Тема «Численные методы оптимизации»

Пусть дана функция $f(x)$, где $x = (x_1, x_2, \ldots, x_n)$ принадлежащая n-мерному пространству R_n, ограниченная снизу на множестве и имеющая непрерывные частные производные во всех его точках.

Требуется найти локальный минимум функции $f(x)$ на множестве допустимых решений R_n, т.е. найти такую точку x^*, что

$$f(x^*) = \min_{x \in R^n} f(x)$$

Метод градиентного спуска с постоянным шагом

Стратегия решения задачи состоит в построении последовательности точек $\{x^k\}$, таких что,

$$f(x^{k+1}) < f(x^k) \quad (k = 0, 1, \dots)$$

Элементы последовательности $\{x^k\}$ вычисляются по правилу

$$x^{k+1} = x^k - t_k \operatorname{grad} f(x^k) \qquad (1)$$

где x^0 - точка начального приближения;

$\operatorname{grad} f(x)$ – градиент скалярной функции $f(x)$;

t_k – величина шага (задается), остается постоянной до тех пор, пока функция убывает в точках последовательности, что контролируется путем проверки выполнения условия

$$f(x^{k+1}) - f(x^k) < 0$$

Построение оптимизирующей последовательности $\{x^k\}$ заканчивается на таком шаге K, при котором выполняется условие

$$\| \operatorname{grad} f(x^K) \| < \varepsilon \qquad (2)$$

где ε - заданное малое число или допустимая погрешность, с которой можно считать, что достигнута стационарная точка.

Напомним, что для функции многих переменных $f(x_1, x_2, \ldots, x_n)$ или в сокращенной записи $f(x)$, необходимым условием экстремума является равенство нулю всех частных производных (условие стационарной точки)

$$\frac{\partial f}{\partial x_i} = 0 \; (i = 1,2,\ldots, n) \qquad (3)$$

или $\operatorname{grad} f(x) = \left(\frac{\partial f}{\partial x_1}, \quad \frac{\partial f}{\partial x_2}, \quad \ldots, \frac{\partial f}{\partial x_n}\right) = 0$

Поэтому условие (2) является численным приближением условий (3)

Заметим, что число итераций не должно превышать некоторого заданного их числа, то есть $K < M$. При превышении данного числа M, алгоритм останавливается и рекомендуется выбрать другую начальную точку.

Для остановки процесса вычислений можно также воспользоваться критерием, основанным на критерии Коши существования предела, то есть если при некотором малом ε_1 выполняются оценки

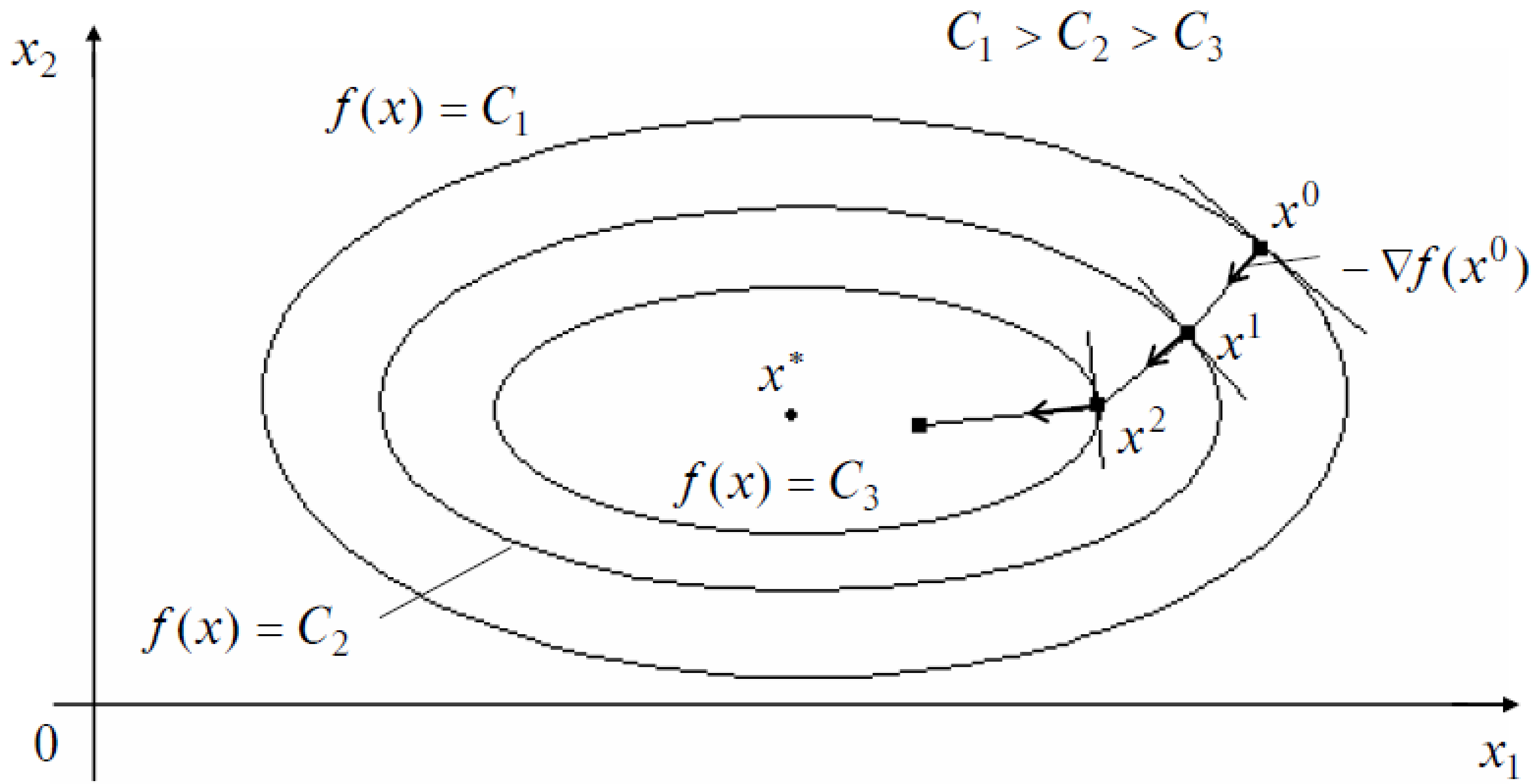
$$\| x^K - x^{K+1} \| < \varepsilon_1 \text{ и } | f(x^K) - f(x^{K+1}) | < \varepsilon_1$$

То последовательность $\{x^k\}$ сходится и ее значение на K шаге отличается от предельного значения x^* не более чем на ε_1

$$\| x^K - x^* \| < \varepsilon_1$$

Заметим, что не всякая стационарная точка является точкой экстремума (min в нашем случае). Поэтому для определения наличия экстремума в точке x^* необходимы дополнительные действия.

На рисунке дана графическая иллюстрация сходимости метода градиентного спуска при $n = 2$



Метод наискорейшего градиентного спуска

Данный метод является модификацией предыдущего подхода, то есть стратегия поиска минимума состоит в построении последовательности точек $\{x^k\}$, удовлетворяющих условию

$$f(x^{k+1}) < f(x^k) \quad (k = 0, 1, \dots)$$

Элементы последовательности $\{x^k\}$ вычисляются по правилу (1)

$$x^{k+1} = x^k - t_k \operatorname{grad} f(x^k)$$

где x^0 - точка начального приближения; grad $f(x)$ – градиент скалярной функции $f(x)$;

Основное отличие от градиентного спуска с постоянным шагом заключается в том, что величина шага t_k уже не является постоянной величиной.

Величина шага t_k вычисляется для каждой k – итерации из условия

$$\varphi(t_k) = f\left(x^k - t_k gradf(x^k)\right) \to min \tag{4}$$

Эта задача решается обычными методами математического анализа, то есть, чтобы функция одного аргумента $\varphi(t)$ имела минимум в некоторой точке t_k необходимо, чтобы ее производная в этой точке равнялась нулю:

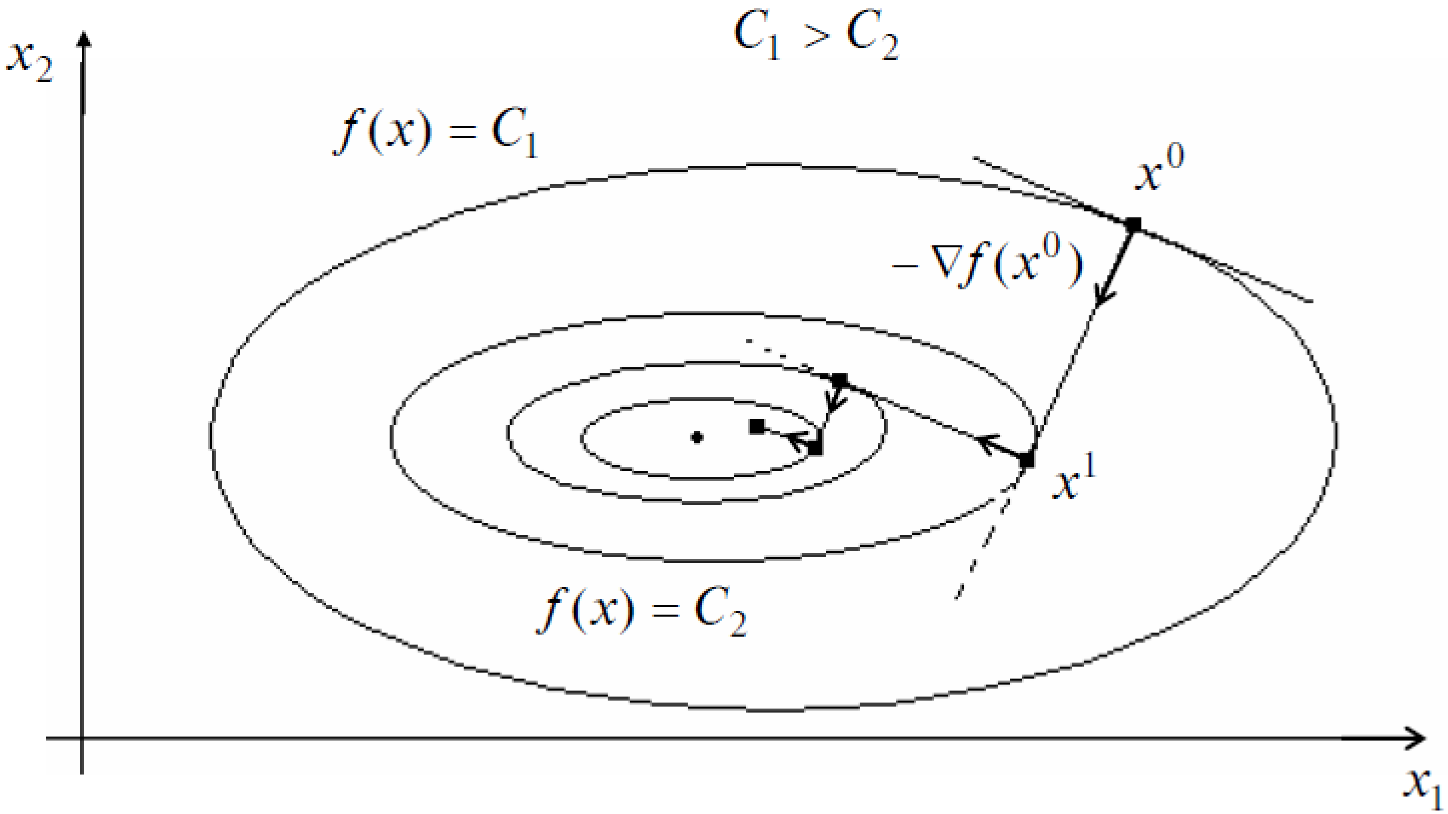
$$\left.\frac{d\varphi}{dt}\right|_{t=t_k} = 0$$

Для того, чтобы в этой точке действительно существовал min требуется

$$\left.\frac{d^2\varphi}{d^2t}\right|_{t=t_k} > 0$$

Заметим, что в некоторых случаях решить одномерную задачу минимизации (4) целесообразнее с помощью численных методов.

На следующем рисунке показана сходимость метода наискорейшего градиентного спуска в случае двух координат



Метод покоординатного спуска

Построение оптимизирующей последовательности точек $\{x^k\}$, сходящихся к точке минимума x^* функции $f(x)$ осуществляется в данном методе по правилу

$$x^{jk+1} = x^{jk} - t_k \left(\frac{\partial f(x)}{\partial x_{k+1}} \right)_{x = x^{jk}} \cdot e_{k+1}$$

где $j = 0,1,\ldots$ – номер цикла вычислений, $k = 0,1,\ \ldots,\ n - 1$ номер итерации внутри цикла, x^{00} – точка начального приближения;

$$e_{\mathrm{k+1}} = (0, \quad 0, \quad \cdots, 1, \quad \ldots \quad 0)$$

единичный вектор, содержащий 1 на $(k + 1)$ позиции и нули на остальных

Шаг t_k выбирается из условия

$$f\left(x^{jk} - t_k \left(\frac{\partial f(x)}{\partial x_{k+1}} \right)_{x = x^{jk}} \cdot e_{k+1} \right) - f\left(x^{jk}\right) < 0$$

Если выбранное условие на текущем значении t_k не выполняется, то шаг уменьшается вдвое

$$t_k \to t_k/2$$

И точка

$$x^{jk} - t_k \left(\frac{\partial f(x)}{\partial x_{k+1}} \right)_{x=x^{jk}} \cdot e_{k+1}$$

пересчитывается. Можно увидеть, что при фиксированном j за одну итерацию с номером k изменяется только одна проекция точки x^{jk}, имеющая номер $k + 1$, а в течение всего цикла с номером j, т.е. начиная с $k = 0$ и кончая $k = n - 1$ изменяются все n проекций точки x^{j0}. После этого точке x^{jn} присваивается номер $x^{j+1,0}$ и она берется за начальную точку для вычислений на $j + 1$ цикле.

Полученные в результате вычислений точки могут быть записаны как элементы последовательности $\{x^l\}$, где $l = n \cdot j + k$, порядковый номер точки, т.е.

$$\left\{x^l\right\} = \{x^0 = x^{00}, x^1 = x^{01}, \ldots, x^n = x^{0n} = x^{10}, x^{n+1} = x^{11}, x^{n+2} = x^{12}, \ldots\}$$

Графическая иллюстрация сходимости метода градиентного спуска при n= 2 дана ниже