Градиентные методы

Решить задачу безусловной минимизации дифференцируемой в E^n функции $f(X)$

$$\min\left\{f(X) \middle| X \in E^n\right\}, \qquad (1)$$

т.е. найти минимальное значение f^* функции $f(X)$ и точку $X^* \in E^n$, в которой это значение достигается.

Задача (1) для достаточно гладкой функции $f(X)$ может быть решена классическим методом. Для этого, используя необходимое условие экстремума

$$f'(X) = 0 \text{ или } \frac{\partial f}{\partial x_i} = 0, i = 1,..., n, \qquad (2)$$

найдем все стационарные точки функции $f(X)$. Далее, проверяя достаточное условие минимума (положительную определенность матрицы ее вторых производных) в каждой из найденных стационарных точек, определим точки локального минимума и, сравнивая значения функции $f(X)$ в них, установим точки глобального минимума.

Однако во многих случаях решение системы (2) весьма затруднительно, даже если функция задана аналитически. Поэтому на практике, как правило, применяют приближенные численные методы.

Для численного решения задачи (1) разработано много алгоритмов, использующих итерационные процедуры типа

$$X^{k+1} = F(X^k, X^{k-1},..., X^0), \forall X^0 \in E^n, \qquad (3)$$

позволяющие при определенных условиях построить минимизирующую последовательность для функции $f(X)$, т.е. последовательность точек $x^k \in E^n$ такую, что

$$\lim_{k\to\infty} f(X^k) = f^*.$$

Вычислительные алгоритмы простейших процедур (3) основаны на рекуррентных формулах вида

$$X^{k+1} = X^k + \alpha_k S^k, k = 0,1,2,..., \quad (4)$$

где S^k (направление поиска точки X^{k+1} из точки X^k) - направление убывания функции $f(X)$ в точке X^k, а число α_k - величина шага в направлении убывания S^k.

От того, как выбирается направление S^k и как определяется величина шага α_k зависят свойства итерационных процедур (4) такие, как поведение функции $f(X)$ на элементах последовательности $\{X^k\}$, сходимость к решению, скорость сходимости, объем требуемых вычислений.

Способ выбора S^k и α_k определяет те или иные численные методы безусловной минимизации.

В данной лабораторной работе изучаются две процедуры градиентного метода, в основе которого известное из математического анализа свойство градиента функции $f(X)$, т.е. вектора $f'(X) = \left(\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n} \right)^T$ указывать направление наискорейшего изменения функции $f(X)$ в точке X.

Метод градиентного спуска

Градиентный спуск — метод нахождения локального минимума или максимума функции с помощью движения вдоль градиента.

Градиент — вектор, своим направлением указывающий направление наибольшего возрастания некоторой величины, значение которой меняется от одной точки пространства к другой (скалярного поля), а по величине (модулю) равный скорости роста этой величины в этом направлении.

Наиболее простой в реализации из всех методов локальной оптимизации. Имеет довольно слабые условия сходимости, но при этом скорость сходимости достаточно мала (линейна). Шаг градиентного метода часто используется как часть других методов оптимизации.

Градиент (антиградиент) дает только направление спуска, но не величину шага. В общем случае один шаг не дает точку экстремума, поэтому процедура спуска должна применяться несколько раз. В точке минимума все компоненты градиента равны нулю. На рисунке 1 представлена графическая интерпретация метода градиентного спуска.

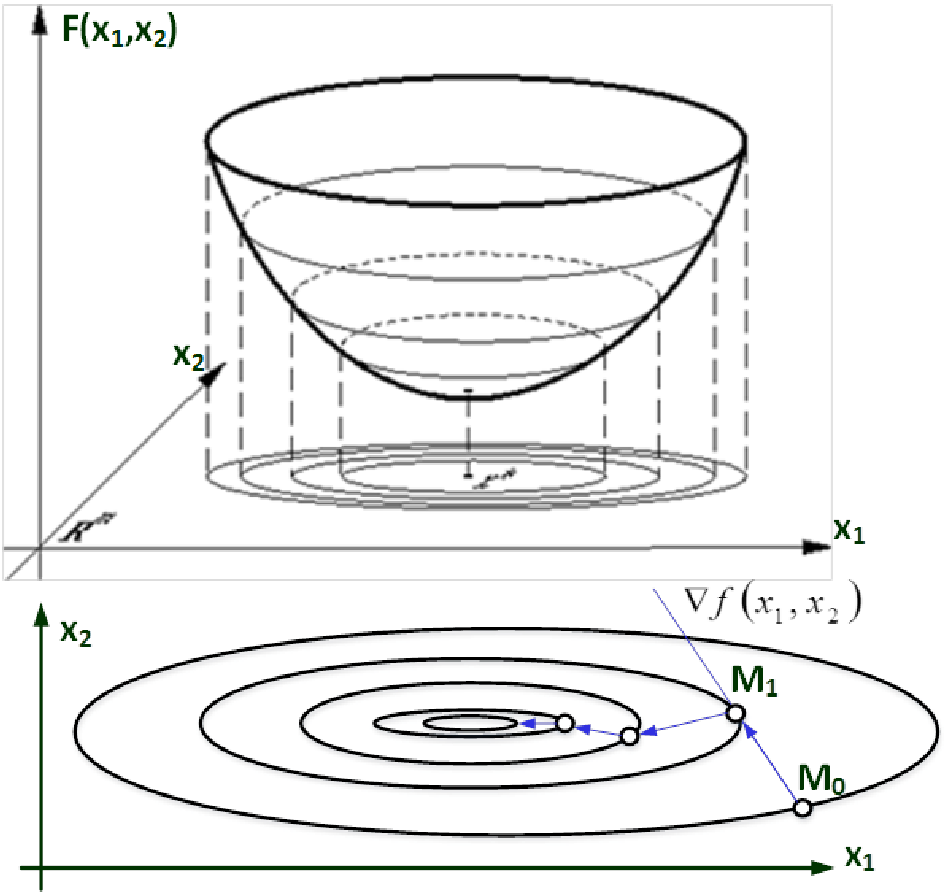

Рис. 1. Геометрическая интерпретация градиентного метода

В соответствии с основной идеей градиентного метода будем строить элементы минимизирующей последовательности $\{X^k\}$ с помощью рекуррентной формулы (4), где в качестве направления S^k выбирается направление антиградиента как направление наискорейшего убывания функции $f(X)$ в малой окрестности точки X^k, т.е. вычислительная схема метода градиентного спуска такова:

$$X^{k+1} = x^k - \alpha_k \cdot f'(X^k), k = 0,1,2,... \quad (5)$$

Величина шага выбирается так, чтобы выполнялось условие

$$f(X^{k+1}) < f(X^k), k = 0,1,2,... \quad (6)$$

В качестве критерия окончания счета на практике используют условия:

$$\left| f(X^{k+1}) - f(X^k) \right| \le \varepsilon_1; \quad (7)$$

$$\left\| X^{k+1} - X^{k} \right\| \le \varepsilon_2 ; \tag{8}$$

$$\left\| f'(X^{k}) \right\| \le \varepsilon_3 , \tag{9}$$

где ε_i - заданные параметры точности.

Опишем алгоритм одного из вариантов рассмотренного метода градиентного спуска.

<u>Шаг 0.</u> Задать параметр точности $\varepsilon > 0$, начальный шаг $\alpha > 0$,параметр алгоритма $\lambda \in (0;1)$, выбрать $X^{0} \in E^{n}$,вычислить значение $f(X^{0})$, положить $k = 0$.

<u>Шаг 1.</u> Найти градиент $f'(X^{k})$ и проверить условие достижения заданной точности $\left\| f'(X^{k}) \right\| \le \varepsilon$. Если оно выполняется, то перейти к шагу 4, иначе - к шагу 2.

<u>Шаг 2.</u> Найти новую точку в направлении антиградиента $X = X^{k} - \alpha \cdot f'(X^{k})$и вычислить $f(X)$. Если $f(X) < f(X^{k})$, то положить $X^{k+1} = X^{k}$, $f(X^{k+1}) = f(X^{k})$, $k = k + 1$, и перейти к шагу 1, иначе - перейти к шагу 3.

<u>Шаг 3.</u> Положить $\alpha = \alpha \cdot \lambda$, где $0 < \lambda < 1$, и перейти к шагу 2.

<u>Шаг 4.</u> Завершить вычисления, положив $X^{\bullet} = X^{k}$, $f^{*} = f(X^{k})$.

Метод наискорейшего спуска

Метод наискорейшего спуска - это итерационный численный метод (первого порядка) решения оптимизационных задач, который позволяет определить экстремум (минимум или максимум) целевой функции.

В соответствии с рассматриваемым методом экстремум (максимум или минимум) целевой функции определяют в направлении наиболее быстрого возрастания (убывания) функции, т.е. в направлении градиента (антиградиента) функции. Градиентом функции в точке $X = \{x_1^*, x_2^*, \dots, x_n^*\}$ называется вектор, проекциями которого на координатные оси являются частные производные функции по координатам.

Градиент в базовой точке $X = \{x_1^*, x_2^*, \dots, x_n^*\}$ строго ортогонален к поверхности, а его направление показывает направление наискорейшего возрастания функции, а противоположное направление (антиградиент), соответственно, показывает направление наискорейшего убывания функции.

Метод наискорейшего спуска является дальнейшим развитием метода градиентного спуска. В общем случае процесс нахождения экстремума функции является итерационной процедурой.

Величина шага выбирается из условия минимума целевой функции $f(x)$ в направлении движения, т. е. в результате решения задачи одномерной оптимизации в направлении градиента или антиградиента.

Таким образом, шаг расчета выбирается такой величины, что движение выполняется до тех пор, пока происходит улучшение функции, достигая, таким образом, экстремума в некоторой точке. В этой точке вновь определяют направление поиска (с помощью градиента) и ищут новую точку оптимума целевой функции и т.д. Таким образом, в данном методе поиск происходит более крупными шагами, и градиент функции вычисляется в меньшем числе точек.

В случае функции двух переменных данный метод имеет следующую геометрическую интерпретацию: направление движения из точки x_k касается линии уровня в точке x_{k+1}. Траектория спуска зигзагообразная, причем соседние звенья зигзага ортогональны друг другу. Это можно увидеть на рисунке 2.

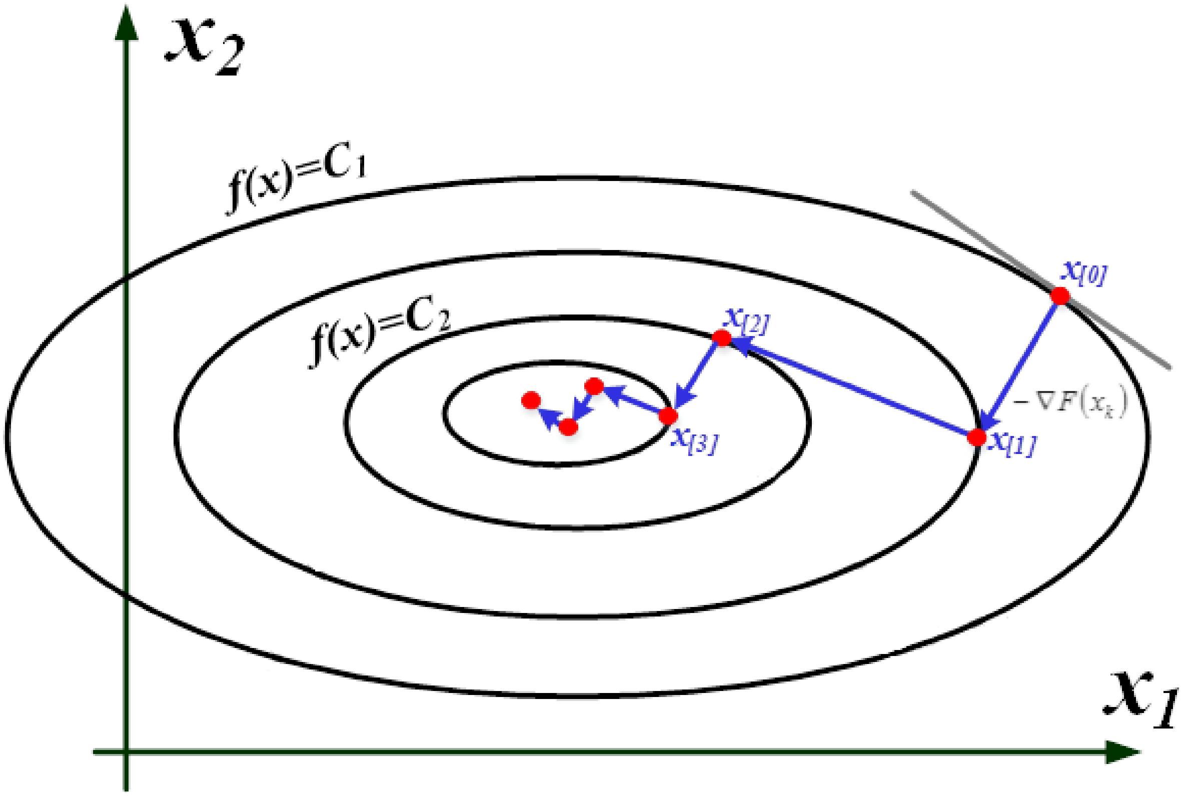

Рис. 2. Траектория движения к точке экстремума

В этом варианте градиентного метода минимизирующая последовательность $\{X^k\}$ также строится по правилу (5). Однако величина шага α_k находится в результате решения вспомогательной задачи одномерной минимизации

$$\min\{\varphi_k(\alpha) \mid \alpha > 0\}, \qquad (10)$$

где $\varphi_k(\alpha) = f(X^k - \alpha \cdot f'(X^k))$. Таким образом, на каждой итерации в направлении антиградиента $S^k = -f'(X^k)$ выполняется исчерпывающий спуск. Для решения задачи (10) можно воспользоваться одним из методов одномерного поиска, изученных в предыдущих лабораторных работах, например, методом поразрядного поиска или методом золотого сечения. Опишем алгоритм метода наискорейшего спуска.

<u>Шаг 0.</u> Задать параметр точности $\varepsilon > 0$, выбрать $X^0 \in E^n$, положить $k = 0$.

<u>Шаг 1.</u> Найти $f'(X^k)$ и проверить условие достижения заданной

точности $\|f'(x^k)\| \le \varepsilon$. Если оно выполняется, то перейти к шагу 4, иначе - к шагу 2.

Шаг 2. Решить задачу (10), т.е. найти α_k. Найти очередную точку $X^{k+1} = X^k - \alpha_k f'(X^k)$, положить $k = k+1$и перейти к шагу 1.

Шаг 3. Завершить вычисления, положив $X^* = X^k$, $f^* = f(X^k)$.

Типовой пример

Минимизировать функцию

$$f(x) = x_1^2 + 4x_2^2 - 6x_1 - 8x_2 + 13. \qquad (11)$$

Вначале решим задачу классическим методом. Запишем систему уравнений (2):

$$\left.\begin{aligned} \frac{\partial f}{\partial x_1} &= 2x_1 - 6 = 0 \\ \frac{\partial f}{\partial x_2} &= 8x_2 - 8 = 0 \end{aligned}\right\}$$

Решив ее, получим стационарную точку $X^* = (3;1)$.

Найдем матрицу вторых производных:

$$f''(X) = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}.$$

Так как согласно критерию Сильвестра эта матрица положительно определена при $\forall X \in E^2$, то найденная точка X^* является точкой минимума функции $f(X)$. Минимальное значение $f^* = f(X^*) = 0$. Таково точное решение задачи (11).

Выполним одну итерацию метода градиентного спуска для (11). Выберем начальную точку $X^0 = (1;0)$, зададим начальный шаг $\alpha = 1$ и параметр $\lambda = 0{,}5$. Вычислим $f(X^0) = 8$.

Найдем градиент функции $f(X)$ в точке X^0:

$$f'(X^0) = \begin{pmatrix} 2x_1^0 & -6 \\ 8x_2^0 & -8 \end{pmatrix} = \begin{pmatrix} -4 \\ -8 \end{pmatrix}. \qquad (12)$$

Определим новую точку $X = X^0 - \alpha \cdot f'(X^0)$, вычислив ее координаты:

$$x_1 = x_1^0 - \alpha \frac{\partial f(X^0)}{\partial x_1} = 1 - \alpha(-4) = 1 + 4\alpha = 5$$

$$x_2 = x_2^0 - \alpha \frac{\partial f(X^0)}{\partial x_2} 0 - \alpha(-8) = 8\alpha = 8. \qquad (13)$$

Вычислим $f(X) = f(X^0 - \alpha \cdot f'(X^0)) = 200$. Так как $f(X) > f(X^0)$, то выполняем дробление шага, полагая $\alpha = \alpha \cdot \lambda = 1 \cdot 0{,}5 = 0{,}5$. Снова вычисляем по формулам (13) $x_1 = 1 + 4\alpha = 3$, $x_2 = 8\alpha = 4$ и находим значение $f(X) = 39$. Так как опять $f(X) > f(X^0)$, то еще уменьшаем величину шага, полагая $\alpha = \alpha \cdot \lambda = 0{,}5 \cdot 0{,}5 = 0{,}25$. Вычисляем новую точку с координатами

$x_1 = 1 + 4 \cdot 0{,}25 = 2$; $x_2 = 8 \cdot 0{,}25 = 2$ и значение функции в этой точке $f(X) = 5$. Поскольку условие убывания $f(X) < f(X^0)$ выполнено, то считаем, что найдена очередная точка минимизирующей последовательности $X^1 = (2;2)$. Первая итерация завершена.

Выполним одну итерацию по методу наискорейшего спуска для (11) с той же начальной точкой $X^0 = (1;0)$. Используя уже найденный градиент (12), находим

$$X = X^0 - \alpha \cdot f'(X^0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \alpha \begin{pmatrix} -4 \\ -8 \end{pmatrix} = \begin{pmatrix} 1 + 4\alpha \\ 8\alpha \end{pmatrix}$$

и строим функцию $\varphi_0(\alpha) = f(X^0 - \alpha f'(X^0)) = (4\alpha - 2)^2 + 4(8\alpha - 1)^2$.

Минимизируя ее с помощью необходимого условия

$$\varphi_0'(\alpha) = 8(4a - 2) + 64(8a - 1) = 0,$$

находим оптимальное значение величины шага $a_0 = \frac{5}{34}$.

Определяем точку минимизирующей последовательности

$$X^1 = X^0 - a_0 \cdot f'(X^0) = \begin{pmatrix} 1 + 4 \cdot \frac{5}{34} \\ 8 \cdot \frac{5}{34} \end{pmatrix} = \begin{pmatrix} \frac{27}{17} \\ \frac{20}{17} \end{pmatrix}$$

Первая итерация завершена. Подсчитав компоненты вектора $f'(X^1)$

$\frac{\partial f(X^1)}{\partial x_1} = 2 \cdot \frac{27}{17} - 6 = -\frac{48}{17}$; $\frac{\partial f(X^1)}{\partial x_2} = 8 \cdot \frac{27}{17} - 8 = -\frac{24}{17}$, можно убедиться, что скалярное произведение

$$< f'(X^0)\ ,\ f'(X^1) > (-4)(-\frac{48}{17}) + (-8)\frac{24}{17} = 0\ ,$$

что характерно для наискорейшего спуска.