

# Лекция 16

Луя Yaroshevskiy

13 мая 2023 г.

## Содержание

<b>1</b>	<b>Квазиньютоновские методы</b>	<b>1</b>
1.1	Общий вид релаксационной последовательности	1
1.1.1	Вычисление матриц $G_k$	1
1.1.2	Параметр $\alpha_k$ в 1	2
1.2	Свойства метода ДФП	3
1.3	Метод Бroyдена-Флетчера-Шенно (БФС-метод)	3
1.4	Метод Пауэлла	4
1.5	Способы построения $G_k$	4
<b>2</b>	<b>Сравнение методов</b>	<b>4</b>
2.1	Сравнение по скорости	4
2.2	Оценка эффективности	4

## 1 Квазиньютоновские методы

- объединение достоинств
  1. наискорейшего спуска
  2. метода Ньютона
- не требует обращения матрицы  $H$
- сохраняют высокую сходимость итерационной последовательности

### 1.1 Общий вид релаксационной последовательности

$$\begin{aligned}x^k &= x^{k-1} + \alpha_k p^k \\ p^k &= G_k w^k \quad k \in N \\ w^k &= -\nabla f(x^{k-1})\end{aligned}\tag{1}$$

$G_k$  — положительно определенная матрица ( $n \times n$ ) специального вида  
Поскольку  $G_k > 0$ , то  $p^k$  задает направление спуска:

$$\langle \nabla f(x^{k-1}), p^k \rangle = -\langle w^k, G_k w^k \rangle = -\langle G_k w^k, w^k \rangle < 0$$

#### 1.1.1 Вычисление матриц $G_k$

$$\{G_k\} \xrightarrow{k \rightarrow \infty} H^{-1}(x^*)$$

, где  $x^*$  — точка минимума

На первой итерации полагают  $G_1 = I$  — выполняется градиентный спуск

### 1.1.2 Параметр $\alpha_k$ в 1

- задать  $\alpha_k = 1$
- применить дробление шага
- использовать исчерпывающий спуск в направлении  $p^k$  (чаще всего используется на практике)

Поскольку  $\{G_k\} \rightarrow \{H^{-1}(x^*)\}$ , то  $G_k$  на завершающей стадии поиска близка к  $H^{-1}(x^{k-1})$ , используемой в методе Ньютона  $\implies$  квазиньютоновские метода сохраняют высокую скорость сходимости, присущую методу Ньютона. Таким образом  $G_k$  должна быть близка к  $H^{-1}(x^*)$  и  $G_k$  получают путем аппроксимации  $H^{-1}(x^k - 1)$

Выбор удачной аппроксимации может существенно сократить объем вычислений по сравнению с обращением матрицы  $H$ , тем самым упростить процедуру построения направления спуска  $p^k$  — **идея квазиньютоновских методов**

Каждую  $G_{k+1}$  строят, корректируя ранее вычисленную матрицу  $G_k$

$$G_{k+1} = G_k + \Delta G_k \quad k \in \mathbb{N} \quad (2)$$

$\Delta G_k$  — положительно определенная матрица ( $n \times n$ ) — **поправочная матрица**

Способ выбора  $\Delta G_k$  определяет конкретный квазиньютоновский метод. На первой итерации  $G_1 = I$

*Пример.* Пусть

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

, где  $A$  — положительно определенная

Тогда  $\forall x : A = H, \nabla f(x) = Ax + b$

Введем:  $\Delta w^k = w^k - w^{k-1}, \Delta x^k = x^k - x^{k-1}$ , следовательно

$$\Delta w^k = \nabla f(x^{k-1}) - \nabla f(x^k) = A(x^{k-1} - x^k) = -A\Delta x^k$$

Таким образом, для квадратичной функции

$$A^{-1}\Delta w^k = -\Delta x^k, k \in \mathbb{N} \quad (3)$$

Иначе в виду 3 для общего случая потребуются, чтобы матрица  $G_{k+1}$  удовлетворяли условию

$$G_{k+1}\Delta w^k = -\Delta x^k \quad k \in \mathbb{N} \quad (4)$$

— **квазиньютоновское условие**

Для поправочных матриц условие 4 примет вид

$$\Delta G_k \Delta w^k = -\Delta x^k - G_k \Delta w^k$$

Этому условию удовлетворяют матрицы

$$\Delta G_k = -\frac{\Delta x^k y^T}{\langle \Delta w^k, y \rangle} - \frac{G_k \Delta w^k z^T}{\langle \Delta w^k, z \rangle} \quad (5)$$

$y, z \in E_n$  — выбираются произвольно,  $\Delta G_k$  — симметричные матрицы (необходимое условие, чтобы аппроксимация  $H^{-1}$  давала симметричную матрицу)

Пусть в 5  $y = \Delta x^k, z = G_k w^k$  и учитывая 2, получим

$$G_{k+1} = G_k - \frac{\Delta x^k (\Delta x^k)^T}{\langle \Delta w^k, \Delta x^k \rangle} - \frac{G_k \Delta w^k (\Delta w^k)^T G_k^T}{\langle G_k \Delta w^k, \Delta w^k \rangle} \quad (6)$$

Формула 6 в сочетании с исчерпывающим спуском по направлениям векторов  $p^k$  — **метод Давидона-Флетчера-Пауэлла** (ДФП-метод)

- на первой итерации:  $G_1 = I$ , задается  $x^0$

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

$$\alpha_1 = \min_{\alpha} f(x^0 + \alpha p^1)$$

$$x_1 = x^0 + \alpha p^1 \quad \Delta x^1 = x^1 - x^0$$

- $\forall k > 1$

$$\begin{aligned}
 w^k &= -\nabla f(x^{k-1}) \\
 \Delta w^k &= w^k - w^{k-1} \\
 v^k &= G_{k-1} \Delta w^k \\
 G_k &= G_{k-1} - \frac{\Delta x^{k-1} (\Delta x^{k-1})^T}{\langle \Delta w^k, \Delta x^{k-1} \rangle} - \frac{v^k (v^k)^T}{\langle v^k, \Delta w^k \rangle} \\
 p^k &= G_k w^k \\
 \alpha_k &= \min_{\alpha} f(x^{k-1} + \alpha p^k) \\
 x^k &= x^{k-1} + \alpha_k p^k \quad \Delta x^k = x^k - x^{k-1}
 \end{aligned}$$

условие останова  $\|\Delta x^k\| < \varepsilon$

## 1.2 Свойства метода ДФП

**Свойство 1.**  $G_k$  по условию 6 сохраняет положительную определенность: если  $G_k > 0$ , то  $G_{k+1} > 0$  (при условии  $(\Delta x^k)^T \Delta w^k > 0$ )

**Свойство 2.** Если в 6  $G_k$  — симметричная, то  $G_{k+1}$  — симметричная

**Свойство 3.** При минимизации квадратичная функция с положительно определенной матрицей  $A$  метод ДФП сводится к методу сопряженных градиентов, т.к. первые  $n$  векторов  $p^k$  (задающих направление спуска) являются сопряженными относительно матрицы  $A$ . Следовательно ДФП-метод не более чем за  $n$  итераций дает точное решение квадратичной функции

**Свойство 4.** Матрицы  $G_k$ , вычисленные по ДФП-методу, связаны равенством

$$G_k A p^i = p^i \quad i = \overline{1, k} \quad k = \overline{1, n}$$

При  $k = n$  следует, что  $n$  векторов  $p^i$  являются собственными для симметричной матрицы  $G_n A$ , а соответствующие собственные значения равны 1. Следовательно  $G_n A = I$  и  $G_n = A^{-1}$ , т.е.  $G_n$  оказывается обратной к матрице Гессе  $H(x^*) = A$  квадратичной функции в точке  $x^*$

**Свойство 5.** Если  $f(x)$  не является квадратичной, то ДФП-метод не позволяет найти минимум за конечное число итераций.

Чтобы ослабить влияние накапливаемых погрешностей на сходимость итерационной последовательности и уменьшить вероятность появления после очередных  $n$  итераций линейно зависящих направлений спуска, в ДФП-методе применяют процедуру ‘обновления’ алгоритма: через каждые  $n$  итераций в качестве  $G_n$  используют  $G_n = I$

**Свойство 6.** Если  $f(x)$  — квадратичная функция с  $H > 0$ , то ДФП-метод (с точным одномерным поиском) после  $n$  итераций дает

$$H^{-1} = \sum_{i=1}^n \frac{\Delta x^i (\Delta x^i)^T}{(\Delta x^i)^T \cdot \Delta w^i}$$

ДФП-метод очень чувствителен к точности одномерного поиска

## 1.3 Метод Бroyдена-Флетчера-Шенно (БФШ-метод)

На первой итерации  $G_1 = I$ , затем

$$\begin{aligned}
 G_{k+1} &= G_k - \frac{\Delta x^k (\Delta x^k)^T}{\langle \Delta w^k, \Delta x^k \rangle} - \frac{G_k \Delta w^k (\Delta w^k)^T G_k^t}{\rho_k} + \rho_k r^k (r^k)^T \\
 r^k &= \frac{G_k \Delta w^k}{\rho_k} - \frac{\Delta x^k}{\langle \Delta x^k, \delta w^k \rangle} \\
 \rho_k &= \langle G_k \Delta w^k, \Delta w^k \rangle
 \end{aligned} \tag{7}$$

## 1.4 Метод Пауэлла

На первой итерации  $G_1 = I$ , затем

$$G_{k+1} = G_k - \frac{\Delta \tilde{x}^k (\Delta \tilde{x}^k)^T}{\langle \Delta w^k, \Delta \tilde{x}^k \rangle} \quad k \in \mathbb{N} \quad (8)$$

$$\Delta \tilde{x}^k = \Delta x^k + G_k \Delta w^k$$

## 1.5 Способы построения $G_k$

- сохраняет свойство положительно определенности матрицы
- последовательность  $\{G_k\} \xrightarrow[k \rightarrow \infty]{} H^{-1}(x^*)$

## 2 Сравнение методов

Методы:

- нулевого порядка (не используют значения производных)
- первого порядка (наискорейший спуск, сопряженные градиенты, квазиньютоновские методы)
- второго порядка (метод Ньютона и его модификации, метод Марквардта)

### 2.1 Сравнение по скорости

Методы порождают последовательность  $\{x^k\}$ :

$$\lim_{k \rightarrow \infty} x^k = x^*$$

тогда по свойству предела

$$\|x^{k+1} - x^*\| < \|x^k - x^*\|$$

поэтому

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} < 1$$

В этом случае говорят о сходимости метода

*Пример.* Если  $f(x)$  — непрерывно дифференцируемая функция и  $f(x) \rightarrow +\infty$  при  $\|x\| \rightarrow +\infty$ , то для любого  $x^0$  метод наискорейшего спуска сходится к стационарной точке  $f(x)$

- Глобальная сходимость методов сопряженных градиентов обеспечена с использованием рестартов. Тогда их глобальная сходимость следует из глобальной сходимости метода наискорейшего спуска
- Метод Ньютона не обладает свойством глобальной сходимости. Если  $x^0$  далека от  $x^*$ , то метод не сходится  
Но метод Ньютона с одномерным поиском и направлением спуска и метод Марквардта обладают глобальной сходимостью
- Квазиньютоновские методы ДФП и БФШ обладают глобальной сходимостью в случае применения рестартов

### 2.2 Оценка эффективности

Эффективность зависит от числа итераций

Если

$$\lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = \alpha < 1 \quad \alpha \neq 0$$

,то сходимость линейная. Если

$$\lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = 0$$

, то сходимость сверхлинейная. Если  $\exists \gamma > 1$ :

$$\lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^\gamma} < +\infty$$

то сходимость порядка  $\gamma$ . В частности при  $\gamma = 2$  — квадратичная сходимость

- Методы сопряженных градиентов имеют сверхлинейную скорость сходимости по  $n$  шагам.  
Если  $H(x)$  удовлетворяет условию Липшица, то эти методы имеют квадратичную сходимость по  $n$  шагам
- Квазиньютоновские методы ДФП и БФШ имеют сверхлинейную скорость сходимости  
Если  $H(x)$  удовлетворяет условию Липшица, то эти методы имеют квадратичную скорость сходимости  
Это показывает преимущество квазиньютоновских методов перед методами сопряженных градиентов, которые требуют  $\approx$  в  $n$  раз больше итераций для одного и того же асимптотического поведения. Однако это преимущество сильно снижается загрузкой памяти  $\sim n^2$  и объемом промежуточных матричных вычислений  $\sim n^2$
- Метод Ньютона имеет квадратичную локальную сходимость, если  $H(x)$  удовлетворяет условию Липшица  
Следовательно в малой окрестности  $x^*$  метод Ньютона при указанных предположениях целевой функции  $f(x)$  сходится быстрее остальных методов.