**Другие градиентные методы**

Напомним, градиентные методы заключаются в построении релаксационной последовательности:



Градиентные методы различаются между собой способом выбора *αk*.

**1.** **Метод наискорейшего спуска**, который был рассмотрен выше, заключается в выборе

.

Такой способ выбора *αk* является в некотором смысле наилучшим, т.к. он обеспечивает *достижение наименьшего значения функции* вдоль заданного направления. Однако он требует решения на любом шаге *одномерной задачи минимизации*. Эти задачи решаются, как правило, приближенно с помощью численных методов, что приводит к *значительному объему вычислений*. Кроме того, метод может привести к *плохой сходимости* (овраги!).

Другим подходом для построения релаксационной последовательности является попытка определить *αk* до начала вычислений. Какие есть для этого основания?

Допустим, что можно построить оценку для *αk* такую, что для *ε*∈(0,1) выполняется неравенство



Тогда очевидно, что  и соответствующий метод минимизации будет методом спуска.

Справедливы следующие утверждения:

##### **Лемма 1.** Пусть функция C1,1(Rn) и



Тогда для *xk**Rn*, **(0,1) условие (1) выполнено при



**Лемма 2.** Пусть ** дважды дифференцируема и матрица Гессе удовлетворяет условию Липшица и



Тогда для *xk**Rn*, **(0,1) условие (1) выполняется при



**2. Градиентный метод с постоянным шагом.**

В этом методе полагается *αk*≡ *const*. При этом иногда удается добиться выполнения условия (1). Но для этого необходимо знать константы *M* и *D*, что далеко не всегда удается вычислить.

Т.о., метод прост в реализации, но есть проблемы со сходимостью.

**Пример.** Пусть **(*x*) = *x*2

**min = 0; *x\** = 0;

Тогда

 ⇔ метод сходится.

Сходимость *медленная*!

**3. Градиентный метод с убыванием длины шага.**

В ряде методов достаточно потребовать выполнения условий:

 (например, )

На интуитивном уровне объяснение следующее:

* условие сходимости ряда  накладывают, чтобы добиться достаточно быстрой сходимости последовательности *k* к нулю с целью обеспечения сходимости метода в окрестности точки экстремума *x*\*.
* условие расходимости ряда  призвано обеспечить достижение точки экстремума *x*\* даже при неудачном выборе начального приближения *x*0, т.е. при больших расстояниях от *x*0 до *x*\*.

Сходимость *медленная*!

**4. Градиентный метод с дроблением шага.**

Ещё один адаптивный способ выбора коэффициентов *αk*. Выбираются некоторые *constβ *>0 и 0< *λ*< 1 (обычно *λ*= ½). Для коэффициента *α* = ** проверяется выполнение условия . Если оно выполняется, то полагают *k*= **. Если нет, то производится дробление шага, т.е. принимается *α* = **, и т.д. до тех пор, пока не выполнится требуемое неравенство.

Процесс дробления не может продолжаться бесконечно, поскольку −*ϕ*′(*x*) – направление убывания функции. Первое *α*, при котором условие выполнено и принимается за *αk*.

Как показывает следующая лемма, с помощью описанного процесса дробления шага можно добиться выполнения неравенства. (1)

**Лемма 3.** Пусть функция *ϕ* дифференцируема на *Rn*. Тогда для  найдется такое *α*0 > 0, что при ∀*α*∈(0, *α*0] выполнено условие

.

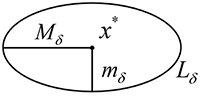
Если необходимое неравенство оказывается выполненным при начальном значении *α*= *β*, то иногда полезно увеличить шаг, взяв *α* = *μβ*, где *μ*> 1. Так можно продолжать до тех пор, пока значения функции не перестанут уменьшаться. Последнее *α*, при котором произошло уменьшение, и берется в этом случае за *αk*.

**5. Овражный метод** – эвристический двухшаговый метод минимизации овражных функций.

*Характеристика степени овражности*:

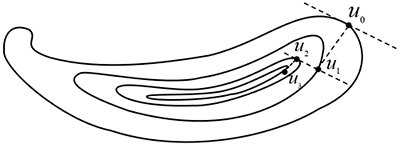
Пусть *x*\* – точка минимума, *δ*> 0.

Рассмотрим поверхность уровня .

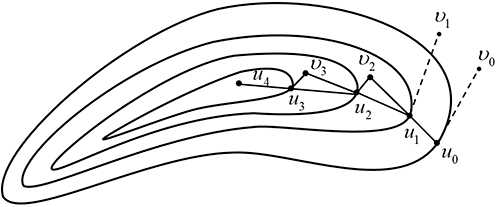
Введем 



**Определение.** Тогда  называется *числом обусловленности точки locmin* .

****Рассмотрим "овражную" функцию (вытянута вдоль некоторых направлений). Если точка лежит на склоне оврага, то направление спуска из этой точки будет почти перпендикулярно к направлению "дна оврага", и в результате приближения {*uk*}, получаемые градиентным методом, будут *поочередно находиться то на одном, то на другом "склоне оврага"*. Если "склоны оврага" достаточно круты, то такие скачки "со склона на склон" точек {*uk*} могут сильно замедлить сходимость градиентного метода.

Для ускорения сходимости можно предложить следующий эвристический прием, называемый *овражным методом*:

Пусть *υ*0, *υ*1 – две произвольные близкие точки. Совершаем из них по одному шагу методом наискорейшего спуска (или ∀ вариант градиентного метода).

Попадаем в окрестность "дна оврага". Соединяя их прямой, делаем большой шаг в полученном направлении, перемещаясь вдоль "дна оврага". Из полученной точки *υ*2, которая находится на "склоне оврага", производят спуск с помощью градиентного метода и определяют следующую точку *u*2 на "дне оврага" и т.д.

Формула метода выглядит следующим образом.



Здесь:

 определяет знак - чтобы спускаться, а не подниматься;

 - определяет направление спуска по дну оврага;

*h -* овражный шаг, выбирается эмпирически и от него многое зависит.

Если *h* – большое, то на крутых склонах точки *υk* могут *слишком далеко удаляться от "дна оврага"* ⇒ большие объемы вычислений для градиентного метода спуска в очередную точку на "дне оврага", кроме этого может произойти выброс точки *υk* из "оврага" и *правильное направление поиска будет потеряно*.

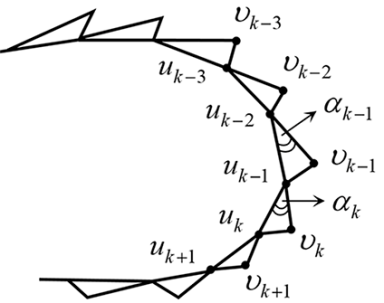
Если *h* – малое, то эффект от применения овражного метода может быть незначительным.

*Эффективность* применения овражного метода может *резко возрасти*, если величину *h* выбирать переменной, реагирующей на "повороты" оврага, с тем, чтобы:

1. быстрее проходить прямолинейные участки на "дне оврага" за счет увеличения овражного шага;
2. на крутых поворотах "оврага" избежать выброса из "оврага" за счет уменьшения овражного шага;
3. добиться min отклонения точки *υk* от дна оврага с целью уменьшения объема вычислений для градиентного метода.

Для правильной реакции на "повороты" оврага надо учитывать "кривизну" оврага.

Причем информацию о кривизне желательно получить по результатам предыдущих шагов.

Один из способов выбора шага:

,

где *k* – угол между векторами , определяемый

условием

,

*c* − *const* > 1 – параметр алгоритма.

 возрастает при возрастании кривизны ⇒ при переходе от участка с меньшей кривизной на участок с большей кривизной имеем

 и наоборот.

На участках с постоянной кривизной



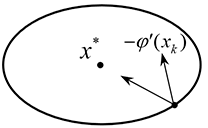
⇒ шаг остается постоянным, который был сформирован при выходе на рассматриваемый участок.

Параметр *с* регулирует чувствительность "метода к изменению кривизны (повороты)" и во многом определяет скорость движения "по оврагу".

# Методы II порядка минимизации функции

(использование вторых производных)

*Общая идея*:



направление спуска по градиентному методу

желаемое направление

(надо "довернуть" направление спуска)

Последовательность {*xk*} будем строить по формулам:

,

где *γk* – длина шага, *Hk* – матрица поворота (*n*×*m*).

Как выбрать матрицу *Hk*?

Если взять квадратичную функцию

,

то хочется сразу попасть в экстремальную точку:



⇒ в качестве "матрицы доворота" надо брать .

*В общем случае*: пусть *ϕ* – дважды дифференцируема в *Rn*, разложим *ϕ*(*x*) в ряд Тейлора в точке

Иначе формулу можно представить в виде:

, где  – квадратичная функция.

Пренебрегаем  и ищем .

;

Пусть *ϕ*″(*xk*) – положительно определена для ∀*xk*∈*Rn* ⇒ существует [*ϕ*″(*xk*)-1].

Решая это уравнение, получим:

 – это и есть метод Ньютона.

Квадратичная функция с положительно определенной *ϕ*″ сильно выпукла, тогда уравнение определяет единственную точку глобального минимума .

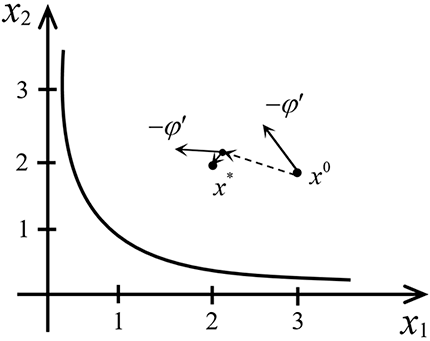
Далее рассмотрим пример использования метода Ньютона для решения задачи минимизации функции.

*Пример.*

Область существования (*ϕ*″(*x*))-1 совпадает с областью положительной определенности матрицы *ϕ*″(*x*), которую мы будем искать с помощью критерия Сильвестра.

*Критерий Сильвестра.*

Симметрическая матрица *A* положительно определена ⇔ если все её главные миноры положительны. При этом главным минором матрицы *A* называется определитель матрицы, построенной из элементов матрицы *A*, стоящих на пересечении строк и столбцов с одинаковыми номерами.

.

Возьмем 









 , и.т.д.

Можно показать, что сходимость при  будет хуже.

*Достоинства метода Ньютона:*

1. Для квадратичной функции сходится за один шаг (метод Ньютона можно рассмотреть, как градиентный метод с преобразованием координат [умножение на *H*-1] таким, что исчезает "овраг", т.е. линии уровня становятся окружностями).
2. Высокая скорость сходимости. Можно показать, что

.

Порядок сходимости:



⇒ важен выбор *q* в алгоритме. Если *q*=10-1, то за один шаг точность результата увеличивается на 2 разряда, а при линейной сходимости – на один разряд.

*Недостатки метода Ньютона:*

1. Локальная сходимость (матрица Гессе должна быть невырожденной). Начальное приближение надо выбирать в окрестности точки локального минимума.
2. Большие вычислительные затраты:

* вычисление матрицы *ϕ*″;
* необходимость обращать её.

*Общие рекомендации*: сначала применять градиентный метод, затем – метод Ньютона.

Например, существует так называемый метод Марквардта-Левенберга:

.

*При больших* *αk* (вдали от *x*\*)матрица и это фактически градиентный метод.

*При малых* *αk* (вблизи от *x*\*) это метод Ньютона.

Имеет место:

**Теорема** (о сходимости метода Ньютона).

Пусть

1) *ϕ* – сильно выпукла на *Rn* с параметром æ.

2) *ϕ*∈*c*2,2, т.е. *ϕ*″ – дважды дифференцируема и *ϕ*″ удовлетворяет условию Липшица:

;

3) Начальное приближение *x*0 удовлетворяет условию

,

т.е. , где 0 < *q* < 1.

Тогда последовательность  сходится к точке минимума *x*\* с квадратичной скоростью:

 (квадратичная сходимость)

1. Несколько слов о норме матрицы:

**Определение.** Норму (*n*×*n*)-матрицы определим следующим образом:

,

Тогда

.

Поскольку *ATA* есть симметричная (*n*×*n*)-матрица, то существует ортогональная матрица  такая, что  – диагональная матрица  ⇒

, где  – наибольшее собственное значение матрицы *ATA*.

 – наименьшее собственное значение матрицы *ATA*.

Для такой нормы выполняются все три условия

1. , если *A* – ненулевая (покомпонентно);
2. ;
3. , где *α* – скаляр.

Кроме того, из определения нормы матрицы следует, что

.

Имеем также

.

2. Отметим ещё раз, что для сходимости метода Ньютона начальная точка *x*0 должна выбираться достаточной близкой к искомой точке *x\**. Это требование в теореме выражено условием 3). Действительно, сильная выпуклость *ϕ* означает:

;

⇒ чем меньше *q*, тем ближе надо выбирать точку *x*0 к *x\** и тем быстрее сходимость.