

НИУ ИТМО
Факультет Информационных Технологий и Программирования
Направление "Прикладная Математика и Информатика"

Лабораторная работа 4

курса "Методы оптимизации"

Верблюжий случай
Сысоев Александр, Зырянова Мария

Санкт-Петербург, 2021

1 Постановка задания

Необходимо разработать и исследовать программы для безусловной минимизации функции многих переменных методом Ньютона и его модификаций.

1. Реализовать классический метод Ньютона, с одномерным поиском и с направлением спуска.
 - (a) Продемонстрировать работу на различных функциях, включая иллюстрации траекторий спуска. Указать количество итераций, необходимых для достижения заданной точности; указывать найденные значения параметров при одномерном поиске.
 - (b) Исследовать работу на двух функциях с заданным начальным приближением. Сравнить результаты с минимизацией методом наискорейшего спуска из лабораторной работы 2. Для каждого метода привести иллюстрации траекторий сходимости.
2. Реализовать метод Бroyдена-Флетчера-Шено и метод Пауэлла. Работу квазиньютоновских методов сравните с наилучшим методом Ньютона (по результатам 1.2) на заданных функциях. Провести исследование влияние выбора начального приближения на результат (не менее трех) и оценить скорость сходимости.
3. Реализовать метод Марквардта двумя вариантами. Результаты работы продемонстрировать на минимизации многомерной функции Розенброка в сравнении с наилучшим методом Ньютона.

2 Рассматриваемые функции

1.

$$f_1(x) = 50x^2 + y^2 + 20x + 20y + 239$$

Минимум достигается в точке $x_{min} = (-0, 2; -10)$, значение минимума -137 .

2.

$$f_2 = (x - 2)^4 + (y - 3)^4 + (z - 4)^4$$

Минимум достигается в точке $x_{min} = (2; 3; 4)$, значение минимума -0 .

3.

$$f_3 = x_1^2 + x_2^2 - 1.2x_1x_2$$

Минимум достигается в точке $x_{min} = (0; 0)$, значение минимума -0 .

4.

$$f_4 = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

Минимум достигается в точке $x_{min} = (1; -1)$ и $x_{min} = (1; 1)$, значение минимума -0 .

5.

$$f_5 = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$$

Минимум достигается в

- $x_{min} = (3; 2);$
- $x_{min} = (-3.779310253377746891890765841; -3.283185991286169412266000514)$
- $x_{min} = (-2.805118086952744853053572398; 3.131312518250572965804300723)$
- $x_{min} = (3.584428340330491744944338239; -1.848126526964403553538300209)$

Значение минимума во всех случаях 0 .

6.

$$f_6 = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)64 + 10(x_1 - x_4)^4$$

Минимум достигается в точке давайте расскажите как мне тут минимум посчитать.

7.

$$f_7 = 100 - \frac{2}{1 + \left(\frac{x_1-1}{2}\right)^2 + \left(\frac{x_2-1}{3}\right)^2} - \frac{1}{1 + \left(\frac{x_1-2}{2}\right)^2 + \left(\frac{x_2-1}{3}\right)^2}$$

Минимум достигается в точке $x_{min} = (1.291643031517492930227373527; 1)$, значение минимума $-97.1531028728543166387604846$.

8. Функция Розенброка:

$$f = \sum_{i=1}^{n-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2$$

Минимум ищите сами

3 Метод Ньютона и его модификации

Метод минимизации Ньютона основывается на следующих рассуждениях.

Пусть дана дважды дифференцируемая целевая функция нескольких переменных $f(x)$, которую необходимо минимизировать. Разложим ее в ряд Тейлора в фиксированной точке x при произвольном приращении аргумента Δx :

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

Пусть $s = \Delta x$, $g(x) = \nabla f(x)$, $H(x) = \nabla^2 f(x)$, пренебрежем слагаемыми второго порядка малости, получим квадратичную функцию:

$$q(x) = f(x) + g(x)^T s + \frac{1}{2} s^T H(x) s$$

Полученная функция достигает минимума при s : $H(x)s = -g(x)$, так как необходимое условие минимума $-\nabla q(x) = 0$:

$$\nabla q(x) = g(x) + H(x)s$$

Тогда вектор перемещения в точку минимума квадратичной функции

$$s = -H(x)^{-1}g(x)$$

Метод Ньютона минимизирует положительно определенную квадратичную функцию за один шаг из любой начальной точки x_0 . Это метод второго порядка, так как используется вычисление гессиана – вторые частные производные.

3.1 Классический метод Ньютона

Классический метод Ньютона является итерационным. Обозначая в текущей точке поиска x_k значения градиента $g_k = \nabla f(x_k)$ и матрицы Гессе $H_k = \nabla^2 f(x_k)$ получаются следующие итерационные формулы метода Ньютона:

$$x_{k+1} = x_k + s_k, \quad H_k s_k = -g_k$$

Итерационный процесс продолжается по мере выполнения условия прерывания с заданной допустимой погрешностью ϵ : $\|x_{k+1} - x_k\| > \epsilon$

Вектор s находится за счет решения системы линейных уравнений LU-разложением, реализация которого была подробно рассмотрена в 3 лабораторной работе. Этот вектор задает и направление, и шаг перехода в следующую точку поиска.

3.2 Метод Ньютона с одномерным поиском

Формулы метода Ньютона основаны на аппроксимации функции квадратичной функцией в малой окрестности текущей точки поиска и могут приводить к возрастанию значений функции. Для улучшения применяется одномерная минимизация функции в направлении поиска s .

На каждой итерации такого модифицированного метода Ньютона из точки x_k выполняется одномерный поиск в направлении метода Ньютона $d_k = s_k$:

$$\lambda_k = \operatorname{argmin} f(x_k + \lambda d_k)$$

Таким образом, итерационный процесс дополняется следующими формулами:

$$H_k d_k = -g_k, \quad r = \operatorname{argmin} f(x_k + \lambda d_k), \quad s = r \cdot d$$

Метод Ньютона с одномерным поиском надежнее исходного метода Ньютона за счет обеспечения убывания значений целевой функции. Однако его эффективность существенно зависит от того, является ли направление поиска направлением спуска.

3.3 Метод Ньютона с направлением спуска

Гарантией убывания значения функции является нахождение такого направления одномерного поиска, которое будет и направлением спуска. Метод Ньютона приводит к ошибочному направлению в точке максимума и в седловых точках.

s_k – направление спуска, если скалярное произведение вектора направления из точки x_k и вектора градиента в этой точке отрицательно (то есть между ними образуется тупой угол). Если же угол острый, то s_k не является направлением убывания, следовательно, разумнее использовать направление антиградиента, так как оно гарантирует убывание функции.

Таким образом, направление спуска задается следующим образом:

$$H_k s_k = -g_k$$

$$d_k = \begin{cases} s_k, & s_k^t g_k < 0 \\ -g_k, & s_k^t g_k \geq 0 \end{cases}$$

Накопленный опыт показал, что первую итерацию разумнее проводить в направлении антиградиента с использованием одномерного поиска. Тогда на первом шаге $d_0 = -g_0$, $r_0 = \operatorname{argmin} f(x_0 + \lambda d_0)$, $s_0 = r_0 \cdot d_0$, $x_1 = x_0 + s_0$. Далее выполняется итерационный процесс, пока $\|s_k\| > \epsilon$:

1. $Hs = -g$
2. Если $s^T \cdot g < 0$, то $d = s$, иначе $d = -g$
3. $r = \operatorname{argmin} f(x + \lambda d)$, $s = rd$
4. $x = x + s$

4 Квазиньютоновские методы

Квазиньютоновские методы основываются на методе Ньютона с одномерным поиском. В методах Ньютона требуется вычисление гессиана, что при больших размерностях пространства требует больших затрат машинного времени, поэтому разумно аппроксимировать матрицу Гессе или обратную к ней с использованием значений градиента. В итоге, получается метод первого порядка.

Гессиан – симметрическая матрица, поэтому и обратная к ней $G(x)$ тоже симметрическая. Зададим ее начальное приближение в виде некоторой симметрической положительно определенной матрицы G_0 . Такое задание обеспечивает соответствующее направление одномерного поиска $d_0 = -G_0 g_0$ из начальной точки x_0 как направление спуска. Следующая точка ищется так:

$$x_1 = x_0 + \lambda_0 d_0, \text{ где } \lambda_0 = \operatorname{argmin}_f(x_0 + \lambda d_0)$$

В последующих итерациях построение аппроксимирующей матрицы основано на свойстве квадратичной функции:

$$g_{k+1} - g_k = H(x_{k+1} - x_k)$$

Пусть $p_k = g_{k+1} - g_k$, $s_k = x_{k+1} - x_k$, тогда $G p_k = s_k$. Если известна аппроксимация для G , то следующее ее приближение в точке x_{k+1} должно удовлетворять уравнению

$$G_{k+1} p_k = s_k$$

Но найти такую матрицу невозможно, так как количество неизвестных в силу симметричности G превышает количество уравнений n . Поэтому используется дополнительное условие – уравнение коррекции:

$$G_{k+1} = G_k + \Delta G_k$$

Различные квазиньютоновские методы отличаются между собой формулами для поправки ΔG_k .

4.1 Метод Бroyдена-Флетчера-Шено

В данном методе аппроксимируется сама матрица Гессе по принципу: $H_{k+1} s_k = p_k$. Для нее поправка имеет вид:

$$\frac{p_k p_k^T}{p_k^T s_k} - \frac{H_k s_k (H_k s_k)^T}{(H_k s_k)^T s_k}$$

Поскольку аппроксимируется гессиан, а не обратная матрица, поэтому направление поиска ищется путем решения СЛАУ: $H_k d_k = -g_k$.

4.2 Метод Пауэлла

$$\Delta G_k = -\frac{u \cdot u^T}{(u^T, p)}$$

где $u_k = s_k - G_k p$, $p = \nabla f(x_{k+1}) - \nabla f(x_k)$.

Направление одномерного поиска $d_k = -G g_k$.

5 Метод Марквардта

Этот метод является комбинацией методов наискорейшего спуска и метода Ньютона. Движение в направлении антиградиента из начальной точки поиска обычно приводит к существенному уменьшению целевой функции. С другой стороны, направления эффективного поиска в окрестности точки минимума определяются по методу Ньютона.

Этот метод и его модификация характеризуются относительной простотой, свойством убывания $f(x)$ при переходе от итерации к итерации, высокой скоростью сходимости в окрестности x^* и отсутствием процедуры одномерного поиска.

5.1 Обычный метод Марквардта

В данном методе используется следующая идея (α – скалярный параметр, E – единичная матрица):

$$(H(x) + \alpha E) s = -g(x)$$

1. При большом значении α матрицей H можно пренебречь, тогда получается уравнение $\alpha E s = -g(x)$, и в этом случае направление вектора s совпадает с антиградиентом (направление наискорейшего спуска).
2. При $\alpha \rightarrow 0$ можно пренебречь αE , тогда s – шаг метода Ньютона.

Таким образом, выбрав начальное значение α_0 достаточно большим, а затем уменьшая его, сначала будут выполняться шаги наискорейшего спуска, а на конечных этапах – метода Ньютона.

В общем виде:

$$x_{k+1} = x_k + s_k, (H_k + \alpha_k E) s_k = -g_k, \alpha_k = \alpha_{k-1} - 1/\beta$$

Но если значение функции в новой точке больше, чем в предыдущей, то требуется корректировка $\alpha := \frac{\alpha}{\beta}$.

5.2 Метод Марквардта с применением разложения Холецкого

Можно начинать с $\alpha_0 = 0$, тогда на каждой итерации должно выполняться условие на $H + \alpha E$ – она должна быть положительно определенной. Если это условие не выполняется, то перебирается $\alpha = \max(1, 2\alpha)$.

Условие положительно определенной формы проверяется с помощью разложения Холецкого: $H + \alpha I = LL^T$, где L – нижнетреугольная матрица. Тогда если такое разложение возможно, то она положительно определена.

Сложность метода в таком случае будет складываться из числа запусков алгоритма Холецкого. Его сложность – $O(n^3)$

6 Выводы

Метод Марквардта позволяет устранить недостатки метода Ньютона – невозможность корректировки длины шага и возможно плохую обусловленность СЛАУ.

Реализация: [GitHub](#).