Optimization

master in fundamental principles of data science

Optimization Problem 5

Conjugate gradient method problem

Author
Vladislav Nikolov Vasilev

Faculty of Mathematics and Computer Science

Academic year 2021-2022

1 Problem description

Solve the linear system

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

using the conjugate-gradient method.

2 Solution

The previously defined linear system of equations can be expressed as a minimization problem of a squared function in which we have to find a vector $\mathbf{x}^*$ that is the minimum of the function, and thus, solution to the original system. Therefore, let us define A and $\mathbf{b}$ as

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

where A is a symmetric positive definite matrix, which implies that there exists a unique solution to this problem.

We can express the previously defined linear system as follows:

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T\mathbf{x}$$

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{2}\begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \\ &= \frac{1}{2}x^2 + y^2 + \frac{3}{2}z^2 - x - y - z \qquad (1) \end{aligned}$$

If we minimize the quadratic function that can be seen in (1), we will find the solution to the linear system. The gradient of this function is

$$\nabla f(\mathbf{x}) = (x - 1,\ 2y - 1,\ 3z - 1) \qquad (2)$$

We are going to solve this problem using the conjugate gradient method. Thus, let us define $\mathbf{x}_0$ as the starting point of the algorithm. In this case, we have chosen the following point:

$$\mathbf{x}_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Now, we can obtain $\nabla f(\mathbf{x}_0)$ and z_1:

$$\nabla f(\mathbf{x}_0) = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \quad z_1 = -\nabla f(\mathbf{x}_0) = \begin{pmatrix} 0 \\ -1 \\ -2 \end{pmatrix}$$

In order to compute the next point, $\mathbf{x}_1$, we have to compute α_1^* first, which can be obtained using the following expression:

$$\alpha_1^* = -\frac{(\mathbf{z}_1)^T \nabla f(\mathbf{x}_0)}{(\mathbf{z}_1)^T A \mathbf{z}_1} = -\frac{\begin{pmatrix} 0 & -1 & -2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}}{\begin{pmatrix} 0 & -1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \\ -2 \end{pmatrix}} = \frac{5}{14}$$

With this value, we can compute now $\mathbf{x}_1$ as follows:

$$\mathbf{x}_1 = \mathbf{x}_0 + \alpha_1^* \mathbf{z}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \frac{5}{14} \begin{pmatrix} 0 \\ -1 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 \\ 9/14 \\ 2/7 \end{pmatrix}$$

In order to compute the next point, $\mathbf{x}_2$, we have to first compute the gradient of the recently obtained point and the new direction:

$$\nabla f(\mathbf{x}_1) = \begin{pmatrix} 0 \\ 2/7 \\ -1/7 \end{pmatrix}$$

$$\mathbf{z}_2 = -\nabla f(\mathbf{x}_1) + \frac{(\nabla f(\mathbf{x}_1))^T \nabla f(\mathbf{x}_1)}{(\nabla f(\mathbf{x}_0))^T \nabla f(\mathbf{x}_0)} \mathbf{z}_1 =$$

$$= -\begin{pmatrix} 0 \\ 2/7 \\ -1/7 \end{pmatrix} + \frac{\begin{pmatrix} 0 & 2/7 & -1/7 \end{pmatrix} \begin{pmatrix} 0 \\ 2/7 \\ -1/7 \end{pmatrix}}{\begin{pmatrix} 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}} \begin{pmatrix} 0 \\ -1 \\ -2 \end{pmatrix} =$$

$$= \begin{pmatrix} 0 \\ -15/49 \\ 5/49 \end{pmatrix}$$

Now, we have to compute α_2^* as follows:

$$\alpha_2^* = -\frac{(\mathbf{z}_2)^T \nabla f(\mathbf{x}_1)}{(\mathbf{z}_2)^T A \mathbf{z}_2} = -\frac{\begin{pmatrix} 0 & -15/49 & 5/49 \end{pmatrix} \begin{pmatrix} 0 \\ 2/7 \\ -1/7 \end{pmatrix}}{\begin{pmatrix} 0 & -15/49 & 5/49 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ -15/49 \\ 5/49 \end{pmatrix}} = \frac{7}{15}$$

With these values, we can now calculate the next point:

$$\mathbf{x}_2 = \mathbf{x}_1 + \alpha_2^* \mathbf{z}_2 = \begin{pmatrix} 1 \\ 9/4 \\ 2/7 \end{pmatrix} + \frac{7}{15} \begin{pmatrix} 0 \\ -15/49 \\ 5/49 \end{pmatrix} = \begin{pmatrix} 1 \\ 1/2 \\ 1/3 \end{pmatrix}$$

If we try to compute the gradient of the recently found point, we can see that

$$\nabla f(\mathbf{x}_2) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0 \tag{3}$$

Since the gradient at this point is 0, this means that the point $\mathbf{x}^* = (1,\ 1/2,\ 1/3)$ is the minimum of the function, and thus, the solution of the linear equation system.