

Методы оптимизации

Оптимизация – выбор наилучшего решения.

Сложность или невозможность отыскания аналитического решения привело к тому, что постепенно стало ясно, что любая задача может считаться решенной, если *указан алгоритм, позволяющий численно построить приближенное решение с требуемой точностью*.

Математическая теория оптимизации включает в себя фундаментальные результаты и численные методы, позволяющие находить наилучший вариант из множества возможных альтернатив *без их полного перебора* и сравнения.

Итак, есть варианты решения задачи, среди которых надо найти лучший. Как оценить какой метод лучше? Надо определить критерий качества.

Критерий качества – функционал, действующий из множества вариантов решения задачи в множества вещественных чисел. Тогда понятие хуже - лучше тождественно больше - меньше. Один вариант лучше другого, если, например, значение функционала меньше, и неопределенность теряется.

У одной и той же задачи часто бывает возможным наличие нескольких функционалов качества. При этом нахождение их экстремума оказывается сложным, и выбраться из этой ситуации можно за счет методов многокритериальной оптимизации.

В этом курсе мы будем заниматься поиском экстремума одного функционала.

Итак, рассмотрим *функционал* $\varphi: X \rightarrow \bar{R}$, где

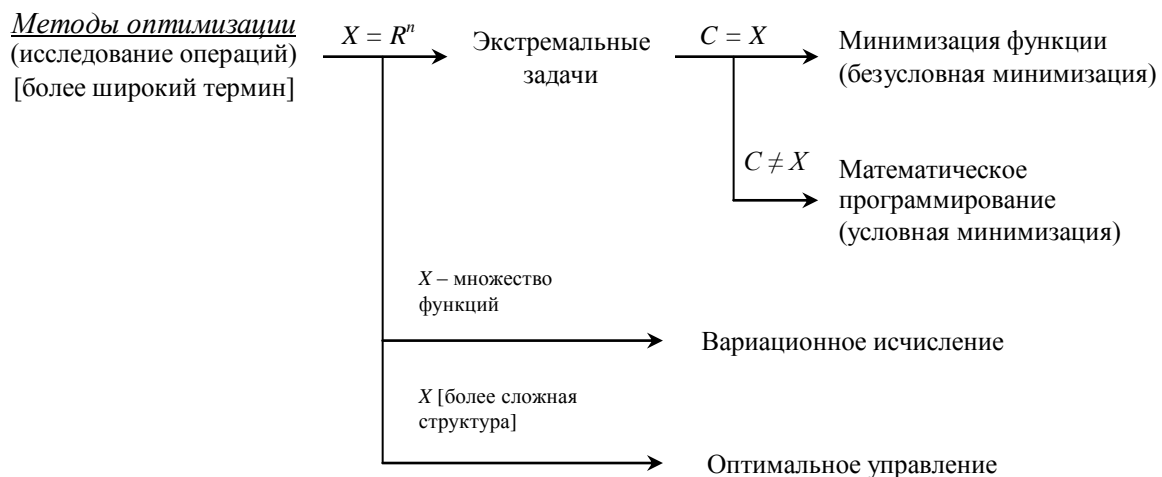
X – множество вариантов или допустимое множество (область определения функционала);

$\bar{R} = R \cup \{-\infty\} \cup \{+\infty\}$ – расширенная вещественная прямая.

Пусть $c \subset X$ – некоторое подмножество X .

Задача: $\varphi(x) \rightarrow \min(\max), x \in c$ называется *экстремальной задачей с ограничением c* .
(экстремум – максимум или минимум)

Терминология и классификация



Этапы решения оптимизационной задачи

Процесс принятия решения в исследовании операций представляет собой сложный процесс, который условно можно разбить на 4 этапа:

1 этап: Построение качественной модели рассматриваемой проблемы, т.е. выделение факторов, которые представляются наиболее важными, установление закономерностей, которым они подчиняются.

2 этап: Построение математической модели, включающей в себя выбор функционала φ (или целевой функции переменных), $\varphi(x) \rightarrow \min(\max)$, формирование ограничений (условий) в виде равенств или неравенств, например

$$X = \{x : f_i(x) \leq a_i, \dots, f_m(x) \leq a_m\}.$$

Этот этап требует привлечения математических знаний и характеризуется, как правило, большим количеством переменных (n и m – велики).

3 этап: Решение математической задачи – выбор метода, реализация его и получение результата (применение ЭВМ, разработка программ, применение существующих СП и т.д.).

4 этап: Анализ полученного результата. Выясняется степень адекватности модели (результаты вычислений) и моделируемого объекта (имитационные данные).

Примеры математических моделей

Вообще, теория математических моделей является предметом специализированного курса и требует знаний в той области, которой принадлежит моделируемый объект. Рассмотрим традиционные примеры, иллюстрирующие применение метода математического моделирования в задачах экономического содержания.

Задача о рационе: Пусть имеется n – число продуктов питания и m – число питательных веществ.

Пусть a_{ij} – содержание j -го вещества в единице i -го продукта;

b_j – минимальная (суточная) потребность (человека) в j -ом веществе;

c_i – стоимость единицы i -го продукта;

x_i – искомое количество (суточное потребление) i -го продукта.

Тогда $\sum_{i=1}^n a_{ij}x_i$ – общее содержание j -го питательного вещества;

$\sum_{i=1}^n c_i x_i$ – стоимость (суточного) рациона.

Задача:

найти $\min \sum_{i=1}^n c_i x_i$ (или тот набор, (x_1, \dots, x_n) при котором достигается минимум)

при условии, что $\sum_{i=1}^n a_{ij} x_i \geq b_j, j = 1, \dots, m, x_i \geq 0, i = 1, \dots, n$

Типичная задача
линейного
программирования

Транспортная задача: Требуется составить план перевозок однородного груза таким образом, чтобы общая стоимость перевозок была минимальной.

Пусть a_i – количество единиц груза в i -ом пункте отправления ($i = 1, \dots, m$);

b_j – потребность в j -ом пункте назначения ($j = 1, \dots, n$) в единицах груза;

c_{ij} – стоимость перевозки единицы груза из i -го пункта в j -ый;

x_{ij} – планируемое количество единиц груза для перевозки из i -го пункта в j -ый.

Тогда $\sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$ – общая (суммарная) стоимость перевозок;

$\sum_{j=1}^n x_{ij}$ – количество груза, вывозимого из i -го пункта;

$\sum_{i=1}^m x_{ij}$ – количество груза, доставляемого в j -ый пункт.

В простейшем случае должны выполняться следующие очевидные условия:

$$\sum_{j=1}^n x_{ij} = a_i, i = \overline{1, m}; \sum_{i=1}^m x_{ij} = b_j, j = \overline{1, n}; \sum_{i=1}^m a_i = \sum_{j=1}^n b_j.$$

Таким образом, математическая формулировка транспортной задачи имеет вид:

Найти:

$$\min \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij},$$

при условиях

$$\sum_{j=1}^n x_{ij} = a_i; \sum_{i=1}^m x_{ij} = b_j; x_{ij} \geq 0, i = \overline{1, m}, j = \overline{1, n}.$$

Задача носит название *замкнутой транспортной модели*, а условие $\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$

является естественным условием разрешимости замкнутой транспортной задачи.

Обозначения и определения

R^n – n -мерное евклидово пространство.

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} - \text{вектор столбец в } R^n;$$

$x^T = (x_1, x_2, \dots, x_n)$ – вектор-строка в R^n ;

$$(x, y) = \sum_{i=1}^n x_i y_i - \text{скалярное произведение, } x, y \in R^n;$$

$$\|x\| = \sqrt{(x, x)} - \text{евклидова норма вектора в } R^n;$$

$A - m \times n$ – матрица, $A^T - n \times m$ – транспонированная матрица;

$Ax \in R^m$ – произведение матрицы ($m \times n$) на вектор ($n \times 1$);

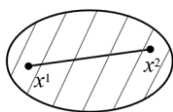
$\varphi(x), f(x), g(x), \dots$ – как правило, вещественные (скалярные) функции, т.е. $\varphi: R^n \rightarrow R$;

$$\varphi'(x^0) = \text{grad } \varphi(x^0) = \left(\begin{array}{c} \frac{\partial \varphi}{\partial x_1} \\ \vdots \\ \frac{\partial \varphi}{\partial x_n} \end{array} \right) \bigg|_{X=X^0} - \text{градиент функции } \varphi \text{ в точке } x^0 \text{ (} n\text{-мерный вектор);}$$

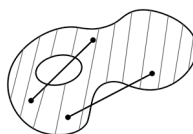
$$\varphi''(x^0) = H(x^0) = \left(\begin{array}{ccc} \frac{\partial^2 \varphi}{\partial x_1^2} & \dots & \frac{\partial^2 \varphi}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 \varphi}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 \varphi}{\partial x_n^2} \end{array} \right) \bigg|_{X=X^0} - \text{матрица Гессе (матрица вторых производных) функции } \varphi \text{ в точке } x^0;$$

Т.к. $\frac{\partial^2 \varphi}{\partial x_i \partial x_j} = \frac{\partial^2 \varphi}{\partial x_j \partial x_i}$, то $H(x^0)$ – есть вещественная симметричная матрица.

Определение. Множество $X \subset R^n$ называется *выпуклым*, если для $\forall x^1, x^2 \in X, \forall \lambda \in [0, 1]$ $\lambda x^1 + (1 - \lambda)x^2 \in X$. Иными словами, множество X выпукло, если оно вместе с любыми своими двумя точками x^1 и x^2 содержит соединяющий их отрезок.



Выпуклое множество



Невыпуклое множество

Примеры.

На числовой прямой R выпуклыми множествами являются всевозможные промежутки, т.е.:

- одноточечные множества;
- интервалы;
- полуинтервалы;
- отрезки;
- полупрямые;
- сама прямая.

В пространстве R^n примерами выпуклых множеств служат:

- само подпространство;
- любое его линейное подпространство;
- одноточечное множество;
- шар;
- отрезок,
- а также следующие множества:

$l_{x^0 h} = \{x \in R^n \mid x = x^0 + \alpha h, \alpha \in R\}$ – прямая, проходящая через $(\cdot) x^0$ в направлении вектора h .

$l_{x^0 h}^+ = \{x \in R^n \mid x = x^0 + \alpha h, \alpha \geq 0\}$ – луч, выходящий из $(\cdot) x^0$ в направлении h .

$H_{p\beta} = \{x \in R^n \mid (p, x) = \beta\}$ – гиперпространство с нормалью p .

$H_{p\beta}^+ = \{x \in R^n \mid (p, x) \geq \beta\}, H_{p\beta}^- = \{x \in R^n \mid (p, x) \leq \beta\}$ – порождаемые ею полупространства.

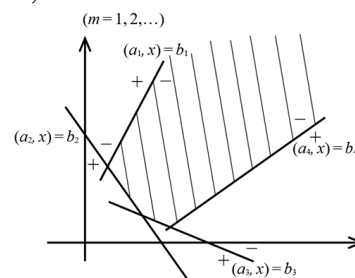
Все перечисленные множества в R^n , кроме шара, являются частными случаями *выпуклого множества вида*:

$$X = \{x \in R^n \mid Ax \leq b\} = \{x \in R^n \mid (a_i, x) \leq b_i, i = 1, \dots, m\},$$

где A – некоторая матрица размера $m \times n$ со строками a_1, \dots, a_m ,

$b \in R^m$ – вектор.

Множества такого вида называют *полиэдральными* или *полиэдрами*. Таким образом, полиэдр – множество решений некоторой системы конечного числа линейных неравенств (пересечение конечного числа полупространств).

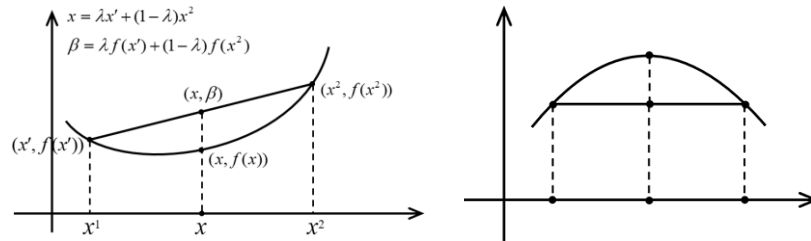


Определение. Функция f , определенная на выпуклом множестве $X \subset R^n$, называется *выпуклой на X* , если

$$f(\lambda x^1 + (1-\lambda)x^2) \leq \lambda f(x^1) + (1-\lambda)f(x^2),$$

при $\forall x^1, x^2 \in X, \lambda \in [0,1]$.

Если для любых $x^1, x^2, x^1 \neq x^2, \lambda \in [0,1]$ неравенство выполняется как строгое, то f называется *строго выпуклой на X* . Функция называется (строго) *вогнутой*, если функция $-f$ (строго) *выпукла*.



Функцию $f(x) = (a, x) + b$, где $a \in R^n, b \in R$ будем называть *линейной*. Ясно, что для неё исходное неравенство выполняется как равенство. Поэтому она *выпукла* и *вогнута* одновременно, но не *строго*.

Минимизация функций

Сама по себе постановка задачи оптимизации проста и естественна: заданы множество X и функция $\varphi(x)$, определенная на X , требуется найти точку минимума или максимума функции φ на X .

Условимся записывать задачу на минимум в виде

$$\varphi(x) \rightarrow \min, x \in X, \text{ где}$$

φ – целевая функция; X – допустимое множество.

Условимся также, что в дальнейшем будем рассматривать задачу на \min , поскольку задача $\min \varphi(x) \Leftrightarrow \max(-\varphi(x))$.

Если допустимое множество $X = R^n$, то задача называется *безусловной* минимизацией, иначе, когда $X \neq R^n$ – задача *условной* минимизации.

Отметим, что само понятие точки минимума неоднозначно и требует уточнения.

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

- 1) точкой глобального минимума функции φ на множестве X , или глобальным решением задачи, если $\forall x \in X \varphi(x) \geq \varphi(x^*)$;
- 2) точкой локального минимума φ на X , или локальным решением задачи, если существует

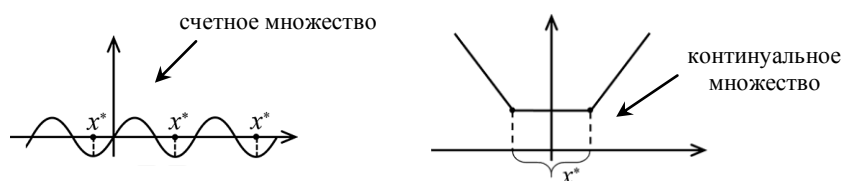
$$\delta > 0: \forall x \in X, \|x - x^*\| < \delta \Rightarrow \varphi(x) \geq \varphi(x^*)$$

(Если $\Delta = \{x: \|x - x^*\| < \delta\}$ – δ -окрестность (\cdot) x^* и x^* – локальный min, то x^* – глобальный min в области $X \cap \Delta$).

Если неравенства в 1) и 2) выполняются как строгие, то говорят, что x^* – точка строгого min в глобальном или локальном смысле.

Ясно, что глобальное решение является локальным, обратное – не верно.

Пример. Глобальных min может быть много:



Для записи того факта, что x^* является точкой глобального min функции φ на X используем запись:

$$\varphi^* = \varphi(x^*) = \min_{x \in X} \varphi(x)$$

или эквивалентная ей запись:

$$\varphi^* = \arg \min_{x \in X} \varphi(x) \text{ — оптимальная точка.}$$

Множество всех точек глобального min φ на X обозначим:

$$\text{Arg min}_{x \in X} \varphi(x) = \{x^* \in X \mid \varphi(x^*) = \varphi^*\}$$

Таким образом, $\arg \min_{x \in X} \varphi(x)$ – это произвольная точка из множества $\text{Arg min}_{x \in X} \varphi(x)$.

В дальнейшем мы часто будем прибегать к геометрической интерпретации задач оптимизации, основанной на понятии линий (или поверхностей) уровня функции φ , т.е. множеств вида:

$$L_\alpha = \{x \in R^n : \varphi(x) = \alpha\} \text{ - такое множество носит название } \textit{поверхность уровня } \alpha.$$

Напомним известный факт из анализа: если функция φ дифференцируема в точке x , то градиент $\varphi'(x)$ ортогонален к проходящей через x линии уровня α и направлен (если $\varphi'(x) \neq 0$) в сторону возрастания функции φ , т.е. поверхность L_α делит R^n на два подпространства:

$$\{x: \varphi(x) > \alpha\} (+) \text{ и } \{x: \varphi(x) < \alpha\} (-).$$

Задача поиска оптимальной точки может быть сформулирована следующим образом: найти $\alpha^* = \min \alpha$ среди тех α , для которых $L_\alpha \cap X \neq \emptyset$. Тогда любая точка $x \in L_{\alpha^*}$ является оптимальной точкой.

Возможны два случая:

- x^* лежит внутри X – рис. 1;
- x^* лежит на границе X – рис. 2.

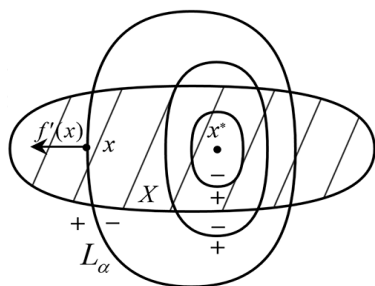


Рис.1

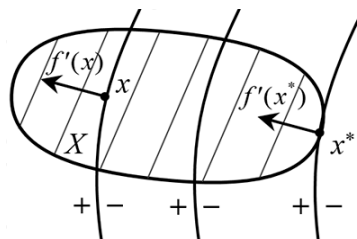


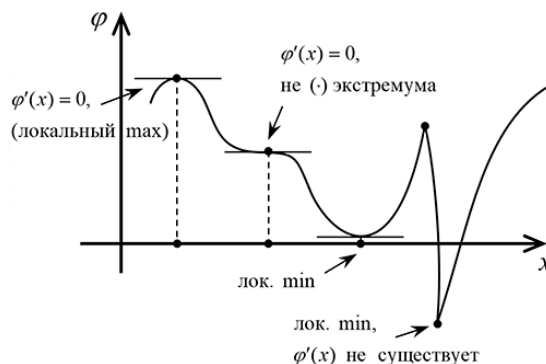
Рис.2

При изучении задач оптимизации в первую очередь возникает вопрос о существовании решения. Напомним в этой связи некоторые результаты из *математического анализа*.

Теорема 1 (Вейерштрасса). Если X – компакт в R^n (т.е. замкнутое ограниченное множество), а φ – непрерывная функция на X , то существует $x^* \in X$: x^* – глобальный минимум φ на X , т.е. *глобальное решение задачи $\varphi(x) \rightarrow \min, x \in X$ существует!*

Теорема 2 (необходимые условия локального минимума). Если φ – дифференцируема в точке $x^* \in X$ и x^* – локальный минимум, то $\varphi'(x^*) = 0$ (градиент равен нулю).

Определение. Точка $\hat{x} \in X$ в $\varphi'(\hat{x}) = 0$, называется *стационарной* (обратное не верно).



Теорема 3 (достаточное условие локального минимума).

Если φ дважды дифференцируема в точке $x^* \in R^n$ и выполняется:

- 1) $\varphi'(x^*) = 0$;
 - 2) матрица Гессе $\varphi''(x^*)$ положительно определена,
- то x^* – (строгий) локальный минимум функции φ .

Определение. Матрица H называется *положительно определенной*, если $\forall h \in R^n, (Hh, h) > 0, h \neq 0$.

Критерий Сильвестра: H положительно определена \Leftrightarrow ее главные миноры положительны.

Приведем несколько теорем для выпуклых задач.

Определение. Задача минимизации $\varphi(x) \rightarrow \min, x \in X$ называется выпуклой, если X – выпуклое множество, φ – выпуклая функция на X .

Справедливы следующие теоремы:

Теорема 4. Если задача минимизации $\varphi(x) \rightarrow \min, x \in X$ выпукла, то любое её локальное решение является также глобальным.

Доказательство. Пусть x^* – локальное решение задачи, т.е. при некотором $\varepsilon > 0$ выполняется условие:

$$\varphi(x^*) \leq \varphi(x) \text{ при } \forall x \in X \cap U_\varepsilon(x^*),$$

где $U_\varepsilon(x^*) = \{x \in R^n \mid \|x - x^*\| \leq \varepsilon\}$ – шар радиуса $\varepsilon > 0$ с центром в x^* .

Для любых точек $x \in X: x \neq x^*$, положим $\lambda = \min \left(\frac{\varepsilon}{\|x - x^*\|}, 1 \right)$.

Тогда $\lambda x + (1 - \lambda)x^* \in X \cap U_\varepsilon(x^*)$. Покажем это.

$$1. \text{ Пусть } \lambda = 1 \Rightarrow \|\lambda x + (1 - \lambda)x^* - x^*\| = \|x - x^*\|,$$

$$\text{Если } \lambda = 1 \Rightarrow \frac{\varepsilon}{\|x - x^*\|} > 1 \Rightarrow \|x - x^*\| < \varepsilon \Rightarrow \text{точка } \lambda x + (1 - \lambda)x^* \in U_\varepsilon(x^*)$$

$$2. \text{ Пусть } \lambda = \frac{\varepsilon}{\|x - x^*\|}$$

$$\Rightarrow \|\lambda x + (1 - \lambda)x^* - x^*\| = \left\| \frac{\varepsilon}{\|x - x^*\|} \cdot x + \left(1 - \frac{\varepsilon}{\|x - x^*\|} \right) \cdot x^* - x^* \right\| = \left\| \frac{\varepsilon x}{\|x - x^*\|} - \frac{\varepsilon x^*}{\|x - x^*\|} \right\| = \varepsilon$$

$$\Rightarrow \text{точка } \lambda x + (1 - \lambda)x^* \in U_\varepsilon(x^*)$$

и, следовательно,

$$\varphi(x^*) \leq \varphi(\lambda x + (1 - \lambda)x^*) \leq \lambda \varphi(x) + (1 - \lambda)\varphi(x^*) \Rightarrow \varphi(x^*) \leq \varphi(x),$$

т.е. x^* – глобальное решение задачи, *ч.т.д.*

Таким образом, для выпуклых задач понятия локального и глобального решений не различаются.

Второе свойство выпуклых задач можно высказать в виде следующего общего принципа: *необходимые условия оптимальности* в том или ином классе задач минимизации при соответствующих предположениях выпуклости *оказываются и достаточным*.

Теорема 5. Пусть функция φ выпукла на R^n и дифференцируема в точке $x^* \in R^n$. Если $\varphi'(x^*) = 0$, то x^* – точка минимума функции на R^n , т.е. решение задачи минимизации $\varphi(x) \rightarrow \min, x \in X$.

Доказательство. Для $\forall x \in X, \lambda \in [0, 1]$ имеем

$$\varphi(\lambda x + (1 - \lambda)x^*) \leq \lambda \varphi(x) + (1 - \lambda)\varphi(x^*).$$

Преобразуя эту формулу и, пользуясь дифференцируемостью функции φ в точке x^* , получаем:

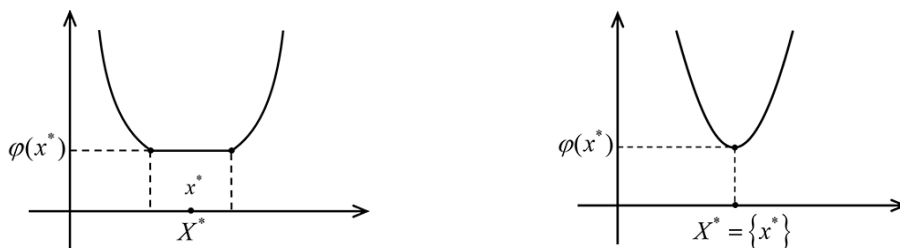
$$\varphi(x) - \varphi(x^*) \geq \frac{\varphi(x^* + \lambda(x - x^*)) - \varphi(x^*)}{\lambda} = \frac{(\varphi'(x^*), \lambda(x - x^*)) + o(\lambda)}{\lambda} = \frac{o(\lambda)}{\lambda};$$

Отсюда предельным переходом при $\lambda \rightarrow 0$ выводим, что $\varphi(x) \geq \varphi(x^*)$, *ч.т.д.* (т.е. для $\forall x \in X \varphi(x) \geq \varphi(x^*)$).

Полученные свойства выпуклых задач имеют важное значение не только в теории, но и в численных методах оптимизации. Дело в том, что большинство существующих численных методов позволяет, вообще говоря, находить лишь локальные решения, а точнее – стационарные точки задачи. Теоремы 4 и 5 говорят о том, что для выпуклой задачи *отыскание стационарной точки* автоматически означает *отыскание решения, причем глобального*.

Укажем ещё одно полезное свойство выпуклых задач.

Теорема 6. Пусть задача минимизации $\varphi(x) \rightarrow \min, x \in X$ выпукла и имеет решение. Тогда множество её решений $X^* = \text{Arg min } \varphi(x)$ выпукло. Если при этом $\varphi(x)$ строго выпукла на X , то решение единственно, т.е. X^* состоит из одной точки.



Доказательство:

1. Пусть $x^1, x^2 \in X^*, \lambda \in [0, 1] \Rightarrow \varphi(x^1) = \varphi(x^2) = \varphi(x^*) = \varphi^*$

При этом

$$\varphi(\lambda x^1 + (1-\lambda)x^2) \leq \lambda\varphi(x^1) + (1-\lambda)\varphi(x^2) = \varphi^* \quad (*)$$

По определению X^* неравенство может выполняться только как равенство, поскольку $\varphi^* = \min$

$$\Rightarrow \lambda x^1 + (1-\lambda)x^2 \in X^*, \text{ т.е. } X^* - \text{выпукло.}$$

2. Пусть φ – строго выпукла. Если предположить, что в X^* существуют две различные точки x^1 и x^2 , то при $\lambda \in [0,1]$ неравенство (*) должно быть строгим, что невозможно, т.к. $\varphi^* = \min$ и получается $< \min$.

Трудности:

1. В случаях, когда функция φ достаточно проста, теоремы 1-3 помогают решить задачу минимизации даже в явном виде. Однако зачастую задача поиска стационарных точек является нетривиальной. А затем – перебор стационарных точек в поисках точки локального минимума, затем – перебор локальных экстремумов в поисках глобального экстремума.
2. Для задач условной минимизации теоремы 1-3 применимы в случае, когда локальное решение x^* – внутренняя точка допустимого множества X . Если же экстремум достигается в угловых точках границы множества условий, то нарушается дифференцируемость \Rightarrow неприменимость методов классического анализа.

Т.о., в большинстве случаев задачу $\min \varphi(x)$ приходится *решать численно с применением ЭВМ и специальных методов минимизации.*

Безусловная минимизация функции

Методы оптимизации функций в R^n делятся на:

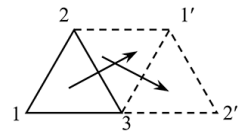
- локальные методы (поиск локального \min , т.е. такой точки x^* , что существует $\delta > 0$,
 $\forall x \in X : \{ \|x - x^*\| \leq \delta \} \Rightarrow \varphi(x^*) \leq \varphi(x)$);
- нелокальные (или прямые) методы (поиск глобального \min для ограничений снизу функции $\varphi(x)$, т.е. если α^* – нижняя грань, то поиск такой точки $x^* : \varphi(x^*) = \alpha^*$). Для этих методов не требуется аналитического задания функции, надо только уметь вычислять ее значение в любой точке. Обычно – для функций сложной структуры.

Нелокальные методы сводятся к уменьшению области, внутри которой находится оптимальная точка. Пример нелокального метода – *симплексный метод*.

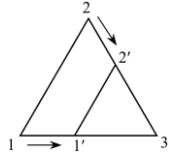
Определение. Симплекс – выпуклое тело в R^n , состоящее из $(n + 1)$ равноудаленных точек – вершин симплекса, отрезок их соединяющий – ребро симплекса, в R^2 – треугольник, в R^3 – тетраэдр.

Неформальное описание симплексного метода: состоит из двух процедур – отражение и сжатие.

– *отражение*: симметричное отражение вершины с наибольшим значением $\varphi(x)$ относительно противоположной грани ["перекатывание симплекса"]. Если $\varphi(x_i') > \varphi(x_i)$, то выбирается другая $(i + 1)$ -я вершина.



Когда заикливание (все $(n + 1)$ -вершины перебрали), то



– *сжатие*: уменьшение размеров симплекса при сохранении вершины с наименьшим значением $\varphi(x)$, затем переход к отражению, и так далее, пока ребро симплекса не станет меньше некоторого числа: $\|x_i - x_j\| < \varepsilon$.

Достоинства: с большой вероятностью метод не распознает локальный минимум ("не остановится").

Локальные методы основаны на построении *релаксационной* последовательности $\{x_i\}$ такой, что $\varphi(x_i) \geq \varphi(x_{i+1})$ и $x_i \xrightarrow{i \rightarrow \infty} x^* = \arg \min \varphi(x)$.

Поэтому *релаксационные* методы называют также *методами спуска*.

Классификация релаксационных методов

С одной стороны,

- *одношаговые* методы: $x_{i+1}(x_i)$ – каждый шаг $(i + 1)$ зависит только от предыдущей точки x_i и значения функции $\varphi(x_i)$;
- *двухшаговые* методы: $x_{i+1}(x_i, x_{i-1})$ – зависимость от двух предыдущих точек;
- и т.д.;

С другой стороны,

- *методы нулевого порядка*: если используются только значения минимизируемой функции $\varphi(x)$;
- *методы первого порядка*: если используются только значение $\varphi(x)$ и $\varphi'(x)$;
- *методы второго порядка*: если используются значения $\varphi(x)$, $\varphi'(x)$ и $\varphi''(x)$;
- etc;

Градиентные методы (методы первого порядка)

Итак, будем рассматривать задачу:

$$\varphi(x) \rightarrow \min, x \in X \equiv R^n \text{ (безусловная минимизация),}$$

предполагая, что функция $\varphi(x)$ непрерывно дифференцируема на R^n , т.е. $\varphi(x) \in C^1(R^n)$.

По определению дифференцируемой функции

$$\varphi(x+h) - \varphi(x) = (\varphi'(x), h) + o(h), \quad (1)$$

$$\text{где } \lim_{\|h\| \rightarrow 0} o(h) \|h\|^{-1} = 0.$$

Если $\varphi'(x) \neq 0$, то при достаточно малых $\|h\|$ главная часть приращения для φ будет определяться дифференциалом функции $d\varphi(x) = (\varphi'(x), h)$. Оценим величину $d\varphi(x)$. Справедливо неравенство Коши-Буняковского:

$$-\|\varphi'(x)\| \cdot \|h\| \leq (\varphi'(x), h) \leq \|\varphi'(x)\| \cdot \|h\|,$$

причем, если $\varphi'(x) \neq 0$, то правое неравенство превращается в равенство, только при $h = -\alpha \varphi'(x)$, а левое только при $h = \alpha \varphi'(x)$, где $\alpha = \text{const} \geq 0$.

Отсюда ясно, что при $\varphi'(x) \neq 0$ направление *наибыстрейшего возрастания* функции $\varphi(x)$ в точке x совпадает с *направлением градиента* $\varphi(x)$, а направление *наибыстрейшего убывания* – с направлением *антиградиента* $-\varphi'(x)$.

Это свойство градиента лежит в основе ряда итерационных методов минимизации функций. Один из таких – *градиентный*. Он предполагает, как, впрочем, и все остальные итерационные методы, наличие априорной точки начального приближения.

Предположим, что начальная точка x_0 уже выбрана, тогда градиентный метод заключается в построении последовательности $\{x_k\}$ по правилу:

$$x_{k+1} = x_k - \alpha_k \varphi'(x_k), \quad \alpha_k > 0, \quad k = 0, 1, \dots \quad (2)$$

α_k – величина шага, x_k – направление спуска.

Если $\varphi'(x_k) \neq 0$, то шаг $\alpha_k > 0$ можно выбрать так, чтобы получить релаксационную последовательность: $\varphi(x_{k+1}) < \varphi(x_k)$. Действительно, подставляя (2) в (1), имеем:

$$\varphi(x_{k+1}) - \varphi(x_k) = \alpha_k \left[-\|\varphi'(x_k)\|^2 + o(\alpha_k) \cdot \alpha_k^{-1} \right] < 0,$$

при всех достаточно малых $\alpha_k > 0$.

Если $\varphi'(x_k) = 0$, то x_k – стационарная точка. В этом случае процесс (2) прекращается и проводятся дополнительные исследования поведения функции в окрестности точки x_k для выяснения того, достигается ли в точке x_k минимум функции $\varphi(x)$ или не достигается.

Существуют различные *способы выбора величины шага* α_k в методе (2). В зависимости от способа выбора α_k можно получить различные варианты градиентного метода.

Метод наискорейшего спуска

На луче $\{x \in R^n : x = x_k - \alpha \varphi'(x_k), \alpha \geq 0\}$, направленном по антиградиенту, введем функцию одной переменной

$$\psi(\alpha) = \varphi(x_k - \alpha \varphi'(x_k)), \alpha \geq 0$$

и определим α_k из условий

$$\alpha_k = \arg \min_{\alpha \geq 0} \varphi(x_k - \alpha \varphi'(x_k)).$$

Другими словами α_k выбирается так, чтобы $\varphi(x_{k+1})$ в заданном направлении была наименьшей для чего на любом шаге необходимо решать задачу одномерной минимизации функции $\psi(\alpha)$, например, с помощью $\psi'(\alpha) = 0$.

Пример. Рассмотрим задачу

$$\varphi(x) = x_1^2 + 2x_2^2 \rightarrow \min$$

$$\text{с начальной точкой } x^0 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \varphi(x^0) = 6.$$

Из общих соображений ясно, что $\varphi_{\min} = 0$ при $x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

1-й шаг:

$$\varphi'(x) = \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix}; \quad \varphi'(x^0) = \begin{pmatrix} 4 \\ 4 \end{pmatrix}$$

Ищем

$$x^1 = x^0 - \alpha \varphi'(x^0) = \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \alpha \begin{pmatrix} 4 \\ 4 \end{pmatrix} = \begin{pmatrix} 2-4\alpha \\ 1-4\alpha \end{pmatrix}.$$

Функция $\psi(\alpha)$ имеет следующий вид:

$$\psi(\alpha) = \varphi(x^1) = (2-4\alpha)^2 + 2(1-4\alpha)^2.$$

Решаем уравнение $\psi'(\alpha) = 0$, т.е.

$$2(2-4\alpha) \cdot (-4) + 4(1-4\alpha) \cdot (-4) = 0;$$

$$4-8\alpha+4-16\alpha=0; \Rightarrow 24\alpha=8 \Rightarrow \alpha=\frac{1}{3}; \Rightarrow x^1 = \begin{pmatrix} 2-\frac{4}{3} \\ 1-\frac{4}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}.$$

2-й шаг:

$$\varphi'(x^1) = \begin{pmatrix} \frac{4}{3} \\ -\frac{4}{3} \end{pmatrix}; \quad x^2 = x^1 - \alpha \varphi'(x^1) = \begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix} - \alpha \begin{pmatrix} \frac{4}{3} \\ -\frac{4}{3} \end{pmatrix};$$

$$\psi(\alpha) = \varphi(x^2) = \left(\frac{2}{3} - \frac{4}{3}\alpha\right)^2 + 2\left(-\frac{1}{3} + \frac{4}{3}\alpha\right)^2.$$

Решаем уравнение $\psi'(\alpha) = 0 \Rightarrow$

$$2\left(\frac{2}{3} - \frac{4}{3}\alpha\right) \cdot \left(-\frac{4}{3}\right) + 4\left(\frac{4}{3}\alpha - \frac{1}{3}\right) \cdot \left(\frac{4}{3}\right) = 0; \Rightarrow$$

$$-\frac{4}{3} + \frac{8}{3}\alpha + \frac{16}{3}\alpha - \frac{4}{3} = 0; \Rightarrow \frac{24}{3}\alpha = \frac{8}{3}; \Rightarrow \alpha = \frac{1}{3}; x^2 = \begin{pmatrix} \frac{2}{3} - \frac{4}{9} \\ -\frac{1}{3} + \frac{4}{9} \end{pmatrix} = \begin{pmatrix} \frac{2}{9} \\ \frac{1}{9} \end{pmatrix}.$$

3-й шаг:

$$\varphi'(x^2) = \begin{pmatrix} \frac{4}{9} \\ \frac{4}{9} \end{pmatrix}; x^3 = x^2 - \alpha \varphi'(x^2) = \begin{pmatrix} \frac{2}{9} - \alpha \cdot \frac{4}{9} \\ \frac{1}{9} - \alpha \cdot \frac{4}{9} \end{pmatrix}$$

$$\psi(\alpha) = \varphi(x^3) = \left(\frac{2}{9} - \frac{4}{9}\alpha\right)^2 + 2\left(\frac{1}{9} - \frac{4}{9}\alpha\right)^2$$

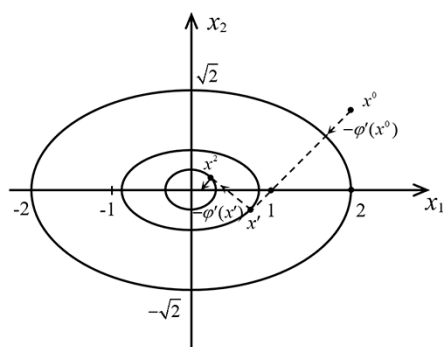
Решаем уравнение $\psi'(\alpha) = 0 \Rightarrow$

$$2\left(\frac{2}{9} - \frac{4}{9}\alpha\right) \cdot \left(-\frac{4}{9}\right) + 4\left(\frac{1}{9} - \frac{4}{9}\alpha\right) \cdot \left(-\frac{4}{9}\right) = 0;$$

$$\frac{4}{9} - \frac{8}{9}\alpha + \frac{4}{9} - \frac{16}{9}\alpha = 0; \Rightarrow \frac{8}{9} = \frac{24}{9}\alpha; \Rightarrow \alpha = \frac{1}{3}; x^3 = \begin{pmatrix} \frac{2}{9} - \frac{4}{27} \\ \frac{1}{9} - \frac{4}{27} \end{pmatrix} = \begin{pmatrix} \frac{2}{27} \\ -\frac{1}{27} \end{pmatrix}, \text{ и т.д.}$$

Представим решение задачи графически:

Из графического представления можно сделать вывод, что имеет место:



\Rightarrow а) сходимость к истинной точке минимума $x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

б) взаимная перпендикулярность градиентов

Свойства метода наискорейшего спуска

1. На любом шаге направление спуска меняется на ортогональное. Действительно, α_k ищется из условия $\psi'(\alpha) = 0 \Rightarrow$

$$\left. \frac{\partial \varphi(x_k - \alpha \varphi'(x_k))}{\partial \alpha} \right|_{\alpha=\alpha_k} = (\varphi'(x_k - \alpha_k \varphi'(x_k)), -\varphi'(x_k)) = -(\varphi'(x_{k+1}), \varphi'(x_k)) = 0$$

2. Точка x_{k+1} лежит на луче, исходящем из точки x_k и касательным к поверхности уровня $L\varphi(x_{k+1})$. Действительно, с одной стороны, несомненно, что $x_{k+1} \in L = \{x : \varphi(x) = \varphi(x_{k+1})\}$. С другой стороны, градиент $\varphi'(x_{k+1})$ ортогонален касательной к поверхности уровня $L\varphi(x_{k+1})$, поэтому по свойству 1 направление спуска касательно к поверхности $L\varphi(x_{k+1})$.

Иначе. $\varphi'(x_{k+1})$ ортогонален направлению спуска \Rightarrow луч, проходящий из точки x_k – касательной к поверхности $L = \{x : \varphi(x) = \varphi(x_{k+1})\}$.

Проблемы (общие для релаксационных методов).

- а) Имеет ли последовательность $\{x_k\}$ предел в смысле сходимости по норме: существует $\hat{x} : \lim_{k \rightarrow \infty} \|x_k - \hat{x}\| = 0$?
- б) Является ли этот предел аргументом, составляющим минимум функции φ $\hat{x} = \arg \min \varphi = x^*$?
- в) Какова скорость сходимости $\|x_k - x^*\|$ или $\varphi(x_k) - \varphi(x^*)$?
- г) Каковы вычислительные затраты.

Исследование метода наискорейшего спуска для квадратичной функции

Рассмотрим квадратичную функцию

$$\varphi(x) = \frac{1}{2}(Ax, x) - (b, x),$$

где A – симметричная, положительно определенная матрица.

Можно показать, что A – симметричная положительно определенная матрица $\Leftrightarrow \varphi$ – строго выпукла.

$\varphi'(x) = Ax - b$, т.е. $x^* = A^{-1} \cdot b$ – стационарная точка.

Попробуем записать метод наискорейшего спуска для квадратичной функции.

Итак,

$$\psi(\alpha) = \varphi(x_k - \alpha \varphi'(x_k)), \quad \alpha \geq 0$$

$$\psi(\alpha) = \varphi(x_k - \alpha(Ax_k - b)) = [\dots] = \varphi(x_k) - \alpha(Ax_k - b, Ax_k - b) + \frac{\alpha^2}{2}(A(Ax_k - b), Ax_k - b)$$

(w)

$$\psi'(\alpha) = -\|Ax_k - b\|^2 + \alpha(A(Ax_k - b), Ax_k - b) = 0 \Rightarrow \alpha_k = \frac{\|Ax_k - b\|^2}{(A(Ax_k - b), Ax_k - b)} > 0,$$

т.к. A – положительно определена, и значит для нее справедливо: $(Ah, h) > 0 \forall h \in R^n \neq 0$.

Для определения скорости сходимости оценим отношение

$$\frac{\varphi(x_{k+1}) - \varphi(x^*)}{\varphi(x_k) - \varphi(x^*)}$$

Имеем:

$$\varphi(x_{k+1}) = \psi(\alpha_k) = \varphi(x_k) - \frac{\|Ax_k - b\|^4}{(A(Ax_k - b), Ax_k - b)} + \frac{\|Ax_k - b\|^4}{2(A(Ax_k - b), Ax_k - b)} = \varphi(x_k) - \frac{\|\varphi'(x_k)\|^4}{2(A\varphi'(x_k), \varphi'(x_k))}$$

С другой стороны,

$$\varphi(x_k) - \varphi(x^*) = \frac{1}{2}(Ax_k, x_k) - \frac{1}{2}(Ax^*, x^*) - (b, x_k - x^*) = \frac{1}{2}(Ax_k - b, x_k - A^{-1}b) = \frac{1}{2}(A^{-1}\varphi'(x_k), \varphi'(x_k))$$

Для простоты дальнейших изложений предположим, что матрица A приведена к диагональному виду (т.е. выполнено преобразование координат) так, что $A = \text{diag}(\lambda_1, \dots, \lambda_n)$, где λ_i – собственные числа матрицы A .

- Собственные числа симметричной положительно определенной матрицы всегда положительны.
- Для симметричной матрицы существует ортогональная матрица $(T^T = T^{-1})$ T такая, что $T^T A T$ – диагональная матрица $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$.

Если $l = \min \lambda_i, L = \max \lambda_i$, то

$$\begin{aligned} (A\varphi'(x), \varphi'(x)) &\leq L \|\varphi'(x)\|^2 \\ (A^{-1}\varphi'(x), \varphi'(x)) &\leq \frac{1}{l} \|\varphi'(x)\|^2 \end{aligned}$$

Тогда

$$\frac{\varphi(x_{k+1}) - \varphi(x^*)}{\varphi(x_k) - \varphi(x^*)} = 1 - \frac{\|\varphi'(x_k)\|^4}{(A\varphi'(x_k), \varphi'(x_k))(A^{-1}\varphi'(x_k), \varphi'(x_k))} \leq 1 - \frac{l}{L} = \frac{L-l}{L}.$$

Если ввести обозначение $q \stackrel{\text{def}}{=} \frac{L-l}{L} < 1$, то

$$\varphi(x_k) - \varphi(x^*) \leq \text{const} \cdot q^k$$

Это называется *геометрической скоростью сходимости* (сходимость геометрической прогрессии).

Рассмотрим величину

$$\Delta_k \stackrel{\text{def}}{=} \|x_k - x^*\|.$$

Верхний предел $\overline{\lim}_{k \rightarrow \infty} \frac{\ln \Delta_{k+1}}{\ln \Delta_k}$ называется *порядком сходимости метода*.

В нашем случае квадратичной функции

$$\text{const} \cdot q^k \geq \varphi(x_k) - \varphi(x^*) = \frac{1}{2}(Ax_k - b, x_k - x^*) = \frac{1}{2}(A(x_k - x^*), x_k - x^*) \geq \frac{l}{2} \|x_k - x^*\|^2.$$

Поэтому

$$\begin{aligned} \|x_k - x^*\| &\leq \text{const} \cdot q^{\frac{k}{2}} \\ \Rightarrow \overline{\lim}_{k \rightarrow \infty} \frac{\ln \Delta_{k+1}}{\ln \Delta_k} &= \overline{\lim}_{k \rightarrow \infty} \frac{\frac{1}{2} \ln q + \ln \Delta_k}{\ln \Delta_k} = 1 \end{aligned}$$

\Rightarrow получили сходимость с порядком 1 или *линейную сходимость*. Бывает порядок больше 1 – *сверхлинейная сходимость*.

При исследовании метода наискорейшего спуска для квадратичной функции получили, в частности, следующие результаты:

- а) $\varphi(x_k) - \varphi(x^*) \leq \text{const} \cdot q^k, q < 1$
- б) $\Delta_k \stackrel{\text{def}}{=} \|x_k - x^*\|, \overline{\lim}_{k \rightarrow \infty} \frac{\ln \Delta_{k+1}}{\ln \Delta_k} = 1$

Определение.

Пусть $\varphi(x_k) \rightarrow \varphi(x^*)$ при $k \rightarrow \infty$.

Последовательность $\varphi(x_k)$ сходится к $\varphi(x^*)$ *линейно* (с линейной скоростью, со скоростью геометрической прогрессии), если существуют такие константы $q \in (0, 1)$ и k_0 , что

$$\|\varphi(x_{k+1}) - \varphi(x^*)\| \leq q \|\varphi(x_k) - \varphi(x^*)\|, \text{ при } k \geq k_0.$$

Последовательность $\varphi(x_k)$ сходится к $\varphi(x^*)$ *сверхлинейно*, если

$$\|\varphi(x_{k+1}) - \varphi(x^*)\| \leq q_{k+1} \|\varphi(x_k) - \varphi(x^*)\|, q_k \rightarrow 0+, \text{ при } k \rightarrow \infty.$$

Последовательность $\varphi(x_k)$ сходится к $\varphi(x^*)$ с *квадратичной скоростью*, если существуют такие константы $c \geq 0$ и k_0 , что

$$\|\varphi(x_{k+1}) - \varphi(x^*)\| \leq c \|\varphi(x_k) - \varphi(x^*)\|^2, \text{ при } k \geq k_0.$$

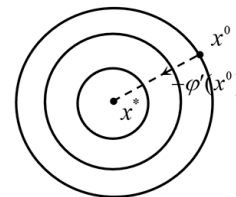
Вообще, порядок сходимости, равный 1, означает, что значение величины Δ_k убывает, в основном, по закону геометрической прогрессии. Порядок сходимости, равный 2 (квадратичная сходимость) означает, что при достаточно больших k $\Delta_{k+1} \sim \Delta_k^2$. В этом случае, если к тому же Δ_k – малая величина, например, $a \cdot 10^{-p}$ при $0.1 < a < 1$, то Δ_{k+1} равно $a^2 \cdot 10^{-2p}$, т.е. фактически удваивается число нулей после запятой.

Частные случаи:

- 1) Пусть $l = L$, т.е. матрица $A = LI = II$ – пропорциональна единичной окружности (линии уровня – окружности).

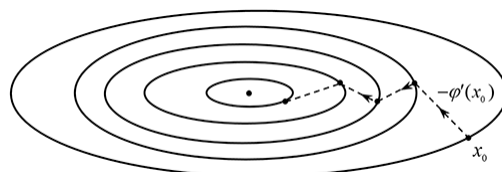
Тогда:

$$x_{k+1} = x_k - \frac{\|lx_k - b\|^2}{l\|lx_k - b\|^2} (lx_k - b) = \frac{b}{l} = x^*$$



$\Rightarrow f(x_{k+1}) = f(x^*)$ метод сходится за один шаг.

- 2) $l \leq L$: сходимость может быть еле заметной ($q \sim 1$), а графически это означает, что линии уровня функции сильно вытянуты и функция имеет так называемый "овражный" характер. Это означает, что небольшое изменение некоторых переменных приводит к резкому изменению значений функции – эта группа переменных характеризует "склон оврага", а по остальным переменным, задающим направление "дна оврага", функция меняется незначительно.

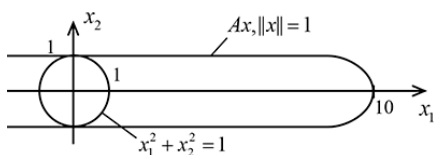


Число $cond \stackrel{def}{=} \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{L}{l}$ называется числом обусловленности матрицы $\Rightarrow cond \geq 1$.

Матрица называется *хорошо обусловленной*, если $cond \sim 1$ и наоборот.

Вообще, число обусловленности геометрически можно трактовать как меру искажения отображения матрицей A единичной сферы. Действительно, $cond(A)$ есть отношение наибольшего к наименьшим расстояниям между точками на единичной сфере после её отображения матрицей A . Чем больше $cond(A)$, тем больше искажение единичной сферы при её преобразовании в эллиптическую форму – пусть $A = diag(10, 1)$.

Вывод: Метод наискорейшего спуска быстро сходится для хорошо обусловленных матриц и наоборот.



Почему так много внимания уделяли квадратичной функции?

В окрестности Iosmin любую функцию можно приблизить квадратичной, и всё сказанное выше про матрицу A будет справедливым для матрицы Гесса $H(x^*)$, которая заменяет A в рассмотренном выше примере.



Геометрически: Линии уровня становятся замкнутыми и по мере приближения к x^* всё более напоминают эллипс.

Общий случай.

Определение 1. Функция φ на множестве $X \in R^n$ удовлетворяет условию Липшица, если существует $L > 0: \forall u, v \in X \quad \|\varphi(u) - \varphi(v)\| \leq L\|u - v\|$. Если градиент функции φ существует, непрерывен и удовлетворяет условию Липшица, то обозначается $\varphi \in C^{1,1}$.

Определение 2. Функция φ называется сильно выпуклой с параметром $\alpha < 0$, если $\forall u, v \in X, \varphi(u) \geq \varphi(v) + (\varphi'(v), u - v) + \alpha\|u - v\|^2$.

Теорема (о сходимости метода наискорейшего спуска). Рассмотрим задачу $\varphi(x) \rightarrow \min, x \in R^n$. Пусть $\varphi \in C^{1,1}(R^n)$ и φ – сильно выпуклая с параметром α . Тогда при любом начальном приближении для последовательности $\{x_k\}$, построенной по методу наискорейшего спуска, справедливы соотношения:

1) $x_k \rightarrow x^* = \arg \min \varphi(x)$

2) $\varphi(x_k) - \varphi(x^*) \leq \text{const} \cdot q^k, q = 1 - \frac{2\alpha}{L}, q \in [0, 1)$

Замечания.

1. Для квадратичной функции $\varphi(x) = \frac{1}{2}(Ax, x) - (b, x)$:

а) постоянная Липшица L есть наибольшее собственное число матрицы A :

$$\|\varphi'(u) - \varphi'(v)\| = \|Au - Av\| \leq L\|u - v\|;$$

б) она сильно выпукла с параметром $\frac{l}{2}$. Действительно,

$$\begin{aligned} \varphi(u) - \varphi(v) &= \frac{1}{2}(Au, u) - \frac{1}{2}(Av, v) - (b, u - v) = \\ &= \frac{1}{2}(A(u - v), u - v) + (\varphi'(v), u - v) \geq \frac{l}{2}\|u - v\|^2 + (\varphi'(v), u - v) \end{aligned}$$

2. Эквивалентные ограничения сильной выпуклости:

а) φ – сильно выпукла $\Leftrightarrow \xi(u) = \varphi(u) = \alpha\|u\|^2$ – выпукла (это означает, что φ имеет "квадратичный запас" выпуклости);

б) пусть $\varphi \in C^1$, φ – сильно выпукла $\Leftrightarrow (\varphi'(u) - \varphi'(v), u - v) \geq 2\alpha\|u - v\|^2$;

в) пусть $\varphi \in C^2$, φ – сильно выпукла $\Leftrightarrow (\varphi''(u)x, x) \geq 2\alpha\|x\|^2, \forall x$, т.е. $\varphi''(u) \geq 2\alpha I$ [в смысле положительной определенности разности матриц]. С другой стороны, из

условия Липшица $\varphi''(u) \leq LI$, поэтому для сильно выпуклой $\varphi \in C^2$ существует двойная оценка матрицы Гессе: $2\alpha I \leq \varphi''(u) \leq LI$.

Покажем, что $2\alpha \leq L$. С одной стороны, из б) имеем

$$(\varphi'(u) - \varphi'(v), u - v) \geq 2\alpha \|u - v\|^2$$

С другой стороны,

$$\|\varphi'(u) - \varphi'(v)\| \leq L \cdot \|u - v\|$$

$$\Rightarrow 2\alpha \|u - v\|^2 \leq (\varphi'(u) - \varphi'(v), u - v) \leq \|\varphi'(u) - \varphi'(v)\| \cdot \|u - v\| \leq L \cdot \|u - v\|^2 \Rightarrow 2\alpha \leq L, \text{ ч.т.д.}$$

Выпуклость:

- $\varphi(u) \geq \varphi(v) + (\varphi'(v), u - v)$.

Строгая выпуклость:

- $\varphi(u) > \varphi(v) + (\varphi'(v), u - v)$.

Сильная выпуклость:

- $\varphi(u) \geq \varphi(v) + (\varphi'(v), u - v) + \alpha \|u - v\|^2$ для $\forall u, v \in \mathbb{R}^n$

Графическое представление дифференциальных критериев выпуклости.

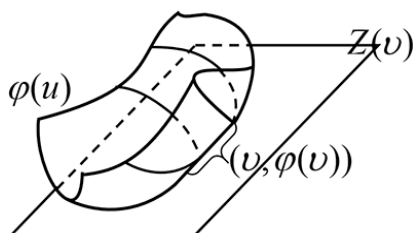


График *выпуклой* функции расположен не ниже касательной плоскости $Z = \varphi(v) + (\varphi'(v), u - v)$, проходящей через произв. точку поверхности $(v, \varphi(v))$.

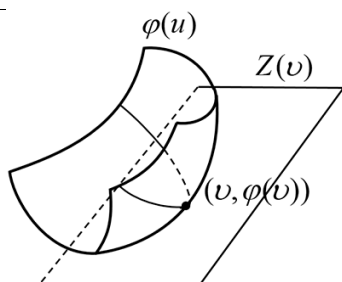
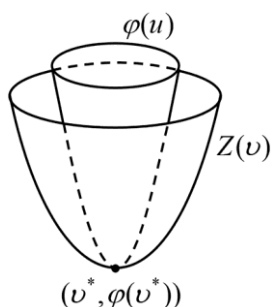


График *строго выпуклой* функции имеет единственную общую точку с этой плоскостью.



Пусть $v^* - (\cdot) \min \Rightarrow \varphi'(v^*) = 0 \Rightarrow \varphi(v^*) + \alpha \|u - v^*\|^2$.

Поверхность $Z = \varphi(v^*) + \frac{2\alpha}{2} \|u - v^*\|^2$ — это параболоид вращения с вершиной в точке $(v^*, \varphi(v^*))$.

\Rightarrow График *сильно выпуклой* функции расположен внутри некоторого параболоида вращения.

Другие градиентные методы

Напомним, градиентные методы заключаются в построении релаксационной последовательности:

$$\{x_k\}: x_{k+1} = x_k - \alpha_k \varphi'(x_k)$$

Градиентные методы различаются между собой способом выбора α_k .

1. Метод наискорейшего спуска, который был рассмотрен выше, заключается в выборе

$$\alpha_k = \arg \min \varphi(x_k - \alpha \varphi'(x_k)).$$

Такой способ выбора α_k является в некотором смысле наилучшим, т.к. он обеспечивает *достижение наименьшего значения функции* вдоль заданного направления. Однако он требует решения на любом шаге *одномерной задачи минимизации*. Эти задачи решаются, как правило, приближенно с помощью численных методов, что приводит к *значительному объему вычислений*. Кроме того, метод может привести к *плохой сходимости* (овраги!).

Другим подходом для построения релаксационной последовательности является попытка определить α_k до начала вычислений. Какие есть для этого основания?

Допустим, что можно построить оценку для α_k такую, что для $\varepsilon \in (0,1)$ выполняется неравенство

$$\varphi(x_{k+1}) - \varphi(x_k) \leq -\varepsilon \cdot \alpha_k \cdot \|\varphi'(x_k)\|^2 \quad (1)$$

Тогда очевидно, что $\varphi(x_{k+1}) < \varphi(x_k)$ и соответствующий метод минимизации будет методом спуска.

Справедливы следующие утверждения:

Лемма 1. Пусть функция $\varphi \in C^{1,1}(R^n)$ и

$$\|\varphi'(x) - \varphi'(x')\| \leq M \|x - x'\|, x, x' \in R^n, M > 0$$

Тогда для $\forall x_k \in R^n, \varepsilon \in (0,1)$ условие (1) выполнено при

$$0 < \alpha_k \leq \frac{1-\varepsilon}{M}$$

Лемма 2. Пусть φ дважды дифференцируема и матрица Гессе удовлетворяет условию Липшица и

$$(\varphi''(x)h, h) \leq D \cdot \|h\|^2, x, h \in R^n, D > 0$$

Тогда для $\forall x_k \in R^n, \varepsilon \in (0,1)$ условие (1) выполняется при

$$0 < \alpha_k \leq \frac{2(1-\varepsilon)}{D}$$

2. Градиентный метод с постоянным шагом.

В этом методе полагается $\alpha_k \equiv \text{const}$. При этом иногда удается добиться выполнения условия (1). Но для этого необходимо знать константы M и D , что далеко не всегда удается вычислить.

Т.о., метод прост в реализации, но есть проблемы со сходимостью.

Пример. Пусть $\varphi(x) = \alpha x^2$

$$\varphi_{\min} = 0; x^* = 0;$$

Тогда

$$x_{k+1} = x_k - \alpha \cdot 2\alpha x_k = (1 - 2\alpha\alpha) \cdot x_k \Rightarrow |1 - 2\alpha\alpha| < 1 \Leftrightarrow \text{метод сходится.}$$

Сходимость медленная!

3. Градиентный метод с убыванием длины шага.

В ряде методов достаточно потребовать выполнения условий:

$$\alpha_k > 0, k = 0, 1, \dots; \sum_{k=0}^{\infty} \alpha_k = \infty; \sum_{k=0}^{\infty} \alpha_k^2 < \infty \quad (\text{например, } \alpha_k = \frac{c}{k+1})$$

На интуитивном уровне объяснение следующее:

- условие сходимости ряда $\sum_{k=0}^{\infty} \alpha_k^2$ накладывают, чтобы добиться достаточно быстрой сходимости последовательности α_k к нулю с целью обеспечения сходимости метода в окрестности точки экстремума x^* .
- условие расходимости ряда $\sum_{k=0}^{\infty} \alpha_k$ призвано обеспечить достижение точки экстремума x^* даже при неудачном выборе начального приближения x^0 , т.е. при больших расстояниях от x^0 до x^* .

Сходимость медленная!

4. Градиентный метод с дроблением шага.

Ещё один адаптивный способ выбора коэффициентов α_k . Выбираются некоторые $\text{const} \beta > 0$ и $0 < \lambda < 1$ (обычно $\lambda = 1/2$). Для коэффициента $\alpha = \beta$ проверяется выполнение условия $\varphi(x_k - \alpha \varphi'(x_k)) \leq \varphi(x_k)$. Если оно выполняется, то полагают $\alpha_k = \alpha$. Если нет, то производится дробление шага, т.е. принимается $\alpha = \lambda \beta$, и т.д. до тех пор, пока не выполнится требуемое неравенство.

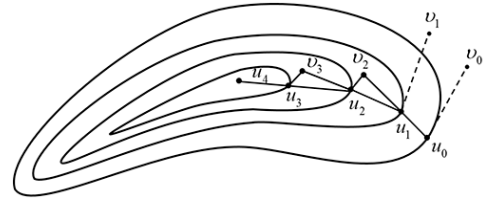
Процесс дробления не может продолжаться бесконечно, поскольку $-\varphi'(x)$ — направление убывания функции. Первое α , при котором условие выполнено и принимается за α_k .

Как показывает следующая лемма, с помощью описанного процесса дробления шага можно добиться выполнения неравенства. (1)

Лемма 3. Пусть функция φ дифференцируема на R^n .

Тогда для $\forall x_k \in R^n, \varepsilon \in (0,1)$ найдется такое $\alpha_0 > 0$,
что при $\forall \alpha \in (0, \alpha_0]$ выполнено условие

$$\varphi(x_k - \alpha \varphi'(x_k)) - \varphi(x_k) \leq -\varepsilon \alpha \|\varphi'(x_k)\|^2.$$



Если необходимое неравенство оказывается выполненным при начальном значении $\alpha = \beta$, то иногда полезно увеличить шаг, взяв $\alpha = \mu\beta$, где $\mu > 1$. Так можно продолжать до тех пор, пока значения функции не перестанут уменьшаться. Последнее α , при котором произошло уменьшение, и берется в этом случае за α_k .

5. Овражный метод – эвристический двухшаговый метод минимизации овражных функций.

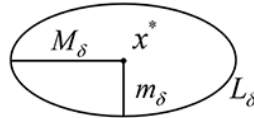
Характеристика степени овражности:

Пусть x^* – точка минимума, $\delta > 0$.

Рассмотрим поверхность

$$L_\delta = \{x : \varphi(x) = \varphi(x^*) + \delta\}.$$

Введем $m_\delta = \min_{x \in L_\delta} \|x - x^*\|$



$$M_\delta = \max_{x \in L_\delta} \|x - x^*\|$$

Определение. Тогда $r = \lim_{\delta \rightarrow 0} \frac{M_\delta}{m_\delta}$ называется *числом обусловленности точки $locmin$* .

Рассмотрим "овражную" функцию (вытянута вдоль некоторых направлений). Если точка лежит на склоне оврага, то направление спуска из этой точки будет почти перпендикулярно к направлению "дна оврага", и в результате приближения $\{u_k\}$, получаемые градиентным методом, будут *поочередно находиться то на одном, то на другом "склоне оврага"*. Если "склоны оврага" достаточно круты, то такие скачки "со склона на склон" точек $\{u_k\}$ могут сильно замедлить сходимость градиентного метода.



Для ускорения сходимости можно предложить следующий эвристический прием, называемый *овражным методом*:

Пусть v_0, v_1 – две произвольные близкие точки. Совершаем из них по одному шагу методом наискорейшего спуска (или \forall вариант градиентного метода).

Попадаем в окрестность "дна оврага". Соединяя их прямой, делаем большой шаг в полученном направлении, перемещаясь вдоль "дна оврага". Из полученной точки v_2 , которая находится на "склоне оврага", производят спуск с помощью градиентного метода и определяют следующую точку u_2 на "дне оврага" и т.д.

Формула метода выглядит следующим образом.

$$v_{k+1} = u_k - \text{sign}(\varphi(u_k) - \varphi(u_{k-1})) \cdot \frac{h}{\|u_k - u_{k-1}\|} \cdot (u_k - u_{k-1})$$

Здесь:

$\text{sign}(\varphi(u_k) - \varphi(u_{k-1}))$ определяет знак - чтобы спускаться, а не подниматься;

$(u_k - u_{k-1})$ - определяет направление спуска по дну оврага;

h - овражный шаг, выбирается эмпирически и от него многое зависит.

Если h – большое, то на крутых склонах точки v_k могут *слишком далеко удаляться от "дна оврага"* \Rightarrow большие объемы вычислений для градиентного метода спуска в очередную точку на "дне оврага", кроме этого может произойти выброс точки u_k из "оврага" и *правильное направление поиска будет потеряно*.

Если h – малое, то эффект от применения овражного метода может быть незначительным.

Эффективность применения овражного метода может *резко возрасти*, если величину h выбирать переменной, реагирующей на "повороты" оврага, с тем, чтобы:

- а) быстрее проходить прямолинейные участки на "дне оврага" за счет увеличения овражного шага;
- б) на крутых поворотах "оврага" избежать выброса из "оврага" за счет уменьшения овражного шага;
- в) добиться \min отклонения точки u_k от дна оврага с целью уменьшения объема вычислений для градиентного метода.

Для правильной реакции на "повороты" оврага надо учитывать "кривизну" оврага.

Причем информацию о кривизне желательно получить по результатам предыдущих шагов.

Один из способов выбора шага:

$$h_{k+1} = h_k \cdot c^{\cos \alpha_k - \cos \alpha_{k-1}}, k = 2, 3, \dots,$$

где α_k – угол между векторами $v_k - u_{k-1}$, $u_k - u_{k-1}$, определяемый условием

$$\cos \alpha_k = \frac{(v_k - u_{k-1}, u_k - u_{k-1})}{\|v_k - u_{k-1}\| \cdot \|u_k - u_{k-1}\|},$$

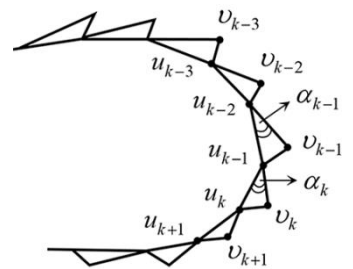
$c - \text{const} > 1$ – параметр алгоритма.

α_k возрастает при возрастании кривизны \Rightarrow при переходе от участка с меньшей кривизной на участок с большей кривизной имеем

$$\cos \alpha_k - \cos \alpha_{k-1} < 0 \Rightarrow h_{k+1} < h_k \text{ и наоборот.}$$

На участках с постоянной кривизной

$$\cos \alpha_k - \cos \alpha_{k-1} \square 0$$



\Rightarrow шаг остается постоянным, который был сформирован при выходе на рассматриваемый участок.

Параметр c регулирует чувствительность "метода к изменению кривизны (повороты)" и во многом определяет скорость движения "по оврагу".

Методы II порядка минимизации функции (использование вторых производных)

Общая идея:

Последовательность $\{x_k\}$ будем строить по

$$x_{k+1} = x_k - \gamma_k H_k \varphi'(x_k),$$

где γ_k – длина шага, H_k – матрица поворота ($n \times m$).

Как выбрать матрицу H_k ?

Если взять квадратичную функцию

$$\varphi(x) = \frac{1}{2}(Ax, x) - (b, x), \varphi'(x) = Ax - b, x^* = A^{-1}b,$$



то хочется сразу попасть в экстремальную точку:

$$x_{k+1} = x_k - (x_k - A^{-1}b) = x_k - A^{-1}(Ax_k - b) = x_k - A^{-1} \cdot \varphi'(x_k) \Rightarrow x_{k+1} = x_k - A^{-1} \cdot \varphi'(x_k)$$

\Rightarrow в качестве "матрицы доворота" надо брать $\gamma_k H_k = A^{-1}$.

В общем случае: пусть φ – дважды дифференцируема в R^n , разложим $\varphi(x)$ в ряд Тейлора в точке x_k :

$$\varphi(x) = \varphi(x_k) + (\varphi'(x_k), x - x_k) + \frac{1}{2}(\varphi''(x_k)(x - x_k), (x - x_k)) + o(x - x_k)^2.$$

Иначе формулу можно представить в виде:

$$\varphi(x) = \bar{\varphi}(x) + o(\|x - x_k\|^2), \text{ где } \bar{\varphi}(x) - \text{квадратичная функция.}$$

Пренебрегаем $o(\|x - x_k\|^2)$ и ищем $\min \bar{\varphi}(x)$.

$$x_{k+1} = \arg \min_x \bar{\varphi}(x) \Rightarrow \bar{\varphi}'(x) = \varphi'(x_k) + \varphi''(x_k)(x - x_k) = 0;$$

Пусть $\varphi''(x_k)$ – положительно определена для $\forall x_k \in R^n \Rightarrow$ существует $[\varphi''(x_k)^{-1}]$.

Решая это уравнение, получим:

$$x_{k+1} = x_k - [\varphi''(x_k)]^{-1} \text{grad} \varphi(x_k) - \text{это и есть метод Ньютона.}$$

Квадратичная функция с положительно определенной φ'' сильно выпукла, тогда уравнение определяет единственную точку глобального минимума $\bar{\varphi}(x)$.

Далее рассмотрим пример использования метода Ньютона для решения задачи минимизации функции.

Пример.

$$\varphi(x) = x_1^3 + x_2^3 - 6x_1x_2 \rightarrow \min, x \in R^2$$

$$\varphi'(x) = \begin{pmatrix} 3x_1^2 - 6x_2 \\ 3x_2^2 - 6x_1 \end{pmatrix}; \quad H(x) = \begin{pmatrix} 6x_1 & -6 \\ -6 & 6x_2 \end{pmatrix}$$

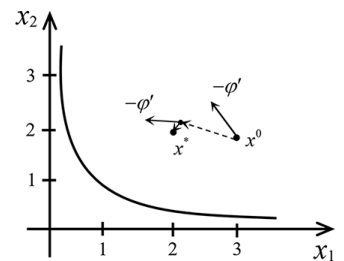
Область существования $(\varphi''(x))^{-1}$ совпадает с областью положительной определенности матрицы $\varphi''(x)$, которую мы будем искать с помощью критерия Сильвестра.

Критерий Сильвестра.

Симметрическая матрица A положительно определена \Leftrightarrow если все её главные миноры положительны. При этом главным минором матрицы A называется определитель матрицы, построенной из элементов матрицы A , стоящих на пересечении строк и столбцов с одинаковыми номерами.

Возьмем $x^0 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$

$$\begin{cases} x_1 > 0 \\ 36x_1x_2 - 32 > 0 \end{cases} \Leftrightarrow \begin{cases} x_1 > 0 \\ x_1x_2 > 1 \end{cases}$$



$$\varphi'(x^0) = \begin{pmatrix} 15 \\ -6 \end{pmatrix}; \quad H(x^0) = 6 \begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}$$

$$H^{-1}(x^0) = \frac{1}{30} \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$$

$$x^1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix} - \frac{1}{30} \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 15 \\ -6 \end{pmatrix} = \begin{pmatrix} 2,2 \\ 2,1 \end{pmatrix}$$

$$\varphi'(x^1) = \begin{pmatrix} 1,92 \\ 0,03 \end{pmatrix}; \quad H(x^1) = 6 \begin{pmatrix} 2,2 & -1 \\ -1 & 2,1 \end{pmatrix}; \quad H^{-1}(x^1) = \frac{1}{21,72} \begin{pmatrix} 2,1 & 1 \\ 1 & 2,2 \end{pmatrix}$$

$$x^2 = \begin{pmatrix} 2,2 \\ 2,1 \end{pmatrix} - \frac{1}{21,72} \begin{pmatrix} 2,1 & 1 \\ 1 & 2,2 \end{pmatrix} \begin{pmatrix} 1,92 \\ 0,03 \end{pmatrix} = \dots \approx \begin{pmatrix} 2,04 \\ 2,01 \end{pmatrix}, \text{ и т.д.}$$

Можно показать, что сходимость при $x^0 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ будет хуже.

Достоинства метода Ньютона:

1. Для квадратичной функции сходится за один шаг (метод Ньютона можно рассмотреть, как градиентный метод с преобразованием координат [умножение на H^{-1}] таким, что исчезает "овраг", т.е. линии уровня становятся окружностями).
2. Высокая скорость сходимости. Можно показать, что

$$\|x_k - x^*\| \leq \text{const} q^{2^k}, \quad 0 < q < 1.$$

Порядок сходимости:

$$\lim_{k \rightarrow \infty} \frac{\ln \|x_{k+1} - x^*\|}{\ln \|x_k - x^*\|} = \lim_{k \rightarrow \infty} \frac{\text{const} + 2^{k+1} \ln q}{\text{const} + 2^k \ln q} = 2$$

\Rightarrow важен выбор q в алгоритме. Если $q = 10^{-1}$, то за один шаг точность результата увеличивается на 2 разряда, а при линейной сходимости – на один разряд.

Недостатки метода Ньютона:

1. Локальная сходимость (матрица Гессе должна быть невырожденной). Начальное приближение надо выбирать в окрестности точки локального минимума.
2. Большие вычислительные затраты:
 - вычисление матрицы φ'' ;
 - необходимость обращаться её.
 -

Общие рекомендации: сначала применять градиентный метод, затем – метод Ньютона.

Например, существует так называемый метод Марквардта-Левенберга:

$$x_{k+1} = x_k - \gamma_k (\varphi''(x_k) + \alpha_k I)^{-1} \varphi'(x_k), \alpha_k > 0.$$

При больших α_k (вдали от x^*) матрица $\varphi''(x) + \alpha_k I \sim I$, при $\alpha_k \rightarrow \infty$ и это фактически градиентный метод.

При малых α_k (вдали от x^*) это метод Ньютона.

Имеет место:

Теорема (о сходимости метода Ньютона).

Пусть

- 1) φ – сильно выпукла на R^n с параметром ε .
- 2) $\varphi \in C^{2,2}$, т.е. φ'' – дважды дифференцируема и φ'' удовлетворяет условию Липшица:

$$\exists L > 0: \forall u, v \in R^n \|\varphi''(u) - \varphi''(v)\| \leq L \|u - v\|;$$

- 3) Начальное приближение x^0 удовлетворяет условию

$$\|\varphi'(x^0)\| \leq \frac{8\varepsilon^2}{L},$$

$$\text{т.е. } \|\varphi'(x^0)\| = \frac{8\varepsilon^2}{L} \cdot q, \text{ где } 0 < q < 1.$$

Тогда последовательность $x_{k+1} = x_k - [\varphi''(x_k)]^{-1} \varphi'(x_k)$ сходится к точке минимума x^* с квадратичной скоростью:

$$\|x_k - x^*\| \leq \frac{4\varepsilon}{L} q^{2^k}, x^* = \arg \min_x \varphi(x) \text{ (квадратичная сходимость)}$$

1. Несколько слов о норме матрицы:

Определение. Норму $(n \times n)$ -матрицы определим следующим образом:

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|},$$

Тогда

$$\|A\|^2 = \max_{x \neq 0} \frac{x^T A^T A x}{x^T x}.$$

Поскольку $A^T A$ есть симметричная $(n \times n)$ -матрица, то существует ортогональная матрица $T (T^T = T^{-1})$ такая, что $T^T (A^T A) T$ – диагональная матрица $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) \Rightarrow$

$\|A\| = \lambda_1$, где λ_1^2 – наибольшее собственное значение матрицы $A^T A$.

$\|A^{-1}\| = \lambda_n^{-1}$, λ_n^2 – наименьшее собственное значение матрицы $A^T A$.

Для такой нормы выполняются все три условия

- 1) $\|A\| > 0$, если A – ненулевая (покомпонентно);
- 2) $\|A + B\| \leq \|A\| + \|B\|$;
- 3) $\|\alpha A\| = |\alpha| \cdot \|A\|$, где α – скаляр.

Кроме того, из определения нормы матрицы следует, что

$$\|Ax\| \leq \|A\| \cdot \|x\|.$$

Имеем также

$$\|AB\| \leq \|A\| \cdot \|B\|.$$

2. Отметим ещё раз, что для сходимости метода Ньютона начальная точка x^0 должна выбираться достаточной близкой к искомой точке x^* . Это требование в теореме выражено условием 3). Действительно, сильная выпуклость φ означает:

$$2\varphi \|x^0 - x^*\|^2 \leq (\varphi'(x^0) - \varphi'(x^*), x^0 - x^*) \leq \|\varphi'(x^0)\| \cdot \|x^0 - x^*\| = \frac{8\varphi^2}{L} q \|x^0 - x^*\|; \|x^0 - x^*\| \leq \frac{4\varphi}{L} q;$$

\Rightarrow чем меньше q , тем ближе надо выбирать точку x^0 к x^* и тем быстрее сходимость.

Достоинства и недостатки градиентных методов и метода Ньютона

Метод	Достоинства	Недостатки
Градиентный метод	1. Глобальная сходимость, т.е. слабые требования на начальные приближения точки x_0 и к $f(x)$; 2. Относительная простота вычислений	1. Медленная сходимость (геометрическая скорость сходимости, порядок сходимости $d=1$); 2. Необходимость вычисления длины шага.

Метод Ньютона	1. Быстрая сходимость.	1. Локальная сходимость (начальная точка должна быть близка к x^*); 2. Большой объем вычислений (на любом шаге требуется вычислять матрицу вторых производных и обращать её). 3. Жесткие требования на саму функцию (непрерывная вторая производная).
---------------	------------------------	--

Порядок применения методов

- 1) На 1-м этапе – методы 1-ого порядка, т.к. они обеспечивают глобальную сходимость.
- 2) На 2-м этапе (когда приращения невелики) – выгодно применять методы 2-ого порядка.

Перечисленные методы (градиентные и Ньютона) являются классическими.

Можно предложить методы более высокого порядка, тогда естественно ожидать, что порядок сходимости будет равен p .

Как уже отмечалось, к недостаткам метода Ньютона относятся:

- локальная сходимость;
- большой объем вычислений;
- жесткие требования на гладкость функции.

В силу названных причин применение классического метода Ньютона не всегда приводит к успеху.

Многочисленные модификации направлены на то, чтобы, *сохраняя* основное достоинство метода Ньютона – *быструю сходимость*, *уменьшить трудоемкость и ослабить* требования на выбор *начального приближения*.

Метод Ньютона с регулировкой шага

Рассмотрим метод

$$x_{k+1} = x_k + \alpha_k h_k, \alpha_k > 0, h_k = -(\varphi''(x_k))^{-1} \varphi'(x_k),$$

это метод Ньютона с регулировкой шага.

При $\alpha_k \equiv 1$ мы получили классический метод Ньютона.

Выбор α_k обычно – из условия \min функции, вдоль заданного направления, или методом дробления шага, обеспечивающего выполнение условия $\varphi(x_{k+1}) < \varphi(x_k)$.

Можно показать, что подобные методы регулировки шага *сходятся* при *любой начальной* точке $x^0 \in R^n$, причем скорость сходимости будет либо *сверхлинейна*, либо *квадратичная* в зависимости от требований, которым удовлетворяет функция φ .

Таким образом, с помощью регулировки длины шага преодолевается недостаток метода, связанный с необходимостью отыскания хорошего начального приближения.

Однако, трудоемкость вычислений при этом не исчезает. Более перспективным в этом плане оказывается другой подход, при котором строится аппроксимация матрицы $(\varphi''(x_k))^{-1}$ на основе информации о значениях градиентов $\varphi'(x_k), \varphi'(x_{k+1}), \dots$

Квазиньютоновы методы

Пусть функция φ дважды дифференцируема. Рассмотрим метод

$$x_{k+1} = x_k - \alpha_k H_k \cdot \varphi'(x_k) \quad (1)$$

α_k – шаг, H_k – матрица.

- а) Если H_k – единичная, имеем градиентный метод.
- б) Если $H_k = (\varphi''(x_k))^{-1}$, то это метод Ньютона (с точностью до шага).
- в) Если $H_k \approx H_k(\varphi'(x_i), i=1, 2, \dots, k) \approx (\varphi''(x_k))^{-1}$, то имеем метод, который объединяет достоинства обоих методов.

Заметим, что

$$\varphi'(x_k) - \varphi'(x_{k+1}) = \varphi''(x_{k+1})(x_k - x_{k+1}) + o(\|x_k - x_{k+1}\|).$$

Полагая невырожденной матрицу $\varphi''(x_{k+1})$, отсюда с точностью до членов более высокого порядка малости по сравнению с $\|x_k - x_{k+1}\|$ имеем:

$$\varphi''(x_{k+1})^{-1}(\varphi'(x_{k+1}) - \varphi'(x_k)) \approx x_{k+1} - x_k.$$

Рассмотрим квадратичную функцию $\varphi(x) = \frac{1}{2}(Ax, x) - (b, x)$. Для нее

$$\varphi'(x_{k+1}) = A(x_{k+1}) - b, \quad \varphi''(x_k) = A,$$

и приближенное равенство обращается в точное:

$$\varphi''(x_{k+1})^{-1}(\varphi'(x_{k+1}) - \varphi'(x_k)) = x_{k+1} - x_k.$$

Поэтому естественно потребовать, чтобы для матрицы H_{k+1} , приближающей $(\varphi''(x_{k+1}))^{-1}$, выполнялось условие:

$$H_{k+1}(\varphi'(x_{k+1}) - \varphi'(x_k)) = x_{k+1} - x_k \quad (*)$$

Это условие называется *квазиньютоновским*. Оно лежит в основе целого ряда методов аппроксимации $(\varphi'')^{-1}$. Соответствующие методы минимизации, для которых на любом шаге выполняется квазиньютоновское условие, также называются квазиньютоновскими.

Пусть приближения к $(\varphi'')^{-1}$ пересчитываются шаг от шага по формуле $H_{k+1} = H_k + \Delta H_k$.

Различные квазиньютоновские методы различаются способом вычисления "добавки" ΔH_k таким образом, чтобы удовлетворялось соотношение (*).

1. Метод Дэвидона-Флетчера-Пауэлла

Обозначим:

$$\left. \begin{aligned} q_k &= \varphi'(x_{k+1}) - \varphi'(x_k) \\ r_k &= x_{k+1} - x_k \end{aligned} \right\} (*) \Rightarrow H_{k+1} \cdot q_k = r_k$$

Метод заключается в построении релаксационной последовательности по следующему правилу:

$$\begin{aligned} x_{k+1} &= x_k - \alpha_k H_k \varphi'(x_k) \\ H_{k+1} &= H_k + \underbrace{\left[\frac{r_k \cdot r_k^T}{(r_k, q_k)} - \frac{(H_k q_k)(H_k q_k)^T}{(H_k q_k, q_k)} \right]}_{\Delta H_k} \end{aligned} \quad (1.1)$$

Длина шага α_k в квазиньютоновых методах выбирается так же, как в методе наискорейшего спуска:

$$\alpha_k = \arg \min_{\alpha \geq 0} \varphi(x_k - \alpha H_k \varphi'(x_k))$$

Как правило, начальное значение $H_0 = I$. Вообще, если H_0 – симметричная матрица, то H_k – симметричная матрица для любого k .

2. Метод Бroyдена-Флетчера-Шанно

Имеем

$$(H_{k+1})^{-1} r_k = q_k.$$

Если поставить задачу уточнять обратную матрицу, т.е. $G_k = (H_k)^{-1}, G_{k+1} = G_k + \Delta G_k$ тогда:

$$H_{k+1} = H_k + \left[1 + \frac{(q_k, H_k q_k)}{(r_k, q_k)} \right] \cdot \frac{r_k r_k^T}{(r_k, q_k)} - \frac{r_k q_k^T H_k + H_k q_k r_k^T}{(r_k, q_k)} \quad (1.2)$$

(этот метод более устойчив к ошибкам округления)

Можно доказать, что для квадратичной функции $\varphi(x) = \frac{1}{2}(Ax, x) + (b, x)$, где A – симметричная, положительно определенная матрица, оба метода (1.1) и (1.2) при любом начальном приближении $x_0 \in R^n$ генерируют одну и ту же последовательность точек x_1, x_2, \dots, x_n , причем

$$H_n = (\varphi''(x_n))^{-1} = A^{-1}, \quad x_n = x^* = -A^{-1}b = \arg \min_{x \in R^n} \varphi(x),$$

т.е. квазиньютоновские методы позволяют найти \min квадратичной функции за n -шагов.

Для неквадратичной функции, это не так. Однако можно показать, что при соответствующих предположениях $H_k - (\varphi''(x_k))^{-1} \rightarrow 0, x_n \rightarrow x$, причем скорость сходимости сверхлинейна.

Так, например, пусть φ – дважды непрерывно дифференцируемая функция, сильно выпукла на R^n .

Тогда при любом начальном приближении $x_0 \in R^n$ последовательность точек $\{x_k\}$, определяемая формулами (1.1.) и (1.2), сходится к x^* .

Если, при этом, для всех x : $\varphi(x) \leq \varphi(x_0)$, справедливы неравенства

$$\left\| \frac{\partial^2 \varphi}{\partial x_i \partial x_j}(x) - \frac{\partial^2 \varphi}{\partial x_i \partial x_j}(x^*) \right\| \leq M \|x - x^*\|, \quad i, j = 1, \dots, n,$$

то x_k сходится к x^* сверхлинейно

$$\left(\|x_{k+1} - x^*\| \leq q_{k+1} \|x_k - x^*\|, q_k \rightarrow 0+, k \rightarrow \infty \right).$$

Замечания о квазиньютоновских методах:

- 1) Это двухшаговые методы.
- 2) Для квадратичных функций сходятся за n -шагов.
- 3) Обладают следующими преимуществами:
 - небольшая вычислительная сложность;
 - более глобальная сходимость, чем в методе Ньютона;
 - сверхлинейная скорость сходимости.