1. Метод градиентного спуска

Градиентные методы довольно часто используются для решения задач многомерной безусловной оптимизации. Пусть стоит задача найти минимум функции:

$$f(\theta) \to \min_{\theta}, \tag{1}$$

где θ – **вектор** переменных. Алгоритм метода градиентного спуска состоит в последовательном движении в направлении наискорейшего спуска, то есть в направлении антиградиента $-\nabla_\theta f(\theta)$.

<u>Метод градиентного спуска</u>

Вход: функция $f(\theta)$, начальное приближение θ^0, шаг градиентного спуска η, число итераций n, ϵ.

1. Определить $\nabla_\theta f(\theta)$.
2. На каждом шаге $t = 1, 2, \ldots, n$:
   $\theta^{(t+1)} = \theta^{(t)} - \eta \nabla_\theta f\left(\theta^{(t)}\right)$.

Возможные критерии останова:

a. $\| \theta^{(t+1)} - \theta^{(t)} \| \leq \epsilon$,

b. $\| f\left(\theta^{(t+1)}\right) - f\left(\theta^{(t)}\right) \| \leq \epsilon$.

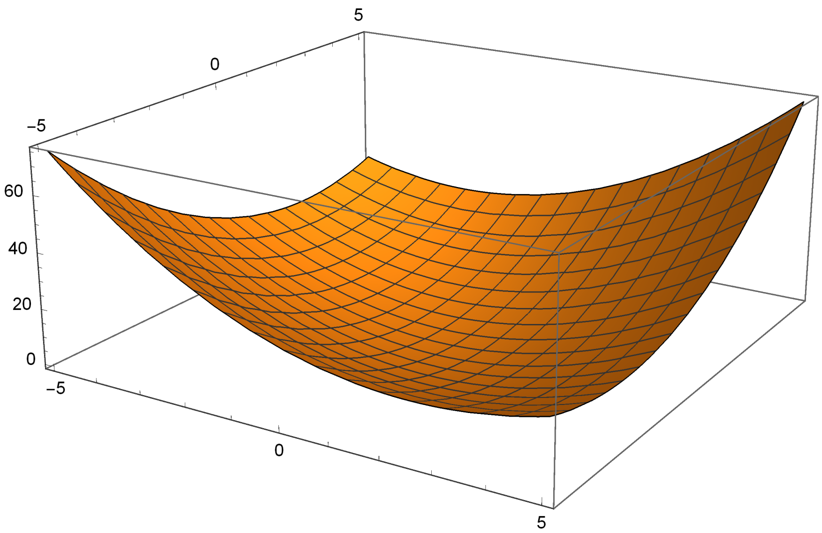
Задание 1

Методом градиентного спуска найти минимум функции $f(\theta_1, \theta_2) = \theta_1^2 + \theta_1 \theta_2 + \theta_2^2$. В качестве начального приближения взять $\theta^{(0)} = \{1, 1\}$. Отобразить вектор шага градиентного спуска на каждом шаге.

```
In[30]:= Clear[f]
         f[θ1_, θ2_] := θ1^2 + θ1 θ2 + θ2^2

In[∘]:= Plot3D[f[θ1, θ2], {θ1, -5, 5}, {θ2, -5, 5}, ImageSize → 300]
```

Out[∘]=



Если взять чересчур большим шаг градиентного спуска, то есть риск «перескочить» минимум.

Задание 2

Добавить в реализацию алгоритма изменение размера шага градиентного спуска на каждой итерации:

1. Обратно пропорционально номеру итерации:

$$\theta^{(t+1)} = \theta^{(t)} - \eta^{(t)}\, \nabla_\theta f\big(\theta^{(t)}\big),$$

$$\eta^{(t+1)} = \frac{\eta^{(0)}}{t}.$$

2. Недостатком первого способа является то, что с увеличением номера итерации дальнейшие шаги будут чересчур маленькими, и есть риск не дойти до минимума. Чтобы этого избежать, можно использовать экспоненциально затухающий шаг:

$$\eta^{(t+1)} = \eta^{(0)}\, e^{\frac{1-t}{t}}.$$

Задание 3

Выбирать шаг градиентного спуска можно таким образом, чтобы значение функции $f\big(\theta^{(t+1)} - \eta^{(t+1)}\, \nabla_\theta f\big(\theta^{(t+1)}\big)\big)$ было наименьшим. Для нахождения оптимального шага $\eta^{(t+1)}$ используются любые методы одномерной оптимизации. Полученный алгоритм имеет название **метод наискорейшего спуска**.

2. Метод импульсов

Приведенные выше методы не учитывают характер и форму целевой функции. Метод импульсов помогает ускорить градиентный спуск в нужном направлении.

Согласно методу импульсов (методу моментов) точка обладает массой, соответствующей текущим значениям вектора переменных, а значит в тот момент, когда точка начнет движение в сторону, противоположную направлению градиента функции, у нее появится ненулевая скорость. Если точка пришла в новое положение с некоторой ненулевой скоростью, то ускорение направлено по градиенту и точка не может резко изменить направление движения. Идея метода состоит в том, чтобы на каждом шаге учитывать направление движения на предыдущем шаге. Обозначим направление движения на предыдущем шаге как $v^{(t-1)}$, тогда:

$$v^{(t)} = \gamma\, v^{(t-1)} + \eta\, \nabla_\theta f\big(\theta^{(t)}\big), \tag{2}$$

$$\theta^{(t+1)} = \theta^{(t)} - v^{(t)}, \tag{3}$$

где $\gamma < 1$ – параметр, определяющий скорость изменения направления движения.

Задание 4

Найти минимум функции: $f(\theta_1,\, \theta_2) = (1.5 - \theta_1 + \theta_1\, \theta_2)^2 + (2.25 - \theta_1 + \theta_1\, \theta_2^2)^2 + (2.625 - \theta_1 + \theta_1\, \theta_2^3)^2$

1) методом градиентного спуска: $\big(\theta_1^{(0)},\, \theta_2^{(0)}\big) = (0.7,\, 1.4)$, $\eta = 0.01$, $\left|\big(\theta_1^{(t+1)},\, \theta_2^{(t+1)}\big) - \big(\theta_1^{(t)},\, \theta_2^{(t)}\big)\right| < 10^{-7}$, $n = 10^3$;

2) методом импульсов: $\gamma = 0.9$.

Сравнить результаты и количество итераций.

3. Метод Нестерова

Согласно методу Нестерова градиент необходимо считать не в текущей точке, а в той, в которую она придет после шага градиентного спуска. Таким образом алгоритм заглядывает вперед по вектору обновления:

$$v^{(t)} = \gamma\, v^{(t-1)} + \eta\, \nabla_\theta f\big(\theta^{(t)} - \gamma\, v^{(t-1)}\big),$$

$$\theta^{(t+1)} = \theta^{(t)} - v^{(t)}.$$

Значение параметра γ рекомендуют выбирать равным 0.9.

Задание 5

Найти минимум функции из задания 4 методом Нестерова с теми же параметрами.