Ah^i + b, h^k \rangle + \frac{1}{2}\lambda_k^2 \langle Ah^k, h^k \rangle = \\ &= f(x^k) + \lambda_k \langle Ax^0 + b, h^k \rangle + \lambda_k \langle \sum_{i=0}^{k-1} \alpha_i Ah^i, h^k \rangle + \frac{1}{2}\lambda_k^2 \langle Ah^k, h^k \rangle = \\ &= f(x^k) + \lambda_k \langle Ax^0 + b, h^k \rangle + \frac{1}{2}\lambda_k^2 \langle Ah^k, h^k \rangle \end{aligned} \tag{32}$$

а отсюда следует справедливость (31).

Для любой точки $x \in X_m$ имеем

$$f(x) = f(x^0 + \sum_{k=0}^{m-1} \lambda_k h^k =$$

$$= \frac{1}{2}\langle A(x^0 + \sum_{k=0}^{m-1} \lambda_k h^k), x^0 + \sum_{k=0}^{m-1} \lambda_k h^k\rangle + \langle b, x^0 + \sum_{k=0}^{m-1} \lambda_k h^k\rangle = \tag{33}$$

$$= f(x^0) + \sum_{k=0}^{m-1} \left(\lambda_k \langle Ax^0 + b, h^k\rangle + \frac{1}{2}\lambda_k^2 \langle Ah^k, h^k\rangle\right)$$

Отсюда, с учетом (31) для любой точки $x \in X_m$ *справедливо* $f(x) = f(x^0) + \sum_{k=0}^{m-1} \left(f(x^k + \lambda_k h^k) - f(x^k)\right)$. *Учитывая изложенное, получим*

$$\min_{x \in X_m} f(x) = \min_{\lambda^0, \ldots, \lambda^{m-1}} f(x^0 + \sum_{k=0}^{m-1} \lambda_k h^k) =$$

$$= \min_{\lambda^0, \ldots, \lambda^{m-1}} \left(f(x^0) + \sum_{k=0}^{m-1} \left(f(x^k + \lambda_k h^k) - f(x^k)\right)\right) =$$

$$= f(x^0) + \sum_{k=0}^{m-1} \left(\min_{\lambda_k} f(x^k + \lambda_k h^k) - f(x^k)\right) = \tag{34}$$

$$= f(x^0) + \sum_{k=0}^{m-1} \left(f(x^k + \alpha_k h^k) - f(x^k)\right) =$$

$$= f(x^0) + \sum_{k=0}^{m-1} \left(f(x^{k+1}) - f(x^k)\right) = f(x^m)$$

Таким образом справедливо $f(x^m) = \min_{x \in X_m} f(x)$.