- 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):*

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

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

Тогда

$$\lim_{k\to\infty} \|\overrightarrow{g_k}\| = 0;\ (M_k \to M^*)$$

т.е. метод спуска обладает сходимостью (как правило — линейной).

В основном соответствующие методы спуска отличаются выбором очередного направления $\overrightarrow{p_k}$ и шага h_k:

Метод "наискорейшего" спуска. Рассмотрим линейную аппроксимацию целевой функции $\Phi(\overrightarrow{x})$ в окрестности точки $\overrightarrow{x_k}$. Опираясь на формулу Тейлора:

$$\Phi(\overrightarrow{x_k} + \overrightarrow{p}) = \Phi(\overrightarrow{x_k}) + (grad\Phi(\overrightarrow{x_k}), \overrightarrow{p}) + o(\|\overrightarrow{p}\|),$$

с определенной точки зрения (локально!) естественно искать напрвление, по которому $\frac{\partial\Phi}{\partial p} \equiv \overrightarrow{g_k} \cdot \overrightarrow{p}$ наибольшее по модулю отрицательное число. Это направление в первом порядке по $\|\overrightarrow{p}\|$ обеспечивает наибольшее убывание функции Φ.

Итак, необходимо найти направление $\vec{p}$

$$\begin{cases} \min(\overrightarrow{g_k} \cdot \overrightarrow{p}) \\ \|\overrightarrow{p}\| = 1 \end{cases} \quad — \quad \begin{array}{c}\text{задача на}\\ \text{условный}\\ \text{экстремум}\\ \text{для } \vec{p}\end{array}$$

Решение полученной задачи зависит от вида рассматриваемой нормы. Если выбрать C-энергетическую норму $\|\overrightarrow{p}\|^2 = (C\overrightarrow{p}, \overrightarrow{p})$, где $C > 0$ и симметрична, тогда направление $\overrightarrow{p}$ (с точностью до нормировочной $Const$)

$$\overrightarrow{p} = -C^{-1} \cdot \overrightarrow{g_k}.^{*1)}$$

Для евклидовой нормы — $C \equiv E$ и $p = -\overrightarrow{g_k}$, что приводит нас к *методу наискорейшего спуска.*

$$\begin{cases} \overrightarrow{x}_{k+1} = \overrightarrow{x_k} - h_k\overrightarrow{g_k} \\ h_k:\ \varphi(h_k) = \min\limits_h \Phi(\overrightarrow{x_k} - h\overrightarrow{g_k}) \end{cases} \tag{16}$$

Замечания:

1) При таком выборе $\overrightarrow{p_k}$ и h_k (16) траектория спуска перпендикулярна линии уровня $\Phi(x_k)$ в точке x_k.

2) По сходимости *наискорейший спуск* не лучше, чем координатный спуск, т.е. он обладает лишь линейной сходимостью.

3) Анализ сходимости наискорейшего спуска на квдратичной функции с симметричной и положительно определенной матрицей (что характерно для гессиана в окрестности невырожденного минимума)

$$\Psi(x) = \frac{1}{2}(Ax, x) + (\overrightarrow{b}, x) + C:\ A > 0,\ A^T = A$$

*1)Показать!

дает лишь линейную сходимость. Поскольку $A > 0, A^T = A$ следовательно все собственные значения матрицы A положительны $\forall i \quad \lambda_i(A) > 0$. Сходимость метода наискорейшего спуска характеризуют величиной

$$\varkappa = \frac{\lambda_{max}(A)}{\lambda_{min}(A)} = \|A\| \cdot \|A^{-1}\| = CondA$$

$$\Psi(x_{k+1}) - \Psi(x^*) \simeq \left(\frac{\varkappa - 1}{\varkappa + 1}\right)^2 (\Psi(\overrightarrow{x_k}) - \Psi(x^*)) . \tag{17}$$

Полученная оценка скорости сходимости, например для $\varkappa = 100$ (хорошая обусловленность матрицы A) даёт $q \approx 0,96(!)$ и нужны сотни итераций для уменьшения погрешности на порядок.

Расчетные формулы наискорейшего спуска (16) в этом случае принимают вид:

$$\overrightarrow{g} = A\overrightarrow{x} + \overrightarrow{b};\ hess\Psi = A \quad \Rightarrow \quad \overrightarrow{p_k} = -\overrightarrow{g_k},$$

$$\psi(h) = \Psi(\overrightarrow{x} + h\overrightarrow{p}_k) = \Psi(\overrightarrow{x}) + h(Ax + b, \overrightarrow{p}_k) + \frac{h^2}{2}(A\vec{p}_k, \vec{p}_k),$$

$$\frac{\partial\psi}{\partial h} = 0 \Leftrightarrow h_k = \left\{ \begin{array}{c} \text{получить} \\ \text{самостоятельно} \\ \text{расчетные} \\ \text{формулы} \end{array} \right\} . \tag{18}$$

Тем не менее:

1) Необходимо бесконечное число итераций для нахождения экстремума даже в случае квадратичной функции.

2) Метод наискорейшего спуска не рекомендуется как серьезная минимизационная процедура. Дело в том, что свойство наискорейшего спуска является лишь *локальным* свойством, поэтому необходима частая смена направлений спуска и относительно малый шаг движения по каждому направлению, что и приводит в итоге к неэффективной вычислительной процедуре (например в случае разрешимого оврага).

3) Метод наискорейшего спуска невозможно адаптировать для использования информации о вторых производных $\Phi(\overrightarrow{x})$.

3.6 Методы второго порядка

Ньютоновские методы. Эта группа методов основана на более точной аппроксимации целевой функции в окрестности точки $\overrightarrow{x_k}$

$$\Phi(\overrightarrow{x_k} + \overrightarrow{p}) = \underbrace{\Phi(\overrightarrow{x_k}) + \overrightarrow{g_k} \cdot \overrightarrow{p} + \frac{1}{2}(G_k\overrightarrow{p}, \overrightarrow{p})}_{\Psi(\overrightarrow{p})} + o\left(\|\overrightarrow{p}\|^2\right) .$$

Минимизируемая функция $\Psi(\overrightarrow{p})$. Соответствующее направление и шаг берут из условия минимума $\Psi(\overrightarrow{p})$:

$$\left.\begin{array}{rl} grad\Psi = 0 \quad \Leftrightarrow \quad G_k\overrightarrow{p} + \overrightarrow{g_k} = 0; \quad \Leftrightarrow & \underbrace{\overrightarrow{p_k} = -G_k^{-1}\cdot\overrightarrow{g_k}}_{\text{Ньютоновское направление}} \\ & \overrightarrow{x}_{k+1} = \overrightarrow{x_k} + \overrightarrow{p_k} = x_k - G_k^{-1}\cdot\overrightarrow{g_k} \end{array}\right\} \qquad (19)$$

- Для квадратичной целевой функции $\Psi(\overrightarrow{p})$ метод (19) решает задачу минимизации за одну(!) итерацию.
- В окрестности невырожденного экстремума имеет *квадратичную* сходимость (гессиан $G_k > 0$ и симметричен).
- Ньютоновское направление – это направление *наискорейшего* спуска в G-энергетической метрике
$$\|\overrightarrow{p}\| = \sqrt{(G\overrightarrow{p}, \overrightarrow{p})}.$$
- Существенным является то, что на каждом шаге необходимо решать систему линейных уравнений (19) для определения *ньтоновского направления* очередной итерации.
- При модификации метода Ньютона, когда гессиан фиксируется на определенное число итераций G_{k_0} — в методе Ньютона-Рафсона — существенен алгоритмический выйгрыш, но при этом обеспечена лишь линейная сходимость метода.

Метод сопряженных градиентов. Методы *координатного спуска* или *наискорейшего спуска* требовали даже для минимизации квадратичной функции бесконечного числа итераций.

Опираясь на тейлоровское разложение в окрестности невырожденного экстремума x^* выгодно строить методы спуска, которые, по крайней мере, эффективны для квадратичных функций.

Такими методами, не требующими решения СЛАУ (19) на каждом итерационном шаге для определения направления спуска, являются методы *сопряженных направлений*.

Для квадратичной функции $\Psi(\overrightarrow{x})$:

$$\Psi(x) = \frac{1}{2}(Ax, x) + (b, x) + c, \quad A > 0, \quad A^T = A$$

они позволяют не более чем за n шагов спуска получить её минимум. Напомним:

Симметричная положительноопределенная матрица $A > 0$, $A^T = A$ – позволяет ввести "A-энергетическю" норму вектора

$$\|x\|_A = \sqrt{(Ax, x)}$$

и соответствующеее скалярное произведение

$$(x, y)_A = (Ax, y) = (x, Ay).$$

Определение *Векторы, ортогональные в A-энергетическом смысле, называются сопряженнымы относительно матрицы A.*

$$x \underset{A}{\perp} y \quad \Leftrightarrow \quad (x,y)_A = (Ax,y) = (x,Ay) = 0.$$

Сопряженные векторы обладают рядом "хороших" свойств:

1) *Если* $\{x_i\}_k$ *– система сопряженных векторов и* $k \le n$*, то эта сисетма векторов — линейно независима.*

   Действительно, пусть $\overrightarrow{x_1} = \sum\limits_{i=2}^{k} \alpha_i \overrightarrow{x_i}$ – ненулевая комбинация остальных векторов. Тогда

   $$(x_1, Ax_1) = (x_1, A\sum_{i=2}^{k} \alpha_i \overrightarrow{x_i}) = \sum_{i=2}^{k} \alpha_i (x_1, Ax_i) \equiv 0$$

   но $A > 0$ и следовательно $\overrightarrow{x_1}$ нулевой вектор, что невозможно ■

2) Если число векторов в рассматриваемой системе $k = n$, то $\{x_i\}_n$ – сопряженный базис. Можно считать его сопряженным ОНБ, т.е. $(x_i, x_j)_A = \delta_{ij}$. Разложим направление $\overrightarrow{p}$ по ОНБ $\{x_i\}_n$ и рассмотрим квадратичную функцию на этом направлени

   $$\Psi(\overrightarrow{x} + \overrightarrow{p}) = \Psi(\overrightarrow{x}) + (Ax + b, \overrightarrow{p}) + \frac{1}{2}(A\overrightarrow{p}, \overrightarrow{p}) = \left| p = \sum \alpha_i \overrightarrow{x_i} \right| =$$

   $$= \Psi(\overrightarrow{x}) + (Ax + b, \sum_i \alpha_i \overrightarrow{x_i}) + \frac{1}{2}\left(A\sum_i \alpha_i \overrightarrow{x_i}, \sum_k \alpha_k \overrightarrow{x_k} \right) =$$

   $$= \underbrace{\sum_i \left\{ \frac{1}{2}\alpha_i^2 + \alpha_i (Ax + b, x_i) \right\}}_{n \text{ независимых слагаемых}} + \Psi(\overrightarrow{x}); \qquad (*)$$

   Движение по каждому из сопряженных направлений x_i изменяет только одно слагаемое в сумме $(*)$ и, тем самым, за не более, чем n шагов приводит к минимуму функции Ψ.

Существуют различные способы построения сопряженных относительно A направлений, в частности – метод *сопряженных градиентов* (метод Флетчера-Ривса) – приводит к одной из наиболее эффективных процедур многомерной численной минимизации.

Рассмотрим снова квадратичную аппроксимацию $\Psi(x)$ целевой функции $\Phi(x)$ в окрестности точки $\overrightarrow{x_k}$:

$$\Phi(\overrightarrow{x_k} + \overrightarrow{p}) = \underbrace{\Phi(x_k) + (grad\Phi(x_k), \overrightarrow{p}) + \frac{1}{2}(hess\Phi(x_k)\overrightarrow{p}, \overrightarrow{p})}_{\Psi_k(\overrightarrow{p})} + o\left(\|\overrightarrow{p}\|^2\right).$$

На каждом *цикле* итерационных шагов для построения *сопряженного базиса* будем использовать одну и ту же матрицу $G_k \equiv hess\Phi(x_k)$. При этом мы будем считать, что находимся в достаточно малой окрестности точки минимума x^*, где $G(x_k) > 0$.

В методе *сопряженных градиентов* совокупность сопряженных относительно $G \equiv G(x_k)$ направлений строится следующим образом. Опишем процедуру построения одного цикла минимизации, содержащего n шагов и точно минимизирующего $\Psi_k(\overrightarrow{p})$.

$$
\begin{array}{ll}
\text{Цикл движения} & M_k \equiv \overset{(1)}{M_k} \xrightarrow[\overrightarrow{p_1}]{} \overset{(2)}{M_k} \xrightarrow[\overrightarrow{p_2}]{} \cdots \xrightarrow[\overrightarrow{p}_{n-1}]{} \overset{(n)}{M_k} \xrightarrow[\overrightarrow{p_n}]{} M_{k+1} \\
\text{1-ый шаг:} & \overrightarrow{p_1} = -\overrightarrow{g_1};\ \overset{(2)}{x_k} = \overset{(1)}{x_k} + h_1 \overrightarrow{p_1};\ h_1 : \psi(h_1) = \min\limits_h \Psi_k \left(\overset{(1)}{x_k} + h \overrightarrow{p_1} \right); \\
\text{2-ой шаг:} & \overrightarrow{p_2} = -\overrightarrow{g_2} + \alpha_1 \overrightarrow{p_1};\ \alpha_1 = \frac{(g_2, g_2)}{(g_1, g_1)};\ \overrightarrow{p_2} \perp \overrightarrow{p_1} \text{ отн-но } G_k
\end{array}
\tag{20}
$$

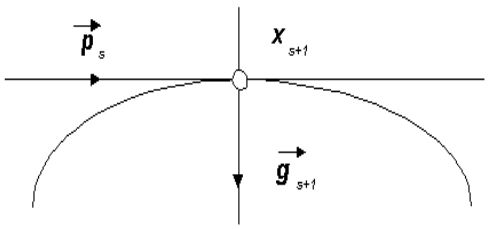
Пусть $\overrightarrow{p_1}, \ldots, \overrightarrow{p_s}$ G_k-сопряженная система векторов

$$
\left.
\begin{array}{ll}
S+1\text{-ый шаг:} & \overrightarrow{p}_{S+1} = -\overrightarrow{g}_{S+1} + \alpha_S \overrightarrow{pS} \text{ (все остальные} \alpha_i = 0) \\
& \alpha_S = \frac{g^2_{S+1}}{g^2_S};\ \text{(из сообр. } (\overrightarrow{p}_{S+1}, \overrightarrow{pS})_{G_k} = 0) \\
& \vec{x}_{S+2} = \overrightarrow{x}_{S+1} + h_{S+1} \cdot \overrightarrow{p}_{S+1} \\
& h_{S+1} :\ \psi(h_{S+1}) = \min\limits_h \Psi_k \left(\overrightarrow{x}_{S+1} + h \overrightarrow{p}_{S+1} \right)
\end{array}
\right\}
$$

Покажем, что (20) определяет систему сопряженных относительно G_k векторов движения $\{\vec{p}_s\}_n$.

а) Проверить самостоятельно 2-й шаг;

б) 1: $\overrightarrow{g}_{S+1}$ ортогонально всем предыдущим $\overrightarrow{p_j}$ при $j \le s$, ибо спускаясь на предыдущем, S-ом шаге, мы пришли в точку $\vec{x}_{S+1} = \vec{x}_S + h_S \overrightarrow{pS}$ вдоль направления $\overrightarrow{pS}$.



Но эта точка — $\overrightarrow{x}_{S+1}$ — точка "минимума", т.е. $\overrightarrow{g}_{S+1} \perp \overrightarrow{pS}$, $(\overrightarrow{g}_{S+1}, \overrightarrow{pS}) = 0$. Если проследить "вглубь" траектории, то

$$
\vec{x}_{S+1} = \vec{x}_S + h_S \vec{p}_S = \vec{x}_{S-1} + h_{S-1} \vec{p}_{S-1} + h_S \vec{p}_S = \cdots = \vec{x}_{j+1} + \sum_{j+1}^{S} h_i \overrightarrow{p_i}, \quad 1 \le j \le S-1.
$$

Тогда

$$
G x_{S+1} = G x_{j+1} + \sum_{j+1}^{S} h_i G \overrightarrow{p_i}.
$$

Добавим слева и справа по $\overrightarrow{b} \equiv \overrightarrow{g}(M_k)$, и учтём, что $G_k \equiv G$; $Gx + b \equiv \overrightarrow{g}(x)$. Таким образом

$$\overrightarrow{g}_{S+1} = \overrightarrow{g}_{j+1} + \sum_{j+1}^{S} h_i G p_i.$$

Тогда

$$\overrightarrow{g}_{S+1} \cdot \overrightarrow{p_j} = \underbrace{\overrightarrow{g}_{j+1} \cdot \overrightarrow{p_j}}_{=0 \text{ для этого шага}} + \sum_{j+1}^{S} h_i \cdot \underbrace{(Gp_i, p_j)}_{=0 \text{ в силу индукции}} \quad \Rightarrow$$

$$\Rightarrow \quad (\vec{g}_{S+1}, \vec{p}_j) = 0,\ 1 \le j \le S-1;\ S.$$

2: Покажем, что вектор $\vec{g}_{S+1}$ ортогонален всем градиентам $\vec{g}_j, j = \overline{1,S}$. Имеем

$$\overrightarrow{p_j} = -g_j + \alpha_{j-1}\overrightarrow{p}_{j-1} \quad \Leftrightarrow \quad \underbrace{\overrightarrow{g}_{S+1} \cdot \overrightarrow{p_j}}_{=0} = -(g_{S+1}, g_j) + \underbrace{\alpha_{j-1} \cdot (\overrightarrow{g}_{S+1}, \overrightarrow{p}_{j-1})}_{=0}$$

т.о.

$$(\overrightarrow{g}_{S+1}, \overrightarrow{g_j}) = 0,\ j = \overline{1,S}.$$

3: Рассмотрим очередное направление:

$$\overrightarrow{p}_{S+1} = -\overrightarrow{g}_{S+1} + \alpha_S \overrightarrow{pS}; \quad \alpha_S = \frac{g_{S+1}^2}{g_S^2}$$

и покажем, что $\overrightarrow{p}_{S+1}$ сопряжено всем $\overrightarrow{p_j}$, $j \le S$. Оно сопряжено, по крайней мере, со всеми $\overrightarrow{p}$ до предыдущего, т.е. $(\overrightarrow{p}_{S+1}, \overrightarrow{p_j})_{G_k} = 0$, $j = \overline{1, S-1}$. Действительно, поскольку $j \le S-1$, то

$$(\overrightarrow{p}_{S+1}, G\overrightarrow{p_j}) = \left(-\overrightarrow{g}_{S+1} + \alpha_S \overbrace{\overrightarrow{pS}, G\overrightarrow{p_j}}^{\text{сопряжены}}\right) = -\left(g_{S+1}, G\frac{x_{j+1} - x_j}{h_j}\right) =$$

$$= -\left(g_{S+1}, \frac{(Gx_{j+1} + b_k) - (Gx_j + b_k)}{h_j}\right) = -\left(g_{S+1}, \frac{g_{j+1} - g_j}{h_j}\right) \equiv 0.$$

Предыдущее направление:

$$(\overrightarrow{p}_{S+1}, G\overrightarrow{pS}) = -(g_{S+1}, Gp_S) + \alpha(p_S, Gp_S) =$$

$$= -\left(g_{S+1}, G\frac{x_{S+1} - x_S}{h_S}\right) + \alpha_S\left(-g_S + \alpha_{S-1}\overrightarrow{p}_{S-1}, \frac{g_{S+1} - g_S}{h_S}\right) =$$

$$= -\frac{g_{S+1}^2}{h_S} + \frac{g_{S+1}^2}{\not{g}_S^2}\frac{\not{g}_S^2}{h_S} = 0 \ \blacksquare$$

Метод Флетчера-Ривса обладает квадратичной сходимость в достаточно малой окрестности точки $\vec{x}^*$. Рестарт в точке M_k осуществляется по антиградиенту $(-\overrightarrow{g_k})$.

Это один из наиболее эффективных методов численной минимизации функций многих переменных.

§4. Задача минимизации функционала

4.1 Постановка задачи

Если любому $y(x) \in Y$ поставлено в соответствие число , то говорят, что на множестве функций Y задан функционал $\Phi[y(x)]$.

В задаче минимизации функционала требуется: *найти* $y^*(x) \in Y$, *на которой функционал достигает своей точной нижней грани (абсолютный экстремум)*

$$y^* : \quad \Phi^* \equiv \Phi[y^*(x)] = \inf_Y \Phi[y(x)]. \tag{21}$$

В такой постановке (21) называется задачей *минимизации по аргументу*, в отличие от

$$\Phi^* \equiv \inf_Y \Phi[y(x)] \tag{21'}$$

задачи *минимизации значений* функционала.

Не всякий функционал и не на всяком множестве имеет минимум. Скажем, если функционал неограничен снизу на заданном множестве, или соответствующее множество некомпактно в себе, или если функционал разрывен и т.д. Мы не будем исследовать постановку задачи (21), а будем предполагать, что (21) поставлена корректно, то есть её решение $y^*(x)$ на Y существует, единственно и устойчиво относительно малых возмущений входных данных.

Постановка задачи (21) возникает, как правило, когда сама модель сформулирована соответствующим образом, например рассматривается функционал "действия" (или нечто похожее)

$$\Phi[y(x)] = \int_a^b F(x, y, y', ..., y^{(p)})\, dx. \tag{*}$$

Обычно к задаче (21) приводит использование вариационных методов решения "операторного" уравнения $A[y(x)] = f(x)$. Рассмотренный нами *метод наименьших квадратов* даёт задачу минимизации функционала "невязки" для этого уравнения

$$\Phi[y(x)] = \| Ay - f \|^2 = \int_a^b (Ay(x) - f(x))^2 \rho(x) dx, \quad \rho(x) > 0.$$

В случае *некорректно* поставленной задачи $A[y(x)] = f(x)$ её *регуляризация* приводит к задаче минимизации *сглаживающего функционала Тихонова*

$$M_\alpha[y(x)] = \| Ay - f \|^2 + \alpha\Omega[y(x)],$$

где $\Omega[y(x)]$ — функционал со свойствами нормы (то есть с его помощью на Y вводится структура нормированного пространства и множество $\Omega[y] \le Const$ компактно в Y в введенной метрике). Тогда решение задачи минимизации $y_\alpha(x)$ при определённом способе согласования параметра регуляризации α с априорной информацией о