Лабораторная работа 3

Градиентные методы минимизации функций многих переменных

Постановка задачи: Требуется найти безусловный минимум функции n переменных $f(x)=f(x_1,x_2,...,x_n)$, т.е. такую точку $x^* \in E_n$, что $f(x^*)=\min\limits_{x\in E_n} f(x)$. Предполагается, что целевая функция $f(x)$ - дифференцируема в E_n и возможно вычисление ее производных в произвольной точке E_n.

В данной работе рассматриваются итерационные процедуры минимизации вида

$$x^{k+1}=x^k+\alpha_k p^k, \quad k=0,1,... \tag{1}$$

где направление убывания p^k определяется тем или иным способом с учетом информации о частных производных функции $f(x)$, а величина шага $\alpha_k>0$ такова, что

$$f(x^{k+1})<f(x^k), \quad k=1,2,... \tag{2}$$

Метод градиентного спуска

Стратегия поиска: Положим в (1) на каждом шаге $p^k=-\nabla f(x^k)$. Если $\nabla f(x^k)\neq 0$, то условие $(\nabla f(x^k),p^k)<0$, очевидно, выполнено. Следовательно, направление вектора p^k является направлением убывания функции $f(x)$, причем в малой окрестности точки x^k направление p^k обеспечивает наискорейшее убывание этой функции. Поэтому найдется такое $\alpha_k>0$, что будет выполняться условие (2).

Алгоритм:

1.Задать параметр точности $\varepsilon>0$, начальный шаг $\alpha>0$, выбрать $x\in E_n$. Вычислить $f(x)$.

2. Вычислить $\nabla f(x)$ и проверить условие достижения точности: $\|\nabla f(x)\|<\varepsilon$. Если оно выполнено, вычисления завершить, положив $x^*=x$, $f^*=f(x)$. Иначе – перейти к шагу 3.

3. Найти $y=x-\alpha\cdot\nabla f(x)$ и $f(y)$. Если $f(y)<f(x)$, то положить $x=y$, $f(x)=f(y)$ и перейти к шагу 2, иначе – к шагу 4.

4. Положить $\alpha=\alpha/2$ и перейти к шагу 3.

Замечание. Вблизи стационарной точки функции $f(x)$ величина $\|\nabla f(x)\|$ становится малой. Это часто приводит к замедлению сходимости последовательности $\{x^k\}$. Поэтому в основной формуле (1) иногда полагают $p^k=-\nabla f(x^k)/\|\nabla f(x^k)\|$, используя вместо антиградиента вектор единичной длины в этом же направлении.

Метод наискорейшего спуска

Стратегия поиска: Метод является модификацией метода градиентного спуска. Здесь также полагают $p^k=-\nabla f(x^k)$, но величина шага α_k из (1) находится в результате решения задачи одномерной оптимизации

$$\Phi_k(\alpha) \to \min, \quad \Phi_k(\alpha) = f(x^k - \alpha \nabla f(x^k)), \quad \alpha > 0, \tag{3}$$

т.е. на каждой итерации в направлении антиградиента $-\nabla f(x^k)$ совершается исчерпывающий спуск.

Алгоритм:

1. Задать параметр точности $\varepsilon > 0$, выбрать $x \in E_n$. Вычислить $f(x)$.

2. Вычислить $\nabla f(x)$ и проверить условие достижения точности: $\|\nabla f(x)\| < \varepsilon$. Если оно выполнено, вычисления завершить, полагая $x^* = x$, $f^* = f(x)$. Иначе – перейти к шагу 3.

3. Решить задачу одномерной оптимизации (3) для $x^k = x$, т.е. найти α^*. Положить $x = x - \alpha^* \cdot \nabla f(x)$ и перейти к шагу 2.

Замечание. Решение задачи (3) можно получить одним из рассмотренных для одномерной минимизации способов. Проверять выполнение условия (2) в этом методе необязательно. Если функция $f(x)$ квадратична, т.е. $f(x) = \frac{1}{2}(Ax, x) + (b, x) + c$, то для величины шага исчерпывающего спуска можно получить точную формулу $\alpha_k = -\frac{(\nabla f(x^k), p^k)}{(Ap^k, p^k)} = -\frac{(Ax^k + b, p^k)}{(Ap^k, p^k)}$, положив $p^k = -\nabla f(x^k)$.

Метод сопряженных градиентов

Стратегия поиска: В методах градиентного спуска в итерационной процедуре (1) в качестве направления убывания функции $f(x)$ использовалось направление антиградиента: $p^k = -\nabla f(x^k)$. Однако такой выбор направления убывания не всегда бывает удачным. В частности, для плохо обусловленных задач минимизации направление антиградиента в точке x^k может значительно отличаться от направления к точке минимума x^*. В результате траектория приближения к точке минимума имеет зигзагообразный характер. В методе сопряженных градиентов применяется другой подход. Используется итерационный процесс

$$x^{k+1} = x^k + \alpha_k \cdot p^k, \quad k = 0,1,...; \quad x^0 \in E_n, \quad p^0 = -\nabla f(x^0),$$

в котором величина шага α_k находится из условия исчерпывающего спуска по направлению p^k.

$$f(x^k + \alpha_k\, p^k) = \min_{\alpha > 0} f(x^k + \alpha\, p^k), \quad k = 0,1,...,$$

Далее, после вычисления очередной точки x^{k+1}, $k = 0,1,...$, новое направление поиска p^{k+1} находится по формуле, отличной от антиградиента:

$$p^{k+1} = -\nabla f(x^{k+1}) + \beta_k\, p^k, \quad k = 0,1,...,$$

где коэффициенты β_k выбираются так, чтобы при минимизации квадратичной функции $f(x)$ с положительно определенной матрицей A получалась последовательность A- ортогональных векторов $p^0, p^1, ...$.

Для величины β_k можно получить формулу

$$\beta_k = \frac{\|\nabla f(x^{k+1})\|^2}{\|\nabla f(x^k)\|^2}, \quad k = 1,2,...$$

Выражение для коэффициента β_k не содержит в явном виде матрицу A квадратичной формы. Поэтому метод сопряженных градиентов применяют для минимизации любых, а не только квадратичных функций.

Если функция $f(x)$ квадратична, т.е. $f(x)=\frac{1}{2}(Ax,x)+(b,x)+c$, то для величины шага исчерпывающего спуска можно получить точную формулу

$$\alpha_k=-\frac{(\nabla f(x^k),p^k)}{(Ap^k,p^k)}.$$

При реализации метода сопряженных градиентов применяется практический прием – через каждые N шагов производят обновление метода, полагая $\beta_{m\cdot N}=0$, $m=1,2,\ldots$. Номера $m\cdot N$ называют моментами рестарта. Часто полагают $N=n$ - размерности пространства E_n. Если $N=1$, то получается частный случай метода сопряженных градиентов – метод наискорейшего спуска.

Алгоритм:

1.Задать точность вычислений ε, выбрать начальное приближение x^0 и частоту обновления N.

2.Положить $k=0$ (k - номер итерации). Вычислить значение $p^0=-\nabla f(x^0)$.

3.Вычислить значение α_k из решения задачи одномерного исчерпывающего спуска $f(x^k+\alpha_k\,p^k)=\min\limits_{\alpha>0} f(x^k+\alpha\,p^k)$

4.Вычислить точку $x^{k+1}=x^k+\alpha_k\cdot p^k$ и градиент $\nabla f(x^{k+1})$.

5.Проверить критерий окончания поиска $\left|\nabla f(x^{k+1})\right|<\varepsilon$. Если критерий выполнен, перейти к шагу 8.

6.Если $k+1=N$, то положить $x^0=x^{k+1}$, $k=0$ и перейти к шагу 2 (рестарт).

7.Вычислить коэффициент $\beta_k=\dfrac{\left\|\nabla f(x^{k+1})\right\|^2}{\left\|\nabla f(x^k)\right\|^2}$ и найти новое направление поиска $p^{k+1}=-\nabla f(x^{k+1})+\beta_k\,p^k$. Положить $k=k+1$ и перейти к шагу 3.

8.Выбрать приближенно $x^*=x^{k+1}$, $f(x^*)=f(x^{k+1})$. Поиск завершен.

Метод Ньютона

Стратегия поиска: Пусть 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 порядка $o(\|\Delta x\|^2)$ может быть описано квадратичной функцией

$$\Phi_k(x)=f(x^k)+\nabla f(x^k)^T\Delta x+\frac{1}{2}\Delta x^T H(x^k)\Delta x. \qquad (4)$$

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

$$\nabla\Phi_k(x)=H(x^k)\cdot(x-x^k)+\nabla f(x^k)=0. \tag{5}$$

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 (4) с положительно определенной матрицей $H(x^k)$ сильно выпукла и уравнение (5) определяет единственную точку глобального минимума функции $\Phi_k(x)$. Умножим слева обе части равенства (5) на матрицу $\left[H(x^k)\right]^{-1}$ и найдем точку минимума x^{k+1} квадратичной функции $\Phi_k(x)$, аппроксимирующей $f(x)$ в окрестности точки $x=x^k$:

$$x^{k+1}=x^k-\left[H(x^k)\right]^{-1}\cdot\nabla f(x^k),\quad k=0,1,... \tag{6}$$

Итерационный процесс (6), начатый из произвольной точки $x^0\in E_n$, называется <u>методом Ньютона минимизации функции многих переменных</u> и является обобщением метода Ньютона в одномерном случае.

<u>**Квазиньютоновские методы**</u>

Построить квазиньютоновский алгоритм в соответствии с п.5.8. Главы 5 пособия или воспользоваться одним из следующих квазиньютоновских алгоритмов из дополнительной литературы: метод Давидона – Флетчера – Пауэлла (ДФП); метод Бройдена – Флетчера – Шенно (БФШ); метод Мак-Кормика.

<u>**Задания**</u>

1. Реализовать в среде MATLAB метод градиентного спуска, метод наискорейшего спуска и метод сопряженных градиентов.

В методе наискорейшего спуска и в методе сопряженных градиентов наряду с аналитическим выражением для величины шага исчерпывающего спуска для квадратичной функции реализовать решение задач одномерной минимизации методом поразрядного поиска.

2. Протестировать работу реализованных методов на примере овражной функции

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

при a = 1, 100, 500, 1000. При $\varepsilon=10^{-3}$ и $\varepsilon=10^{-5}$ сравнить скорость работы методов при различных значениях параметра a по числу итераций и по числу вызовов совокупности значений функций и производных.

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): 23-9×2=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. Графически отобразить линии уровня выбранной функции. Сравнить эффективность методов градиентного, наискорейшего спуска, а так же метода сопряженных градиентов для задачи п.2 при a=250 и тестовой функции п.3 по числу итераций. *Объяснить полученные результаты.*

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 $x^0 = (-1,\ 1)^T$.
Для решения задачи использовать метод Ньютона, один из квазиньютоновских методов (воспользоваться одним из следующих квазиньютоновских алгоритмов из дополнительной литературы: метод Давидона – Флетчера – Пауэлла (ДФП); метод Бройдена – Флетчера – Шенно (БФШ); метод Мак-Кормика) и метод сопряженных градиентов. Определить, сколько итераций потребуется каждому методу для того, чтобы разность между численным и точным решением $x^* = (1,\ 1)^T$ была меньше ε. *Сравнить эффективность методов.*

6. Исследовать работу метода сопряженных градиентов в зависимости от частоты обновлений N на примере функции Розенброка. *Какое значение можно назвать оптимальным?*

7. На примере функции Химмельблау

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

рассмотреть особенности применения градиентных 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 необходимо сохранить для использования в лабораторной работе 4.

9.Сдать лабораторную работу преподавателю, *ответив предварительно на все следующие контрольные вопросы.*

Контрольные вопросы к Лабораторной работе 3

1).Функции какого вида называются квадратичными функциями n переменных?

2).Чему равны градиент и гессиан квадратичной функции?

3).Каким свойством обладает квадратичная функция с положительно определенной матрицей A?

4).При каких a, b, c функция $f(x) = ax_1^2 + bx_1x_2 + cx_2^2$ будет выпукла?

5).Выписать матрицу A квадратичной функции $f(x) = x_1^2 + 3x_3^2 + 2x_1x_2 - x_2x_3$ $+ 2x_2 + x_3$.

6).Какая последовательность $\{x^k\}$, $k = 0,1,2...$ называется минимизирующей?

7).Привести пример минимизирующей последовательности, не сходящейся к точке минимума.

8).Что такое скорость сходимости минимизирующей последовательности? Какие скорости сходимости Вы знаете?

9).Когда говорят, что в итерационном процессе $x^{k+1} = x^k + \alpha_k p^k$, $k = 0,1,...$ производится исчерпывающий спуск?

10).Какие направления дифференцируемой в точке x^k функции $f(x)$ называются направлениями убывания? Каков геометрический смысл направления убывания?

11).Какова скорость сходимости метода градиентного спуска для квадратичной функции $f(x)$ с положительно определенной симметрической матрицей A, где $0 < l < L$ - ее наименьшее и наибольшее собственные значения?

12).Когда говорят, что сильно выпуклая функция $f(x)$ имеет "овражный характер"? Какие задачи минимизации называются хорошо обусловленными, а какие – плохо обусловленными?

13).В чем состоят преимущества и недостатки метода наискорейшего спуска по сравнению с методом градиентного спуска?

14).Каков главный недостаток градиентных методов?

15).В чем состоит идея метода сопряженных градиентов? Чем этот метод отличается от методов градиентного и наискорейшего спуска?

16).Какова скорость сходимости метода Ньютона для дважды дифференцируемой выпуклой функции $f(x)$ многих переменных? Какова трудоемкость этого метода?

17).Чем отличаются классический и обобщенный методы Ньютона для сильновыпуклой дважды дифференцируемой функции многих переменных?

18).Сформулировать общий принцип построения квазиньютоновских методов. Какую скорость сходимости следует ожидать от квазиньютоновских методов? Какова их трудоемкость?