

## Лабораторная работа 4.

### Метод Ньютона и квазиньютоновские методы минимизации функций многих переменных.

**Постановка задачи:** Требуется найти безусловный минимум функции  $n$  переменных  $f(x) = f(x_1, x_2, \dots, x_n)$ , т.е. такую точку  $x^* \in E_n$ , что  $f(x^*) = \min_{x \in E_n} f(x)$ .

Предполагается, что целевая функция  $f(x)$  дважды дифференцируема в  $E_n$  и возможно вычисление ее производных в произвольной точке  $E_n$ .

#### Метод Ньютона

**Стратегия поиска:** Пусть целевая функция  $f(x)$  дважды дифференцируема в  $E_n$ . Тогда, с помощью градиента и матрицы Гессе, для нее можно записать разложение в ряд по формуле Тейлора в окрестности точки  $x^k$

$$f(x) = f(x^k) + \nabla f(x^k)^T \Delta x + \frac{1}{2} \Delta x^T H(x^k) \Delta x + o(\|\Delta x\|^2)$$

где  $o(\|\Delta x\|^2)$  - сумма всех членов разложения, имеющих порядок выше второго,  $\Delta x^T H(x^k) \Delta x$  - квадратичная форма. Отсюда следует, что поведение функции  $f(x)$  с точностью до величины  $o(\|\Delta x\|^2)$  может быть описано квадратичной функцией:

$$\Phi_k(x) = f(x^k) + \nabla f(x^k)^T \Delta x + \frac{1}{2} \Delta x^T H(x^k) \Delta x$$

Минимизируем функцию  $\Phi_k(x)$  вместо целевой функции  $f(x)$ . Найдем точку минимума  $x^{k+1}$  функции  $\Phi_k(x)$  из условия стационарности градиента  $\nabla \Phi_k(x) = 0$

$$\nabla \Phi_k(x) = H(x^k)(x - x^k) + \nabla f(x^k) = 0 \quad (1)$$

Пусть матрица Гессе  $H(x)$  положительно определена при всех  $x \in E_n$  и, следовательно, невырождена ( $\det H(x) > 0$ ). Тогда существует обратная матрица  $[H(x)]^{-1}$ . Отметим, что квадратичная функция  $\Phi_k(x)$  с положительно определенной матрицей  $H(x^k)$  сильно выпукла, и уравнение (1) определяет единственную точку глобального минимума функции  $\Phi_k(x)$ . Умножим слева обе части равенства (1) на матрицу  $[H(x^k)]^{-1}$  и найдем точку минимума  $x^{k+1}$  квадратичной функции  $\Phi_k(x)$ , аппроксимирующей  $f(x)$  в окрестности точки  $x = x^k$

$$x^{k+1} = x^k + p^k, \quad (2)$$

где

$$p^k = -H^{-1}(x^k) \nabla f(x^k) \quad (3)$$

- направление минимизации на  $k + 1$  шаге.

Итерационный процесс (2), начатый из произвольной точки  $x^0 \in E_n$ , называется методом Ньютона минимизации функции многих переменных и является обобщением метода Ньютона в одномерном случае.

Очевидно, для квадратичной функции с положительно определенной матрицей применение метода Ньютона обеспечивает получение точки глобального минимума ровно за один шаг из любой точки  $x^0$ .

### **Алгоритм:**

1. Задать точность вычислений  $\varepsilon$ , выбрать начальное приближение  $x^0$ .
2. Положить  $k = 0$  ( $k$  - номер итерации). Вычислить значение  $p^0 = -H^{-1}(x^0)\nabla f(x^0)$ .
3. Вычислить точку  $x^{k+1} = x^k + p^k$ , градиент  $\nabla f(x^{k+1})$  и гессиан  $H(x^{k+1})$ .
4. Проверить критерий окончания поиска  $\|\nabla f(x^{k+1})\| < \varepsilon$ . Если критерий выполнен, перейти к шагу 6.
5. Найти новое направление поиска  $p^{k+1} = -H^{-1}(x^{k+1})\nabla f(x^{k+1})$ . Положить  $k = k + 1$  и перейти к шагу 3.
6. Выбрать приближенно  $x^* = x^{k+1}$ ,  $f(x^*) = f(x^{k+1})$ . Поиск завершен.

### **Дополнительные сведения о методе Ньютона**

#### **1. Минимизация неквадратичных функций**

При минимизации неквадратичной выпуклой функции применение метода Ньютона обеспечивает, как правило, быструю сходимость последовательности  $\{x^k\}$  к точке минимума  $x^* \in E_n$ . На каждом шаге итерационного процесса используется информация о поведении функции в окрестности точки  $x^k$ , содержащаяся в значениях не только первых, но и вторых ее частных производных. Поэтому при прочих равных условиях следует ожидать более быструю сходимость метода Ньютона по сравнению с градиентными методами.

Если график целевой функции имеет овражную структуру, то вектор  $p^k$  из (3) может составлять с осью оврага меньший угол, чем антиградиент. Эта особенность алгоритмов метода Ньютона делает их эффективнее алгоритмов метода градиентного спуска при минимизации овражных функций.

#### **2. Монотонное убывание**

Вдали от решения параметр длины шага  $p^k$ , характеризующий обычный метод Ньютона, может не обеспечивать монотонного убывания минимизирующей последовательности (2), и в этом смысле классический метод Ньютона не является методом спуска. Даже сходящаяся минимизирующая последовательность при использовании метода Ньютона не всегда обеспечивает монотонное убывание целевой функции  $f(x)$ , т.е. неравенство  $f(x^{k+1}) < f(x^k)$  для некоторых  $k$  может нарушаться. Этот недостаток устранен в обобщенном методе Ньютона

$$x^{k+1} = x^k - \alpha_k H^{-1}(x^k) \nabla f(x^k), \quad (4)$$

где величина  $\alpha_k > 0$  находится на каждом шаге из условия исчерпывающего спуска по направлению  $p^k = -H^{-1}(x^k) \nabla f(x^k)$ . Можно показать, что если целевая функция является сильно выпуклой и ее матрица Гессе  $H(x)$  для любых точек  $x, y \in E_n$  удовлетворяет неравенству  $\|H(x) - H(y)\| \leq L\|x - y\|$ , где  $L = \text{const} > 0$ , то при произвольном выборе начальной точки  $x^0 \in E_n$  обобщенный метод Ньютона (4) обладает квадратичной скоростью сходимости.

### 3. Скорость сходимости

При выборе достаточно хорошего начального приближения минимизирующая последовательность для сильно выпуклой дважды дифференцируемой функции сходится к точке минимума с квадратичной скоростью. Если начальная точка выбрана недостаточно близко к искомой точке минимума, или же целевая функция не является сильно выпуклой, то последовательность (2) может расходиться. Высокая скорость сходимости метода достигается за счет того, что он использует информацию о вторых производных целевой функции. Соответственно, итерация метода Ньютона существенно более трудоемка, чем, например, итерации градиентных методов.

### 4. Трудоемкость итерации

На каждой итерации необходимо вычислять и обращать матрицу порядка  $n$ , что в случае большой размерности пространства является достаточно трудоемкой операцией. На практике обычно не вычисляют матрицу, обратную к положительно определенной матрице  $H^k$ , а вектор  $p^k$  находят из решения системы линейных алгебраических уравнений (СЛАУ):  $H^k p^k = -\nabla f(x^k)$ . Решение такой системы требует  $O(n^3)$  действий на каждой итерации.

### 5. Итог

*Главным достоинством* метода Ньютона является высокая скорость сходимости. При выборе достаточно хорошего начального приближения минимизирующая последовательность для сильно выпуклой дважды дифференцируемой функции сходится к точке минимума с квадратичной скоростью. Это очень высокая скорость сходимости.

*Существенный недостаток* метода Ньютона – локальная сходимость. Если точка  $x^0$  выбрана недостаточно близко к точке  $x^*$ , то последовательность  $\{x^k\}$  может расходиться. Отметим, что даже сходящаяся последовательность  $\{x^k\}$  метода Ньютона не всегда обеспечивает монотонное убывание.

*Другой существенный недостаток* метода Ньютона состоит в том, что на каждой итерации метода нужно вычислять матрицу из вторых производных целевой функции и решать соответствующую систему линейных уравнений для

определения вектора  $p^k$ . В целом, итерация метода существенно более трудоемка, чем соответствующая итерация градиентных методов.

### Квазиньютоновские методы

Квазиньютоновские методы представляют собой итерационные процедуры минимизации вида

$$x^{k+1} = x^k + \alpha_k p^k, \quad (5)$$

где направление убывания определяется формулой

$$p^k = -A^k \nabla f(x^k) \equiv -A^k g^k, \quad (6)$$

а величина шага  $\alpha_k > 0$  находится в результате решения задачи одномерной минимизации  $f(x^k + \alpha p^k) \rightarrow \min$ .

**Стратегия поиска:** Эффективность минимизации целевой функции методом Ньютона обусловлена учетом информации о кривизне целевой функции, содержащейся в матрице Гессе и позволяющей строить локально точные квадратичные модели минимизируемой функции  $f(x)$ . В квазиньютоновских методах данные относительно кривизны целевой функции  $f(x)$  накапливаются на основе наблюдения за изменением градиента  $g^k \equiv \nabla f(x^k)$  во время итераций спуска. В этом заключается принципиальное отличие квазиньютоновских методов от метода Ньютона. Теория квазиньютоновских методов опирается на возможность аппроксимации кривизны нелинейной функции без явного формирования ее матрицы Гессе. Все квазиньютоновские алгоритмы включают в себя процедуру минимизации (5) и (6). Конкретный алгоритм определяется способом выбора матриц  $A^k$ .

Выше было указано, что для сильно выпуклой целевой функции алгоритм метода Ньютона обладает квадратичной скоростью сходимости. Поэтому можно ожидать, что квазиньютоновские методы будут иметь достаточно высокую скорость сходимости, если на каждой  $k+1$ -ой итерации матрица  $A^{k+1}$  будет выбрана близкой к матрице, обратной матрице Гессе  $H^{-1}(x^k)$  в точке  $x^k$ . Используя при конструировании матрицы  $A^{k+1}$  аппроксимацию матрицы  $H^{-1}(x^k)$  с учетом о градиенте целевой функции в точке  $x^k$ , можно существенно упростить процедуру нахождения направления спуска на  $k+1$ -ой итерации. Именно эти соображения лежат в основе построения алгоритмов квазиньютоновских методов.

Будем искать поправочную матрицу  $U^k$ , которая определяет способ получения  $A^{k+1}$  из  $A^k$  по формуле пересчета  $A^{k+1} = A^k + U^k$ . Матрица  $A^k$  служит средством отображения информации о кривизне целевой функции, накопленной на предыдущих шагах. Желательно при этом, чтобы  $A^{k+1} \xrightarrow[k \rightarrow \infty]{} (H^*)^{-1}$ , где  $H^* = H(x^*)$ .

Для краткости введем обозначения  $s^k = x^{k+1} - x^k$ ,  $y^k = g^{k+1} - g^k$ .

Используя разложение Тейлора функции  $g^{k+1}$  в окрестности точки  $x^k$ , получим:  $g^{k+1} = g^k + H^k s^k + O(|s^k|^2)$ , или, иначе  $y^k = H^k s^k + O(|s^k|^2)$ . Отсюда следует, что если целевая функция  $f(x)$  квадратичная, то  $y^k = H^k s^k$ . Так как значения  $y^k$  и  $s^k$  мы получаем только после одномерной минимизации для вычисления шага, то в общем случае матрица  $A^k$  не удовлетворяет условию  $A^k y^k = s^k$ . Будем искать такую матрицу  $A^{k+1}$ , которая бы удовлетворяла этому условию:  $A^{k+1} y^k = s^k$ . Последнее равенство принято называть квазиньютоновским условием. Один шаг дает информацию о кривизне целевой функции  $f(x)$  вдоль одного-единственного направления, поэтому мы вправе надеяться, что удастся обойтись поправочной матрицей  $U^k$  малого ранга. Существуют различные пути достижения этого условия. Один из них будет рассмотрен ниже.

### **Общий алгоритм для квазиньютоновских методов:**

1. Задать точность вычислений  $\varepsilon$ , выбрать начальное приближение  $x^0$  и матрицу  $A^0$ .
2. Положить  $k = 0$  ( $k$  - номер итерации). Вычислить значение  $p^0 = -A^0 \nabla f(x^0)$ .
3. Вычислить значение  $\alpha_k$  из решения задачи одномерного исчерпывающего спуска  $f(x^k + \alpha_k p^k) = \min_{\alpha > 0} f(x^k + \alpha p^k)$ , либо другим методом.
4. Вычислить точку  $x^{k+1} = x^k + \alpha_k p^k$  и градиент  $\nabla f(x^{k+1})$ .
5. Проверить критерий окончания поиска  $\|\nabla f(x^{k+1})\| < \varepsilon$ . Если критерий выполнен, перейти к шагу 7.
6. Вычислить матрицу  $A^{k+1}$  по формуле для выбранного метода и найти новое направление поиска  $p^{k+1} = -A^{k+1} \nabla f(x^{k+1})$ . Положить  $k = k + 1$  и перейти к шагу 3.
7. Выбрать приближенно  $x^* = x^{k+1}$ ,  $f(x^*) = f(x^{k+1})$ . Поиск завершен.

### **Дополнительные сведения о квазиньютоновских методах.**

#### **1. Виды матрицы $A^k$**

Ввиду симметричности матрицы Гессе удобно строить ее квазиньютоновские приближения также в симметричном виде. Для этого следует строить формулы пересчета, которые гарантируют сохранение симметрии.

Самая простая возможность сохранить симметрию – построить матрицу  $A^{k+1}$  по формуле  $A^{k+1} = A^k + a u u^T$ , т. е. добавить симметрическую матрицу ранга 1. В случае с поправками ранга 1 требование симметрии определяет  $U^k$  однозначно. Квазиньютоновское условие принимает вид  $A^k y^k + a u u^T y^k = s^k$ . Перепишем это

выражение иначе:  $au(\underbrace{u^T y^k}_{\text{скаляр}}) = s^k - A^k y^k$ . Следовательно, вектор  $u$  должен быть параллелен разности  $s^k - A^k y^k$ . Положим  $u = s^k - A^k y^k$  и определим величину  $a$  из выражения  $au^T y^k = 1$ . В этом случае симметричная формула ранга 1 принимает вид:

$$A^{k+1} = A^k + \frac{(s^k - A^k y^k)(s^k - A^k y^k)^T}{(s^k - A^k y^k)^T y^k}.$$

Отметим два наиболее существенных недостатка этого способа пересчета матрицы  $A^{k+1}$ , которые имеют место даже в случае квадратичной целевой функции:

1. В общем случае положительная определенность матрицы  $A^k$  не гарантируется.
2. Знаменатель может обращаться в ноль, тогда метод вообще будет не определен. Также численные трудности возникнут, когда знаменатель будет равен маленькому числу.

Более гибкую формулу можно получить, используя для матрицы  $U^k$  поправку ранга 2. Формула пересчета примет вид  $A^{k+1} = A^k + auu^T + bvv^T$ . Квазиньютоновское условие в этом случае запишется в виде  $A^k y^k + auu^T y^k + bvv^T y^k = s^k$ .

Из формулы, приведенной выше, однозначно определить вектора  $u$  и  $v$  невозможно. Одним из вариантов является следующий выбор:  $u = s^k$  и  $v = A^k y^k$ . При этом получим следующий вид квазиньютоновского условия:

$$A^k y^k + as^k u^T y^k + bA^k y^k v^T y^k = s^k$$

или

$$A^k y^k (1 + bv^T y^k) = s^k (1 - au^T y^k)$$

Для тождественного выполнения последнего равенства необходимо положить  $1 + bv^T y^k = 0$ ,  $1 - au^T y^k = 0$ . Получаем формулу ДФП:

$$A^{k+1} = A^k + \frac{1}{u^T y^k} uu^T - \frac{1}{v^T y^k} vv^T \text{ или, учитывая, что } u = s^k \text{ и } v = A^k y^k, \text{ окончательно:}$$

$$A^{k+1} = A^k + \frac{s^k (s^k)^T}{(s^k)^T y^k} - \frac{A^k y^k (y^k)^T A^k}{(y^k)^T A^k y^k}.$$

Пусть  $B^{k+1} = (A^{k+1})^{-1}$ , тогда  $s^k = A^{k+1} y^k$ , и  $y^k = B^{k+1} s^k$ . Оба выражения имеют одну и ту же форму записи и получаются друг из друга заменой  $A^{k+1}$  на  $B^{k+1}$ , а также  $y^k$  на  $s^k$ . С учетом этого, выражение пересчета матрицы  $A^{k+1}$  может быть преобразовано к формуле, пересчитывающей матрицу  $B^{k+1}$ . Для ДФП соответственно получим  $B^{k+1} = B^k + \frac{y^k (y^k)^T}{(y^k)^T s^k} - \frac{B^k s^k (s^k)^T B^k}{(s^k)^T B^k s^k}$ .

Преобразуем полученное выражение, используя матрицу  $A$ , а так же формулу Шермана-Моррисона. В результате получим выражение для пересчета матрицы  $A$

$$A^{k+1} = A^k + \frac{1 + (y^k)^T A^k (y^k)}{(y^k)^T s^k} \cdot \frac{s^k (s^k)^T}{(s^k)^T y^k} - \frac{s^k (y^k)^T A^k + A^k y^k (s^k)^T}{(y^k)^T s^k}.$$

Эта формула пересчета носит название Бroyдена-Флетчера-Гольдфарба-Шанно.

Обе формулы ДФП и БФГШ получены с помощью поправки шага 2 и построены с помощью использования векторов  $s^k$  и  $A^k y^k$ , поэтому можно полагать, что комбинация этих формул будет иметь те же самые свойства. Этот вывод приводит нас к целому семейству формул:  $A^\varphi = (1 - \varphi)A_{DFP} + \varphi A_{BFGS}$ .

## 2. Точность одномерной минимизации

**Теорема:** при минимизации положительно определенной квадратичной функции квазиньютоновскими методами семейства Бroyдена, если шаг в направлении поиска  $\alpha_k$  выбирается с помощью точной одномерной минимизации, последовательность получаемых точек  $\{x^k\}$  совпадает с последовательностью  $\{x^k\}$ , которая получается при использовании метода сопряженных градиентов, если выбраны  $A^0 = E$  и одинаковые начальные приближения.

Доказательство теоремы можно найти в литературе. Отсюда следует, что при минимизации квадратичной функции величина параметра  $\varphi$  не имеет значения, и методы ДФП и БФГШ дают одинаковый результат. Более того, можно показать, что это утверждение остается справедливым не только для квадратичной функции. Выбор значения  $\varphi$  влияет на результат, если только одномерная минимизация неточна. Вообще говоря, сами матрицы  $A^{k+1}$  для различных  $\varphi$  будут отличаться друг от друга на всех итерациях, и если для одной из них отклонение от матрицы Гессе окажется «малым», то для другой оно наверняка будет «большим». Таким образом, если одномерная минимизация неточна, это может сильно повлиять на результат вычислений. Принята точка зрения, что метод ДФП обладает высокой чувствительностью к точности одномерного поиска. Теоретическая эквивалентность различных формул семейства Бroyдена при точной одномерной минимизации еще не означает, что они эквивалентны с практической точки зрения. При точной одномерной минимизации количество вычислений функции на этапах одномерного поиска, скорее всего, окажутся различными. Причина этого следующая: величина первого пробного шага в процедуре поиска  $\alpha_k$  при различных матрицах  $A_\varphi^k$  будет разной, в одном случае получится хорошая оценка искомой величины  $\alpha_k$ , а в другом – плохая. Соответственно, и вся процедура вычисления  $\alpha_k$  в одном случае проработает быстрее, в другом – медленнее. Таким образом, даже в случае точной одномерной минимизации методы ДФП и БФГЖ не эквивалентны.

### 3. Другой выбор $\alpha_k$ .

Для решения на каждом шаге одномерной задачи оптимизации могут использоваться различные методы. Кроме того, можно пользоваться менее трудоемкими способами выбора параметра длины шага, например, правилом Армихо, правилом постоянного параметра, правилом Вулфа, правилом Голдстейна.

На практике в случае неквадратичных функций параметры длины шага в квазиньютоновских методах часто определяют с помощью правила Вулфа, поскольку, как и в случае одномерной минимизации, правило гарантирует положительную определенность генерируемых матриц  $A^k$ . Алгоритм правила Вулфа был рассмотрен в предыдущей лабораторной работе.

### 4. Скорость сходимости

При минимизации квадратичной функции квазиньютоновские методы сходятся не более чем за  $n$  шагов.

Если целевая функция является сильно выпуклой, то алгоритмы метода Ньютона обладают квадратичной скоростью сходимости. Поэтому можно ожидать, что алгоритмы квазиньютоновских методов будут иметь достаточно высокую скорость сходимости, если на каждой итерации матрица  $A^k$  будет выбрана близкой к матрице  $H^{-1}(x^k)$ . Рассмотренные квазиньютоновские методы обладают, как правило, сверхлинейной скоростью сходимости.

### Сравнение квазиньютоновских методов с методом сопряженных градиентов и с методом Ньютона

Метод сопряженных градиентов имеет 3 важных свойства:

- 1) минимизирует квадратичную функцию от  $n$  переменных за  $n$  шагов;
- 2) не требует вычисления второй производной;
- 3) не требует хранения или инвертирования матриц.

Квазиньютоновские алгоритмы обладают свойствами 1) и 2). Компенсацией за отсутствие свойства 3) является то, что для большинства квазиньютоновских алгоритмов не требуется точная одномерная минимизация.

Действительно, квазиньютоновские методы не так чувствительны к точности одномерной минимизации. Это подтверждается практикой. Некоторым объяснением этого факта может служить соображение о том, что квазиньютоновский метод использует положительно определенную матрицу  $A^k$ , следовательно, последовательность  $\{x^k\}$  формируется в направлении спуска.

Если используемая одномерная минимизация точна, то квазиньютоновский метод не только генерирует сопряженные направления, но и аппроксимирует обратную матрицу Гессе. В результате, вблизи точки минимума с положительно определенной матрицей Гессе квазиньютоновский метод начинает воспроизводить метод Ньютона, тем самым приобретая более быструю сходимость. Важно, что это



свойство не зависит от начальной матрицы, поэтому нет необходимости периодически делать рестарт, как в методе сопряженных градиентов.

Заметим, что на  $k$ -ой итерации метода сопряженных градиентов требуется  $O(n)$  операций для того, чтобы вычислить направление спуска  $p^k$  и следующее приближение  $x^{k+1}$ . Квазиньютоновские методы требуют приблизительно то же количество действий для вычисления целевой функции и ее градиента, а также  $O(n^2)$  операций для вычисления матрицы  $A^k$  и следующей точки  $x^{k+1}$ . В том случае, если для нахождения функции и ее градиента требуется количество вычислений, большее либо сравнимое с  $O(n^2)$ , то получается, что квазиньютоновские методы требуют ненамного больше вычислений на одну итерацию, чем метод сопряженных градиентов. В задачах, где для нахождения целевой функции и ее градиента требуется выполнить гораздо меньше операций, чем  $O(n^2)$ , предпочтительным является метод сопряженных градиентов. Так, для задач оптимального управления, где обычно число  $n$  является большим, вычисление функции и ее градиента часто требует  $O(n)$  итераций. Для решения таких задач метод сопряженных градиентов весьма эффективен.

Необходимо отметить, что как квазиньютоновские методы, так и метод сопряженных градиентов требуют меньше вычислений на одну итерацию, чем классический метод Ньютона, для которого нужно находить значения целевой функции, ее градиента и гессиана, а также выполнять  $O(n^3)$  операций для вычисления направления спуска. С другой стороны, этот недостаток частично уравнивается более высокой скоростью сходимости метода Ньютона.

Потенциальные преимущества квазиньютоновских методов по сравнению с методом Ньютона:

- Требуется вычислять только первые производные (а в методе Ньютона необходимо вычислять и вторые производные)
- положительно определенная матрица  $A^k$  обеспечивает направление спуска (в то время как матрица Гессе может быть положительно не определена)
- требуется  $O(n^2)$  действий на итерацию (в отличие от  $O(n^3)$  действий в методе Ньютона)

**Итог:** Достоинство квазиньютоновских алгоритмов состоит в том, что при их применении нет необходимости обращать матрицу Гессе целевой функции и в то же время удается сохранить высокую скорость сходимости, присущую методу Ньютона и его модификациям. Недостатком этих методов является необходимость хранения в памяти ЭВМ матриц  $A^k$ , что при решении задач высокой размерности может создать определенные трудности.

Квазиньютоновские методы чрезвычайно популярны среди пользователей методов оптимизации в связи с тем, что они сочетают высокую скорость сходимости с невысокой трудоемкостью итерации.

## Задания

1. Реализовать в среде MATLAB метод Ньютона и квазиньютоновские методы ДФП и БФГШ.

В методе Ньютона и в квазиньютоновских методах для вычисления величины шага исчерпывающего спуска при решении задач одномерной минимизации для квадратичных функций использовать точное аналитическое выражение.

Обратить внимание, что при оценке эффективности метода наиболее важным критерием, при достижении заданной точности, является количество вычисленных значений целевой функции и ее производной.

2. Протестировать работу реализованных методов на примере овражной функции

$$f(x) = x_1^2 + a x_2^2,$$

при  $a = 1, 250, 1000$ . Задавая  $\varepsilon = 10^{-3}$  и  $\varepsilon = 10^{-5}$ , сравнить скорость работы методов при различных значениях параметра  $a$  по числу вычислений целевой функции и ее производной.

3. Выбрать для выполнения работы тестовую функцию, номер которой соответствует номеру Вашего компьютера. Например, для компьютера №3 это будет функция 3), для компьютера №13 – функция 4):  $13-9=4$ ; для компьютера №23 это будет функция 5):  $23-9 \times 2=5$ .

1)  $f(x) = 64x_1^2 + 126x_1x_2 + 64x_2^2 - 10x_1 + 30x_2 + 13$

2)  $f(x) = 129x_1^2 - 256x_1x_2 + 129x_2^2 - 51x_1 - 149x_2 - 27$

3)  $f(x) = 254x_1^2 + 506x_1x_2 + 254x_2^2 + 50x_1 + 130x_2 - 111$

4)  $f(x) = 151x_1^2 - 300x_1x_2 + 151x_2^2 + 33x_1 + 99x_2 + 48$

5)  $f(x) = 85x_1^2 + 168x_1x_2 + 85x_2^2 + 29x_1 - 51x_2 + 83$

6)  $f(x) = 211x_1^2 - 420x_1x_2 + 211x_2^2 - 192x_1 + 50x_2 - 25$

7)  $f(x) = 194x_1^2 + 376x_1x_2 + 194x_2^2 + 31x_1 - 229x_2 + 4$

8)  $f(x) = 45x_1^2 - 88x_1x_2 + 45x_2^2 + 102x_1 + 268x_2 - 21$

9)  $f(x) = 99x_1^2 + 196x_1x_2 + 99x_2^2 - 95x_1 - 9x_2 + 91$

4. Сравнить эффективность метода Ньютона и квазиньютоновских методов ДФП и БФГЖ для задачи п.2 при  $a=250$ , а также этих методов и метода сопряженных градиентов для тестовой функции п.3. *Объяснить полученные результаты.*

*Результаты расчетов сохранить для сравнения с результатами, которые будут получены в лабораторной работе №5.*

#### 5. Минимизировать функцию Розенброка

$$f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2$$

с точностью  $\varepsilon = 10^{-3}$  и  $\varepsilon = 10^{-5}$ , выбрав начальную точку  $x^0 = (-1, 2)^T$ .

Для решения задачи использовать квазиньютоновские методы ДФП и БФГШ. Параметр точности одномерного поиска установить постоянным, например,  $10^{-6}$ . Определить, сколько вычислений функций и ее производной потребуется методу для того, чтобы разность между численным решением и точным решением  $x^* = (1, 1)^T$  была меньше  $\varepsilon$ .

6. Продолжить изучение квазиньютоновских методов ДФП и БФГШ на примере минимизации функции Розенброка.

6.1 Для решения задачи одномерной минимизации использовать как метод поразрядного поиска, так и метод золотого сечения. Параметр точности одномерного поиска менять (например, задавать  $10^{-4}$ ,  $10^{-6}$ ,  $10^{-8}$ ). *Как зависит точность нахождения решения основной задачи от точности одномерной минимизации?*

6.2. Реализовать нахождение шага спуска с помощью правила Вулфа. Результат сильно зависит от начального значения, поэтому попробуйте несколько различных начальных значений.

6.3. *Сравнить работу программы (точность решения основной задачи, количество вычислений целевой функции и ее производной) при использовании методов поразрядного поиска, золотого сечения и правила Вулфа (для каждого метода рассматривайте его наилучший результат).*

*Результаты расчетов п.5 и п.6 сохранить для сравнения с результатами, которые будут получены прямыми методами в Л.Р. №5.*

#### 7. На примере функции Химмельблау

$$f(x) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$$

рассмотреть особенности применения квазиньютоновских методов для минимизации многомодальных функций. В качестве начального приближения взять точки  $(0, 0)$  и  $(-5, 0)$ . Как зависит работа рассматриваемых алгоритмов от выбора начального приближения?

8. Результаты работы представлять в доступной форме (например, графически или с помощью таблиц), объяснять полученные результаты. Сдать лабораторную работу преподавателю, *ответив предварительно на все следующие контрольные вопросы.*

### Контрольные вопросы к Лабораторной работе 4.

1. В чем состоит идея метода Ньютона?
2. За сколько итераций с помощью метода Ньютона можно найти точку минимума функции  $f(x_1, x_2) = 100x_1^2 + x_2^2$ ?
3. Можно ли с помощью метода Ньютона найти точку максимума целевой функции? Произвольную стационарную точку?
4. Какова скорость сходимости метода Ньютона в окрестности точки минимума для дважды дифференцируемой выпуклой функции  $f(x)$  многих переменных?
5. Какова трудоемкость одной итерации метода Ньютона?
6. В чем заключается обобщенный метод Ньютона минимизации функции многих переменных?
7. В чем состоит отличие классического и обобщенного методов Ньютона при минимизации сильновыпуклой дважды дифференцируемой функции многих переменных?
8. Показать, что если матрица Гессе  $H(x^k) > 0$  для целевой функции  $f(x)$  в точке  $x^k$  и  $f'(x^k) \neq 0$ , то направление  $p^k = -H^{-1}(x^k)\nabla f(x^k)$  является направлением убывания функции  $f(x)$  в точке  $x^k$ .
9. Сформулировать общий принцип построения квазиньютоновских методов.
10. Какую скорость сходимости следует ожидать от квазиньютоновских методов?
11. Какова их трудоемкость?
12. В чем заключаются потенциальные преимущества квазиньютоновских методов перед классическим методом Ньютона?
13. Сформулировать относительные достоинства и недостатки метода сопряженных градиентов и квазиньютоновских методов.
14. Какой порядок арифметических действий необходимо выполнять на одной итерации метода сопряженных градиентов и квазиньютоновских методов? Метода Ньютона?
15. Какие методы минимизации эффективны в случае большой размерности пространства  $E_n$ ?