ГЛАВА VI

МЕТОДЫ ОПТИМИЗАЦИИ

§1. Постановка задачи оптимизации. Необходимые и достаточные условия экстремума

Говоря о задачах оптимизации выделяют несколько общих моментов:

- Определяют некоторую "скалярную" (что важно для нас) меру *качества* — *целевую функцию* "Φ".
- Определяют набор *независимых* переменных и формулируются условия, которые характеризуют их приемлемые значения (размерность задачи и её ограничения).
- Решение оптимизационной задачи — это приемлемый набор значений переменных, которому отвечает *оптимальное* значение целевой функции.

Под *оптимальностью* (в нашем рассмотрении) обычно понимают *минимальность* целевой функции.

Пусть $x \in \mathcal{M}$ — элемент метрического пространства $\mathcal{M}$ и с помощью ограничений выделено множество $\mathcal{X} \subseteq \mathcal{M}$.

Говорят, что целевая функция $\Phi(x)$ *имеет локальный минимум на элементе* $x^* \in \mathcal{X}$, *т. е.* $x^* \in loc\min\limits_{\mathcal{X}} \Phi(x)$, *если существует некоторая конечная* ϵ-*окрестность точки* x^* — *шар* $K_\epsilon(x^*)$, *такая, что*

$$\Phi(x^*) < \Phi(x), \quad \forall x \in K_\epsilon(x^*), \quad 0 < \rho(x, x^*) < \varepsilon.^{*1)} \tag{1}$$

У функции $\Phi(x)$ может быть несколько локальных минимумов — множество $loc\min\limits_{\mathcal{X}} \Phi(x)$. Если же в этом множестве существует точка $x^* \in loc\min\limits_{X} \Phi$, в которой достигается *наименьшее* значение функции

$$\Phi(x^*) = \inf_{\mathcal{X}} \Phi(x), \tag{2}$$

то говорят о достижении в т. x^* *абсолютного* минимума[*2)].

[*1)] В случае (1) говорят о *строгом минимуме* (в смысле неравенства), тогда как $\Phi(x^*) \leq \Phi(x)$ при $\rho(x, x^*) < \epsilon$ говорят о нестрогом минимуме.

[*2)] Мы ограничимся именно рассмотрением такого случая, когда глобальный экстремум функции совпадает с одним из её локальных минимумов.

Относительно целевой функции $\Phi(x)$ естественно требовать её непрерывности, хотя и не всегда; а относительно множества $\mathcal{X}$ — *компактности* и *замкнутости* этого множества. Напомним:
Множество $\mathcal{X}$ *— компактно, если из каждого его бесконечного и ограниченного подмножества можно выделить сходящуюся последовательность точек.*
Множество $\mathcal{X}$ *— замкнуто, если предел любой сходящейся последовательности точек* $\{x_n\}$ *из* $\mathcal{X}$ *принадлежит* $\mathcal{X}$.

В частности при $\mathcal{X} = \mathcal{M}$ само $\mathcal{M}$ должно быть *банаховым* пространством.

Мы ограничимся рассмотрением задачи (1) о *локальном* экстремуме. Задача (2) решается выбором наименьшего из соответствующих локальных минимумов.

Второе существенное ограничение — это рассмотрение задачи минимизации без ограничений, т. е. $\mathcal{X} = \mathcal{M}$:

1) $\mathcal{X} = R^1$ — задача минимизации функции одного переменного;

2) $\mathcal{X} = R^n$ — задача минимизации функциии n переменных;

3) $\mathcal{X}$ — гильбертово пространство и задача о минимизации функционала (скалярная целевая функция).

С решением задачи (1) в предположении соответсвующей гладкости целевой функции $\Phi(x)$ связывают *необходимое* условие экстремума (Эйлера)

$$\left.\frac{\delta\Phi}{\delta x}\right|_{x^*} = 0. \tag{3}$$

Для случая одного переменного это условие приводит к одному нелинейному уравнению

$$\Phi'(x) = 0.$$

В случае n-мерной задачи мы получаем систему нелиненых уравнений

$$\frac{\partial\Phi}{\partial x_k}(x_1, ..., x_n) = 0.$$

В случае задачи минимизации функционала $\Phi(x)$ уравнение (3), как правило, дифференциальное или интегро-дифференциальное уравнение. Например, для функционала

$$\Phi(x) = \int_a^b F(t, x(t), \dot{x}(t)dt)$$

получаем

$$\delta\Phi = 0 \quad \Leftrightarrow \quad \begin{cases} \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}}\right) - \frac{\partial F}{\partial x} = 0 \\ + \text{краевые условия} \end{cases} \quad \begin{matrix}\text{уравнение} \\ \text{Эйлера-Лагранжа}\end{matrix}$$

Численное решение задачи (3) — отдельная самостоятельная проблема (частично нами решенная). Как правило здесь используются итерационные методы, обладающие своими достоинствами и недостатками. Нас же будут интересовать в основном методы безусловной минимизации (1), не связанные прямо с решением необходимого условия (3).

§2. Минимум функции одного переменного

2.1 Постановка задачи одномерной минимизации

Рассмотрим задачу безусловной минимизации функции одного переменного:
Требуется найти т. $x^* \in R$ такую, что

$$\Phi(x^*) = \min \Phi(x) \quad \Leftrightarrow \quad x^* \in loc \min_{x \in R} \Phi(x). \tag{4}$$

Если функция $\Phi(x) \in C^{(2)}(R)$ дважды непрерывно дифференцируема, то известны необходимое и достаточное условия минимума:

$$\begin{array}{ccc} \begin{array}{c}\text{необходимое}\\ \text{условие}\\ \text{экстремума}\end{array} & \qquad & \begin{array}{c}\text{достаточное}\\ \text{условие}\\ \text{экстремума}\end{array} \\ \Phi'(x^*) = 0 & & \Phi'(x^*) = 0 \\ \Phi''(x^*) \geq 0 & & \Phi''(x^*) > 0 \end{array} \tag{5}$$

(Взятые по отдельности — это соответствующие условия оптимальности точки x^* первого и второго порядков как необходимые, так и достаточные).

В таком случае, при нахождении в достаточно малой окрестности точки x^*, разложение целевой функции в ряд Тейлора с центром в точке x^* имеет вид

$$\Phi(x^* + h) = \Phi(x^*) + \underbrace{\Phi'(x^*)\, h}_{\equiv 0 \text{в силу (5)}} + \frac{1}{2!}\Phi''(x^*)\, h^2 + o(h^2).$$

Мы говорим о *невырожденности минимума* в точке x^*, если $\Phi''(x^*) \neq 0$, тем самым, согласно (5), $\Phi''(x^*) > 0$. В дальнейшем будем предполагать это условие выполненным.

Подчеркнем еще раз, что мы пытаемся рассмотреть способы минимизации непосредственно задачи (4), а не решение задачи (5) из необходимого условия экстремума. Хотя, конечно, это тесно связанные проблемы.

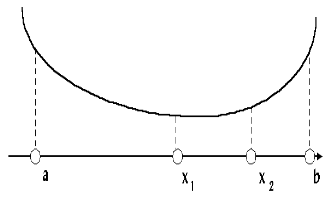
2.2 Методы минимизации нулевого порядка

Под методами минимизации *нулевого порядка* подразумевают группу методов не использующих явно производные целевой функции.

Предположим, что точки a и b определяют, возможно и достаточно грубо, интервал, где расположено значение точки минимума x^* задачи (4). Если считать, что внутри этого интервала функция $\Phi(x^*)$ *унимодальна*, т. е. имеет единственный минимум, то одна из возможностей построения последовательности стягивающихся отрезков $x^* \in [x_{k-1}, x_k]$, локализующих x^* возможна на основании:
а)*метода дихотомии* — (разобрать самостоятельно!);
б)*метода "золотого сечения"*.

Пусть на $[a, b]$ даны две внутренние точки x_1 и x_2. Сравним значение целевой функции $\Phi(x^*)$ в точках $\{a; x_1; x_2; b\}$ и выберем из них наименьшее. Пусть это $\Phi(x_1)$.

Тогда минимум функции $\Phi(x)$ — точка x^* — расположен на одном из соседних с точкой x_1 отрезков и отрезок $(x_2, b]$ можно из рассмотрения удалить.

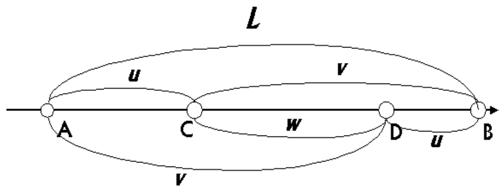


Теперь, на отрезке $[a, x_2]$, одна точка (x_1) — уже есть. Добавим следующую — x_3 и повторим отбор.

Как выгодно размещать точки? В методе *"золотого сечения"* поступают так: Напомним,

Точка C делит отрезок AB по правилу "золотого сечения", если отношение длины всего отрезка к длине его большей доли такое же, как и отношение длины большей доли к длине меньшей его доли.

$$\begin{cases} \frac{L}{v} = \frac{v}{u} \\ u + v = L. \end{cases}$$



Обозначив отношение длин отрезков через τ, найдем:

$$\begin{cases} \frac{AB}{CB} = \frac{L}{v} = \frac{v}{u} = \tau > 1 \\ \qquad u + v = L \end{cases} \quad \Leftrightarrow \quad \tau^2 - \tau - 1 = 0 \quad \Leftrightarrow \quad \tau = \frac{1+\sqrt{5}}{2} \approx 1,62. \tag{6}$$

Если далее на большем отрезке отложить меньший отрезок $u = BD$ от точки D (или от точки C, то точка D делит BC по правилу золотого сечения, ибо $\frac{CB}{BD} = \frac{v}{u} = \tau$, причем BD — "большая" доля. Заметим, что $AD = CB$ всегда.

Итак, пусть длины L_k частичных отрезков, локализующих экстремум x^*, строились по правилу "золотого сечения". Это позволяет написать

$$\frac{L}{L_1} = \frac{L_1}{L_2} = \ldots = \frac{L_{k-1}}{L_k} = \frac{L_k}{L_{k+1}} = \ldots = \tau = \frac{1+\sqrt{5}}{2}$$
$$L_{k-1} = L_k + L_{k+1}.$$

Перемножая, найдем

$$\frac{L}{L_1} \cdot \frac{L_1}{L_2} \cdot \ldots \cdot \frac{L_{k-1}}{L_k} = \tau^k \quad \Leftrightarrow \quad L_k = \frac{L}{\tau^k}. \tag{7}$$

Таким образом последовательность длин частичных отрезков $L_k \to 0$, при $k \to \infty$. Мы получили, что итерационная процедура (6-7) — процедура первой степени точности. Сходимость метода "золотого сечения" линейная (со скоростью геометрической прогрессии).

Замечание. Формулы (6-7) дают достаточно медленную сходимость. Много раз вычисляется целевая функция $\Phi(x)$, но мало используется информация о самих значениях функции $\Phi(x)$ на предыдущих шагах метода, только сравнение "$>$" или "$<$" при выборе очередной точки последовательности $\{L_k\}$.

2.3 Методы более высокого порядка (метод парабол)

Использование информации о значениях $\Phi(x)$ или её производных позволяет аппроксимировать $\Phi(x)$ многочленом в окрестности точки x_k, при этом а) если $\Phi(x) \in C^{(2)}$, то из формулы Тейлора

$$\Phi(x) = \underbrace{\Phi(x_k) + \Phi'(x_k)(x - x_k) + \frac{1}{2!}\Phi''(x_k)(x - x_k)^2}_{\Psi(x)} + o\left((x - x_k)^2\right)$$

В качестве x_{k+1} берется точка экстремума квадратичной функции $\Psi(x)$ т. е. $\Psi'(x_{k+1}) = 0$. Получим

$$\Phi'(x_k) + (x - x_k)\Phi''(x_k) = 0 \quad \Leftrightarrow \quad x_{k+1} = x_k - \frac{\Phi'(x_k)}{\Phi''(x_k)}. \tag{8}$$

Мы получили итерационный процесс, совпадающий с методом Ньютона поиска корня уравнения $\Phi'(x) = 0$. Поэтому (8) обеспечивает не хуже, чем "квадратичную" сходимость в достаточно малой окрестности невырожденного экстремума x^*. Но на каждой итерации необходимо вычислять производные $\Phi'(x_k)$ и $\Phi''(x_k)$ целевой функции.

б) Если не прибегать к вычислению $\Phi'(x_k)$ и $\Phi''(x_k)$, то по трем точкам x_{k-2}, x_{k-1}, x_k и соответствующим значениям$\Phi(x_{k-2})$, $\Phi(x_{k-1})$ и $\Phi(x_k)$ можно построить интерполяционный многочлен Ньютона (парабола)

$$N_2(x) = \Phi(x_k) + (x - x_k)\Phi(x_k, x_{k-1}) + (x - x_k)(x - x_{k-1})\Phi(x_k, x_{k-1}, x_{k-2})$$

и вычислить x_{k+1} как точку экстремума $N_2(x)$, т.е. координату "вершины" параболы.

$$N_2'(x) = 0 \Leftrightarrow \Phi(x_k, x_{k-1}) + (2x - (x_k + x_{k-1}))\Phi(x_k, x_{k-1}, x_{k-2}) = 0$$

$$x_{k+1} = \frac{x_k + x_{k-1}}{2} - \frac{\Phi(x_k, x_{k-1})}{\Phi(x_k, x_{k-1}, x_{k-2})} = \left\{ \begin{matrix} \text{получить} \\ \text{самостоятельно} \\ \text{окончательную} \\ \text{расчётную} \\ \text{формулу} \end{matrix} \right\}. \tag{9}$$

Замечания:

В обоих случаях обязательна проверка условия $\Phi(x_{k+1}) < \Phi(x_k)$ для исходной целевой функции.

Сходимость метода парабол (9) выше чем линейная, но не квадратичная:

$$|x_{k+1} - x^*| \le C\,|x_k - x^*|^{\sim 4/3}, \quad \lim_{k\to\infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|} = 0.$$

§3. Минимум функции многих переменных

3.1 Постановка задачи. Необходимые и достаточные условия экстремума

1) Если целевая функция $\Phi(x) \equiv \Phi(x_1, \ldots, x_n)$, $x \in R^n$, то минимизация $\Phi(x)$ приводит к задаче:

$$x^* \in loc \min_{R^n} \Phi(x); \qquad \Phi(x^*) = \min_{K_\varepsilon(x^*)} \Phi(x). \tag{10}$$

Введем в рассмотрение градиент и гессиан функции Φ:

$$\vec{g}(x) = grad\Phi = \vec{\nabla}\Phi = \left\{ \frac{\partial \Phi}{\partial x_1}, \ldots, \frac{\partial \Phi}{\partial x_n} \right\},$$

$$G(x) = \operatorname{hess}\Phi(x) = \left\| \frac{\partial^2 \Phi}{\partial x_i \partial x_j} \right\| \quad \text{— симметричная матрица вторых производных } \Phi(x_1, \ldots, x_n).$$

Тогда разложение функции $\Phi(\vec{x})$ в ряд Тейлора в окрестности точки $\vec{x}$ при $\Delta\vec{x} = h\vec{p}$; $\|\vec{p}\| = 1$; $h = \Delta x$ имеет вид:

$$\Phi(\vec{x} + h\vec{p}) = \Phi(\vec{x}) + d\Phi(\vec{x}) + \frac{1}{2} d^2\Phi(\vec{x}) + o\left(\|\delta x\|^2\right) =$$

$$= \Phi(\vec{x}) + (grad\Phi, \vec{p})\, h + \frac{h^2}{2} \underbrace{(\operatorname{hess}\Phi(\vec{x}) \cdot \vec{p}, \vec{p})}_{(G\vec{p}, \vec{p})} + o\left(h^2\right).$$

Величина $(\vec{g}(\vec{x}), \vec{p}) \equiv \frac{\partial \Phi}{\partial p}$ — производная Φ в точки $\vec{x}$ по направлению $\vec{p}$; $(G\vec{p}, \vec{p}) = (\operatorname{hess}\Phi\vec{p}, \vec{p}) \equiv K(\vec{p})$ — "кривизна" поверхности $u = \Phi(\vec{x})$ в точке $\vec{x}$ по направлению $\vec{p}$.

2)Необходимые и достаточные условия минимума для дважды дифференцируемой функции $\Phi(x_1, \ldots, x_n)$. Напомним

необходимое условие экстремума	достаточное условие экстремума	
$\|\vec{g}(x^*)\| = 0$ $\operatorname{hess}\Phi(x^*) \geq 0$	$\|\vec{g}(x^*)\| = 0$ $\operatorname{hess}\Phi(x^*) > 0^{*1)}$	(11)

*1) матрица $A > 0$, если $\forall x \neq 0 (Ax, x) > 0$ — положительно определенная квадратичная форма.

3.2 Квадратичная функция аргумента $\vec{x}$

Опираясь на тейлоровское разложение естественно в качестве удобной апроксимации гладкой функции $\Phi(x)$ в окрестности некоторой точки (в том числе и точки возможного экстремума) использовать квадратичную функцию $\Psi(\vec{x})$:

$$\Psi(\vec{x}) = \frac{1}{2}(A\vec{x}, \vec{x}) + \left(\vec{b}, \vec{x}\right) + c,$$

где A — симметричная, невырожденная матрица $A = A^T$, $detA \neq 0$.
Установим вид градиента $\vec{\nabla}\Psi$ и гессиана $G = \text{hess}\,\Psi$ функции $\Psi(\vec{x})$:

$$\Psi(\vec{x} + h\vec{p}) = \frac{1}{2}(A(\vec{x} + h\vec{p}), \vec{x} + h\vec{p}) + \left(\vec{b}, \vec{x} + h\vec{p}\right) + c =$$

$$= \left\{\frac{1}{2}(Ax, x) + (b, x) + c\right\} + h\left\{\frac{1}{2}(A\vec{x}, \vec{p}) + \frac{1}{2}(A\vec{p}, \vec{x}) + \left(\vec{b}, \vec{p}\right)\right\} + \frac{h^2}{2}(A\vec{p}, \vec{p}) =$$

$$= \Psi(\vec{x}) + h\left(A\vec{x} + \vec{b}, \vec{p}\right) + \frac{h^2}{2}(A\vec{p}, \vec{p}) \text{ , т.е.}$$

$$grad\Psi = A\vec{x} + \vec{b}; \quad \text{hess}\,\Psi(\vec{x}) = A \text{ — } const. \tag{12}$$

Стационарная точка для $\Psi(\vec{x})$ удовлетворяет условию:

$$grad\Psi(x^*) = 0 \quad \Leftrightarrow \quad Ax^* + b = 0 \quad \Leftrightarrow \quad Ax^* = -b \text{ — СЛАУ} \tag{13}$$

Решение системы (13) зависит от ранга матрицы A. В случае совместной системы решение может быть и неединственным.

В окрестности стационарной точки $\vec{x}^*$:

$$\Psi(\vec{x}) = \Psi(\vec{x}^* + h\vec{p}) = \Psi(\vec{x}^*) + \frac{h^2}{2}(A\vec{p}, \vec{p}).$$

И поведение квадратичной функции определяется только свойствами матрицы A. Если A — симметричная невырожденная матрица, то существует ортонормированный базис (ОНБ) из собственных векторов матрицы A. Пусть $\{\lambda_i, \vec{x}_i\}$ — собственные значения и собственные векторы матрицы A, $\{\vec{x}_i\}$ — ОНБ. Разложим направление $\vec{p}$ по базису $\{\vec{x}_i\}$ — $\vec{p} = \sum\limits_{i=1}^{n} \alpha_i \vec{x}$, тогда

$$\Psi(\vec{x}^* + h\vec{p}) = \Psi(\vec{x}^*) + \frac{h^2}{2}\left(\sum_i \alpha_i\lambda_i\vec{x}_i, \sum_i \alpha_j\vec{x}_j\right) = \Psi(\vec{x}^*) + \frac{h^2}{2}\sum_i \lambda_i\alpha_i^2. \tag{14}$$

Характер изменения $\Psi(\vec{x})$ при движении вдоль $\vec{x}_k$ полностью определяется знаком λ_k. Если $A > 0$, то все $\lambda_i > 0$ и x^* — точка минимума.

3.3 Рельеф поверхности целевой функции $\Phi(x)$. Поверхности уровня

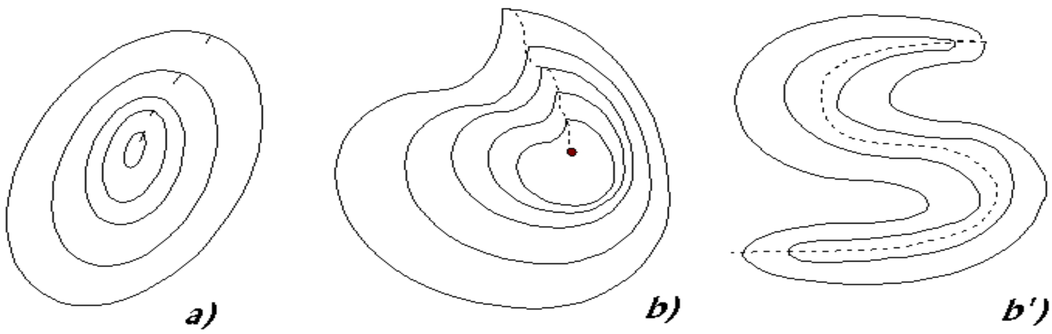
Трудности и проблемы задачи минимизации, характерные для общего случая, столь же ясно проявляются и при рассмотрении минимизации функции двух переменных $\Phi(x,y)$. Геометрию поверхности $z = \Phi(x,y)$ представляют с помощью "плоских" линий уровня

$$L_0 = \{(x,y): \quad \Phi(x,y) = \Phi(x_0,y_0) = \Phi_0 = const\},$$

являющихся проекциями на плоскость OXY сечения поверхности $z = \Phi(x,y)$ плоскостью $z_0 = \Phi_0$.

Выделяют три основных типа рельефа поверхности.

а) *котловинный* — линии уровня похожи на концентричекские эллипсы с главными осями параллельными собственным векторам $\text{hess}\,\Phi(x,y)$ В малой окрестности невырожденного минимума (x^*,y^*) $\text{hess}\,\Phi(x,y) > 0$ и рельеф поверхности именно *котловинный*.



б) *овражный* — если линия уровня кусочно-гладкая, то геометрическое место точек (ГМТ) излома по всем линиям уровня называют *истинным оврагом* (если угол излома направлен в сторону возрастания функции) или *истинным гребнем* (если угол излома направлен в сторону убывания функции).

Однако чаще приходится иметь дело с *разрешимыми* оврагами и гребнями (ГМТ наибольшей кривизны — рисунок b')). Например, одна из стандартных тестовых функций многомерной минимизации (функция Розенброка)

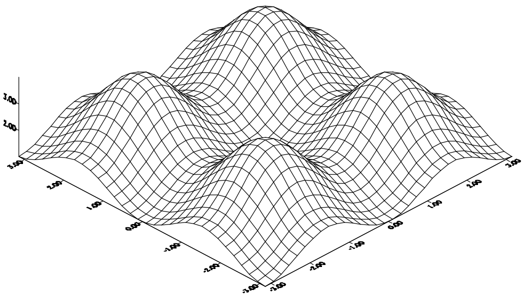
$$\Phi(x,y) = 100\left(y - x^2\right)^2 + (1-x)^2$$

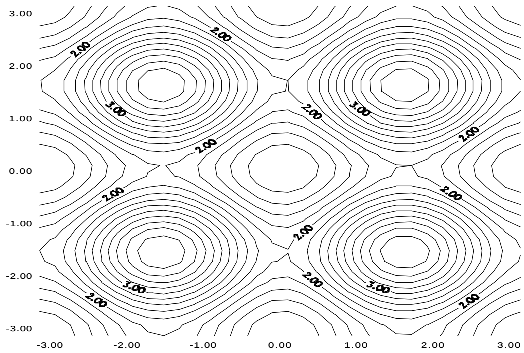
обладает пологим серповидным ("банановидным") ущельем и имеет абсолютный минимум в точке $x^*(1,1)$.

в) *неупорядоченный тип рельефа* — характеризуется наличием многих максимумов, минимумов и седловин. Приведем в качестве примера функцию

$$\Phi(x,y) = \left(1 + \sin^2 x\right)\left(1 + \sin^2 y\right)$$

с достаточно неупорядоченным рельефом:





Если рассматривать дифференцируемую в каждой точке функцию $\Phi(\vec{x})$, то её производная по направлению $\vec{p}$

$$\frac{\partial \Phi}{\partial p} = (grad\Phi, \vec{p}) = \vec{g} \cdot \vec{p}$$

обладает характерными свойствами на поверхности уровня

–производная по направлению градиента — максимальна;

–вдоль линии уровня $\frac{\partial \Phi}{\partial p}$ равна нулю и градиент $\vec{g}$ перпендикулярен линии уровня в каждой точке.

3.4 Спуск по координатам

Все методы минимизации сводятся к построению траектории спуска $\{M_k\}$ вдоль которой целевая функция убывает:

$$\Phi(M_{k+1}) < \Phi(M_k)$$

(или не возрастает).

Опишем *координатный спуск.* Выберем нулевое приближение $M_0(\overset{(0)}{x}_1, ..., \overset{(0)}{x}_n)$ и зафиксируем все значения координат, кроме первой. Тогда $\Phi(\vec{x})$

$$\Phi(x_1, \overset{(0)}{x}_2, ..., \overset{(0)}{x}_n) \equiv \varphi_1(x_1)$$

становится функцией одного переменного.

Используя методы минимизации функции одного переменного, найдем точку её минимума $\overset{(1)}{x}_1$ и совершим шаг из M_0 в $M_0^{(1)}(\overset{(1)}{x}_1, \overset{(0)}{x}_2, ..., \overset{(0)}{x}_n)$.

На k-м шаге спуска: Из точки $M_0^{(k-1)}(\overset{(1)}{x}_1, ..., \overset{(1)}{x}_{k-1}, \overset{(0)}{x}_k, ..., \overset{(0)}{x}_n)$ спускаемся по x_k минимизируя

$$\varphi_k(x_k) \equiv \Phi\left(\overset{(1)}{x}_1, ..., \overset{(1)}{x}_{k-1}, x_k, \overset{(0)}{x}_{k+1}, ..., \overset{(0)}{x}_n\right),$$

$$\overset{(1)}{x}_k: \quad \varphi(\overset{(1)}{x}_k) = \min_{x_k} \varphi_k(x_k) \tag{15}$$

в точку

$$M_0^{(k)}\left(\overset{(1)}{x}_1, ..., \overset{(1)}{x}_k, \overset{(0)}{x}_{k+1}, ..., \overset{(0)}{x}_n\right).$$

И так до тех пор, пока не выполним один цикл спуска по координатам. Последнюю точку спуска назовем $M_1 \equiv M_0^{(n)}(\overset{(1)}{x}_1, ..., \overset{(1)}{x}_n) \equiv M_1(\overset{(0)}{x}_1, ..., \overset{(0)}{x}_n)$. Траектория $\{M_k\}$ — траектория спуска, поскольку

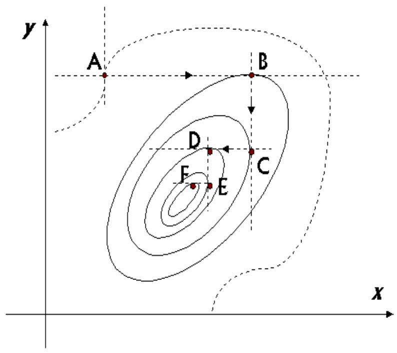
$$\Phi\left(M_k\right) \leq \Phi\left(M_{k-1}\right).$$

В силу ограниченности снизу значений $\Phi(x)$ значением $\Phi\left(x^*\right) \equiv \Phi^*$ (мы предполагаем, что экстремум существует), то

$$\Phi_k \geq \Phi^* \Rightarrow \lim_{k\to\infty} \Phi_k = \tilde{\Phi} \geq \Phi^* \; (!)$$

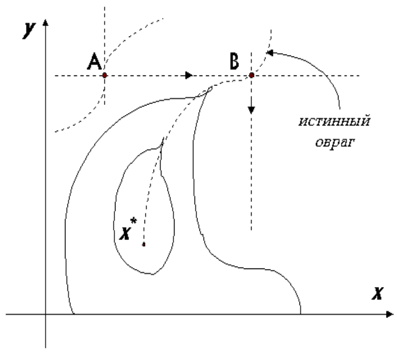
Будет ли здесь равенство, т.е. сходится ли спуск по координатам к минимуму и как быстро, зависит от функции $\Phi\left(\vec{x}\right)$ и выбранного начального приближения $\vec{x}_0$ (оно должно попасть в область влияния локального экстремума).

Рассмотрим трактовку координатного спуска на примере функции двух переменных:



Двигаясь по прямой AB мы пересекаем линии уровня $(x, y) = const$, при этом $\Phi\left(x, y\right)$ либо возрастает, либо убывает в зависимости от направления движения. Только в одной точке B, где данная прямая касается линии уровня, функция $\Phi\left(x, y\right)$ имеет минимальное значение в данном направлении (экстремум по x или по y). Найдя такую точку, завершаем спуск по данному направлению.

Заметим, что в координатном спуске соответствующие направления взаимно ортогональны.



Если в рельефе наличествует "истинный" овраг, то спуск (в данном случае первый же спуск в точку B) приводит к попаданию на "дно" оврага. А поскольку он ориентирован достаточно произвольно, то дальнейший спуск может оказаться невозможным.

Хотя минимум еще и не достигнут!

Если же $\Phi\left(\vec{x}\right)$ достаточно гладкая функция и минимум невырожден, $\operatorname{hess}\Phi\left(\vec{x}^*\right) > 0$, то в окрестности $\vec{x}^*$ рельеф котловинный и координатный спуск ведет нас к локальному минимуму при произвольном начальном приближении $\vec{x}_0$ в этой окрестности.

Рассмотрим достаточные условия сходимости координатнрого спуска на примере функции двух переменных:

Теорема 1. *Пусть D - множество уровня, ограниченное линией уровня $\Phi(x, y) = \Phi_0$, т.е.*

$$D = \left\{(x, y): \quad \Phi(x, y) \leq \Phi(x_0, y_0)\right\},$$

замкнутая ограниченная область и в D функция $\Phi\left(x, y\right)$ дважды дифференцируема, причем

$$\Phi_{xx} \geq a > 0;\; \Phi_{yy} \geq b > 0;\; |\Phi_{xy}| \leq c \text{ и } ab > c^2. \;^{*1)} \qquad (15)$$

Тогда траектория координатного спуска $\{M_k\}$ (14) из произвольной точки $M_0 \in D$ сходится к локальному минимуму x^* в области D.

Доказательство. Докажем сходимость $grad\Phi(M_k)$ на траектории спуска $\{M_k\}$. Проследим за изменением $|\Phi_x|$ и $|\Phi_y|$ на траектории спуска $\{M_k\}$. Поскольку $\Phi(x,y)$ вдоль траектории спуска не возрастает, то все точки $M_k \in D_0$. Пусть предыдущий цикл спусков закончился в точке A, тогда

$$\Phi_y(A) = 0, \quad |\Phi_x(A)| = U \neq 0.$$

Попав в точку экстремума B на прямой AB получим следующие компоненты градиента

$$|\Phi_y(B)| = V \neq 0, \quad \Phi_x(B) = 0.$$

Теперь нетрудно получить, что

$$\begin{cases} U = |\Phi_x(B) - \Phi_x(A)| = |\Phi_{xx}(\xi)| \cdot |x_B - x_A| \geq a \cdot \rho(A,B) \\ V = |\Phi_y(B) - \Phi_y(A)| = |\Phi_{xy}(\eta)| \cdot |x_B - x_A| \leq c \cdot \rho(A,B) \end{cases} \Rightarrow \quad Uc \geq Va.$$

Спустившись далее по направлению BC в точку экстремума C, найдём

$$\begin{cases} V = |\Phi_y(C) - \Phi_y(B)| = |\Phi_{yy}(\xi)| \cdot |y_C - y_B| \geq b \cdot \rho(B,C) \\ W = |\Phi_x(C) - \Phi_x(B)| = |\Phi_{xy}(\eta)| \cdot |y_C - y_B| \leq c \cdot \rho(B,C) \end{cases} \Rightarrow \quad Vc \geq Wb.$$

Окончательно, за один цикл спуска, получаем

$$W \leq \frac{c}{b} V \leq \frac{c^2}{ab} U = q \cdot U,$$

причём, в силу условий теоремы (15), $q < 1$.

Итак, за один цикл спусков $|\Phi_x|$ уменьшился в q раз. Аналогично, со сдвигом на 1/2 цикла, $|\Phi_y|$ уменьшится в q раз. Выполнив n циклов координатного спуска получим, что

$$|\Phi_x|_{(n)} \leq q^n |\Phi_x|_{(0)} \implies |\Phi_x| \xrightarrow[n\to\infty]{} 0 \text{ и } |\Phi_y| \xrightarrow[n\to\infty]{} 0.$$

Далее, в окрестности точки экстремума x^* компоненты градиента можно разложить по формуле Тейлора

$$\left.\begin{aligned} \Phi_x(M) &= \underbrace{\Phi_x(M^*)}_{\equiv 0} + \tfrac{\partial \Phi_x}{\partial x}(M^*) \cdot \Delta x + \tfrac{\partial \Phi_x}{\partial y}(M^*) \cdot \Delta y + \cdots \\ \Phi_y(M) &= \underbrace{\Phi_y(M^*)}_{\equiv 0} + \tfrac{\partial \Phi_y}{\partial x}(M^*) \cdot \Delta x + \tfrac{\partial \Phi_y}{\partial y}(M^*) \cdot \Delta y + \cdots \end{aligned}\right| \quad \begin{aligned} \Delta x &= x - x^* \\ \Delta y &= y - y^*. \end{aligned}$$

Пренебрегая в разложении слагаемыми высших порядков, получаем линейную систему относительно приращений координат Δx и Δy. По условию теоремы (15) гессиан $G(M^*) > 0$, тем самым полученная система совместна и можно выразить Δx и Δy через линейную комбинацию компонент градиента в точке $M = M_{(n)}$. При этом $\Delta x, \Delta y \to 0$ на траектории $\{M_k\}$, $M_k \to M^*$.

Итак:

---

*1) $G(x,y) \geq d > 0$ в D. Используя критерий Сильвестра можно сформулировать многомерный аналог этого условия.

- Вблизи точки экстремума M^* сходимость координатного спуска и по координатам, и по градиенту *линейная* (достаточно медленная, что с практической точки зрения плохо);
- по "циклам" спусков можно делать ускорения по методу Эйткена;
- При попадании траектории спуска в разрешимый овраг расчет практически невозможен (слишком медленная сходимость при произвольной ориентации оврага относительно координатных осей). Поэтому выгоднее использовать методы, обладающие повышенным порядком точности.

3.5 Градиентные методы минимизации

В общем случае для траектории спуска $\{M_k\}\ :\ \Phi_{k+1} < \Phi_k$ при минимизации достаточно гладких функций можно сформулировать *достаточные* условия сходимости соответствующего метода спуска, характеризующие изменение функции Φ и её градиента $\vec{g} = grad\Phi$ на траектории $\{M_k\}$.
Пусть очередной шаг совершается вдоль направления $\overrightarrow{p_k}$ и приводит нас в точку M_{k+1}:

$$\overrightarrow{x}_{k+1} = \overrightarrow{x_k} + \overrightarrow{p_k} h_k.$$

Шаг h_k выбирается из условия минимальности $\Phi(M)$ вдоль $\overrightarrow{p_k}$

$$h_k :\ \varphi(h_k) = \min_h \varphi(h) = \min_h \Phi(\overrightarrow{x_k} + h\overrightarrow{p_k}).$$

Сформулируем достаточные условия сходимости метода спуска.
Теорема 2. *Пусть*

1) $\Phi(\overrightarrow{x})$ *– дважды дифференцируемая функция;*

2) множество уровня

$$D(\Phi(\overrightarrow{x_0})) = \{\overrightarrow{x}\ :\ \Phi(\vec{x}) \le \Phi(\vec{x}_0)\}$$

ограничено и замкнуто;

3) на каждой итерации

а) направление $\overrightarrow{p_k}$ *– "существенное направление спуска":*

$$\exists \beta < 0, \quad \overrightarrow{p_k}\,\overrightarrow{g_k} \le \beta < 0$$

б) $\Phi(x)$ *"существенно убывает" (т.е. выбрано соответствующее ограничение на шаг):*

$$\exists \mu_1, \quad \mu_2 :\ 0 < \mu_1 \le \mu_2 \le 1$$

$$-\mu_1 h_k \overrightarrow{g_k} \cdot \overrightarrow{p_k} \le \Phi_k - \Phi_{k+1} \le -\mu_2 h_k \underbrace{\overrightarrow{g_k} \cdot \overrightarrow{p_k}}_{\text{отриц. число}}$$