Численные методы оптимизации

Список литературы:

1. *Банди Б.* Основы линейного программирования. - М.: Радио и связь, 1989. - 51 с.
2. *Бейко И., Бублик Б., Зинько П.* Методы и алгоритмы решения задач оптимизации. - К.: Вища шк., 1983. - 512 с.
3. *Васильев Ф.* Численные методы решения экстремальных задач. - М.: Наука, 1988. - 522 с
4. *Гилл Ф., Мюррей У., Райт М*. Практическая оптимизация. - М.: Мир, 1985. - 481 с.
5. *Пантелеев А., Летова Т.* Методы оптимизации в примерах и задачах. - М.: Высш.шк., 2002. -544 с.
6. *Поляк Б.*, Введение в оптимизацию. - М.: Наука, 1983 — 384 с.
7. *Пшеничный Б.* Выпуклый анализ т экстремальные задачи. - М.: Наука, 1984 — 320 с.
8. *Пшеничный Б., Данилин Ю.* Чсленные методы в экстремальных задачах. - М.: Наука, 1975. 319 с.
9. *Сухарев А., Тимохов А., Федоров В.* Курс методов оптимизации. - М.: Наука, 1986. - 326 с.
10. *Химмельбрау Д.* Прикладное нелинейное программирование. - М.: Мир, 1975. - 398 с.

Понятие о численных методах оптимизации.

Рассмотрим оптимизационную задачу:

$$\begin{cases} f(x) \to min \\ x \in X \end{cases}, \qquad \textbf{(1)}$$

Иногда удается, опираясь на условия оптимальности или на геометрическую интерпретацию, получить решение задачи (1) в явном виде. Но в большинстве случаев задачу (1) приходится решать численно, с помощью компьютера.

При этом можно свести задачу оптимизации к некоторой другой задаче, а затем использовать разработанные для нее численные методы. Так, например, для решения задачи минимизации на R^n дифференцируемой функции f можно воспользоваться каким-либо численным методом решения системы уравнений $f'(x)=0$.

Однако, как правило, наиболее эффективными являются методы, разработанные специально для задач оптимизации.

Мы будем рассматривать численные методы для решения безусловных и условных многомерных задач.

Любой численный метод решения задачи оптимизации основан на точном или приближенном вычислении ее характеристик:

- значений целевой функции ;
- значений функций, задающих допустимое множество;
- значений производных функции.

На основании полученной информации строится приближение к решению задачи - искомой точке x* или, если такая точка не единственная, к множеству точек минимума. Иногда строится приближение к минимальному значению целевой функции $f^*=\min\limits_{x\in X} f(x)$.

Какие характеристики следует выбирать для вычислений, решается в зависимости от свойств минимизируемой функции, ограничений и имеющихся возможностей по хранению и обработке информации. Так для минимизации не дифференцируемой функции нельзя использовать алгоритм, в котором нужно вычислять в произвольной точке градиент функции; для машины с малым объемом памяти нельзя использовать алгоритм, требующий вычисления на каждом шаге и хранения в памяти матрицы вторых производных.

Алгоритмы, использующие лишь информацию о значениях минимизируемой функции, называются алгоритмами нулевого порядка; алгоритмы, использующие также информацию о значениях первых производных - алгоритмами 1-го порядка, о вторых производных- второго порядка.

Работа алгоритма состоит из двух этапов:

на 1-м этапе вычисляются предусмотренные алгоритмом характеристики задачи;

на 2-м этапе по полученной информации строится приближение к решению.

Для задач оптимизации на втором этапе выбор способа построения приближения не вызывает затруднений: например, для метода спуска, в котором на каждом шаге происходит переход в точку с меньшим, чем предыдущее, значением функции $f(x^{k+1})<f(x^k)$, за приближение к точке минимума обычно выбирается точка последнего вычисления. Поэтому для задания алгоритма достаточно указать способ выбора точек вычисления.

Если все точки выбираются одновременно до начала вычислений, тогда такой алгоритм называется **пассивным**.

Однако для решения большинства задач точки вычисления выбираются поочередно, т.е. точка x^{i+1} выбирается тогда, когда уже выбраны точки предыдущих вычислений $x^1,\dots,x^i$ и в каждой из них произведены предусмотренные алгоритмом вычисления, результаты которых обозначаются через $y_1,\dots,y_i$. Такие алгоритмы называются **последовательными**. Таким образом последовательный алгоритм определяется точкой $x^1\in X$ и набором отображений вида:

$$\overline{x}^{i+1}:\{\overline{x}^1,\dots,\overline{x}^i,y_1,\dots,y_i\}\to X,\, i\geq 1$$

при этом

$$\overline{x}^{i+1}=\overline{x}^{i+1}(\overline{x}^1,\dots,\overline{x}^i,y_1,\dots,y_i)$$

На практике обычно выбор точки очередного вычисления зависит лишь от точки предыдущего вычисления и от полученного результата:

$$\overline{x}^{i+1}=\overline{x}(\overline{x}^i,y^i)=\overline{x}(\overline{x}^i,l_1y_1+\dots+l_iy_i)\ .$$

Правило получения точки x^{k+1} из точки x^k называется <u>итерацией</u> алгоритма или <u>шагом</u>.

Итерацию любого алгоритма для решения задачи (1) можно записать в виде :

$$x^{k+1}=x^k+\alpha_k h^k, \alpha_k \in R, k=0,1,2,3,... \qquad \textbf{(2)}$$

При этом конкретный алгоритм определяется заданием точки x^0, правилом выбора векторов h^k и чисел α_k на основании полученной в результате вычислений информации, а также условии остановки.

Вектор h^k определяет направление (к+1)-го шага метода оптимизации, а коэффициент α_k - длину этого шага. Обычно название метода минимизации определяется способом выбора h^k, а его различные варианты связываются с различными способами выбора α_k.

Трудность, решения задачи оптимизации определяется числом ее переменных, видом целевой функции, видом и числом ограничений. Часто методы разработанные для одного вида задач, оказываются полезными для решения более сложных задач (так, например, алгоритмы **одномерной** оптимизации применяются при решении **многомерных**; многие алгоритмы **условной** оптимизации используют алгоритмы **безусловной** оптимизации или являются модификацией; методы решения задач линейного программирования используются для нелинейных задач).

Конечношаговыми называются методы, гарантирующие отыскание решения за конечное число шагов. Конечношаговые методы удается построить лишь для некоторых специальных задач оптимизации - например, для задач **линейного** и **квадратичного** программирования.

Для **бесконечношаговых** методов достижение решения гарантируется лишь в пределе.

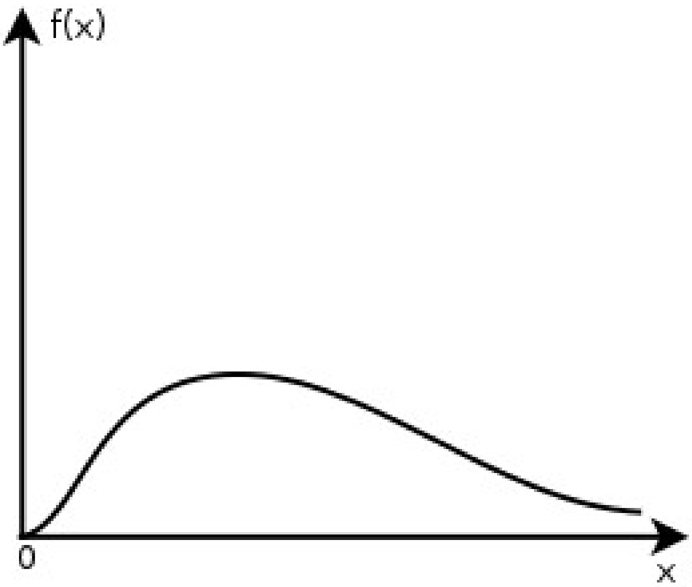
<u>Сходимость методов оптимизации</u>

Будем говорить, что <u>метод (2) сходится</u>, если:

$x^k \to x^*$ при $k \to \infty$, **(3)**

где x^* - решение задачи (1).

Если $f(x^k) \to f(x^*)$, $k \to \infty$, то тогда также говорят, что метод (2) сходится <u>по функции.</u> Последовательность x^k при этом называется <u>минимизирующей.</u> Минимизирующая последовательность может и не сходится к точке минимума.



Так для функции, график которой на рисунке, минимизирующая последовательность $x^k = K$ не сходится к точке минимума $x^* = 0$.

В случае , если точка минимума *x* не единственна, под сходимостью метода понимается сходимость последовательности x^k к множеству X^* точек минимума функции f.

Эффективность метода характеризуется **скоростью сходимости.**

Пусть $x^k \to x^*$ при $k \to \infty$.

1) Говорят, что последовательность x^k сходится к x^* **линейно**, если существуют также константы $q \in (0;1)$ и k_0, что :

$$\|x^{k+1} - x^*\| \le q\|x^k - x^*\| \text{ при } k \ge k_0 \qquad \textbf{(4)}$$

Такая скорость также называется **скоростью геометрической прогрессии**.

2) Говорят , что x^k сходится к x^* **сверхлинейно** или со **сверхлинейной скоростью схлдимости** , если

$$\|x^{k+1} - x^*\| \le q_k \|x^k - x^*\| , \quad q_k \to 0^+ \text{ при } k \to \infty . \qquad \textbf{(5)}$$

3) Скорость сходимости называется **квадратичной**, если

$$\exists c, k_0: \quad \|x^{k+1} - x^*\| \le c\, \|x^k - x^*\|^2 , \quad \forall k \ge k_0 \qquad \textbf{(6)}$$

Иногда пользуются другими соотношениями.

Рассмотрим (4):

$$\|x^1 - x^*\| \le q\|x^0 - x^*\| = qc_1$$

$$\|x^2 - x^*\| \le q\|x^1 - x^*\| = q^2 c_1 \text{ и т.д.,}$$

Итак, (4) эквивалентно $\|x^{k+1} - x^*\| \le q^{k+1} c_1$ **(7)**

при $k \ge k_0$, $q \in (0;1)$.

Рассмотрим (5):

$$\|x^1 - x^*\| \le q_1 \|x^0 - x^*\| = q_1 c_2 ,$$

$$\|x^2 - x^*\| \le q_1 \|x^1 - x^*\| = q_1 q_2 c_2, \ldots$$

$$\|x^{k+1} - x^*\| \le q_1 q_2 \ldots q^{k+1} c_2 , \quad q_k \to 0^+ \text{ при } k \to \infty . \qquad \textbf{(8)}$$

Если предположить, что справедливо (6) и при некотором

$q \in (0;1)$ $\|x^{k_0} - x^*\| \le \frac{1}{c} q^{2^{k_0}}$, то

$$\|x^{k_0+1} - x^*\| \le C \|x^{k_0} - x^*\|^2 \le C \frac{1}{C^2} (q^{2^{k_0}})^2 = \frac{1}{C} q^{2^{k_0+1}}$$

$$\|x^{k_0+2} - x^*\| \le C \|x^{k_0+1} - x^*\|^2 \le C \frac{1}{C^2} (q^{2^{k_0+1}})^2 = \frac{1}{C} q^{2^{k_0+2}}$$

...

$$\|x^{k_0+1} - x^*\| \le \frac{1}{C} q^{2^{k+1}} = \tilde{C} q^{2^{k+1}} \qquad \textbf{(9)}$$

Для характеристики сходимости последовательности $f(x^k) \to f(x^*)$ используются аналогичные термины: линейная, сверхлинейная и квадратичная сходимость.

Для дифференцируемых на $\mathbb{R}^n$ функций из сходимости по аргументу следует сходимость по функции.

Пусть в задаче (1)

1) множество X совпадает с $\mathbb{R}^n$

2) функция $f(x)$ имеет положительно определенный гессиан $\ddot{f}(x^*)$

3) в такой окрестности точки x выполняется условие:

$$m\|x - x^*\|^2 \le |f(x) - f(x^*)| \le M\|x - x^*\|^{2,} \qquad \textbf{(10)}$$

где $m = \lambda_{min} + \varepsilon$, $M = \lambda_{max} + \varepsilon$, $\varepsilon > 0$, λ_{min} и λ_{max} - соответственно наименьшее и наибольшее собственные значения гессиана $\ddot{f}(x^*)$.

При этих условиях сходимость последовательности $x^k \to x^*$ <u>линейна в смысле абсолютного значения</u> $|f(x^k) \to f(x^*)|$ <u>тогда и только тогда, когда сходимость</u> $x^k \to x^*$ в смысле нормы $\|x^k - x^*\|$ линейна.

То же самое относится и к сходимости более высокого порядка. Если же перечисленные условия не выполняются, абсолютной эквивалентности между двумя понятиями скорости сходимости нет.

Большинство теорем о скорости сходимости методов оптимизации доказывается о предположении выпуклости целевой функции, оценки скорости сходимости выводятся при еще более сильном предположении - сильной выпуклости целевой функции.

Для невыпуклых задач численные методы позволяют отыскать лишь локальные решения или стационарные точки. Задача отыскания глобалъного решения в общем случае очень сложна даже для функции одной переменной. Получение же достаточно точного решения многомерных задач глобальной оптимизации с помощью существующих методов в настоящее время невозможно.

Под решением оптимизационной задачи будем подразумевать точку x^*, в которой выполняются необходимые условия экстремума. Например, в безусловной оптимизации под этими условиями будем понимать стационарность точки x^*, а в случае условной оптимизации - выполнение условий Куна-Таккера в этой точке.

Минимальное требование, которое мы будем налагать при построении алгоритма - это требование, чтобы все предельные точки последовательности, генерируемые алгоритмом из любой начальной точки x^0, были решением оптимизационной задачи в определенном выше смысле.

Если алгоритм обладает этим свойством, то будем говорить, что алгоритм глобально сходится.

Бесконечношаговый метод нужно дополнить условием остановки.

Остановка может производиться по достижению заданной точности решения, при выполнении условий (7) - (9). Однако константы, фигурирующие в этих условиях, обычно неизвестны; при решении реальной задачи трудно установить точность.

На практике обычно используют следующие **условия остановки**:

$$\|x^{k+1}-x^k\|\le\varepsilon_1 \quad \textbf{(11)}$$

$$\|f(x^{k+1})-f(x^k)\|\le\varepsilon_2 \quad \textbf{(12)}$$

$$\|f'(x^{k+1})\|\le\varepsilon_3 \quad \textbf{(13)}$$

До начала вычислений выбирается одно из условий (11) - (13) и малое положительное число ε_i . Вычисления прекращаются после *(k+1)*-го шага, если впервые оказывается выполненным выбранное условие остановки. Бывает, требуют выполнения двух или трех условий.

Вместо критериев, основанных на понятии абсолютной погрешности, используют другие, основанные на понятии относительной погрешности:

$$\|x^{k+1}-x^k\|\le\varepsilon(1+\|x^{k+1}\|)$$

$$\|f(x^{k+1})-f(x^k)\|\le\varepsilon(1+|f(x^{k+1})|)$$

$$\|f'(x^{k+1})-f'(x^k)\|\le\varepsilon(1+|f'(x^{k+1})|) \ .$$

Введем важное для дальнейшего понятие направления убывания и соответствующий ему класс методов оптимизации.

Направление убывания и методы спуска

Большинство методов минимизации, которые мы будем рассматривать, относятся к числу методов спуска. В методах спуска направление движения к минимуму на каждом шаге выбирается из числа направлений убывания минимизируемой функции.

Говорят, что вектор h задает направление убывания функции f в точке x, если $f(x+\alpha h)<f(x)$ **(1)**

при всех достаточно малых $\alpha>0$.

Сам вектор h также называется направлением убывания.

Множество всех направлений убывания функции f в точке x обозначается через $U(x,f)$. Таким образом, если любой достаточно малый сдвиг из точки x в направлении вектора h приводит к уменьшению значения функции f, то $h\in U(x,f)$.

Заменив в неравенстве (1) знак на противоположный, получим направление возрастания.

Достаточными и необходимыми условиями убывания являются следующие:

Лемма 1. Пусть функция f дифференцируема в точке $x\in\mathbb{R}^n$.

1) Если вектор h удовлетворяет условию

$(f'(x),h)<0$, **(2)**

то $h\in U(x,f)$.

2) Если $h\in U(x,f)$, то $(f'(x),h)\le 0$. **(3)**

Доказательство:

1) Пусть выполнено условие (2). Тогда

$$f(x+ah)-f(x)=(f'(x),\alpha h)+o(\alpha)=\alpha((f'(x),h)+\frac{o(\alpha)}{\alpha})<0$$

при всех достаточно малых α, то есть $h\in U(x,f)$.

2) Пусть $h\in U(x,f)$ и $(f'(x),h)>0$. Тогда, применяя те же рассуждения, получим, что при всех достаточно малых α $f(x+\alpha h)-f(x)>0$, то есть $f(x+\alpha h)>f(x)$, и h - направление возрастания.

Противоречие**.**

Лемма доказана.

Геометрически условие (2) означает, что вектор h составляет тупой угол с градиентом $f'(x)$.

Метод $x^{k+1}=x^k+\alpha_k h^k$, $\alpha_k\in\mathbb{R}$, $k=0,1,2,3,...$ **(4)**

называется <u>методом спуска</u>, если вектор h^k задает направление убывания функции f в точке x^k :

$h^k\in U(x,f)$, $k=0,1,2,3,...$

а число $\alpha_k>0$ и таково, что $f(x^{k+1})<f(x^k)$, $k=0,1,2,3,...$.

Простейшим примером метода спуска является градиентный метод, в котором

$h^k=-f'(x^k)$

[Если $f'(x^k)\neq 0$, то $(f'(x^k),h^k)=(f'(x^k),-f'(x^k))=-\|f'(x^k)\|^2<0$, поэтому $-f'(x^k)\in U(x^k,f)$ в силу Леммы1]

Методы выбора длины шага

Эффективнее всего выбирать α_k из следующего условия:

$$f(x^k+\alpha_k h^k)=\min_{\alpha} f(x^k+\alpha h^k) \ , \qquad \textbf{(5)}$$

где минимум берется по $\alpha>0$. Такой метод наилучший, так как он обеспечивает достижение наименьшего значения функции вдоль заданного направления. Однако он требует на каждом шаге решения одномерной задачи минимизации.

В простейших случаях величину α_k удается найти в явном виде.

Пример

$$f(x)=\frac{1}{2}(Ax,x)+(b,x) \ , \quad x\in\mathbb{R}^n$$

А - симметрическая, положительно определенная матрица. Тогда

$$f(x^k+\alpha_k h^k)=\frac{1}{2}(Ax^k+\alpha_k h^k, x^k+\alpha_k h^k)+(b, x^k+\alpha_k h^k)=$$

$$=\frac{1}{2}(Ah^k,h^k)\alpha^2+(Ax^k+b,h^k)\alpha+\frac{1}{2}(Ax^k+b,x^k)$$

Полученный квадратный трехчлен достигает минимума по

$\alpha\in\mathbb{R}$ в точке $\alpha_k=-\dfrac{(Ax^k+b,h^k)}{(Ah^k,h^k)}$.

Если h^k направление убывания, то с учетом Леммы 1 и того, что $f'(x^k)=Ax^k+b$, имеем:

$$\alpha_k = -\frac{(Ax^k + b, h^k)}{(Ah^k, h^k)} = -\frac{(f'(x^k), h^k)}{(Ah^k, h^k)} > 0 \Rightarrow$$
$$\Rightarrow f(x^k + \alpha_k h^k) = \min_{\alpha} f(x^k + \alpha h^k) = \min_{\alpha \in \mathbb{R}} f(x^k + \alpha h^k) \qquad \textbf{(6)}$$

<u>Выбор</u> α_k . Приведем еще один способ вычисления α_k .

При выборе длины шага из условия минимизации функции вдоль заданного направления приходится производить дополнительные вычисления характеристик целевой функции f в точках, отличных от x^0, x^1, ... ,x^k.

Приведем формулы для α_k , где будут использоваться лишь значения $f'(x^k)$ и некоторые константы, характеризующие глобальные свойства функции f. Эти формулы выбраны с целью обеспечить при соответствующих предположениях о f выполнение неравенства

$$f(x^k + \alpha_k h^k) - f(x^k) \le \varepsilon \alpha_k (f'(x^k), h^k) \qquad \textbf{(7)}$$

где , $\varepsilon \in (0;1)$ направление h^k таково, что $(f'(x^k), h^k) < 0$ и, следовательно, $h^k \in U(x^k, f)$. Неравенство (2) необходимо для обоснования сходимости многих методов минимизации, из него следует, что $f(x^{k+1}) < f(x^k)$ и, таким образом, метод является методом спуска.

Лемма 2.

Пусть функция f дифференцируема на $\mathbb{R}^n$, а ее градиент f' удовлетворяет условию Липшица:

$$\|f'(x) - f'(y)\| \le M \|x - y\| \ , \quad x \in \mathbb{R}^n \ , \quad M > 0 \ . \qquad \textbf{(8)}$$

Тогда для произвольных $x \in \mathbb{R}^n$, $\varepsilon \in (0;1)$ и h^k, удовлетворяющего неравенству $(f'(x^k), h^k) < 0$, условие (7) выполнено при $0<\alpha_k<\dfrac{(1-\varepsilon)(f'(x^k),h^k)}{M\|h^k\|^2}$ **(9)**

Доказательство.

По формуле Лагранжа при некотором $\Theta\in(0;1)$ имеем:

$$f(x^k+\alpha_k h^k)-f(x^k)\le(f'(x^k+\Theta\alpha_k h^k),\alpha_k h^k)=$$

$$=(f'(x^k),\alpha_k h^k)+(f(x^k+\Theta\alpha_k h^k)-f(x^k),\alpha_k h^k)\le$$

$$\le(f'(x^k),\alpha_k h^k)+\|f(x^k+\Theta\alpha_k h^k)-f(x^k)\|\cdot\|\alpha_k h^k\|\le$$

$$\le\alpha_k((f'(x^k),h^k)+M\cdot\|\Theta\alpha_k h^k\|\cdot\|h^k\|)\le$$

$$\le\alpha_k((f'(x^k),h^k)+M\alpha_k\cdot\|h^k\|^2)\le$$

$$\le\alpha_k((f'(x^k),h^k)-(1-\varepsilon)(f'(x^k),h^k))=\alpha_k\varepsilon(f'(x^k),h^k)$$

Лемма 3.

Пусть функция f дважды дифференцируема на R^n, а ее матрица вторых производных удовлетворяет условию:

$$(f''(x)h,h)\le D\|h\|^2 \quad x,h\in\mathbb{R}^n,\quad D>0 \tag{10}$$

Тогда для произвольных , $x^k\in\mathbb{R}^n$, $\varepsilon\in(0;1)$ и h^k , удовлетворяющего неравенству $(f'(x^k),h^k)<0$ условие (7)

выполнено при $0<\alpha_k<-\dfrac{2(1-\varepsilon)(f'(x^k),h^k)}{D\|h^k\|^2}$ **(11)**

Доказательство.

По формуле Тейлора при некотором $\Theta\in[0;1]$ имеем

$$f(x^k+\alpha_k h^k)-f(x^k)=(f'(x^k),\alpha_k h^k)+\frac{1}{2}(f''(x^k+\Theta\alpha_k h^k)\alpha_k h^k,\alpha_k h^k)\le$$

$$\le\alpha_k[(f'(x^k),h^k)+\frac{1}{2}D\alpha_k\|h^k\|^2]$$

К сожалению константы M и D часто неизвестны; в таких случаях способы выбора a_k по Лемме 2 и Лемме 3 неосуществимы. Однако можно указать способ выбора длины шага, который не требует знания M и D, но тем не менее обеспечивает выполнение условия (7).

В некоторых методах коэффициенты a_k можно выбирать постоянными $\alpha_k = \alpha > 0$, при этом удается добиться выполнение условия (7). Например, если $h^k = -f'(x^k)$, то (9) и (11) приобретают вид:

$$0 < \alpha_k \le \frac{(1-\varepsilon)}{M} \ ; \quad 0 < \alpha_k \le \frac{2(1-\varepsilon)}{D} \ .$$

Поэтому условие $f(x^k - \alpha_k f'(x^k)) - f(x^k) \le -\varepsilon \alpha_k \|f'(x^k)\|^2$ выполнено при $\alpha = \alpha_k \in (0; \frac{(1-\varepsilon)}{M})$ и $\alpha = \alpha_k \in (0; \frac{2(1-\varepsilon)}{D})$ соответственно.

Есть еще способ выбора коэффициентов α_k , называемый дроблением шага. Если h^k - направление убывания, то дробление шага выполняется так:

выбираются некоторые константы $\beta > 0$, $0 < \lambda < 1$ (часто $\lambda = \frac{1}{2}$). Для коэффициента $\alpha = \beta$ проверяется условие:

$$f(x^{k+1}) \qquad \textbf{(*)}$$

Если оно выполнено, полагают $\alpha_k = \alpha$. Если нет , то производится дробление шага, то есть полагают $\alpha = \lambda\beta$ и вновь проверяют условие (*). Процесс дробления , то есть умножения текущего значения α на λ , продолжается до тех пор, пока (*) не окажется выполненным. Этот процесс не может быть бесконечным, поскольку h^k - направление убывания. Первое α, при котором условие выполнено, принимается за α_k.

Как показывает следующая Лемма, при $(f'(x^k), h^k) < 0$ с помощью описанного процесса дробление шага можно добиться и выполнения условия (7).

Лемма 4.

Пусть функция f дифференцируема на $\mathbb{R}^n$. Тогда для произвольных

$x^k \in \mathbb{R}^n$, $\varepsilon \in (0;1)$ и h^k, удовлетворяющего неравенству $(f'(x^k), h^k) < 0$ найдется $\alpha_0 > 0$, что при всех $\alpha \in (0;\alpha_0]$ выполнено условие:

$$f(x^k - \alpha_k f'(x^k)) - f(x^k) \le -\varepsilon \alpha_k \|f'(x^k)\|^2 \qquad \textbf{(**)}$$

Доказательство.

Имеем

$$f(x^k + \alpha h^k) - f(x^k) = (f'(x^k), \alpha h^k) + o(\alpha) =$$

$$= \varepsilon \alpha (f'(x^k), h^k) + \alpha [(1-\varepsilon)(f'(x^k), h^k) + \frac{o(\alpha)}{\alpha}]$$

В силу условия леммы при достаточно малых α

$(1-\varepsilon)(f'(x^k), \alpha h^k) < 0$. Отсюда следует (**).

Предположим, что $h^k = -f'(x^k)$ и выполняется либо условие Леммы 2, либо Леммы 3, однако константы M и D неизвестны. Тогда в силу выше сказанного (*) заведомо окажется выполненным, если в результате процесса дробления шага получим

$$\alpha \le \frac{(1-\varepsilon)}{M} \quad \text{или} \quad \alpha \le \frac{2(1-\varepsilon)}{D} \; .$$

Поэтому для коэффициентов α_k, построенных по методу дробления шага с проверкой (**) , справедливы неравенства:

$$\alpha_k \ge \overline{\alpha} = min\{\beta, \lambda \frac{(1-\varepsilon)}{M}\} \tag{12}$$

$$\alpha_k \ge \tilde{\alpha} = min\{\beta, \lambda \frac{2(1-\varepsilon)}{d}\} \tag{13}$$

Численные методы безусловной оптимизации.

Пусть дана задача безусловной оптимизации:

$$f(x) \to min, x \in \mathbb{R}^n$$

Рассмотрим классические методы минимизации - градиентный метод и метод Ньютона. Они оба основаны на идее замены минимизируемой функции в окрестности очередной точки x^k первыми членами ее разложения в ряд Тейлора. В градиентном методе берут линейную часть разложения, в методе Ньютона - квадратичную часть.

Градиентный метод.

Мы рассматривали общую схему методов спуска, в которых последовательность приближений $x^0, x^1, x^2, \ldots$ к точке минимума выбирается по правилу

$x^{k+1} = x^k + \alpha_k h^k$, где h^k - направление убывания функции f в точке x^k, то есть $h^k \in U(x, f)$, а α_k - параметр, регулирующий длину шага вдоль h^k.

В <u>градиентном методе</u> h^k берется равным антиградиенту функции *f* в точке x^k, то есть

$$h^k = -f'(x^k) \ .$$

Итак, в градиентном методе

$$x^{k+1} = x^k - \alpha_k f'(x^k), \quad \alpha_k > 0, \quad k = 0,1,2,\ldots \qquad \textbf{(1)}$$

Такой выбор h^k определяется следующими соображениями.

Пусть *f* дифференцируема на $\mathbb{R}^n$. Тогда

$$f(x) = f(x^k) + (f'(x^k), x - x^k) + o(\|x - x^k\|)$$

Для определения точки x^{k+1} минимизируем по *x* на шаре $U = \{x : \|x - x^k\| \le r\}$ функцию $(f'(x^k), x - x^k)$, являющуюся линейной частью приращения $f(x) - f(x^k)$. При $x \in U$ в силу неравенства Коши-Буняковского имеем:

$$(f'(x^k), x - x^k) \ge -\|f'(x^k)\| \cdot \|x - x^k\| \ge -r\|f'(x^k)\|$$

С другой стороны для (та $x = x^k - r\dfrac{f'(x^k)}{\|f'(x^k)\|}$ к как шар с центром в точке x^k).

$$(f'(x^k), x - x^k) = (f'(x^k), -r\frac{f'(x^k)}{\|f'(x^k)\|}) = -r\|f'(x^k)\|$$

Следовательно, минимум функции достигается в точке

$x = x^k - r\dfrac{f'(x^k)}{\|f'(x^k)\|}$. Таким образом при фиксированной величине шага, то есть величине, $\|x^{k+1} - x^k\|$ минимум линейной части разложения функции *f* в окрестности точки x^k в ряд Тейлора достигается, если направление шага $h = x^{k+1} - x^k$ совпадает с направлением антиградиента

$-f'(x^k)$.

В градиентном методе используются различные способы выбора длины шага. Если длина шага выбирается из условия минимизации функции вдоль направления антиградиента, то получаем вариант градиентного метода, который называется методом наискорейшего спуска:

$$f(x^k-\alpha_k f'(x^k))=\min_{\alpha>0} f(x^k-\alpha_k f'(x^k))$$

$$x^{k+1}=x^k-\alpha_k f'(x^k), \quad \alpha_k>0, \quad k=0,1,2,\ldots$$

Рассмотрим подробнее другой вариант градиентного метода: выбор длины шага осуществляется с помощью метода дробления.

Будем считать, что коэффициенты a_k удовлетворяют условию:

$$f(x^k-\alpha_k f'(x^k))-f(x^k)\le-\varepsilon\alpha_k\|f'(x^k)\|^2 , \quad \varepsilon\in(0;1) \qquad \textbf{(2)}$$

И если функция f дифференцируема и удовлетворяет неравенству $\|f'(x)-f'(y)\|\le M\|x-y\|$, **(3)**

при $x,y\in\mathbb{R}^n$, $M>0$, то $\alpha_k\ge\overline{\alpha}>0$, **(4)**

а если f дважды дифференцируема и

$$(f''(x)h,h)\le D\|h\|^2 , \qquad \textbf{(5)}$$

при $x,h\in\mathbb{R}^n$, $D>0$, то $\alpha_k\ge\tilde{\alpha}>0$ **(6)**

Сходимость градиентного метода

Если минимизируемая функция невыпукла, градиентный метод обеспечивает лишь сходимость к множеству стационарных точек функции f.

<u>Теорема 1.</u> Пусть *f* дифференцируема и ограничена снизу на $\mathbb{R}^n$, а ее градиент удовлетворяет условию Липшица (3). Тогда <u>при произвольной начальной точке</u> $x^0 \in \mathbb{R}^n$ для метода (1),(2),(4)

$$\lim_{k\to\infty} \|f'(x^k)\| = 0$$

Доказательство. В силу (2)

$$f(x^{k+1}) - f(x^k) \le -\varepsilon\alpha_k \|f'(x^k)\|^2 \le 0 \qquad \textbf{(7)}$$

то есть последовательность $\{f(x^k)\}$ не возрастает. Так как функция *f* ограничена снизу, то последовательность $\{f(x^k)\}$ сходится, и, следовательно,

$$f(x^k) - f(x^{k+1}) \to 0 \quad \text{при} \quad k\to\infty$$

Тогда из (7) и (4)

$$\|f'(x^k)\|^2 \le \frac{f(x^k) - f(x^{k+1})}{\varepsilon\alpha_k} \le \frac{f(x^k) - f(x^{k+1})}{\varepsilon\overline{\alpha}} \quad \Rightarrow \quad \|f'(x^k)\| \to 0$$

при $k\to\infty$.

Теорема 1 показывает, что любая предельная точка x^* последовательности $\{x^k\}$, генерируемой градиентным методом (1),(2),(4) является стационарной точкой минимизируемой функции *f*, то есть $f'(x^*) = 0$.

Если функция f выпукла, то любая стационарная точка есть точкой минимума.

Элементы теории сильно выпуклых функций

Определение. Функция *f*, определенная на выпуклом множестве $X \subset \mathbb{R}^n$ называется <u>сильно выпуклой</u> с константой

$\Theta > 0$ на X , если

$$f(\lambda x_1 + (1-\lambda) x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2) - \Theta \lambda (1-\lambda) \|x_1 - x_2\|^2$$

при всех $x_1, x_2 \in X$, $\lambda \in [0;1]$.

В некоторой литературе дается такое определение:

$$f\left(\frac{x_1+x_2}{2}\right) \le \frac{1}{2}(f(x_1)+f(x_2)) - \Theta \|x_1 - x_2\|^2$$

Θ – произвольная малая константа.

Простейшим примером сильно выпуклой функции является функция $f(x) = \|x\|^2$ на $\mathbb{R}^n$, для нее неравенство выполняется при $\Theta = 1$.

Сильно выпуклая функция, очевидно, строго выпукла, но не наоборот. Например, x^4 не является сильно выпуклой.

Критерии выпуклости функции:

1. Функция f сильно выпукла с константой $\Theta \ge 0 \Leftrightarrow$

$$f(x) - f(x^*) \ge (f'(x^*), x - x^*) + \Theta \|x - x^*\|^2, \forall x, x^* \in X \subset \mathbb{R}^n \qquad \textbf{(**)}$$

и $f(x^*) - f(x) \ge (f'(x), x^* - x) + \Theta \|x^* - x\|^2$

2. Имеется критерий сильной выпуклости в терминах вторых производных: если f дважды непрерывно-дифференцируемая функция на выпуклом множестве $X \subset \mathbb{R}^n$, $int\, X \ne \emptyset$, то она будет сильно выпуклой с константой $\Theta \ge 0$ на X в том и только том случае, если $(f''(x^*)h, h) \ge 2\Theta \|h\|^2$ при всех $x^* \in X, h \in \mathbb{R}^n$ **(*)**

(при $\Theta = 0$ данное условие означает неотрицательную определенность матрицы f'' в любой точке $x^* \in X$. При $\Theta > 0$ оно означает нечто большее, чем просто положительная определенность. Это условие еще называют так: матрица f'' сильно положительная.)

Примеры:

а) $f(x)=x^4$ не является сильно выпуклой на $\mathbb{R}$, так как, но на множестве $X=[\alpha;+\infty]$ при $\alpha>0$ она сильно выпукла.

б) $f(x)=e^x$ также не является сильно выпуклой на $\mathbb{R}$, так как $f''(x)=e^x\to 0$ при $x\to-\infty$, но на множестве $X=[\alpha;+\infty]$ она сильно выпукла, где $\alpha\in R$. Этот пример говорит о том, что из положительной определенности матрицы $f''(x^*)$ всюду на X еще не следует сильная выпуклость функции.

Для квадратичной функции имеет место следующее **утверждение:**

Если A - симметрическая матрица размера nxn, $b\in\mathbb{R}^n$, то квадратичная функция $f(x)=(Ax,x)+(b,x)$ сильно выпукла в том и только том случае, если A положительно определена: $f''=2A$. A положительно определена в том случае и только том случае, если $\lambda_{min}>0$ и

$$(Ax,x)\geq\lambda_{min}\|x\|^2\Rightarrow$$
$$\Rightarrow(f''(x^*)h,h)=(2Ah,h)\geq 2\lambda_{min}\|h\|^2\Rightarrow$$
$$\Rightarrow\Theta=\lambda_{min}$$

В случае <u>сильно выпуклой</u> функции градиентный метод сходится к точке минимума x функции f со скоростью геометрической прогрессии.

<u>**Теорема 2.**</u> Пусть функция f <u>дважды дифференцируема</u> и <u>сильно выпукла</u> на $\mathbb{R}^n$, а ее матрица вторых производных удовлетворяет условию (5): $(f''(x)h,h)\leq D\|h\|^2$, $x,h\in\mathbb{R}^n$,

$D > 0$.

Тогда <u>при произвольной начальной точке</u> $x_0 \in \mathbb{R}^n$ __ последовательность $\{x_k\}$, определенная формулами (1),(2),(6), сходится к точке минимума x^* функции f со скоростью геометрической прогрессии:

$$f(x^k) - f(x^*) \le q^k (f(x^0) - f(x^*)) \quad \textbf{(8)}$$

$$\|x^k - x^*\| \le C(\sqrt{q})^k \quad \textbf{(9)}$$

где $C > 0$, $q \in (0;1)$ константы.

Доказательство.

Из теории сильно выпуклых функций точка минимума x существует и единственна.

Из (*) следует, что $(f''(x^*)h, h) \ge d\|h\|^2$, $x, h \in \mathbb{R}^n$, **(10)**

где $d = 2\Theta > 0$. Так как $f'(x^*) = 0$, из формулы Тейлора

1) $f(x) - f(x^*) = \frac{1}{2}(f''(x^* + \xi(x - x^*))(x - x^*), (x - x^*))$

2) $\frac{d}{2}\|x - x^*\|^2 \le f(x) - f(x^*) \le \frac{D}{2}\|x - x^*\|^2$ **(11)**

Применяя (**) при $\Theta = \frac{d}{2}$ и неравенство Коши-Буняковского, получаем:

$$f(x^*) - f(x) \ge (f'(x), x - x^*) + \frac{d}{2}\|x^* - x\|^2 \ge$$

3) $\ge -\|f'(x)\|\|x - x^*\| + \frac{d}{2}\|x - x^*\|^2$ **(12)**

Отсюда и левого неравенства из (11) имеем:

$$\frac{d}{2}\|x - x^*\|^2 \le \|f'(x)\| \cdot \|x - x^*\| - \frac{d}{2}\|x - x^*\|^2 \Rightarrow$$

(переносим влево $\frac{d}{2}\|x-x^*\|^2$ и сокращаем на $\|x-x^*\|$)

4) $\|x-x^*\|\le\frac{\|f'(x)\|}{d}$ **(α)**

Правая часть (11) дает $\frac{D}{2}\|x-x^*\|^2\ge f(x)-f(x^*)$ или

$$\|x-x^*\|^2\ge\frac{2}{D}(f(x)-f(x^*))\Rightarrow$$

5) $\Rightarrow-\|x-x^*\|^2\le-\frac{2}{D}(f(x)-f(x^*))$ **(β)**

Из (12) имеем:

$$f(x^*)-f(x)\le\|f'(x)\|\cdot\|x-x^*\|-\frac{d}{2}\|x-x^*\|^2\le(\alpha,\beta)$$

$$\le\frac{f'(x)}{d}-\frac{d}{D}(f(x)-f(x^*))$$

Откуда

$$\|f'(x)\|^2\ge d(1+\frac{d}{D})(f(x)-f(x^*)) \qquad \textbf{(γ)}$$

Вернемся к (7):

$$f(x^{k+1})-f(x^k)\le-\varepsilon\alpha_k\|f'(x^k)\|^2$$

Тогда

$$f(x^{k+1})-f(x^*)\le f(x^k)-\varepsilon\alpha_k d(1+\frac{d}{D})\cdot(f(x^k)-f(x^*))-f(x^*)=$$

$$=(1-\varepsilon\alpha_k d(1+\frac{d}{D}))\cdot(f(x)-f(x^*))\le q\cdot(f(x^k)-f(x^*)) \qquad \textbf{(δ)}$$

где $q=1-\varepsilon\tilde{\alpha}d(1+\frac{d}{D})$.

x^* - единственная точка минимума (то есть $f(x^{k+1})-f(x^*)>0$, $f(x^k)-f(x^*)>0$ для $\forall x^{k+1}, x^k x^*$)

$\Rightarrow q>0$, так как $1-\varepsilon\tilde{\alpha}d(1+\frac{d}{D})>0\Rightarrow0<\varepsilon\tilde{\alpha}d(1+\frac{d}{D})<1\Rightarrow q<1$

Таким образом из полученного неравенства следует (8).

Так как из (11)

$\frac{d}{2}\|x-x^*\|^2\le f(x)-f(x^*)$, то

$\|x-x^*\|\le\sqrt{\frac{2}{d}}\cdot\sqrt{f(x)-f(x^*)}\le C(\sqrt{q})^k$, где

$C=\sqrt{\frac{2}{d}}\cdot\sqrt{f(x^0)-f(x^*)}$, так как

$\|x^0-x^*\|\le\sqrt{\frac{2}{d}}\cdot\sqrt{f(x^0)-f(x^*)}$,

$$\|x^1-x^*\|\le\sqrt{\frac{2}{d}}\cdot\sqrt{f(x^1)-f(x^*)}\le\sqrt{\frac{2}{d}}\cdot\sqrt{q\cdot(f(x^0)-f(x^*))}=$$

$$=\sqrt{\frac{2}{d}}\cdot\sqrt{q}\cdot\sqrt{f(x^0)-f(x^*)}\ ;$$

$$\|x^2-x^*\|\le\sqrt{\frac{2}{d}}\cdot\sqrt{f(x^2)-f(x^*)}\le\sqrt{\frac{2}{d}}\cdot\sqrt{q\cdot(f(x^1)-f(x^*))}\le$$

$$\le\sqrt{\frac{2}{d}}\cdot\sqrt{q\cdot q\cdot(f(x^0)-f(x^*))}=\sqrt{\frac{2}{d}}\cdot(\sqrt{q})^2\cdot\sqrt{f(x^0)-f(x^*)}$$

....

$$\|x^k-x^*\|\le(\sqrt{d})^k\cdot\sqrt{\frac{2}{d}}\cdot\sqrt{f(x^0)-f(x^*)}=C\cdot(\sqrt{q})^k$$

Обсуждение градиентного метода

Найдем минимальное значение q для метода дробления шага. Подставляя $\tilde{\alpha}=\lambda\cdot\frac{2(1-\varepsilon)}{D}$ в выражение для q, найдем минимум по ε:

$$q=1-\varepsilon\lambda\cdot\frac{2(1-\varepsilon)}{D}d(1+\frac{d}{D})=1-\varepsilon(1-\varepsilon)z=1-\varepsilon z+\varepsilon z^2$$

$$z=\frac{2\lambda d}{D}\left(1+\frac{d}{D}\right)$$

$$q'_{\varepsilon}=-z+2\varepsilon z=0\Rightarrow\varepsilon=\frac{1}{2}$$

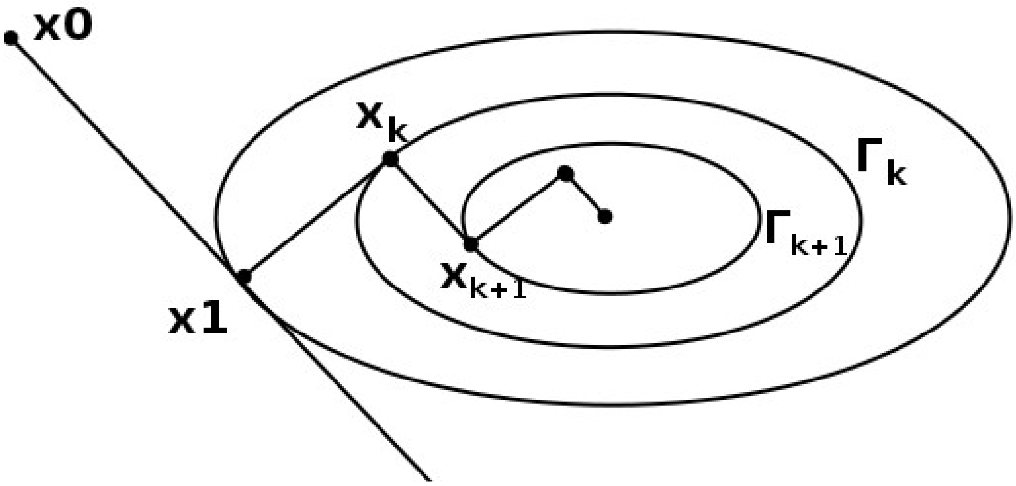
$$q_{min}=1-\frac{\lambda}{2}\cdot\frac{d}{D}\left(1+\frac{d}{D}\right)$$

Нижняя грань по λ, достигается при $\lambda\to 1$ ($\lambda\in(0,1)$). Если коэффициенты α_k выбираются из условия минимизации функции вдоль направления антиградиента, то метод наискорейшего спуска также сходиться со скоростью геометрической прогрессии сознаменателем $q_1=\frac{D-d}{D+d}$ (Пшеничный, Данилин)

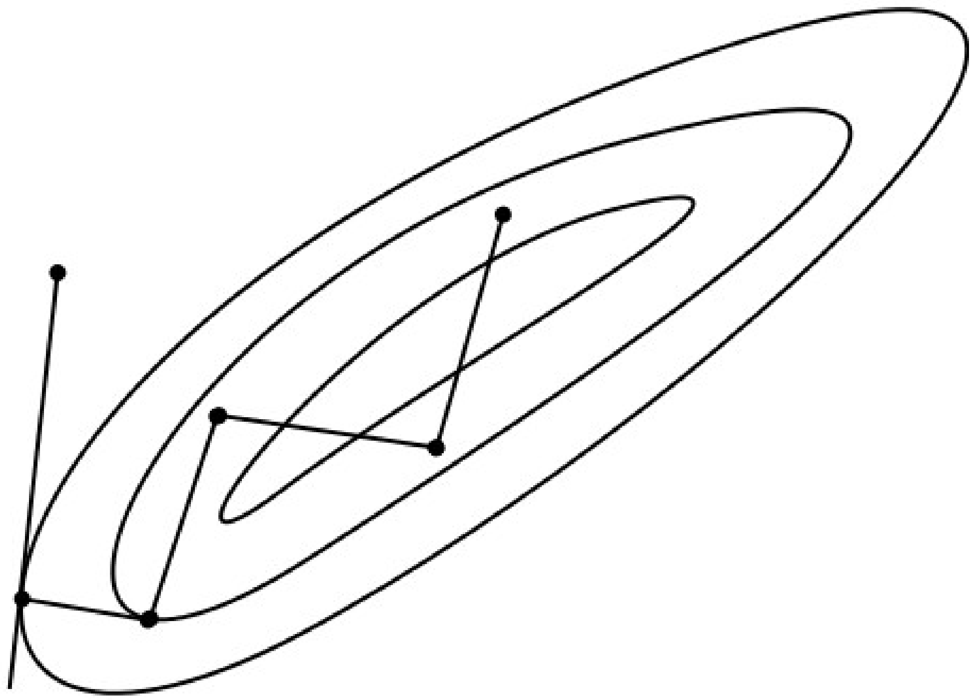
В обеих формулах величина q мала лишь при условии, что d и D незначительно отличаются друг от друга. В качестве d и D часто берут наименьшее и наибольшее собственные значения матрицы $f''(x)$ (если *f* - сильно выпукла и *x* - фиксировано).

Если же $\frac{d}{D}\ll 1$ (при плохо обусловленной матрице), то *q* близко к единице, и градиентный метод медленно сходится. Геометрически это выглядит так: в этом случае линии уровня функции *f* имеют "овражную" структуру, и направление вектора $-f'(x^k)$ может сильно отклоняться от направления в точку минимума.

Метод имеет следующий геометрический смысл: точка x^{k+1} , что определяется условием $x^{k+1}=x^k-\alpha_k f'(x^k),\alpha_k>0$, лежит на луче $l_k=\{x:x=x^k-\alpha_k f'(x^k),\alpha_k\geq 0\}$ в точке его касания линии уровня $\Gamma_{k+1} = \{ x \in \mathbb{R}^n : f(x) = f(x^{k+1}) \}$, а сам луч l_k перпендикулярен к линии уровня Γ_k, то есть $\Gamma_k = x \in \mathbb{R}^n : f(x) = f(x^k)$



Из рисунков видно, что чем ближе линии уровня *f(x)=const* к <u>окружности</u>, тем лучше сходится метод.



Вследствие этого движение к минимуму в градиентном методе носит явно выраженный зигзагообразный характер.

Происходит, как говорят, "рыскание" метода, что является недостатком.

Если "склоны оврага" достаточно круты, то такие скачки со склона на склон сильно замедляют сходимость метода.

При поиске минимума "овражной" функции для ускорения сходимости метода применяют прием, называемый овражным методом.

В начале поиска задаются две точки v^0 и v^1, из которых производят спуск по градиентному методу и получают две точки x^0 и x^1 на дне оврага. Затем получают

$$v^2 = x^1 - \frac{(x^1 - x^0)}{\|(x^1 - x^0)\|} \cdot h \cdot sign(f(x^1) - f(x^0))$$

где h - положительная константа, овражный шаг. Из точки v^2 , которая, вообще говоря, на склоне оврага, производят спуск при помощи градиентного метода и получают точку x^2 на дне оврага. Если уже известны точки $x^0, x^1, ..., x^k$ $(k \geq 2)$, то из точки

$$v^{k+1} = x^k - \frac{(x^k - x^{k-1})}{\|(x^k - x^{k-1})\|} \cdot h \cdot sign(f(x^k) - f(x^{k-1}))$$

совершают спуск с помощью градиентного метода и находят следующую точку x^{k+1} на дне оврага.

Это один из способов ускорять сходимость градиентного метода. Есть и другие.

Особенно наглядно можно себе представить это, рассматривая в качестве примера квадратичную функцию $f(x, y) = \frac{1}{2}(\frac{x^2}{a^2} + \frac{y^2}{b^2})$ в $\mathbb{R}^2$. Матрица вторых производных этой функции имеет постоянные элементы $\begin{pmatrix} \frac{1}{a^2} & 0 \\ 0 & \frac{1}{b^2} \end{pmatrix}$,

поверхность уровня ее - эллипсы , $\frac{1}{2}(\frac{x^2}{a^2}+\frac{y^2}{b^2})=C$ точка минимума совпадает с центром эллипсов. Собственные числа матрицы f'' соответственно $\frac{1}{a^2}$ и $\frac{1}{b^2}$. Чем больше отношение $\frac{a^2}{b^2}$ отличается от единицы, тем сильнее вытянуты линии уровня вдоль одной из осей OX или OY и тем больше шагов в направлении антиградиента нужно сделать при движении из произвольной точки (x_0, y_0) для того, чтобы попасть в достаточно малую окрестность точки минимума.

Медленная сходимость градиентных методов не позволяет решать с их помощью сложные задачи минимизации. Градиентные методы используются в комбинации с другими, с более высокой скоростью сходимости, на начальной стадии решения задачи, когда точка x^k находиться далеко от минимума и шаг вдоль антиградиента позволяют достичь существенного убывания функции.

Вместе с тем несомненным достоинством градиентных методов является их простота и возможность использовать для минимизации весьма различных по характеру функций.

Аналогом градиентного метода для выпуклых недифферинцируемых функций является субрадиентный метод. В нем точка x^{k+1} выбирается по правилу:

$x^{k+1} = x^k - \alpha_k h^k, \alpha_k > 0, k = 0, 1, 2, \ldots$, где $h \in \partial f(x^k)$, то есть h^k является субградиентом функции f в точке x^k. При этом h^k выбирается из $\partial f(x^k)$ произвольным образом. Идея этого метода принадлежит Н.З.Шору.

Метод Ньютона и его модификации

Данные методы являются наиболее эффективными для решения задач безусловной оптимизации. Метод Ньютона методом 2-го порядка, т.е. использует вычисление 2-х производных минимизируемой функции.

Предположим, что функция f строго выпукла и дважды дифференцируема на $\mathbb{R}^n$, причём матрица $f''(x)$ невыроджена на $\mathbb{R}^n$. В методе Ньютона последовательность $x^0, x^1, x^2, \ldots$ строится исходя из следующих соображений: f - дважды дифференцируемая, выпукла, - f'' невыродженая на $\mathbb{R}^n$

Для дважды дифференцируемой функции имеем:

$$f(x) - f(x^k) = (f'(x^k), x - x^k) + \frac{1}{2}(f''(x^k)(x - x^k), x - x^k) + o(\|x - x^k\|^2)$$

Для определения следующей точк x^{k+1} и минимизируется функция $f_k(x)$, являющаяся квадратичной частью приращения $f(x) - f(x^k)$, то есть решается задача:

$$f_k(x) = (f'(x^k), x - x^k) + \frac{1}{2}(f''(x^k)(x - x^k), x - x^k) \to min, x \in \mathbb{R}^n$$

Ясно, что $f_k''(x)=f''(x^k)$. Так как необходимым и достаточным условием выпуклой функции является неотрицательная определённость матрицы ее вторых производных, а f - выпукла, отсюда следует, что f_k выпукла. Условие минимума имеет вид: $f_k''(x)=f'(x^k)+f''(x^k)(x-x^k)=0$

Решая полученную сумму линейных уравнений и принимая найденную точку минимума за x^{k+1} , получаем:

$$x^{k+1}=x^k+h^k, h^k=-[f''(x^k)]^{-1}\cdot f'(x^k) \qquad \textbf{(1)}$$

Данное соотношение определяет метод Ньютона минимизации функции $f, \alpha_k=1$

Если $f'(x^k)\neq 0$ и f - строго выпуклая функция, то направление, определяемое вектором h^k является направлением убывания функции f , так как в силу строгой $(f'(x),h)=-(f''(x)h,h)<0$ выпуклости f .

$$-f''(x)h=-f''(x)\cdot[-f''(x)]^{-1}\cdot f'(x)=f'(x)$$

Квадратичная функция в малой окрестности некоторой точки аппроксимирует минимизируемую функции гораздо более точно по сравнению с линейной функцией. Поэтому естественно ожидать, по крайней мере, если точка x находится в достаточно малой окрестности решение x^* , что движение из точки x вы направлени $h=-[f''(x)]^{-1}\cdot f'(x)$ и позволит достичь более существенного убывания функции и получить более точное приближение к решению, чем движение в направлении , $-f'(x)$ используемое в градиентном методе. Таким образом можно ожидать, что итерационный процесс(1) окажется более эффективным по сравнению с методом наискорейшего спуска, то есть скорость сходимости при использовании алгоритма (1) будет выше, чем в градиентном методе.

Теорема 1.

Пусть функция f дважды дифференцируема, сильно выпукла с константой q>0 на $\mathbb{R}^n$ и удовлетворяет условию:

$$\|f''(x)-f''(y)\|\leq M\|x-y\|, \quad x,y\in\mathbb{R}^n, M>0 \qquad \textbf{(2)}$$

а начальная точка x^0 такова, що $\|f'(x^0)\|\leq\frac{8\Theta^2}{M}$, то есть

$$\|f'(x^0)\|=\frac{8\Theta^2 q}{M}, \text{ где } q\in(0;1) \qquad \textbf{(3)}$$

Тогда последовательность (1) сходится к точке минимума x^* с квадратичной скоростью, то есть

$$\|x^k-x^*\|\leq\frac{4\Theta q^{2^k}}{M}$$

Доказательство.

Так как условие сильной выпуклости функции f эквивалентно тому, что

$$(f''(x)h,h)\geq 2\Theta\|h\|^2 \quad x,h\in\mathbb{R}^n \qquad \textbf{(4)}$$

отсюда следует, что матрица $f''(x)$ положительно определена, отсюда следует, невырождена при всех $x\in\mathbb{R}^n$.

Если f - сильно выпуклая функция, то имеет место свойство:

$$(f'(x)-f'(x^*),x-x^*)\geq 2\Theta\|x-x^*\|^2 \quad x,x^*\in\mathbb{R}^n$$

Потому, учитывая ,что $f'(x)=0$

$$\|x^k-x^*\|^2\leq\frac{1}{2\Theta}(f'(x^k)-f'(x^*),x^k-x^*)\leq\frac{1}{2\Theta}\|f'(x^k)\|\cdot\|x^k-x^*\|$$

то есть

$$\|x^k - x^*\| \le \frac{1}{2\Theta}\|f'(x^k)\| \qquad \textbf{(5)}$$

Оценим величину $\|f'(x^k)\|$. Интегрируем от 0 до 1 по t соотношение:

$$\frac{\partial}{\partial t} f'(x^k + t(x^{k+1} - x^k)) = f''(x^k + t(x^{k+1} - x^k))(x^{k+1} - x^k)$$

И вычитая из обеих частей полученного равенства $f''(x^k)(x^{k+1} - x^k)$, получаем:

$$\int_0^1 \frac{\partial}{\partial t} f'(x^k + t(x^{k+1} - x^k))\,dt = f'(x^k + t(x^{k+1} - x^k))|_0^1 =$$

$$= f'(x^k + x^{k+1} - x^k) - f'(x^k) = f'(x^{k+1}) - f'(x^k)$$

$$f'(x^{k+1}) - f'(x^k) - f''(x^k)(x^{k+1} - x^k) =$$

$$= \int_0^1 \frac{\partial}{\partial t}[f''(x^k + t(x^{k+1} - x^k)) - f''(x^k)](x^{k+1} - x^k)\,dt =$$

$$\|f'(x^{k+1}) - f'(x^k) - f''(x^k)(x^{k+1} - x^k) = \| \le \frac{M}{2}\|x^{k+1} - x^k\|^2$$

Подставив сюда $(x^{k+1} - x^k) = -[f''(x^k)]^{-1} f'(x^k)$ получим:

$$f'(x^{k+1}) - f'(x^k) - f''(x^k)(x^{k+1} - x^k) =$$

$$= f'(x^{k+1}) - f'(x^k) + f''(x^k)[f''(x^k)]^{-1} f'(x^k) =$$

$$= f'(x^{k+1}) - f'(x^k) + f'(x^k) = f'(x^{k+1})$$

таким образом

$$\|f'(x^{k+1})\| \le \frac{M}{2}\|[f''(x^k)]^{-1}\|^2 \|f'(x^k)\|^2 \qquad \textbf{(6)}$$

Оценим $\|[f''(x^k)]^{-1}\|$. Полагая $h = [f''(x^k)]^{-1} y$, получаем из (4)

$$f''(x)\cdot([f''(x)]^{-1} y, [f''(x)]^{-1} y) \ge 2\Theta \|[f''(x)]^{-1} y\|^2 \quad \text{или}$$

$$\|[f''(x)]^{-1} y\|^2 \le \frac{1}{2\Theta}(y, [f''(x)]^{-1} y) \le \frac{1}{2\Theta}\|y\| \cdot \|[f''(x)]^{-1} y\| \quad ,$$

откуда:

$$\|[f''(x)]^{-1}\| = \max_{y \neq 0} \frac{\|[f''(x)]^{-1} y\|}{\|y\|} \leq \frac{1}{2\Theta}$$

Тогда из (6) следует, что

$$\|f'(x^{k+1})\| \leq \frac{M}{2} \cdot \frac{1}{4\Theta^2} \|f'(x^k)\|^2 = \frac{M}{8\Theta^2} \|f'(x^k)\|^2$$

Это справедливо для всех $k = 0, 1, 2, \ldots$ Потому из (3) получаем, что

$$\|f'(x^k)\| \leq \frac{M}{8\Theta^2} \|f'(x^{k-1})\|^2 \leq \frac{M}{8\Theta^2} \cdot \left(\frac{M}{8\Theta^2}\right)^2 \|f'(x^{k-2})\|^4 \leq \ldots \leq$$

$$\leq \left(\frac{M}{8\Theta^2}\right)^{2^k - 1} \|f'(x^0)\|^{2^k} = \left(\frac{M}{8\Theta^2}\right)^{2^k - 1} \cdot \left(\frac{8\Theta^2}{M}\right)^{2^k} \cdot q^{2^k} = \frac{8\Theta^2}{M} \cdot q^{2^k}$$

Из (5) следует, что

$$\|x^k - x^*\| \leq \frac{1}{2\Theta} \cdot \frac{8\Theta^2}{M} \cdot q^{2^k} = \frac{4\Theta}{M} \cdot q^{2^k}$$

Таким образом сходимость метода Ньютона доказана лишь для достаточно хорошего начального приближения x^0. Условие (3), трудно проверяемо, константы, фигурирующие в (3) обычно не известны. Сложность отыскания нужного начального приближения является недостатком метода. Он также трудоемкий, так как на каждом шаге нужно вычислять и обращать матрицу вторых производных функции.

Модификации метода Ньютона направлены на то, чтобы сохранить его основное достоинство – быструю сходимость, но уменьшить трудоемкость и ослабить требования на выбор начального приближения.

Метод Ньютона с регулировкой шага

(или обобщенный метод Ньютона)

Рассмотрим метод

$$x^{k+1}=x^k+\alpha_k h^k;\alpha_k>0;h_k=-(f''(x^k))^{-1}\cdot f'(x^k) \qquad \textbf{(7)}$$

(При $\alpha_k=1$ он совпадает с классическим методом Ньютона)

Если элементы матрицы $(f''(x^k))^{-1}$ обозначить через $\varphi_{ij}(x^k), i, j=1,2,\dots,n,$ i-индекс строки, то метод (7) можно записать в координатной форме:

$$x_i^{k+1}=x_i^k-\alpha_k\sum\varphi_{ij}\frac{\partial f(x^k)}{\partial x_j}, i=1,2,\dots,n,$$

Метод (7) можно представить также в виде:

$$f''(x^k)h^k=-f'(x^k),\ \ x^{k+1}=x^k+\alpha_k h^k$$

или в координатной форме:

$$\sum\frac{\partial^2 f(x^k)}{\partial x_i\partial x_j}h_j^k=\frac{\partial f(x^k)}{\partial x_i};\ \ x_i^{k+1}=x^k+\alpha_k h_i^k, i=1,\dots,n$$

Следовательно, для определения вектора h^k можно решать <u>систему линейных</u> <u>уравнений</u> вместо того, чтобы обращать матрицу вторых производных.

Мы будем рассматривать два варианта обобщённого метода Ньютона (то же – «с регулировкой шага»), в которых используются <u>разные способы выбору</u> параметра α. Первый способ заключается в следующем:

1) полагаем $\alpha=1$ и вычисляем точку $x=x_k+\alpha h_k$;

2) вычисляем $f(x)=f(x_k+\alpha h_k)$;

3) производим проверку неравенства:

$$f(x) - f(xk) \le \varepsilon\alpha(f'(x^k), h^k)\ , \quad 0 < \varepsilon < \frac{1}{2} \qquad \textbf{(8)}$$

4) если это неравенство выполняется, то значение $\alpha = 1$ берем в качестве искомого $\alpha_k = 1$. В противном случае производим дробление а до тех пор, пока неравенство (8) не выполнится. А ε - произвольно выбранная константа, одна и та же при всех $k = 0, 1, 2, \ldots$

Как видно, такой способ выбора длины шага строится аналогично способу выбора α_k в методе наискорейшего спуска.

В другом варианте метода (7) значение α_k должно доставлять минимум функции в направлению движения:

$$f(x^k - a_k(f''(x^k))^{-1} f'(x^k)) = \min_{\alpha > 0} f(x^k - a_k(f''(x^k))^{-1} f'(x^k)) \qquad \textbf{(9)}$$

Теорема о свойствах метода

Рассматриваемый метод может применяться лишь для минимизации таких функций, у которых существует обратимая матрица вторых производных, причём она должна быть ограниченной. Такими свойствами обладают сильно выпуклые дважды непрерывно-дифференцируемые функции. Поэтому будем предполагать, что функция $f(x)$ удовлетворяет следующим условиям:

$$d\|h\|^2 \leqslant (f''(x)h, h) \leqslant D\|h\|^2, d > 0, \forall\ x, h \in \mathbb{R}^n \qquad \textbf{(10)}$$

(Такие функции ограничены снизу и у них существует точка минимума)

Теорема 2.

Если для минимизации функции $f(x)$, удовлетворяющей условию (10), используется метод (7), в котором выбор параметра α_k производится из условия (8), то <u>независимо от выбора начальной точки</u> x^0 последовательность $\{x^k\}$ сходится к точке минимума x^* со <u>сверхлинейной</u> скоростью:

$$\|x^{n+k}-x^*\|\leqslant Cq_N\ldots q_{N+k}, \quad ,$$

где константы $\Upsilon\leqslant N, C<\infty, q_{N+k}<1$ при любом $k\geqslant 0,\ q_i\to 0$ при $i\to\infty$.

Предположим теперь, что матрица $f''(x)$ удовлетворяет ещё и условию Липшица (помимо (10)), то есть

$$\|f''(x)-f''(y)\|\leqslant M\|x-y\|,\ M>0,\ x,y\in\mathbb{R}^n \qquad \textbf{(11)}$$

В этом случае будет справедлива

<u>Теорема 3.</u>

Если функция $f(x)$ такая, что выполняются условия (10) и (11),то последовательность (7), в которой α_k выбираются из условия (8) <u>независимо от выбора начальной точки x^0</u> сходится к решению с <u>квадратичной скоростью</u> сходимости, то есть справедлива оценка:

$$\|x^{k+1}-x^*\|\leqslant\frac{M}{d}\|x^k-x^*\|^2 \text{ или}$$

$$\|x^{N+k}-x^*\|\leqslant\frac{d}{M}q_N^{2^k},\ 0<q<1, N<\infty$$

Когда α_k выбирают из условия (9), скорость сходимости будет также <u>сверхлинейной</u> в случае, когда минимизируемая функция $f(x)$ удовлетворяет условию(10) и <u>квадратичной</u>, если выполняется ещё и условие (11).

Сравнение двух способов регулировки длины шага, связанных с проверкой условий (8) или (9), говорит в пользу <u>первого</u>, так как он оказывается менее трудоемким по количеству вычислений функции.

Квазиньютоновские методы

Предположим, что функция f дважды дифференцируема.

Рассмотрим метод: $\begin{array}{l} x_{k+1}=x_k+\alpha_k h_k, \\ h_k=-H_k f'(x_k) \end{array}$ **(1)**

Матрицу H^k будем выбирать таким образом, чтобы она в некотором смысле аппроксимировала матрицу $(f''(x^k))^{-1}$. Заметим, что

$$f'(x^k)-f'(x^{k-1})=f''(x^{k+1})(x^k-x^{k+1})+o(\|x^k-x^{k+1}\|) \quad \textbf{(2)}$$

Предполагая, что $f''(x^{k+1})$ невырождена, с точностью до членов более высокого порядка малости по сравнению с $\|x^k-x^{k+1}\|$ имеем:

$$(f''(x^k))^{-1}(f'(x^{k+1})-f'(x^k))\approx x^{k+1}-x^k \quad \textbf{(3)}$$

При этом, если f (x) - квадратичная функция,

$$f(x)=\frac{1}{2}(Ax,x)+(b,x)$$

A - симметричная положительно определенная матрица, то $f'(x)=Ax+b$, $f''(x)=A$ и (3)

$$A^{-1}(Ax^{k+1}+b-Ax^k-b)=A^{-1}A(x^{k+1}-x^k)=x^{k+1}-x^k$$

превращается в точное равенство:

$$(f''(x^k))^{-1}\Delta y^k=\Delta x^k,$$

где $\Delta y^k=f'(x^{k+1})-f'(x^k), \Delta x^k=x^{k+1}-x^k$

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

$$H_{k+1} \cdot \Delta y^k = \Delta x^k \qquad \textbf{(4)}$$

Это условие носит название квазиньютоновского; методы минимизации, для которых на каждом шаге выполняется условие (4), называются квазиньютоновскими.

Основную сложность при реализации метода Ньютона представляет собой вычисление матрицы вторых производных минимизируемой функции. Следовательно, алгоритмы, которые бы превосходили метод Ньютона по своей эффективности, не должны содержать вычисление 2-х производных, сохраняя тем самым скорость сходимости метода Ньютона.

Возникает вопрос: нельзя ли для построения последовательных приближений к решению строить направления h^k , близкие к направлениям, получаемым в методе Ньютона, используя для этой цели лишь первую производную минимизируемой функции?

Пусть приближение $(f'')^{-1}$ пересчитывается от шага к шагу по формуле:

$$H_{k+1} = H_k + \Delta H_k \qquad \textbf{(4')}$$

Попробуем указать какую-нибудь матрицу ΔH_k , обеспечивающую выполнение условия (4). Для этого перепишем (4) в виде (подставим (4') в (4)):

$$(H_k + \Delta H_k)\Delta y^k = \Delta x^k \Rightarrow \Delta H_k * \Delta y^k = \Delta x^k - H_k \Delta y^k$$

Ясно, что этому равенству удовлетворяет матрица (ранга 1), заданная формулой:

$$\Delta H_k = \frac{1}{(z^k, \Delta y^k)} (\Delta x^k - H_k \Delta y^k) z^k , \qquad \textbf{(5)}$$

где z^k - произвольный вектор такой, что $(z^k, \Delta y^k) \neq 0$.

Выбрав $z^k = \Delta x^k - H_k \Delta y^k$, получаем из (5) следующую формулу для пересчета матриц H_k :

$$H_{k+1} = H_k + \Delta H_k = H_k + \frac{(\Delta x^k - H_k \Delta y^k), (\Delta x^k - H_k \Delta y^k)}{(\Delta x^k - H_k \Delta y^k, \Delta y^k)} \qquad \textbf{(6)}$$

при $(\Delta x^k - H_k \Delta y^k, \Delta y^k) \neq 0.$ Есть и другие формулы для H_{k+1} . В качестве H_0 можно выбрать любую положительно определенную симметричную матрицу. На практике часто выбирается единичная матрица.

Длина шага в квазиньютоновских методах чаще всего выбирается из условия минимизации функции вдоль заданного направления

$$f(x^k + \alpha_k h^k) = \min_{\alpha} f(x^k + \alpha_k h^k) \qquad \textbf{(7)}$$

Иногда берут $\alpha_k = 1$ как в классическом методе Ньютона или α_k выбирается в процессе дробления шага. Оказывается, что для квадратичной функции

$$f(x) = \frac{1}{2}(Ax, x) + (b, x),$$

где A - положительно определенная матрица, метод (1), (6), (7) при любом начальном приближении $x^0 \in \mathbb{R}^n$ генерирует последовательность точек $x_1, x_2, \ldots, x_n,$ причем

$$H_k = (f''(x^n))^{-1} = A^{-1}, x^n = x^* = -A^{-1} b = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x),$$

т.е. квазиньютоновский метод позволяет найти минимум квадратичной функции за n шагов.

Для неквадратичных функций это, конечно, не так. Однако можно показать, что при соответствующих предположениях

$$H_k - (f''(x^n))^{-1} \to 0,\ x^k \to x^*,$$

причем скорость сходимости сверхлинейна. Так, например, если f -дважды непрерывно дифференцируемая функция, сильно выпуклая на R^n, то при любом начальном приближении последовательность точек $\{x^k\}$, определяемая формулами (1), (7) и некоторый аналог (6) сходится к x^*. Если при этом для всех x, таких, что $f(x^k) < f(x^0)$ справедливы неравенства

$$\left| \frac{\partial^2 f(x)}{\partial x_i \partial x_j} - \frac{\partial^2 f(x^0)}{\partial x_i \partial x_j} \right|_{i,j=1,\dots,n} \leqslant M \|x - x^*\|,$$

то x^k сходится к x^* сверхлинейно.

Квазиньютоновские методы называют также методами переменной метрики. Это объясняется тем, что любая симметричная положительно определенная матрица H_k задает скалярное произведение $(u,v)_k = (H_k u, v)$ и связанную с ним метрику. Линейная часть приращения $f(x^k + \Delta x^k) - f(x^k)$ имеет вид

$$(f'(x^k), \Delta x^k) = (H_k \cdot H_k^{-1} f'(x^k), \Delta x^k) = (H_k^{-1} f'(x^k), \Delta x^k)_k,$$

поэтому вектор $H_k^{-1} f'(x^k)$ можно рассматривать как градиент функции *f* в точке x^k в пространстве со скалярным произведением $(\cdot\ ,\ \cdot)_k$. Т.о. метод (1) является обобщением градиентного метода на случай пространства с переменной метрикой.

Квазиньютоновские методы отличает высокая скорость сходимости; при их реализации не нужно выполнять такие трудоемкие операции как вычисление матрицы вторых производных или обращение матрицы. Однако при большой размерности пространства необходимость хранения и пересчета на каждом шаге матриц H_k требует большого объема памяти. Это относиться к недостаткам метода.

Метод Давидона – Флетчера – Пауэлла – пример квазиньютоновского метода, в котором H_{k+1} строится по следующему закону:

1. Пшеничный Б.Н., Данилин Ю.М. “Численные методы в экстремальных задачах” Москва, Наука-1975, 319ст.
2. Гилл, Мюррей, Райт “Практическая оптимизация” Москва, Мир-1985, 509ст.

Метод сопряженных направлений

Рассмотрим квадратичную функцию

$$f(x)=\frac{1}{2}(Ax,x)+(b,x), \qquad \textbf{(1)}$$

где A – симметрическая положительно определенная матрица размера n x n .

Идея метода сопряженных направлений основана на стремлении минимизировать функцию (1) за конечное число шагов. Точнее, в методе сопряженных направлений требуется найти направления $h^{0,} h^{1,} \dots, h^{n-1}$ такие, что последовательность n одномерных минимизаций вдоль этих направлений приводит к отысканию минимума функции (1), т.е.

$$f(x^n)=\min_{x\in\mathbb{R}^n} f(x)$$

при любом $x^0 \in \mathbb{R}^n$, где

$$x^{k+1} = x^k + \alpha_k h^k, \; f(x^k + \alpha_k h^k) = \min f(x^k + \alpha h^k), \; k = 0, 1, \ldots \qquad \textbf{(2)}$$

Оказывается, таким свойством обладает система взаимно сопряженных относительно матрицы A направлений.

<u>**Определение.**</u> Векторы (направления) h' и h'' называются <u>сопряженными</u> (относительно матрицы A), если отличны от нуля и $(Ah', h'') = 0$.

<u>**Определение.**</u> Векторы $h^0, h^1, \ldots, h^{n-1}$ называются <u>взаимно сопряженными</u> (относительно матрицы A), если все они отличны от нуля и $(Ah^i, h^j) = 0, \; i \neq j, \; 0 \leqslant i, j \leqslant k.$

<u>**Лемма 1**</u>. Пусть векторы $h^0, h^1, \ldots, h^{n-1}$ являются взаимно сопряженными. Тогда они линейно независимы.

<u>**Доказательство:**</u> Пусть это неверно, т.е. $h^i = \sum\limits_{i \neq j} \lambda_j (Ah^i, h^j) = 0$ при некотором i.

Тогда $(Ah^i, h^j) = \sum\limits_{j \neq i} \lambda_j (Ah^i, h^j) = 0$, что возможно только при $h^i = 0$, т.к. A положительно определена.

<u>**Теорема.**</u> Если векторы h^k в методе (2) взаимно сопряжены, $k = 0, 1, \ldots, n-1,$ то для функции (1) и произвольной точки $x^0 \in \mathbb{R}^n$

$$f(x^n) = \min_{x \in \mathbb{R}} f(x),$$

т.е. метод (2) позволяет найти точку минимума квадратичной функции (1) не более, чем за n шагов.

Теперь для получения конкретных алгоритмов минимизации достаточно указать способы построения взаимно сопряженных направлений.

Опишем метод <u>сопряженных направлений 1-го порядка.</u>

Строить систему сопряженных направлений будем по правилу

$$h^0 = -f'(x^0) \qquad \textbf{(3)}$$

$$h^k = -f'(x^0) + \beta_{k-1} h^{k-1}, k \geqslant 1$$

Из условия сопряженности векторов h^k и h^{k-1} имеем:

$$0 = (h^k, Ah^{k-1}) = (-f'(x^0), Ah^{k-1}) + \beta_{k-1}(h^{k-1}, Ah^{k-1}),$$

откуда

$$\beta_{k-1} = \frac{(-f'(x^0), Ah^{k-1})}{(h^{k-1}, Ah^{k-1})} \qquad \textbf{(4)}$$

Обоснование того факта, что метод (2), (3), (4) относиться к числу методов сопряженных направлений, дает специальная лемма.

<u>**Лемма 2.**</u> Пусть $x^0 \in \mathbb{R}^n$, точки $x_1, \dots, x_{n-1}$ и векторы $h^{0,} h^{1,} \dots, h^{n-1}$ получены по формулам (2), (3), (4) и $f'(xi)^1 0, i = 0..n-l.$ Тогда векторы $h^{0,} h^{1,} \dots, h^{n-1}$ взаимно сопряжены, а градиенты $f'(x^0), f'(x^1), \dots, f'(x^{n-1})$ взаимно ортогональны. Поэтому метод (2), (3), (4) обеспечивает отыскание точки минимума функции (1) не более, чем за n шагов.

До сих пор предполагалось, что симметричная матрица A <u>положительно определена</u>. Часто встречаются задачи, в которых предполагается лишь, что $(Ax, x) \geqslant 0$ при всех $x \in \mathbb{R}^n,$ т.е. A – неотрицательно определена.

В <u>этом случае</u> рассматриваемый метод также обеспечивает отыскание тоски минимума не более, чем, за n шагов. Если минимальное значение не достигается, то метод позволяет установить этот факт также не более, чем за n шагов. В последнем случае $(Ah^k, h^k)=0$ при некотором $k \leqslant h-1$ и вычисляемый по формуле (4) коэффициент β_k обращается в ∞, т.е. дальнейшее применение метода становится невозможным.

Сформулируем метод для минимизации неквадратичной функции. Для этого преобразуем формулу (4) так, чтобы в ней не фигурировала матрица А.

<u>Лемма3.</u> Справедливы формулы:

$$\beta_{k-1}=\frac{(f'(x^k), f'(x^k)-f'(x^{k-1}))}{\|f'(x^{k-1})\|^2} \quad \textbf{(5)}$$

$$\beta_{k-1}=\frac{\|f'(x^k)\|^2}{\|f'(x^{k-1})\|^2} \quad \textbf{(6)}$$

<u>Доказательство.</u>

$$x^{k+1}=x^k+\alpha_k h^k \Rightarrow Ax^{k+1}=Ax^k+\alpha_k h^k$$

$$f'(x)=Ax+b, f'(x^{k+1})=Ax^{k+1}+b=A(x^k+\alpha_k h^k)+b=$$

$$=Ax^k+b+\alpha_k Ah^k=f'(x^k)+\alpha_k Ah^k, \Rightarrow$$

$$\beta_{k-1}=\frac{(f'(x^k), Ah^{k-1})}{(h^{k-1}, Ah^{k-1})}=\frac{(f'(x^k), f'(x^k)-f'(x^{k-1}))}{(h^{k-1}, f'(x^k)-f'(x^{k-1}))}$$

т.к. по (3)

$$h^{k-1}=-f'(x^{k-1})+\beta_{k-2}h^{k-2}, \quad \text{то}$$

$$h^{k-2}=-f'(x^{k-2})+\beta_{k-3}h^{k-3}$$

$$h^{k-1}=-f'(x^{k-1})+\beta_{k-2}(-f'(x^{k-2})+\beta_{k-3}h^{k-3})=$$

$$=-f'(x^{k-1})-\beta_{k-2}f'(x^{k-2})+\beta_{k-2}\beta_{k-3}h^{k-3}$$

и т.д.

$$\beta_{k-1}=\frac{(f'(x^k),f'(x^k)-f'(x^{k-1}))}{(-f'(x^{k-1})-\beta_{k-2}f'(x^{k-2})-\ldots,f'(x^k)-f'(x^{k-1}))}$$ = в силу ортогональности векторов

$$f'(x^0),\ldots,f'(x^k)=\frac{(f'(x^k),f'(x^k)-f'(x^{k-1}))}{\|f'(x^{k-1})\|^2}=\frac{\|f'(x^k)\|^2}{f'(x^{k-1})^2}$$

Естественно, для неквадратичной функции этот метод перестаёт быть конечным, вырабатываемые им направления $h^{0,} h^{1,} \ldots$ вообще говоря, не являются взаимно сопряжённым относительно какой-либо матрицы.

Обычно в методе сопряжённых направлений для неквадратичной функции вводится процедура “обновления” коэффициентов: $\beta_{n-1},\beta_{2n-1},\beta_{3n-1}$ не вычисляются по формуле (5) или (6), а полагаются равными нулю, при этом $h^i=-f'(x^i), i=1,2n,\ldots$ Такая процедура позволяет уменьшить влияние погрешностей при решении одномерных задач. В этом случае метод сопряжённых направлений принимает вид:

$$x^{k+1}=x^k+\alpha_k h^k, f(x^k+\alpha_k h^k)=\min_{\alpha\geqslant 0} f(x^k+\alpha h^k), k=0,1,\ldots$$

$$h^0=-f'(x^0);\ h^k=-f'(x^k)+\beta_{k-1}h^{k-1}, k\geqslant 1$$

$$\beta_{k-1}=\begin{cases}\dfrac{f'(x^k),f'(x^k)-f'(x^{k-1})}{\|f'(x^{k-1})\|^2}, k\notin\{n,2n,3n,\ldots\}\\ 0, \qquad k\in\{n,2n,3n,\ldots\}\end{cases}$$

Полученный метод носит название метода сопряжённых градиентов и может применятся для минимизации неквадратичной функции.

- При применении этого метода для минимизации ограниченой снизу функции f, градиент которой удовлетворяет условию Липшица, для любой начальной точки $x^0 \in \mathbb{R}^n \|f'(x^k)\| \to 0, k \to \infty$

- Для сильно выпуклой гладкой функции, удовлетворяющей некоторым дополнительным ограничениям, $\{x^k\}$ сходится к точке минимума x^* со сверх линейной скоростью.

Таким образом метод сопряжённых градиентов обладает высокой скоростью сходимости. Трудоёмкость сравнительно не велика, незначительно превосходит метод наискорейшего спуска.

Это позволяет отнести метод сопряжённых градиентов к числу наиболее эффективных алгоритмов первого порядка. Он незначительно уступает по эффективности квазиньютоновским методам, предъявляя в то же время меньшие требования к объёму памяти.