Работа 3. ГРАДИЕНТНЫЕ МЕТОДЫ МИНИМИЗАЦИИ

1. Что такое градиентные методы?

Градиентные методы, как и все «поисковые» методы носят итерационный характер, т.е. включают в свой состав процедуры получения последовательности векторов (или точек) в N-мерном пространстве $x^{(k)}$, k=0,1,2,3,…,**K** таких, что целевая функция f(x) на каждом шаге уменьшается:

$$f(x^{(0)}) > f(x^{(1)}) > \dots > f(x^{(k)}) > f(x^{(k+1)}) > \dots .$$

Более конкретно: **градиентными** называются такие методы минимизации многомерной функции f(x), которые вычисляют очередную точку поиска $x^{(k+1)}$ по итерационной формуле:

$$x^{(k+1)} = x^{(k)} + t_k \cdot [\boldsymbol{s}(x^{(k)})], \qquad (3.1)$$

где $x^{(k)}$ – текущая k-я точка (k=0,1,2,…,**K**) поиска минимума x^* (напомним, что начальная точка поиска $x^{(0)}$ задается пользователем);

t_k - скаляр, определяющий длину k-го шага;

$\boldsymbol{s}(x^{(k)})$ – **направление** k-го шага, которое определяется **градиентом** $\nabla f(x)$, либо **матрицей Гессе** $\nabla^2 f(x)$ исходной функции f(x).

Именно то, что в формуле (3.1) используется **направление** поиска $\boldsymbol{s}(x^{(k)})$, зависящее от градиента, дало название большой группе методов под общим названием **градиентные методы.**

Градиентные методы разделяются на методы **первого и второго порядка.** Способ, которым вычисляется $\boldsymbol{s}(x^{(k)})$ и t_k в (3.1) определяет название метода. Напомним, что вычисление t_k осуществляется путем решения задачи одномерной минимизации $f(x)$ в направлении $\boldsymbol{s}(x^{(k)})$ каким-либо одним из одномерных методов.

В этой работе вы подробно рассмотрите и самостоятельно запрограммируете следующие градиентные методы минимизации:

- метод **оптимального градиентного поиска (метод Коши);**
- метод **Ньютона;**
- метод **Марквардта.**

Кроме этого, вы также сравните вышеуказанные методы с такими, как

- метод **Флетчера-Ривса;**
- метод **Дэвидона-Флетчера-Пауэлла.**

2. Метод оптимального градиентного поиска (метод Коши)

Этот метод **первого порядка** основан на определении **локального направления наискорейшего убывания целевой функции** $f(x)$ **из некоторой произвольной точки** $x^{(k)}$**.**

- Что это за **локальное направление наискорейшего убывания** функции $f(x)$?

Определим его аналитически. Для этого разложим целевую функцию в окрестности точки $x^{(k)}$ в ряд Тейлора до линейного члена включительно, т.е. найдем линейную аппроксимационную функцию:

$$f(x,x^{(k)})=f(x^{(k)})+\nabla f(x^{(k)})^T \Delta x \; + \; \ldots \; , \qquad (3.2)$$

где $\nabla f(x^{(k)})$ - вектор-градиент в точке разложения;

$\Delta x=(x^{(k)}-x)$ - вектор приращения аргумента.

Из (3.2) видно, что **локальное** уменьшение целевой функции $f(x,x^{(k)})$ определяется только вторым слагаемым, т.к. $f(x^{(k)})$ фиксировано.

Тогда наибольшее локальное уменьшение $f(x,x^{(k)})$ связано с выбором такого направления Δx, при котором второе слагаемое, т.е. скалярное произведение в (3.2) будет **максимальным и отрицательным.**

Из свойства скалярного произведения следует, что указанные требования обеспечиваются при условии выбора **антиградиентного** направления, т.е.

$$\Delta x=-\nabla f(x^{(k)}) \, . \qquad (3.2a)$$

Подставляя найденное направление $-\nabla f(x^{(k)})$ в формулу (3.1) получим **формулу градиентного поиска**

$$x^{(k+1)} = x^{(k)} - t_k \nabla f(x^{(k)}) , \; k=0,1,2, \ldots , \qquad (3.3)$$

где $t_k>0$ – **скалярная величина,** задающая **длину шага** поиска, и пока неизвестная.

Определим ее. Говорят, что если t_k является **решением одномерной задачи минимизации**

$$f(x^{(k)}- t_k \nabla f(x^{(k)})) \to \min_{t_k>0} , \qquad (3.4)$$

то в этом случае формулы (3.3) и (3.4) представляют **оптимальный** градиентный метод, или, что то же **- метод Коши**[1]**.**

- **Краткое обсуждение работы оптимального градиентного метода**

Согласно этому методу, **следующая точка поиска** $x^{(k+1)}$ вычисляется так, что:

---

[1] Оптимальный градиентный метод часто называется методом наискорейшего спуска, или методом Коши, который первый использовал аналогичный алгоритм для решения систем линейных уравнений (Коши Огюстен Луи-(1789-1857гг), французский математик)

во-первых, она расположена в направлении **антиградиента** $-\nabla f(x^{(k)})$;

во-вторых, функция $f(\mathbf{x}^{(k+1)})$ в этой точке **принимает минимальное значение**.

Далее итерация повторяется: вновь определяется антиградиент в точке $x^{(k+1)}$ и т.д. Одномерная задача (3.4) вычисления t_k может решаться любым из методов **одномерного поиска** (см. л/р №2).

Важно также отметить, что эпитет «**оптимальный**» в названии метода указывает просто на тот факт, что в качестве следующей точки выбирается точка, лежащая на прямой антиградиента ($x^{(k)} - \nabla f(x^{(k)})$) и являющаяся на этой **прямой точкой минимума («оптимума»)** функции $f(x)$. Значит термин «оптимальный» не говорит о том, что метод Коши обладает какими-либо **иными** оптимальными свойствами, например, способностью осуществлять поиск минимальной точки за минимальное число шагов.

Иногда параметр шага не вычисляется, а просто задается. Градиентный метод, в котором параметр шага t_k берется относительно малой величиной для всех шагов, называют **градиентным методом с малым шагом.** В **градиентных методах с большим шагом** параметр шага t_k назначается больше оптимального значения.

- **Пример 3.1. Оптимальный градиентный метод (метод Коши)**

Рассмотрим квадратичную функцию $f(x)=8x_1^2 + 4x_1x_2 + 5x_2^2$ и используем метод Коши для ее минимизации.

Решение. Компоненты вектор-градиента $\Delta f(x)$ равны

$$\partial f/\partial x_1 = \nabla f_1 = 16x_1 + 4x_2,$$

$$\partial f/\partial x_2 = \nabla f_2 = 10x_2 + 4x_1,$$

т.е. $\Delta f(x)=(16x_1+4x_2,\ 10x_2+4x_1)^T$.

Зададим начальное приближение, например $x^0=(10,10)^T$ и с помощью формулы (3.3) построим новое приближение:

$$x^{(1)} = x^{(0)} - t_0\nabla f(x^{(0)}),$$

решая задачу одномерной минимизации (3.4) относительно скалярного параметра t_0:

$$8(x_1^{(0)}-t_0\nabla f_1)^2+4(x_1^{(0)}-t_0\nabla f_1)(x_2^{(0)}-t_0\nabla f_2)+5(x_1^{(0)}-t_0\nabla f_2)^2 \to \min$$

найдем, что $t_0=0.056$.

Следовательно, из начальной точки $x^{(0)}$ мы переходим в следующую точку поиска:

$$x^{(1)}=(-1.20,\ 2.16)^T.$$

Далее, повторяя итерации, получим данные, которые сведены в табл.3.1.

Таблица 3.1. Результаты минимизации методом Коши

Номер итерации (k)	Составляющая $x_1^{(k)}$	Составляющая $x_2^{(k)}$	Значение функции $f(x^{(k)})$
0	10.0000	10.0000	1700.0000
1	- 1.2403	2.1181	24.2300
2	0.1441	0.1447	0.3540
3	- 0.0181	0.0309	0.0052
4	0.0021	0.0021	0.0000

По данным табл. 3.1 построим на фоне линий равного уровня целевой функции f(x) **три начальных шага траектории Коши** (рис.3.1):

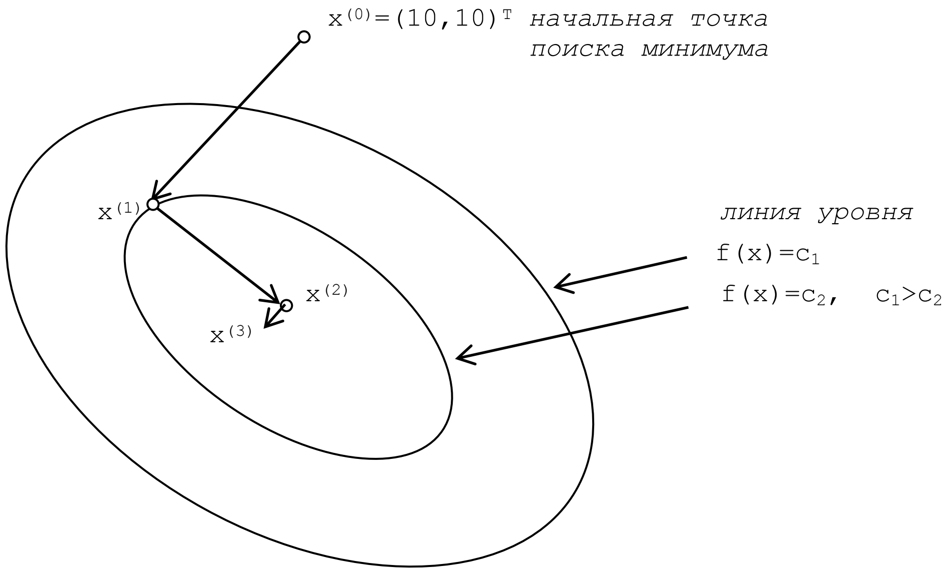

Рис. 3.1. Три начальные итерации (шага) метода Коши поиска минимума из начальной точки x^0

Как видно из рис. 3.1, траектория Коши носит зигзагообразный характер, причем, можно доказать, что шаги траектории **взаимно ортогональны.**

Анализ данных табл. 3.1 показывает, что метод Коши позволяет существенно уменьшить значение целевой функции при движении из точек, расположенных на значительных расстояниях от точки минимума, и поэтому могут использоваться при реализации градиентных методов в качестве начальной процедуры.

- **Представим основные операции алгоритма Коши**

1. Задать M, $x^{(0)}, \varepsilon_1, \varepsilon_2$.
2. Положить $k=0$.
3. Вычислить градиент $\nabla f(x^{(k)})$.
4. Если $\|\nabla f(x^{(0)})\| \le \varepsilon_1$, то перейти к 10.

5. Если M ≤ k, то перейти к 10.
6. Вычислить t_k любым одномерным методом с точностью ε_2.
7. Положить $x^{(k+1)} = x^{(k)} - t_k \nabla f(x^{(k)})$.
8. Если $\|x^{(k+1)} - x^{(k)}\| / \|x^{(k)}\| \le \varepsilon_1$, то перейти к 10.
9. Положить $k=k+1$; перейти к 3.
10. Печать результатов; стоп.

3. Метод Ньютона

Метод основывается на **квадратичной аппроксимации** рядом Тейлора функции $f(x)$ в окрестности точки $x^{(k)}$. Это значит, что в методе Ньютона используется информация о **вторых производных** $f(x)$, т.е. о матрице Гессе. Поэтому он относится к методам **второго порядка.**

Для вывода **итерационной формулы** метода Ньютона разложим $f(x)$ в ряд Тейлора до второго члена включительно. Получим **аппроксимирующую функцию** в окрестности точки $x^{(k)}$:

$$f(x;x^{(k)}) = f(x^{(k)}) + \nabla f(x^{(k)})^T \Delta x + \tfrac{1}{2}\, \Delta x^T\, \nabla^2 f(x^{(k)})\, \Delta x, \qquad (3.5)$$

где $\nabla^2 f(x^{(k)}) = \mathbf{H}(\boldsymbol{x}^{(k)})$ – симметрическая квадратная матрица **вторых частных производных** функции $f(x)$, *или* **матрица Гессе.**

Запомните основную идею, на которой базируется итерационная формула Ньютона:

нужно вычислять последовательность итераций таким образом, **чтобы во вновь получаемой точке $\mathbf{x}^{(k+1)}$ градиент аппроксимирующей функции** (3.5) **обращался в нуль.**

Тогда формально можно записать:

$$\nabla f(x;x^{(k)}) = \nabla f(x^{(k)}) + \nabla^2 f(x^{(k)})\, \Delta x = 0,$$

откуда искомое направление поиска в методе Ньютона равно

$$\Delta x = -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)}) = -H^{-1}(x^{(k)})\, \nabla f(x^{(k)}). \qquad (3.5a)$$

Обратите внимание, что в формуле **направление Ньютона** равно антиградиентному направлению, умноженному на обратную матрицу Гессе. Полезно сравнить направление Коши (3.2a) с направлением Ньютона (3.5a).

Последовательное применение на каждой итерации указанной схемы квадратичной аппроксимации, составляет существо метода Ньютона, т.е. итерации производятся по формуле

$$x^{(k+1)} = x^{(k)} - \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)}), \qquad (3.6)$$

или, введя другое обозначение матрицы Гессе, окончательно получим:

$$x^{(k+1)} = x^{(k)} - H^{-1}(x^{(k)})\, \nabla f(x^{(k)}). \qquad (3.7)$$

Это последняя формула и определяет **метод Ньютона.**

- **Пример 3.2. Минимизация методом Ньютона**

Рассмотрим вновь функцию из предыдущего примера $f(x)=8x_1^2 + 4x_1x_2 + 5x_2^2$ и минимизируем ее методом Ньютона.

Градиент целевой функции Δf(x) равен:

$$\nabla f(x) = (16x_1+4x_2,\ 10x_2+4x_1)^T,$$

а матрица Гессе $\Delta^2 f(x)=H(x)$ такая:

$$\nabla^2 f(x) = H(x) = \begin{bmatrix} 16 & 4 \\ 4 & 10 \end{bmatrix}.$$

Зададим ту же начальную точку $x^0=(10,10)^T$.

Тогда по формуле (3.5) или (3.6) получим следующую точку поиска:

$$x^{(1)}=(10,\ 10)^T - (1/144) \begin{bmatrix} 10 & -4 \\ -4 & 16 \end{bmatrix} (200,\ 140)^T,$$

т.е:

$$x^{(1)}=(10,10)^T-(1/144)(1440,1440)^T=(0,0)^T,$$

что соответствует **точному решению этой задачи.**

Таким образом, метод Ньютона для квадратичной целевой функции за один шаг достигает точки минимума.

Это означает, что данный метод вырабатывает из любой начальной точки **глобальное направление поиска** – но только для квадратичных форм.

4. Модифицированный метод Ньютона

- Заметим, что предыдущий метод Ньютона **не регулирует длину шага.** Хорошо это или плохо? Как показывает практика применения метода для не квадратичных функций, **сходимость** метода к минимуму сильно зависит от выбора начальной точки поиска (вспомните свойства сходимости метода Ньютона для одномерного поиска – см. лабораторную работу №1)

Поэтому сейчас полезно проанализировать **условия сходимости метода.**

По определению известно, что направление поиска ***s**(x)* является направлением уменьшения функции (направлением спуска), если **имеет место неравенство** $\nabla f(x)^T \mathbf{s(x)}<0$. Тогда из формулы метода Ньютона (3.6) следует, что должно выполняться неравенство

$$-\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) < 0\,.$$

Нетрудно видеть, что в случае, когда матрица Гессе $\nabla^2 f(x)$ **положительно знакоопределена,** это условие выполняется, т.е. направление поиска по методу Ньютона оказывается **направлением спуска.**

Однако, если в некоторой точке траектории поиска матрица $\nabla^2 f(x)$ становится **отрицательно знакоопределена**, то указанное направление является **направлением подъема**, а в случае **неопределенности** матрицы Гессе однозначный вывод о сходимости метода сделать нельзя и требуются дополнительные исследования.

- Как уже было отмечено, опыт практической оптимизации показывает, что при минимизации **не квадратичных функций**, метод Ньютона не отличается высокой надежностью.

Это происходит из-за того, что если начальная точка поиска находится достаточно далеко от минимума, то **шаг Ньютона** часто оказывается **чрезмерно большим**. Это может привести к расходящемуся итерационному процессу минимизации.

В **таких ситуациях**[1] нужно применять **модифицированный метод Ньютона,** в котором итерации вычисляются по формуле

$$x^{(k+1)}=x^{(k)}-t_k H^{-1}(x^{(k)})\,\nabla f(x^{(k)}), \qquad (3.8)$$

где t_k–**параметр шага**, аналогичный используемому в методе Коши.

Модифицированный метод Ньютона отличается высокой скоростью сходимости, но на каждой итерации требует вычислений элементов обратной матрицы Гессе.

5. Метод Марквардта: комбинация методов Коши и Ньютона

- Желание использовать методы Коши и Ньютона так, чтобы вдали от точки минимума поиск проводился в антиградиентном направлении, а по мере приближения к минимуму, использовались итерации по ньютоновскому направлению, привело исследователей к созданию **метода Марквардта**, суть которого заключается в следующем.

Если направление поиска $s(x^{(k)})$ в точке $x^{(k)}$ вычислять по формуле

$$s(x^{(k)}) = -(H^{(k)}+\lambda_k \mathbf{I})^{-1}\ \nabla f(x^{(k)}), \qquad (3.9)$$

где $H^{(k)}$ – матрица Гессе точке $x^{(k)}$;

λ_k – некоторая скалярная величина;

I - единичная матрица NxN,

то в зависимости от выбора величины λ_k направление $s(x^{(k)})$ может меняться от антиградиентного до ньютоновского.

Действительно, задав на начальной стадии поиска большое значение параметра $\lambda_0=(10^4 \div 10^6)$, имеем

$$(H^{(0)}+\lambda_0\mathbf{I})^{-1}=(\lambda_0\mathbf{I})^{-1}=(1/\lambda_0)\mathbf{I}.$$

Т.е. большим значениям λ_k в (3.9) соответствует антиградиентное направление или близкое к нему.

С другой стороны, можно показать, что при малых значениях λ_k формула (3.9) будет «вырабатывать» направление Ньютона.

[1] Каким образом графически характеризуется термин «такая ситуация»?

- При программной реализации метода Марквардта управлять выбором значений λ_k рекомендуется путем сравнения значений функции на каждой итерации:

  -если целевая функция на текущей итерации уменьшилась, то следует уменьшить величину λ_k и реализовать очередной шаг;

  -в противном случае следует положить $\lambda_{k+1}=\beta\lambda_k$, где $\beta>0$, и вновь повторить шаг поиска.

Метод Марквардта характеризуется относительной простотой, свойством убывания целевой функции при переходе от итерации к итерации, высокой скоростью сходимости в окрестности точки минимума, а также отсутствием процедуры одномерной минимизации.

Следующий раздел – только по специальному указанию преподавателя.

6. Методы сопряженных градиентов- наиболее эффективная группа методов

Вкратце рассмотрим *методы* ***Флетчера-Ривса и Дэвидона-Флетчера-Пауэлла***, которые составляют группу методов **сопряженных градиентов**.

В *работе* 2 данного курса был рассмотрен метод сопряженных направлений Пауэлла, с помощью которого можно находить сопряженные направления только для квадратичных функций, причем для этого использовалась информация о значениях целевой функции.

Теперь же мы попытаемся минимизировать функцию общего вида, используя значения *квадратичной аппроксимации* исходной функции и значения *компонент градиента*. Кроме того, мы потребуем, чтобы на каждой итерации обеспечивалось уменьшение целевой функции.

Теория методов сопряженных градиентов базируется на следующем свойстве квадратичных форм.

Пусть в пространстве переменных заданы две произвольные несовпадающие точки $x^{(0)}$ и $x^{(1)}$. Как известно, градиент квадратичной функции $q(x)=\frac{1}{2}x^{T}\mathbf{Q}x+b^{T}x+a$ равен

$$\nabla f(x) = \nabla q(x) = \mathbf{Q}x + b = g(x)\,.$$

Обозначение $g(x)$ введено для удобства записи дальнейших выкладок.

Таким образом,

$$g(x^{(0)}) = \mathbf{Q}x^{(0)}+b,$$

$$g(x^{(1)}) = \mathbf{Q}x^{(1)}+b.$$

Запишем **изменение градиента при переходе от точки $x^{(0)}$ к точке $x^{(1)}$**:

$$\Delta g(x) = g(x^{(1)}) - g(x^{(0)}) = \mathbf{Q}(x^{(1)} - x^{(0)}),$$

$$\Delta g(x) = \mathbf{Q}\,\Delta x\,. \qquad (3.10)$$

Это последнее равенство (3.10) и выражает очередное свойство квадратичных функций, которое будет использовано дальше.

Эстенс и Штифель (1952г) предложили эффективный итерационный алгоритм для решения систем линейных уравнений, который по существу представлял собой метод сопряженных градиентов. Они рассматривали левые части линейных уравнений как компоненты градиента квадратичной функции и решали задачу минимизации этой функции.

Напомним, что решение системы N линейных уравнений $\mathbf{A}x=\mathbf{b}$, эквивалентно минимизации квадратичной функции

$$q(x)=(\mathbf{A}x-\mathbf{b})^{\mathrm{T}}(\mathbf{A}x-\mathbf{b}).$$

Позже Флетчер и Ривс обосновали квадратичную сходимость метода и обобщили его для случая не квадратичных функций.

Вывод итерационной формулы метода Флетчера-Ривса проведем в предположении, что целевая функция является квадратичной $q(x)=\frac{1}{2}x^{\mathrm{T}}\mathbf{Q}x+b^{\mathrm{T}}x+a$, а итерации проводятся по основной формуле $x^{(k+1)} = x^{(k)} + t_k s(x^{(k)})$. Это не сужает общности рассуждений, поскольку в случае не квадратичной целевой функции матрицу квадратичной функции $\mathbf{Q}$ необходимо заменить матрицей Гессе $\mathbf{H}$ исходной функции. Направление поиска на каждой итерации определяются с помощью следующих формул:

$$s^{(k)}=-g^{(k)}+\sum_{i=0}^{k-1}\gamma^{(i)}s^{(i)},$$

$$s^{(0)}=-g^{(0)},$$

где $g^{(0)}=\nabla f(x^{(k)})$.

Так как после определения системы направлений проводится последовательный поиск вдоль каждого из направлений, полезно напомнить, что в качестве критерия окончания одномерного поиска обычно используется условие

$$\nabla f(x^{(k+1)})^{T}\, s^{(k)}=0. \qquad (3.11)$$

Значения $\gamma^{(i)}$, $i=1,2,3,\ \dots\ ,k-1$, выбираются таким образом, чтобы направление $s^{(k)}$ было $\mathbf{Q}$-сопряжено со всеми построенными ранее направлениями поиска.

Рассмотрим первое направление

$$s^{(1)}=-g^{(1)}+\gamma^{(0)}s^{(0)}=-g^{(1)}-\gamma^{(0)}g^{(0)}$$

и наложим на него условие сопряженности с $s^{(0)}$

$$s^{(1)\mathrm{T}}\mathbf{Q}s^{(0)}=0,$$

откуда следует, что

$$[\,g^{(1)}+\gamma^{(0)}g^{(0)}\,]^{\mathrm{T}}\mathbf{Q}\ s^{(0)}=0.$$

На начальной итерации

$$s^{(0)}=\Delta x/t^{(0)};$$

следовательно,

$$[\,g^{(1)}+\gamma^{(0)}g^{(0)}\,]^{\mathrm{T}}\mathbf{Q}[\,\Delta x/t^{(0)}\,]=0.$$

Используя выведенное выше свойство (3.10) квадратичных функций, получаем

$$[\, g^{(1)} + \gamma^{(0)} g^{(0)} \,]^{T} \Delta g = 0, \qquad (3.12)$$

откуда

$$\gamma^{(0)} = -(\Delta g^{T} g^{(1)}) / (\Delta g^{T} g^{(0)}) .$$

Из уравнения (3.12) следует, что

$$g^{(1)T} g^{(1)} + \gamma^{(0)} g^{(0)T} g^{(1)} - g^{(1)T} g^{(0)} - \gamma^{(0)} g^{(0)T} g^{(0)} = 0.$$

При соответствующем выборе параметра одномерного поиска $t^{(0)}$ и с учетом формулы (3.11) имеем

$$g^{(1)T} g^{(0)} = 0.$$

Таким образом,

$$\gamma^{(0)} = \| g^{(1)} \|^2 / \| g^{(0)} \|^2 .$$

Далее определим следующее направление поиска

$$s^{(2)} = -g^{(2)} + \gamma^{(0)} s^{(0)} + \gamma^{(1)} s^{(1)}$$

и выберем $\gamma^{(0)}$ и $\gamma^{(1)}$ таким образом, чтобы выполнялись условия

$$s^{(2)T} \mathbf{Q} s^{(0)} = 0 \text{ и } s^{(2)T} \mathbf{Q} s^{(1)} = 0,$$

т.е. условия **Q**-сопряженности направления $s^{(2)}$ с направлениями $s^{(0)}$ и $s^{(1)}$.

С помощью формул (3.10) и (3.11) можно показать, что здесь $\gamma^{(0)} = 0$, а в общем случае $\gamma^{(i)} = 0,\ i = 0, 1, 2, \dots, k-2$, при любом значении k. Отсюда следует, что общая формула для направлений поиска метода Флетчера-Ривса может быть записана в виде:

$$s^{(k)} = -g^{(k)} + [\| g^{(k)} \|^2 / \| g^{(k-1)} \|^2] s^{(k-1)}.$$

Если целевая N-мерная функция квадратичная, то для минимизации потребуется определить ровно N-1 таких направлений и провести ровно N одномерных поисков вдоль этих направлений.

Если же функция общего вида, то количество направлений и одномерных поисков возрастает.

Метод ***Дэвидона-Флетчера-Пауэлла*** (метод ***ДФП***) подобен предыдущему методу, поскольку также основан на свойствах квадратичных функций. Однако метод ДФП вырабатывает ньютоновское направление поиска, причем используя при этом информацию только о первых производных целевой функции. Методы подобного рода достаточно эффективны и составляют целый класс методов, называемых методами *переменной метрики* или *квазиньютоновскими* методами.

В этих методах итерации производятся также по общей формуле (3.1), а направление $s(x^{(k)})$ вычисляется так:

$$s(x^{(k)}) = -\mathbf{A}^{(k)} \nabla f(x^{(k)}) ,$$

где $\mathbf{A}^{(k)}$ - матрица порядка N×N, которая носит название *метрики*[2].

---

[2] Метрика (от греч.-мера, размер)-*расстояние между двумя точками (элементами) a,b множества A*, т.е. действительная числовая функция, удовлетворяющая трем свойствам расстояния. Например, метрика может быть обычным евклидовым расстоянием. Метрики обычно задаются соответствующей нормой.

Согласно принятой классификации метод является *квазиньютоновским*, если в соответствии с ним перемещение пробной точки удовлетворяет следующему условию:

$$\Delta x = \mathbf{H}^{-1} \Delta g \,, \qquad (3.13)$$

где $\mathbf{H}^{-1}$ – обратная матрица в k-й точке.

В квазиньютоновских методах на каждой итерации обратная матрица Гессе целевой функции аппроксимируется с помощью вычисления составляющих вектора-градиента. Обоснование возможности такой аппроксимации заключается в следующем. Запишем для некоторой симметрической матрицы $\mathbf{A}$ следующее рекуррентное соотношение:

$$\mathbf{A}^{(k+1)} = \mathbf{A}^{(k)} + \mathbf{A}_c^{(k)}, \qquad (3.14)$$

где $\mathbf{A}_c^{(k)}$-корректирующая матрица.

Задача заключается в том, чтобы построить матрицу $\mathbf{A}^{(k)}$ таким образом, чтобы последовательность матриц $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$, … , $\mathbf{A}^{(k+1)}$ давала приближение к обратной матрице Гессе целевой функции $\mathbf{H}^{-1} = \nabla^2 f(x^*)^{-1}$ в точке минимума; при этом для получения решения x^* требуется один дополнительный поиск вдоль прямой, если целевая функция - квадратичная.

Как неоднократно подчеркивается в литературе по теории оптимизации, имеются определенные основания полагать, что метод, обеспечивающий нахождение оптимумов квадратичных функций, может привести к успеху при решении задач с нелинейными целевыми функциями общего вида.

Вернемся к важному свойству (3.10) квадратичных форм, которое переписано в виде (3.13), и предположим, что матрица $\mathbf{H}^{-1}$ аппроксимируется по формуле

$$\mathbf{H}^{-1} = \beta \mathbf{A}^{(k)}, \qquad (3.15)$$

где β - скалярная величина.

Наиболее предпочтительным является приближение, удовлетворяющее (3.13), т.е.

$$\Delta x^{(k)} = \mathbf{A}^{(k)} \Delta g^{(k)},$$

где $\Delta x^{(k)} = x^{(k+1)} - x^{(k)}$, $\Delta g^{(k)} = g(x^{(k+1)}) - g(x^{(k)})$; (здесь уместно напомнить, что выше мы приняли обозначение $\nabla f(x) = g(x)$).

Однако ясно, что такую аппроксимацию построить невозможно, поскольку для того, чтобы найти $\Delta g^{(k)}$, необходимо знать матрицу $\mathbf{A}^{(k)}$.

С другой стороны, можно потребовать, чтобы новое приближение удовлетворяло формуле (3.13) :

$$\Delta x^{(k)} = \beta \mathbf{A}^{(k+1)} \Delta g^{(k)} \quad . \qquad (3.16)$$

Подставляя (3.14) в (3.16) получаем уравнение

$$\mathbf{A}_c^{(k)} \Delta g^{(k)} = \tfrac{1}{\beta} \Delta x^{(k)} - \mathbf{A}^{(k)} \Delta g^{(k)},$$

решением которого является следующая матрица

$$\mathbf{A}_C^{(k)} = \tfrac{1}{\beta}\left(\frac{\Delta x^{(k)} y^T}{y^T \Delta g^{(k)}}\right) - \frac{A^{(k)} \Delta g^{(k)} z^T}{z^T \Delta g^{(k)}}, \qquad (3.17)$$

в чем можно убедиться путем непосредственной подстановки.

Здесь y и z – некоторые произвольные векторы, т.е. формула (3.17) определяет некоторое семейство решений.

Если положить $y = \Delta x^{(k)}$ и $z = \mathbf{A}^{(k)} \Delta g^{(k)}$,

то получим формулу искомую, реализующую известный и широко применяемый метод Дэвидона-Флетчера-Пауэлла:

$$\mathbf{A}^{(k)} = A^{(k-1)} + \frac{\Delta x^{(k-1)} \Delta x^{(k-1)T}}{\Delta x^{(k-1)T} \Delta g^{(k-1)}} - \frac{A^{(k-1)} \Delta g^{(k-1)} \Delta g^{(k-1)T} A^{(k-1)}}{\Delta g^{(k-1)T} A^{(k-1)} \Delta g^{(k-1)}}. \qquad (3.18)$$

В теории оптимизации доказывается, что эта рекуррентная формула сохраняет свойства симметрии и положительной определенности матриц. Поэтому, если $\mathbf{A}^{(0)}$ – положительно определенная матрица, то матрицы $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$, … также оказываются симметрическими и положительно определенными при отсутствии ошибок округления; обычно удобно выбирать $\mathbf{A}^{(0)} = \mathbf{I}$. Также можно доказать, что данный метод обеспечивает уменьшение функции на каждой итерации. Метод ДФП в течение ряда лет продолжает оставаться наиболее широко используемым методом оптимизации. Он отличается устойчивостью и успешно применяется при решении самых различных задач, возникающих на практике.

Недостатком методов такого типа является необходимость хранить в памяти компьютера матрицу **A** порядка N×N. Однако более существенным недостатком является то, что в связи с неточностью вычислительных процедур накапливаемая ошибка может привести к тому, что на определенной итерации матрица **A** может оказаться плохо обусловленной и поэтому возникает необходимость возврата к некоторой предыдущей итерации с пересмотренной метрикой. В принципе, при написании надежных программ минимизации данным методом рекомендуется на каждом шаге проверять обусловленность матрицы **A**, хотя такая дополнительная процедура увеличивает трудоемкость вычислений.

На этом заканчивается описание теоретической части работы №3.

7. Практическая часть

Общие замечания

▪ **Напомним,** что основной целью работы является **изучение** практических аспектов градиентных методов минимизации **и их сравнение.**

Также важно провести **сравнение** этих методов с методами **прямого поиска** из предыдущей работы №2 «Алгоритмы прямого поиска в многомерной минимизации».

• Чтобы **корректно сравнивать** методы, в этой работе нужно минимизировать те же три двумерные функции $f_1(x)$, $f_2(x)$ и $f_3(x)$, которые использовались в работе №2:

$$f_1(\mathbf{x}) = 12x_1^2 + 6x_1 x_2 + 2x_2^2 - 2x_1 - x_2 ;$$
$$f_2(\mathbf{x}) = 0.5\ (x_2 - x_1^{\,2})^2 + (1 - x_1)^2 ;$$
$$f_3(\mathbf{x}) = 100\ (x_2 - x_1^{\,2})^2 + (1 - x_1)^2 .$$

Естественно, что для изучения методов нам понадобятся соответствующие **программы**, реализующие поиск минимума соответствующим методом.

Ряд программ **прилагаются** к данной работе как **готовые exe-**файлы, а некоторые – нужно **разработать самостоятельно.**

Самостоятельно нужно запрограммировать 3 следующих метода:

1. градиентный метод с **аналитическим вычислением градиента ∇f** и ручным **заданием параметра длины шага t_k**;
2. метод Коши с **аналитическим градиентом ∇f** и **аналитическим вычислением** параметра длины шага t_k;
3. метод Коши с **аналитическим градиентом ∇f** и с каким-либо **интервальным методом** вычисления параметра шага t_k.

Для изучения методов **Ньютона, Флетчера-Ривса и Дэвидона-Флетчера-Пауэлла** нужно[3] использовать прилагаемые к работе готовые учебные программы:

4. **nt_r, -** метод Ньютона для минимизации функции Розенброка;
5. **nmod_r** – модифицированный метод Ньютона;
6. **fr_r** – метод Флетчера-Ривса;
7. **dfp_r** – метод Дэвидона-Флетчера-Пауэлла.

После выполнения Практической части 7, можно приступить к ответам на контрольные вопросы (часть 8). Обычно, в других лабораторных работах, ответы на контрольные вопросы предшествуют практической части. Мы считаем, что в этой работе контрольные вопросы таковы, что квалифицированные ответы на них проще дать после выполнения части 7.

Теперь переходите к выполнению конкретных пунктов Практической части 7 данной лабораторной работы.

ИЗУЧЕНИЕ МЕТОДА КОШИ

7/1/9. Вначале запрограммируйте **градиентный метод** для минимизации функции $f_1(x)$ с **аналитическим** вычислением градиента ∇f и **ручным** вводом параметра длины шага t_k=const, который остается постоянным для всех итераций данного конкретного поиска. Значения t даны ниже.

Это нестандартная программа, носящая учебно-иллюстративный характер. По существу, она должна отвечать на вопрос:

[3] Не исключается, что обучающийся может разработать свои собственные программы, что априори значительно повышает его рейтинг при сдаче работы преподавателю.

«каковы будут **начальные** шаги траектории (**не менее 5**) минимизации градиентным методом, если параметр длины шага t_k=const **задается не оптимальным**: либо **большим**, либо **малым**, либо остается постоянным на каждом шаге траектории?»

Несмотря на свою «нестандартность», эта программа должна содержать все необходимые процедуры интерактивных программ градиентных методов:

1) реквизиты программы и меню ввода исходных данных,
2) процедуру вычисления направления в виде градиента ∇f,
3) процедуру проверки правила останова по норме ∇f,
4) процедуру оценки параметра шага t_k,(в первой программе эта процедура отсутствует, т.к. t нужно задавать),
5) процедуру вычисления следующей точки поиска,
6) процедура вывода данных поиска минимума.

Теперь становится ясным, что данная программа должна стать основой для самостоятельной разработки следующих программ. Действительно, достаточно ввести соответствующим образом процедуру 4)оценки параметра шага t, можно получить другие требуемые в работе программы. Поэтому обратите **особое внимание** на обсуждаемую программу.

Чтобы построить по 5 шагов требуемых в данном пункте траекторий, полезно выводить данные, которые перечислены в табл. 3.2.(Не надо заносить координаты траектории в таблицу, а сразу переносите их с экрана на график).

Итак, сделав программу градиентного поиска, постройте с ее помощью **по 5 начальных шагов траекторий** минимизации функции f_1 с такими начальными данными:

- из точки $(2,2)^T$ с постоянным параметром шага на каждой итерации: вначале при t=0.02, а затем с t=0.005;
- из точки $(2.1,-1.5)^T$ с постоянным параметром t=0.07 на каждой итерации, а затем с t=0.005;
- на фоне линий уровня f_1 нарисуйте **все 4 траектории** поиска;
- поясните эти траекторий минимизации.

7/2/9. Запрограммируйте метод Коши с аналитическим градиентом и аналитическим параметром шага t_k для минимизации только функции f_1.

- После отладки этой программы метода, минимизируйте функцию f_1 из начальной точки $(2, 2)^T$ с **точностью ε=1e-6**. Результаты работы программы минимизации f_1 сведите в **табл. 3.2**. Чтобы таблица была обозримой, запишите данные **только 4-х** начальных шагов и **конечный результат.**

Таблица 3.2. Результаты минимизации f_1(x)

Итерация k	x_1	x_2	$f_1(x)$	t_k	$\|\nabla f_1(x)\|$
0					
1					
2					
3					
4					
…					
$K_{кон}$					

- **По данным табл. 3.2 на фоне линий равного уровня** функции f_1 постройте **4 начальных** шага минимизации и конечную точку траектории. (Рисунок сделайте как минимум на половине листа формата А4).

Сравните эту траекторию с траекториями предыдущего пункта.

- **На рисунке предыдущего пункта** постройте (примерно) траекторию минимизации **релаксационным** методом. (Используйте ту же начальную точку, что и в табл. 3.2. Поясните различия в траекториях, если таковые имеются.)

- Возьмите какую-либо пару последовательных шагов траектории и **аналитически** проверьте их ортогональность.

7/3/9. Запрограммируйте метод Коши для минимизации функций f_1, f_2, f_3.

- В программах для всех функций используйте **аналитическое вычисление градиента** и любой известный вам **численный метод одномерного поиска** для вычисления параметра шага t_k.

Указание. Перед разработкой программы, изучите блок-схему алгоритма, обратив особое внимание на то, как вычисляется оптимальный параметр шага, т.е. на процедуру **одномерного поиска** в двумерной минимизации. Для этого полезно использовать опыт программирования метода Пауэлла в части организации одномерного поиска (лабораторная работа №2).

7/4/9. Сравнение методов Коши, Ньютона и Флетчера-Ривса

Для выполнения данного пункта работы используйте вашу программу метода Коши предыдущего п.7/3/9 и программы, **прилагаемые** к данной работе:

nt_q - метод Ньютона для квадратичной функции f_1;

nt_r - метод Ньютона для функции Розенброка f_2, f_3;

fr_q, fr_r – метод Флетчера-Ривса для тех же функций $f_1 \div f_3$.

Задаваемые исходные данные вычислительного эксперимента:

- **точность поиска $\varepsilon=10^{-6}$** и **одинакова для всех задач;**
- **только для функции $f_1(x)$** координаты пары начальных точек поиска нужно выбрать из списка и согласовать с преподавателем:

вариант 1: $x^{01}=(2,\ 2)^T$, $x^{02}=(1.4,\ -2.5)^T$

вариант 2: $x^{01}=(0.5,\ -2.8)^T$, $x^{02}=(-2.7,\ 0.5)^T$

вариант 3: $x^{01}=(-2.5,\ -1)^T$, $x^{02}=(2.7,\ 0)^T$

вариант 4: $x^{01}=(-2,\ 2)^T$, $x^{02}=(0.5,\ -2.8)^T$

вариант 5: $x^{01}=(-1,\ 2.75)^T$, $x^{02}=(2,\ 2)^T$

вариант 7: $x^{01}=(2.7,\ 0)^T$, $x^{02}=(-1,\ 2.75)^T$

(*Начальные точки заданы так, что они располагаются приблизительно на одинаковом расстоянии до искомой точки минимума*).

для функций Розенброка $f_2(x)$ и $f_3(x)$ начальные точки заданы в табл. 3.3;

Данные вычислений для удобства сравнения методов необходимо свести в **табл. 3.3:**

Таблица 3.3. Результаты сравнения методов минимизации

Метод минимизации / Целевая функция	**Метод Коши**	**Метод Ньютона**	**Метод Флетчера-Ривса**
Квадратичная форма $f_1(x) = 12x_1^2 + 6x_1x_2 + 2x_2^2 - 2x_1 - x_2$ *Точный минимум* $x^* =$ *Значение в минимуме* $f_1(x^*) =$	1) Нач. точка x^{01} $x^{**} =$ $f(x^{**}) =$ $N_{тр} =$ 2) Нач. точка x^{02} ...	1) Нач. точка x^{01} $x^{**} =$ $f(x^{**}) =$ $N_{тр} =$ 2) Нач. точка x^{02} ...	1) Нач. точка x^{01} $x^{**} =$ $f(x^{**}) =$ $N_{тр} =$ 2) Нач. точка x^{02} ...
Функция Розенброка $f_2(x) = 0.5(x_2 - x_1^2)^2 + (1 - x_1)^2$; *Точный минимум* $x^* =$ *Значение в минимуме* $f_2(x^*) =$	1) Нач. точка $x^0=(-1.2,\ 1)^T$ ... 2) Нач. точка $x^0=(-0.6,\ -0.6)^T$ ...	1) Нач. точка $x^0=(-1.2,\ 1)^T$ ... 2) Нач. точка $x^0=(-0.6,\ -0.6)^T$ ...	1) Нач. точка $x^0=(-1.2,\ 1)^T$ ... 2) Нач. точка $x^0=(-0.6,\ -0.6)^T$ ...
Функция Розенброка $f_3(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$; *Точный минимум* $x^* =$ *Значение в минимуме*	1) Нач. точка $x^0=(-1.2,\ 1)^T$ ... 2) Нач. точка $x^0=(-0.6,\ -0.6)^T$ $x^{**} =$ $f(x^{**}) =$ $N_{тр} =$	1) Нач. точка $x^0=(-1.2,\ 1)^T$ ... 2) Нач. точка $x^0=(-0.6,\ -0.6)^T$ $x^{**} =$ $f(x^{**}) =$ $N_{тр} =$	1) Нач. точка $x^0=(-1.2,\ 1)^T$ ... 2) Нач. точка $x^0=(-0.6,\ -0.6)^T$ $x^{**} =$ $f(x^{**}) =$ $N_{тр} =$

$f_3(x^*)$ =			

Примечание . В табл. 3.3. используются уже известные по работе №2 обозначения, а также и новое: $N_{тр}$ - количество шагов траектории минимизации.

7/5/9. Модифицированный метод Ньютона

- **Модифицированным методом Ньютона** только для функций Розенброка $f_2(x)$ и $f_3(x)$ из точки $x^0=(-1.4,\ 1.4)^T$ с точностью ε=1e-6 найдите точку минимума и постройте траектории поиска.
- Что дает **модификация** метода Ньютона?

Для выполнения данного пункта можно использовать прилагаемую к данной работе программу **NMOD_R**.

ЗАКЛЮЧИТЕЛЬНЫЙ РАЗДЕЛ ПРАКТИЧЕСКОЙ ЧАСТИ

Данный раздел работы является заключительным по изучению и закреплению теории основных методов **безусловной минимизации.**

Поэтому, здесь предлагается решить ряд задач минимизации любым из известных вам методов минимизации.

Точность оценки минимума принять равную $\varepsilon=10^{-6}$.

Для каждой задачи обоснуйте выбранный метод и докажите справедливость полученного решения.

7/6/9. Минимизируйте функцию

$$f_4(x) = [1.5 - x_1 (1- x_2)]^2 +[2.2\ 5 - x_1 (1 - x_2^2)]^2 + [2.625 - x_1 (1-x_2^3)]^2 .$$

7/7/9. Найдите минимум функции

$$f_5(x) = (x_1-2)^2+(x_2-1)^2+\frac{0.04}{-(x_1^2/4)-x_2^2+1}+\frac{1}{0.2}(x_1-2x_2+1)^2 .$$

7/8/9. Найдите минимум функции

$$f_6(x) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2.$$

7/9/9. Решите задачу (Д.Химмельблау, «Прикладное нелинейное программирование». Мир, М. 1975, стр. 206)

Стоимость *C* очищенной нефти, перевозимой морским путем через Малаккский пролив в Японию (**в долларах на килолитр**), зависит от таких основных определяется параметров:

1) стоимости неочищенной нефти;
2) затрат на страхование;
3) таможенных тарифов;
4) затрат на фрахт нефти;
5) затрат на погрузку и разгрузку;

6) платы за морскую стоянку судна;
7) затрат, связанных с подводным перекачиванием и хранением;
8) стоимости площади под цистернами;
9) стоимости очистки и затрат на перевозку продуктов.

Пусть конкретная функция стоимости C задана в таком виде:

$$C = c_c + c_i + c_x + \frac{2.09 \cdot 10^4 \cdot t^{-0.3017}}{360} + \frac{1.064 \cdot 10^6 \cdot at^{0.4925}}{52.47q \cdot 360} +$$
$$+ \frac{4.242 \cdot 10^4 at^{0.7952} + 1.813 \cdot ip(n + 1.2q)}{52.47q \cdot 360} + \frac{4.25 \cdot 10^3 a(nt + 1.2q)}{52.47q \cdot 360} +$$
$$+ \frac{5.042 \cdot 10^3 q^{-0.1899}}{360} + \frac{0.1049 q^{0.671}}{360},$$

где a – фиксированные ежегодные расходы в относительных величинах (0.20);

c_c – цена неочищенной нефти, долл/кл (12.50);

c_i – страховка, долл/кл (0.90);

c_x – таможенные тарифы, долл/кл (0.90);

i - норма процента (0.10);

n - число портов (2);

p - цена земли, долл/м^2 (7000);

q - производительность установки для очистки нефти, баррель/день;

t - объем танкера, кл.

Считая значения, указанные в скобках заданными, вычислите любым известным вам методом

- **минимальную стоимость нефти,**
- **оптимальный объем танкера**, и
- **производительность установки** для очистки нефти (заметим, что 1кл=6.29баррель).

----*----

8. Контрольные вопросы

8/1/5. Проанализируйте на экстремум следующие функции:

$f_1(x,y) = x^2 + y^2$; $f_2(x,y) = x^2 - y^2$;

$f_3(x,y) = x^2 + y^3$; $f_4(x,y) = x^2 - y^3$;

$f_5(x,y) = x^2 + y^4$; $f_6(x,y) = x^2 - y^4$.

Для каждой функции постройте несколько линий уровня.

8/2/5. Рассмотрите функцию $f(x)=-1/(x_1^2+x_2^2+1)$ и по методу Ньютона определите аналитически следующую точку поиска $x^{(1)}$, для заданных двух вариантов начальных точек $x^{(0)}$:

(а) $x^{(0)}=(0.05,\ 0.05)^T$,

(б) $x^{(0)}=(0.0,\ 5)^T$.

8/3/5. Постройте следующие вектора

8/4/5. Докажите аналитически, что любая пара последовательных векторов поиска Коши в табл. 3.1 ортогональна.

--------- *** ---------