

# Лекция 14

Луя Yaroshevskiy

13 мая 2023 г.

## Содержание

<b>1 Минимизация квадратичной функции</b>	<b>1</b>
1.1 Минимизация с использованием исчерпывающего спуска	3
1.2 Метод сопряженных направлений	3
1.2.1 Условие ортогональности	4
1.2.2 Рассмотрим минимизацию	5
1.2.3 Уточнения	6
1.2.4 Общая схема	6
1.3 Модификации	7
1.3.1 Метод сопряженных градиентов	7
1.3.2 Метод Флетчера-Ривса	7
1.3.3 Метод Полака-Рибьера	7
1.4 Выводы	8

## 1 Минимизация квадратичной функции

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

- $A$  — симметричная матрица
- $A = H$  — матрица Гессе  $f(x)$

$$\nabla f(x) = Ax + b$$

Если  $A$  невырождена, то  $f(x)$  в силу необходимого условия экстремума имеет единственную стационарную точку  $x^* = -A^{-1}b$ .  $x^*$  будет точкой наименьшего значения  $f(x)$  тогда и только тогда, когда квадратичная форма  $\langle Ax, x \rangle$  положительно определена

Пусть  $x^* = 0$ , то есть минимум квадратичной функции в начале координат (вектор  $b = 0$ ), тогда будем рассматривать

$$f(x) = \frac{1}{2} \langle Ax, x \rangle \quad x \in E_n \tag{1}$$

$A$  — положительно определена. Тогда квадратичная функция  $\frac{1}{2} \langle Ax, x \rangle$  неотрицательная в  $E_n$  и достигает наименьшего значения 0 в единственной точке  $x^* = 0$

Рассмотрим метод градиентного спуска. Для  $f(x)$  из 1  $\nabla f(x) = Ax$  в точке  $x$ . Пусть  $x^0 \neq 0$ , тогда  $w^1 = -\nabla f(x^0) = -Ax^0$

$$\begin{aligned} x^k &= x^{k-1} + \alpha_k w^k \quad k \in \mathbb{N} \\ x^1 &= x^0 + \alpha_1 w^1 = x^0 - \alpha_1 Ax^0 = (I - \alpha_1 A)x^0 \end{aligned} \tag{2}$$

Из 2 следует:

- точку минимума квадратичной функции можно достичь за одну итерацию, если  $x^0$  — собственный вектор матрицы  $A \implies$  если  $x^0$  — собственный вектор матрицы  $A$ , а  $\lambda_j$  — его собственное значение, то

$$Ax^0 = \lambda_j x^0 \text{ и } (A - \lambda_j I)x^0 = 0$$

Если  $\alpha_1 - \frac{1}{\lambda_j}$

$$x^1 = (I - \frac{1}{\lambda_j} A)x^0 = -\frac{1}{\lambda_j}(A - \lambda_j I)x^0 = 0$$

тогда  $x^1$  совпадает с точкой минимума  $x^* = 0$

В 2-мерном случае  $f(x) = \frac{1}{2} \langle Ax, x \rangle$  — эллиптический параболоид с центром в начале координат. Метод градиентного спуска приведет в точку  $(0, 0)$  за одну итерацию, если начальная точка выбрана на одной из осей эллипсов: радиус-вектор точки является собственным вектором матрицы  $A$ .

- **Частный случай**  $A = \lambda I$  — все собственные значения  $A$  совпадают, а каждый ненулевой вектор  $x \in E_n$  является собственным. Тогда минимум функции  $f(x)$  **1** достигается за одну итерацию при любом выборе  $x^0$

Квадратичную форму **1** невырожденной заменой переменных можно привести к каноническому виду с единичной матрицей. Применим ортогональное преобразование к квадратичной форме, тогда

$$f_1(\xi) = \sum_{j=1}^n \lambda_j \xi_j^2$$

,  $\xi = (\xi_1, \dots, \xi_n)$ ,  $\lambda_j$  — положительные собственные значения  $A$  ( $j = \overline{1, n}$ )

Изменим масштабы переменных, введем замену

$$\eta_j = \sqrt{\lambda_j} \xi_j \quad j = \overline{1, n}$$

тогда

$$f_2(\eta) = \sum_{j=1}^n \eta_j^2 \quad \eta = (\eta_1, \dots, \eta_n)$$

В новых переменных  $\eta_1, \dots, \eta_n$  минимум квадратичной функции методом градиентного спуска достигается за одну итерацию при любом выборе начальной точки  $x^0$

- Недостаток — трудоемкие вычисления, сложности решения задачи на собственные значения  $A$
- Преимущества — полезен метод, когда функция имеет овражный характер

Рассмотрим метод градиентного спуска с  $\alpha_k = \alpha = const$  на всех итерациях

На  $k$ -й итерации

$$x^k = x^{k-1} - \alpha \nabla f(x^{k-1}) = x^{k-1} - \alpha A x^{k-1} = (I - \alpha A)x^{k-1} \quad (3)$$

Заключение о сходимости  $\{x^k\}$  можно сделать на основе теоремы о неподвижной точке. Согласно теореме  $\{x^k\}$  сходится к неподвижной точке  $x^*$  отображения  $f(x)$ , если это отображение является **сжимающим**, то есть подчиняется условию Липшица

$$|f(x) - f(y)| \leq g|x - y| \quad g = const < 1 \quad (4)$$

В **3** отображение — линейный оператор, который является сжимающим отображением, если имеет норму, меньше единицы. Норма = спектральная норма (для симметричной матрицы) =  $|\lambda_{\max}|$

В итоге для сходимости  $\{x^k\}$ :

$$x^k = x^{k-1} - \alpha w^k \quad \alpha > 0$$

согласно **3** и теореме о неподвижной точке достаточно выполнения

$$g(\alpha) = \|I - \alpha A\| < 1$$

Для квадратичной функции с положительно определенной матрицей  $A$  с собственными значениями  $0 < \lambda_1 < \dots < \lambda_n$  матрицы  $I - \lambda A$  имеет собственные значения  $1 - \alpha \lambda_j$ ,  $j = \overline{1, n}$ :

$$1 - \alpha \lambda_n < 1 - \alpha \lambda_{n-1} < \dots < 1 - \alpha \lambda_1 < 1$$

Тогда условие

$$g(\alpha) = \|I - \alpha A\| < 1$$

равносильно

$$\begin{cases} 1 - \alpha \lambda_n > -1 \\ 1 - \alpha \lambda_1 < 1 \end{cases} \implies \alpha \in \left(0, \frac{2}{\lambda_n}\right)$$

из теоремы о неподвижной точке следует, что для  $\{x^k\}$  верна оценка  $|x^k - x^*| \leq g^k |x^0 - x^*|$ ,  $g$  — постоянная Липшица (из 4)

Для ускорения сходимости  $\{x^k\}$   $g$  должно быть как можно меньше. В рассматриваемом случае  $g(\alpha) = \min$ , когда собственные значения  $1 - \alpha\lambda_n$  и  $1 - \alpha\lambda_1$  матрицы  $I - \alpha A$  совпадают по абсолютной величине и противоположны по знаку

$$-(1 - \alpha\lambda_1) = 1 - \alpha\lambda_n$$

, откуда оптимальное  $\alpha$ :

$$\alpha^* = \frac{2}{\lambda_1 + \lambda_n} \leq \frac{2}{\lambda_n}$$

Выбирая  $\alpha = \alpha^*$

$$g^* = g(\alpha^*) = \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} = \frac{\text{cond}(A) - 1}{\text{cond}(A) + 1} \quad (5)$$

$$\text{cond}(A) = \frac{\lambda_n}{\lambda_1}$$

$\text{cond}(A) \gg 1$  — овражная структура,  $\{x^k\}$  сходится медленно

- Если  $A = I$ , то все собственные значения = 1, поэтому  $\text{cond}(A) = 1$  и  $g^* = 0$ , тогда минимум достигается за одну итерацию при любом выборе  $x^0$
- Если  $\alpha \notin \left(0, \frac{2}{\lambda_n}\right)$ ,  $\{x^k\}$  не релаксационная, а метод расходится или заикликивается

## 1.1 Минимизация с использованием исчерпывающего спуска

$$f(x) = \frac{1}{2} \langle Ax, x \rangle$$

происходит с нарушением условия  $\alpha_k \in \left(0, \frac{2}{\lambda_n}\right)$  Исчерпывающий спуск на  $k$ -й итерации: поиск стационарной точки  $\varphi_k(\alpha) = f(x^{k-1} + \alpha w^k)$

$$\varphi_k(\alpha) = \frac{1}{2} \langle A(x^{k-1} + \alpha w^k), x^{k-1} + \alpha w^k \rangle = f(x^{k-1}) + \alpha \langle Ax^{k-1}, w^k \rangle + \frac{\alpha^2}{2} \langle Aw^k, w^k \rangle$$

— функция с положительным коэффициентом при старшей степени, поэтому они имеют единственную стационарную точку

$$\alpha_k = -\frac{\langle Ax^{k-1}, w^k \rangle}{\langle Aw^k, w^k \rangle} = \frac{|w^k|^2}{\langle Aw^k, w^k \rangle} \quad (6)$$

Из 6:

- $\alpha_k = \frac{1}{\lambda_n}$ , если  $x^{k-1}$  собственный вектор  $A$  с собственным значением  $\lambda_n$
- $\alpha_k = \frac{1}{\lambda_1}$ , если  $x^{k-1}$  собственный вектор  $A$  с собственным значением  $\lambda_1$

$$\lambda_1 < \lambda_2 < \dots < \lambda_n \quad \frac{1}{\lambda_1} > \frac{1}{\lambda_2} > \dots > \frac{1}{\lambda_n}$$

Если  $\text{cond}(A) = \frac{\lambda_n}{\lambda_1} > 2$ , то  $\alpha_k \in \left(0, \frac{2}{\lambda_n}\right)$  нарушается при  $\alpha_k = \frac{1}{\lambda_1}$

При  $\alpha = \text{const}$  шаг определяется без учета скорости убывания функции. В случае наискорейшего спуска шаг спуска тем больше, чем медленнее функция убывает

*Примечание.* Для квадратичной функции метод наискорейшего спуска эквивалентен градиентному методу с исчерпывающим спуском (т.к. квадратичная функция является строго выпуклой)

## 1.2 Метод сопряженных направлений

$$x^k = x^{k-1} + \alpha_k p^k \quad k \in \mathbb{N} \quad (7)$$

$\alpha_k$  — шаг,  $p^k$  — вектор спуска  $\in E_n$  (может быть не единичным)

Рассмотрим квадратичную функцию (частный случай)  $f(x) = \frac{1}{2} \langle Ax, x \rangle$ ,  $A$  — положительно определенная матрица.  $f(x)$  можно привести к каноническому виду

$$f_2(\eta) = \sum_{j=1}^n (\eta_j)^2 \quad \eta = (\eta_1, \dots, \eta_n) \quad (8)$$

в переменных  $\eta$  метод градиентного спуска сходится за одну итерацию. Минус такого подхода — вычисление градиента к координатах  $\eta$ , приведение к каноническому виду затратная операция. Необходим другой подход: Линейная невырожденная замена переменных в квадратичной форме  $\sim$  переход в  $E_n$  от одного базиса к другому. Тогда приведение к каноническому виду = выбор базиса в  $E_n$ . Если  $A$  симметричная, положительно определенная матрица, то  $\langle x, y \rangle_A = \langle Ax, y \rangle$  — скалярное произведение в  $E_n$ ,  $\langle Ax, x \rangle$  — квадрат евклидовой нормы вектора  $x$  в евклидовом пространстве с указанным скалярным произведением. Приведение квадратичной формы к каноническому виду = выбор ортонормированного базиса в евклидовом пространстве со скалярным произведением  $\langle x, y \rangle_A$ . Плюсы подхода

- позволяет упростить вид квадратичной формы
- вместо решения задачи на собственные значения можно использовать процедуру построения ортогонального базиса (процесс ортогонализации)

### 1.2.1 Условие ортогональности

$$p^1 \neq 0 \quad p^2 \neq 0$$

относительно скалярного произведения  $\langle x, y \rangle_A$  имеет вид

$$\langle Ap^1, p^2 \rangle = 0$$

Такие вектора называются сопряженными относительно положительно определенной матрицы  $A$ , или  $A$ -ортогональными. Направления, определяемые  $p^1$  и  $p^2$ , сопряженные направления. Рассмотрим  $A$ -ортогональный базис  $p^j$ ,  $j = \overline{1, n}$ . В этом базисе  $f(x) = \frac{1}{2} \langle Ax, x \rangle$  имеет канонический вид

$$f_1(\xi) = \lambda_1 \xi_1^2 + \dots + \lambda_n \xi_n^2 \quad \xi = (\xi_1, \dots, \xi_n)$$

,  $\lambda_j > 0$ ,  $j = \overline{1, n}$  — определяются нормами  $\|p^j\|_A$  с введенным скалярным произведением

$$\lambda_j = \frac{1}{2} \|p^j\|_A^2$$

*Примечание.*  $f_1(\xi)$  — половина квадрата нормы вектора  $(\xi_1, \dots, \xi_n)$ . Если  $p^j = 0, 0, \dots, \underset{j}{1}, \dots, 0 \implies$

$$\lambda_j = \frac{1}{2} \|p^j\|_A^2$$

Функция  $f_1(\xi)$  — сепарабельна:

- исчерпывающий спуск в направлении  $p^j$  = минимизация одного из слагаемых такой функции
- последовательность из  $n$  исчерпывающих спусков в направлении  $p^1, \dots, p^n$  все слагаемые  $\implies$  в точку минимума

Выберем  $x^0$ , исчерпывающий спуск для квадратичной функции  $f(x)$  в направлении  $p^1$  = исчерпывающий спуск для  $f_1(\xi)$  в направлении первого базисного вектора и приведет к обнулению первой координаты точки  $x^0$  в базисе  $p^1, \dots, p^n$ . Следовательно  $x^1 = x^0 - \xi_1^0 p^1$ , где  $\xi_1^0$  — первая координата  $x^0$  в ортогональном базисе  $p^j$ :

$$\xi_1^0 = \frac{\langle Ax^0, p^1 \rangle}{\langle Ap^1, p^1 \rangle}$$

Сравним  $x^1 = x^0 - \xi_1^0 p^1$  и  $x^k = x^{k-1} + \alpha_k p^k$ . Повторим схему для всех  $p^j$

**Теорема 1.1.** Точка минимума квадратичной функции  $f(x) = \frac{1}{2} \langle Ax, x \rangle$  с положительно определенной матрицей  $A$  достигается не более чем за  $n$  итераций спуска, если направления спуска задаются векторами  $p^k \in E_n$ , сопряженными относительно матрицы  $A$ , а параметры  $\alpha_k$ , определяющие шаг спуска в 7, вычисляются по формуле исчерпывающего спуска

$$\alpha_k = - \frac{\langle Ax^{k-1}, p^k \rangle}{\langle Ap^k, p^k \rangle} \quad k = \overline{1, n} \quad (9)$$

*Примечание.* Если  $p^j$  и  $-p^j$  в точке  $x^{k-1}$  не определяют направление спуска, то  $\langle Ax^{j-1}, p_j \rangle = 0$ , значит  $\alpha_j = 0 \implies$  спуск в направлении  $p^j$  не производится, количество итераций  $< n$

Координаты произвольного  $x^0$  в  $A$ -ортогональном базисе можно выразить через скалярное произведение:

$$x^0 = \sum_{i=1}^n \frac{\langle Ax^0, p^i \rangle}{\langle Ap^i, p^i \rangle} p^i \quad (10)$$

Если  $x^*$  — точка минимума квадратичной формы с положительно определенной матрицей  $A$ , то

$$\begin{aligned} x^0 - x^* &= \sum_{i=1}^n \frac{\langle A(x^0 - x^*), p^i \rangle}{\langle Ap^i, p^i \rangle} p^i = \sum_{i=1}^n \frac{\langle \nabla f(x^0), p^i \rangle}{\langle Ap^i, p^i \rangle} p^i \\ x^* &= x^0 - \sum_{i=1}^n \frac{\langle \nabla f(x^0), p^i \rangle}{\langle Ap^i, p^i \rangle} p^i \end{aligned} \quad (11)$$

*Примечание.*

$$\langle Ax^0, p^i \rangle = p^i \langle p^i, Ax^0 \rangle = p^i (p^i)^T Ax^0$$

Тогда 10 примет вид:

$$x^0 = \sum_{i=1}^n \frac{p^i (p^i)^T}{\langle Ap^i, p^i \rangle} Ax^0$$

— верно для любого  $x^0$ , значит линейный оператор в  $E_n$  с матрицей

$$\sum_{i=1}^n \frac{p^i (p^i)^T}{\langle Ap^i, p^i \rangle} A$$

является тождественным и

$$A^{-1} = \sum_{i=1}^n \frac{p^i (p^i)^T}{\langle Ap^i, p^i \rangle} \quad (12)$$

таким образом система векторов  $p^1, \dots, p^n$  сопряженных относительно положительно определенной матрицы  $A$ , позволяет построить  $A^{-1}$ . Тогда 11 перепишем в виде

$$\begin{aligned} x^* &= x^0 - \sum_{i=1}^n \frac{\langle \nabla f(x^0), p^i \rangle}{\langle Ap^i, p^i \rangle} p^i = \\ &= x^0 - \sum_{i=1}^n \frac{p^i (p^i)^T}{\langle Ap^i, p^i \rangle} \nabla f(x^0) = x^0 - A^{-1} \nabla f(x^0) \end{aligned}$$

Для квадратичной функции  $f(x)$  выполнение  $n$  итераций исчерпывающего спуска  $\sim$  одному спуску вида

$$x^* = x^0 - A^{-1} \nabla f(x^0)$$

*Примечание.* Использование векторов  $p^j$  — основа метода сопряженных направлений

Различие в способах построения сопряженных векторов — порождает несколько вариантов метода сопряженных направлений

### 1.2.2 Рассмотрим минимизацию

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

- можно использовать любой базис, проводя процесс ортогонализации относительно скалярного произведения  $\langle x, y \rangle_A$
- более эффективно исходить их системы антиградиентов, тем самым объединяя в процессе спуска процесс ортогонализации

Выберем  $x^0 \in E_n$ :

$$w^1 = -\nabla f(x^0) = -Ax^0 - b$$

$p^1 = w^1$ , если  $|p^1| \neq 0$ , то  $p^1$  — направление спуска, иначе  $x^0 = x^*$

Проводим исчерпывающий спуск в направлении  $p^1 = w^1$  (по формуле 6)

$$\alpha_1 = \frac{|w^1|^2}{\langle Aw^1, w^1 \rangle} = \frac{|p^1|^2}{\langle Ap^1, p^1 \rangle}$$

$$x^1 = x^0 + \alpha_1 p^1$$

Вторая итерация:

$$w^2 = -Ax^1 - b$$

если  $|w^2| = 0$ ,  $x^1 = x^*$ , иначе проводим ортогонализацию  $p^1$  и  $w^2$  относительно скалярного произведения  $\langle x, y \rangle_A$ :

$$p^2 = w^2 - \frac{\langle Ap^1, w^2 \rangle}{\langle Ap^1, p^1 \rangle} p^1 \quad (13)$$

$\langle w^2, p^1 \rangle = 0$  поскольку  $w^2$  и  $p^1$  — антиградиенты на двух последовательных итерация исчерпывающего спуска, а значит  $w^2$  и  $p^1$  — линейно независимы, а  $|p^2| \neq 0$

Вектор  $p^2$  — направление спуска из  $x^1$ , учитывая ортогональность  $w^2$  и  $p^1$ :

$$\langle \nabla f(x^1), p^2 \rangle = -\langle w^2, \beta_1 p^1 + w^2 \rangle = -\langle w^2 w^2 \rangle = -|w^2|^2 < 0$$

$$\beta_1 = -\frac{\langle Ap^1, w^2 \rangle}{\langle Ap^1, p^1 \rangle}$$

Продолжим процесс исчерпывающего спуска вдоль очередного направления, полученного корректировкой антиградиента в текущей точке, получим  $n$  сопряженных направлений спусков и достигнем точки минимума через  $n$  итераций (или раньше)

### 1.2.3 Уточнения

- каждый  $w^k$  ортогонален не только предпоследнему направлению спуска  $p^{k-1}$ , но и всем  $p^i$ ,  $i = \overline{1, k-2}$

$$w^j = -Ax^{j-1} - b \implies w^k - x^{k-1} = -A(x^{k-1} - x^{k-2}) = -\alpha_{k-1} Ap^{k-1}$$

при  $k-1 > j$ :

$$\langle w^k, p^j \rangle = \langle w^{k-1}, p^j \rangle - \alpha_{k-1} \langle Ap^{k-1}, p^j \rangle = \langle w^{k-1}, p^j \rangle$$

следовательно

$$\langle w^k, p^i \rangle = \langle w^{k-1}, p^i \rangle = \dots = \langle w^{i+1}, p^i \rangle = 0$$

Вывод:  $w^k$  и  $w^i$ ,  $k > i$  — ортогональны, т.к.  $w^i$  — линейная комбинация  $p^1, \dots, p^n$ , ортогональных  $w^k$

- антиградиент  $w^k$  сопряжен со всеми  $p^i$ ,  $i < k-1$ :

$$\alpha_i \langle Ap^i, w^k \rangle = \langle w^i - w^{i+1}, w^k \rangle = \langle w^i, w^k \rangle - \langle w^{i+1}, w^k \rangle = 0$$

Из 1, 2 следует, что процесс ортогонализации

$$\begin{aligned} p^k &= w^k - \sum_{i=1}^{k-1} \frac{\langle Ap^i, w^k \rangle}{\langle Ap^i, p^i \rangle} p^i \implies \\ \implies p^k &= w^k - \frac{\langle Ap^{k-1}, w^k \rangle}{\langle Ap^{k-1}, p^{k-1} \rangle} p^{k-1} \end{aligned}$$

*Примечание.* главный плюс при использовании в процессе ортогонализации последовательности антиградиентов

### 1.2.4 Общая схема

для  $k=1$ :  $x^0$

$$w^1 = -Ax^0 - b$$

$$p^1 = w^1$$

$$\alpha_1 = \frac{|p^1|^2}{\langle Ap^1, p^1 \rangle}$$

для  $k > 1$ :

$$\begin{cases} w^k = -Ax^{k-1} - b \\ p^k = w^k - \frac{\langle Ap^{k-1}, w^k \rangle}{\langle Ap^{k-1}, p^{k-1} \rangle} p^{k-1} \\ x^k = x^{k-1} + \alpha_k p^k \end{cases} \quad (14)$$

$\alpha_k$  — определять из условия исчерпывающего спуска, например

$$\alpha_k = \frac{\langle w^k, p^k \rangle}{\langle Ap^k, p^k \rangle} \quad (15)$$

На каждой итерации выполняются

$$\begin{aligned} \langle w^k, w^i \rangle &= 0 \quad k \neq i \\ \langle w^k, p^i \rangle &= 0 \quad k > i \\ \langle Ap^k, p^i \rangle &= 0 \quad k \neq i \end{aligned}$$

Метод сопряженных направлений можно использовать для неквадратичной функции: в 14  $A$  заменить матрицей Гессе  $f(x)$ , тогда получим: При  $k = 1$ :  $x_0$

$$\begin{aligned} w^1 &= -\nabla f(x^0) \\ p^1 &= w^1 \end{aligned}$$

При  $k > 1$ :

$$\begin{cases} w^k = -\nabla f(x^{k-1}) \\ p^k = w^k + \beta_k p^{k-1} \\ x^k = x^{k-1} + \alpha_k p^k \end{cases} \quad (16)$$

$$\beta_k = -\frac{\langle H_k p^{k-1}, w^k \rangle}{\langle H_k p^{k-1}, p^{k-1} \rangle} \quad (17)$$

,  $H$  — матрица Гессе,  $\alpha_k$  — из условия исчерпывающего спуска (например из 15)

### 1.3 Модификации

Для квадратичной функции — без использования  $A$

#### 1.3.1 Метод сопряженных градиентов

$$w^{k-1} - w^k = \alpha_{k-1} Ap^k \quad k > 1$$

$$\langle Ap^{k-1}, w^k \rangle = \frac{\langle w^{k-1} - w^k, w^k \rangle}{\alpha_{k-1}} = -\frac{|w^k|^2}{\alpha_{k-1}} \quad (18)$$

$$\langle Ap^{k-1}, p^{k-1} \rangle = \frac{\langle w^{k-1} - w^k, p^{k-1} \rangle}{\alpha_{k-1}} = \frac{\langle w^{k-1}, p^{k-1} \rangle}{\alpha_{k-1}} \quad (19)$$

Выполнив замену в 14, получим 16, где

$$\beta_k = \frac{|w^k|^2}{\langle w^{k-1}, p^{k-1} \rangle} \quad k = 2, \dots \quad (20)$$

#### 1.3.2 Метод Флетчера-Ривса

$$p^{k-1} = w^{k-1} + \beta_k p^{k-1}$$

$p^{k-1}$  и  $p^{k-2}$  ортогональны  $\implies$

$$\begin{aligned} \langle p^{k-1}, w^{k-1} \rangle &= |w^{k-1}|^2 + \beta_k \langle p^{k-2}, w^{k-1} \rangle = |w^{k-1}|^2 \implies \\ \implies \beta_k &= \frac{|w^k|^2}{|w^{k-1}|^2} \quad k = 2, \dots \end{aligned} \quad (21)$$

#### 1.3.3 Метод Полака-Рибьера

$w^k$  и  $w^{k-1}$  — ортогональны

$$\begin{aligned} |w^k|^2 &= \langle w^k - w^{k-1}, w^k \rangle \implies \\ \implies \beta_k &= \frac{\langle w^k - w^{k-1}, w^k \rangle}{|w^{k-1}|^2} \quad k = 2, \dots \end{aligned} \quad (22)$$

## 1.4 Выводы

20-22 эквивалентны для квадратичных функций, для неквадратичных приводят к разным итерационным процессам. Для неквадратичных функций точку минимума не удается найти за конечное число шагов, а процесс может оказаться расходящимся или заикливающимся.

$\alpha_k$  в общем случае находится численно, решая задачу одномерной минимизации  $\implies$  погрешность на каждой итерации, что снижает скорость сходимости или приводит к расходимости

Чтобы снизить влияние погрешности используют ‘обновление’ алгоритма:

1.  $\beta_k = 0$  через заданное число итераций (моменты рестарта кратные  $n$ ). Это позволяет избежать накопления вычислительных погрешностей и уменьшить вероятность построения после каждых  $n$  итераций линейно зависимых направлений спуска, но приводит к росту общего числа итераций
2. На практике метод работает  $\leq n$  итераций, рестарт может не вызваться ни разу. Альтернативный вариант рестарта — условие Пауэлла:

$$|\langle \nabla f(x^k), \nabla f(x^{k+1}) \rangle| \geq v \|\nabla f(x^{k+1})\|^2$$

— оптимально выбрать  $v = 0.1$

Метод Флетчера-Ривса без рестартов является менее эффективным. Наиболее популярным является метод Полака-Рибьера

*Примечание. Линейный метод Флетчера-Ривса*

- в начальной точке  $x^0$ :

$$w^1 = -\nabla f(x^0) \quad p^1 = w^1$$

- $k > 1$ ,  $\alpha_k$  — линейный поиск

$$x^k = x^{k-1} + \alpha_k p^k$$

$$\beta_k = \frac{|w^k|^2}{|w^{k-1}|^2}$$

$$p^k = w^k + \beta_k p^{k-1}$$

В общем случае метод Флетчера-Ривса не гарантирует того, что  $p^k$  — направление спуска

- существует доказательство того, что если в линейном поиске (одномерная минимизация) используется сильные условия Вульфа с  $c_2 < \frac{1}{2}$ , то  $p^k$  будет направлением спуска (то есть поиск должен быть точнее)
- Сильное условие Вульфа:

$$\Phi_k(\alpha) \leq \Phi_k(0) + c_1 \alpha \Phi_k'(0)$$

$$|\Phi_k'(\alpha)| \leq c_2 |\Phi_k'(0)|$$

используются квадратичные/кубические аппроксимации