**Свойства метода наискорейшего спуска**

1. На любом шаге направление спуска меняется на ортогональное. Действительно, *αk* ищется из условия  ⇒



1. Точка *xk*+1 лежит на луче, исходящем из точки *xk* и касательным к поверхности уровня *Lϕ*(*xk*+1). Действительно, с одной стороны, несомненно, что . С другой стороны, градиент *ϕ*′(*xk*+1) ортогонален касательной к поверхности уровня *Lϕ*(*xk*+1), поэтому по свойству 1 направление спуска касательно к поверхности *Lϕ*(*xk*+1).

*Иначе.* *ϕ*′(*xk*+1) ортогонален направлению спуска ⇒ луч, проходящий из точки *xk* – касательной к поверхности .

*Проблемы* (общие для релаксационных методов).

1. Имеет ли последовательность {*xk*} предел в смысле сходимости по норме: существует ?
2. Является ли этот предел аргументом, составляющим минимум функции *ϕ*?
3. Какова скорость сходимости  или *ϕ*(*xk*) – *ϕ*(*x*\*)?
4. Каковы вычислительные затраты.

**Исследование метода наискорейшего спуска для квадратичной функции**

Рассмотрим квадратичную функцию

,

где *A* – симметричная, положительно определенная матрица.

Можно показать, что *A* – симметричная положительно определенная матрица ⇔ *ϕ* – строго выпукла.

, т.е.  – стационарная точка.

Попробуем записать метод наискорейшего спуска для квадратичной функции.

Итак,



 (w)

,

т.к. *A* – положительно определена, и значит для нее справедливо: (*Ah*, *h*) > 0 ∀*h*∈*Rn*≠ 0.

Для определения скорости сходимости оценим отношение



Имеем:



С другой стороны,



Для простоты дальнейших изложений предположим, что матрица *A* приведена к диагональному виду (т.е. выполнено преобразование координат) так, что , где *λi* – собственные числа матрицы *A*.

* Собственные числа симметричной положительно определенной матрицы всегда положительны.
* Для симметричной матрицы существует ортогональная матрица (*TT* = *T*-1) *T* такая, что *TTAT* – диагональная матрица .

Если , то

,

Тогда

.

Если ввести обозначение , то



Это называется *геометрической скоростью сходимости* (сходимость геометрической прогрессии).

Рассмотрим величину

.

Верхний предел  называется *порядком сходимости метода*.

В нашем случае квадратичной функции

.

Поэтому



⇒ 

⇒ получили сходимость с порядком 1 или *линейную сходимость*. Бывает порядок больше 1 – *сверхлинейная сходимость*.

При исследовании метода наискорейшего спуска для квадратичной функции получили, в частности, следующие результаты:

1. 
2. 

**Определение.**

Пусть *ϕ*(*xk*)→*ϕ*(*x*\*) при *k* → ∞.

Последовательность *ϕ*(*xk*) сходится к *ϕ*(*x\**) *линейно* (с линейной скоростью, со скоростью геометрической прогрессии), если существуют такие константы *q*∈(0,1) и *k*0, что

, при *k*≥ *k*0.

Последовательность *ϕ*(*xk*) сходится к *ϕ*(*x\**) *сверхлинейно*, если

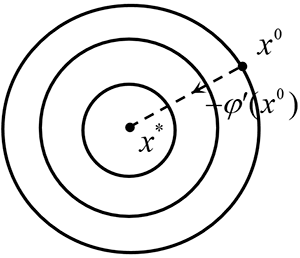
, при *k*→ ∞.

Последовательность *ϕ*(*xk*) сходится к *ϕ*(*x\**) с *квадратичной скоростью*, если существуют такие константы *c*≥ 0 и *k*0, что

, при *k* ≥ *k*0.

Вообще, порядок сходимости, равный 1, означает, что значение величины *k* убывает, в основном, по закону геометрической прогрессии. Порядок сходимости, равный 2 (квадратичная сходимость) означает, что при достаточно больших *k* *k*+1 ~ *k*2. В этом случае, если к тому же *k* – малая величина, например,  при , то *k*+1 равно , т.е. фактически удваивается число нулей после запятой.

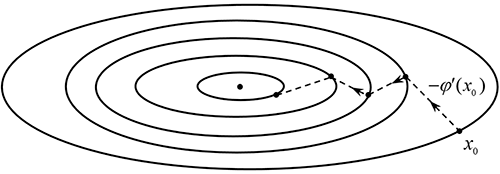
*Частные случаи*:

1. Пусть *l*= *L*, т.е. матрица *A*= ­*LI*= *lI* – пропорциональна единичной окружности (линии уровня – окружности).

Тогда:



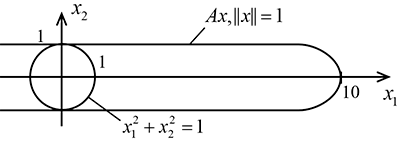
⇒ φ(*xk*+1) = φ(*x*\*) метод сходится за один шаг.

1. *l*≤ *L*: сходимость может быть еле заметной (*q*~ 1), а графически это означает, что линии уровня функции сильно вытянуты и функция имеет так называемый "овражный" характер. Это означает, что небольшое изменение некоторых переменных приводит к резкому изменению значений функции – эта группа переменных характеризует "склон оврага", а по остальным переменным, задающим направление "дна оврага", функция меняется незначительно.

Число  называется числом обусловленности матрицы ⇒ *cond* ≥ 1.

Матрица называется *хорошо обусловленной*, если *cond* ~ 1 и наоборот.

Вообще, число обусловленности геометрически можно трактовать как меру искажения отображения матрицей *A* единичной сферы. Действительно, *cond*(*A*) есть отношение наибольшего к наименьшим расстояниям между точками на единичной сфере после её отображения матрицей *A*. Чем больше *cond*(*A*), тем больше искажение единичной сферы при её преобразовании в эллиптическую форму – пусть *A*= *diag*(10,1).

*Вывод*: Метод наискорейшего спуска быстро сходится для хорошо обусловленных матриц и наоборот.

*Почему так много внимания уделяли квадратичной функции?*

В окрестности locmin любую функцию можно приблизить квадратичной, и всё сказанное выше про матрицу *A* будет справедливым для матрицы Гесса *H*(*x*\*), которая заменяет *A* в рассмотренном выше примере.

*Геометрически*: Линии уровня становятся замкнутыми и по мере приближения к *x*\* всё более напоминают эллипс.

*Общий случай.*

**Определение 1.** Функция ** на множестве *X**Rn* удовлетворяет *условию Липшица*, если существует . Если градиент функции ** существует, непрерывен и удовлетворяет условию Липшица, то обозначается ***C*1,1.

**Определение 2.** Функция ** называется *сильно выпуклой с параметром* , если.

**Теорема** (о сходимости метода наискорейшего спуска). Рассмотрим задачу . Пусть ***С*1,1(*Rn*) и ** – сильно выпуклая c параметром æ. Тогда при любом начальном приближении для последовательности {*xk*}, построенной по методу наискорейшего спуска, справедливы соотношения:

1) 

2) 

**Замечания.**

1. Для *квадратичной функции* :

1. постоянная Липшица *L* есть наибольшее собственное число матрицы *A*:

;

1. она сильно выпукла с параметром . Действительно,





2. Эквивалентные *ограничения сильной выпуклости*:

1. *ϕ* – сильно выпукла ⇔  – выпукла (это означает, что ** имеет "квадратичный запас" выпуклости);
2. пусть ***C*1, *ϕ* – сильно выпукла ⇔ ;
3. пусть ***С*2, ** – сильно выпукла ⇔ , ∀*x*, т.е.  [в смысле положительной определенности разности матриц]. С другой стороны, из условия Липшица , поэтому для сильно выпуклой ***С*2 существует двойная оценка матрицы Гессе: .

Покажем, что . С одной стороны, из б) имеем



С другой стороны,



⇒ , *ч.т.д.*

Выпуклость:

* .

Строгая выпуклость:

* .

Сильная выпуклость:

*  для ∀*u*,*υ*∈*Rn*

*Графическое представление дифференциальных критериев выпуклости.*

|  |  |
| --- | --- |
| 21 | График *выпуклой* функции расположен не ниже касательной плоскости , проходящей через произв. точку поверхности . |
| C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\Content.Word\20.png | График *строго выпуклой* функции имеет единственную общую точку с этой плоскостью. |
| C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\Content.Word\22.png | Пусть *υ*\* – (⋅)min ⇒ . Поверхность  – это параболоид вращения с вершиной в точке .  ⇒ График *сильно выпуклой* функции расположен внутри некоторого параболоида вращения. |