FEDERAL STATE AUTONOMOUS EDUCATIONAL INSTITUTION

OF HIGHER EDUCATION

ITMO UNIVERSITY

Report

on the practical task No. 3

“Algorithms for unconstrained nonlinear optimization. First- and second-order methods”

Performed by

*Vdovkina Sophia*

*Syrchenko Arina*

*Academic group: J4133c*

Accepted by

Dr Petr Chunaev

St. Petersburg

2022

**Goal**

The use of first- and second-order methods (Gradient Descent, Non-linear Conjugate Gradient Descent, Newton’s method and Levenberg-Marquardt algorithm) in the tasks of unconstrained nonlinear optimization.

**Formulation of the problem**

Generate random numbers and . Furthermore, generate the noisy data , where , according to the following rule:

,

where are values of a random variable with standard normal distribution. Approximate the data by the following linear and rational functions:

1. (linear approximant)
2. (rational approximation)

by means of least squares through the numerical minimization (with precision ) of the following function:

To solve the minimization problem, use the methods of Gradient Descent, Conjugate Gradient Descent, Newton’s method and Levenberg-Marquardt algorithm. If necessary, set the initial approximations and other parameters of the methods. Visualize the data and the approximants obtained in a plot separately for each type of approximant so that one can compare the results for the numerical methods used. Analyze the results obtained (in terms of number of iterations, precision, number of function evaluations, etc.) and compare them with those from Task 2 for the same dataset.

**Brief theoretical part**

If it is possible to calculate first- and second-order derivatives of a function, this information can be successfully utilized in solving the optimization problem.

*Gradient Descent*

Gradient descent is a first-order iterative optimization method for differentiable functions. Gradient descent is based on the observation that if the multi-variable function is defined and differentiable in a neighborhood of a point *a*, then *f(x)* decreases fastest in the direction of . It can be written down as the following formula:

Gradient descent algorithm may converge faster, if step size is chosen correctly.

*Conjugate gradient descent*

The conjugate gradient descent is an iterative method for unconditional optimization in a multidimensional space. It works when the function is approximately quadratic near the minimum, which is the case when the function is twice differentiable at the minimum and the second derivative is non-singular there.

This method simulates such a physical quantity as momentum in order to step out of the local minimum. The conjugate gradient method is seen as an evolution of the gradient descent where instead of doing small steps towards the negative gradient, it uses an adjustable step length and performs a line search in this direction until it reaches the minimum.

*Newton’s method*

Newton's method in optimization is applied to the derivative *f ′* of a twice-differentiable function *f* to find roots of constructing a sequence from an initial approximation so that as , where

From the Taylor expansion of *f* near ,

.

We use this quadratic function (with respect to ) as an approximant to *f* in a neighborhood of . The vertex of the corresponding parabola gives us the next point . To find the vertex x-coordinate, we write:

. Incrementing by this gives us a point closer :

It is proved that for the chosen class of *f* one has as .

Newton’s method uses Hessian and its inverse to find the optimum. Also, it requires the function to be convex and have invertible Hessian. Computing the Hessian can be a time demanding task. Quasi-Newton methods are aimed to find an approximation to the Hessian to simplify the task.

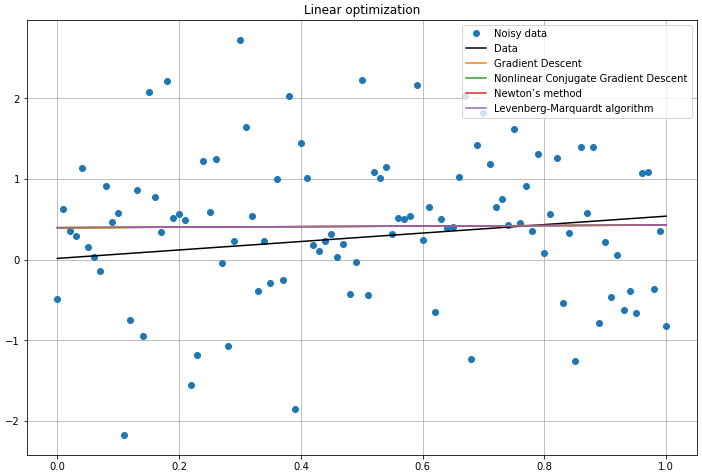
*Levenberg-Marquardt algorithm*

The Levenberg-Marquardt algorithm interpolates between the Gauss–Newton algorithm (GNA) and gradient descent method. In the gradient descent method, the sum of the squared errors is reduced by updating the parameters in the steepest-descent direction. In the Gauss-Newton method, the sum of the squared errors is reduced by assuming that the least squares function is locally quadratic and then it shall find the minimum of the function approximated by the Taylor expansion. The Levenberg-Marquardt method uses gradient-descent method when parameters are far from the optimal value. The Gauss-Newton method is used when parameters are close to the optimal value.

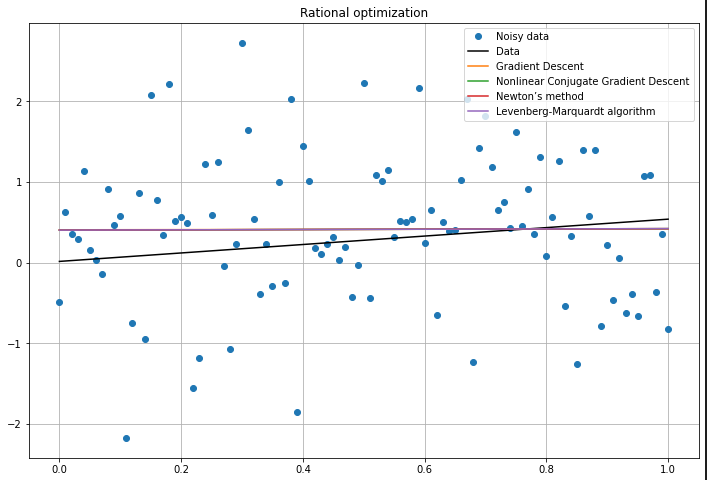
**Results**

To solve the optimization problem Gradient Descent, Conjugate Gradient Descent, Newton’s method and Levenberg-Marquardt algorithm were used.

Data was generated by the rules provided in the task, pictures 1 – 2 respectively show each of the approximating functions (linear and rational) using proposed methods.

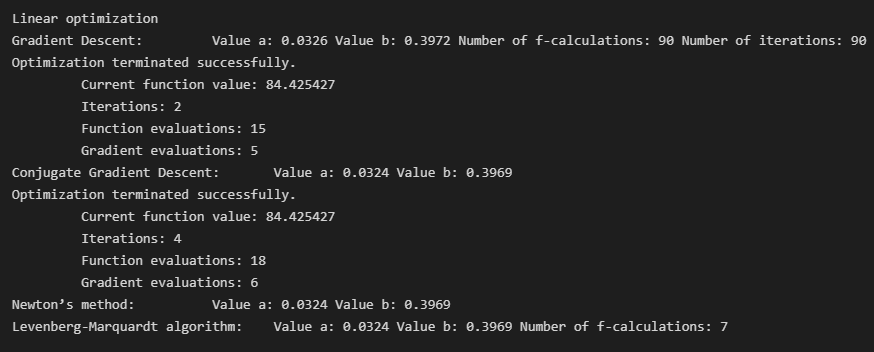


Picture 1 – linear optimization

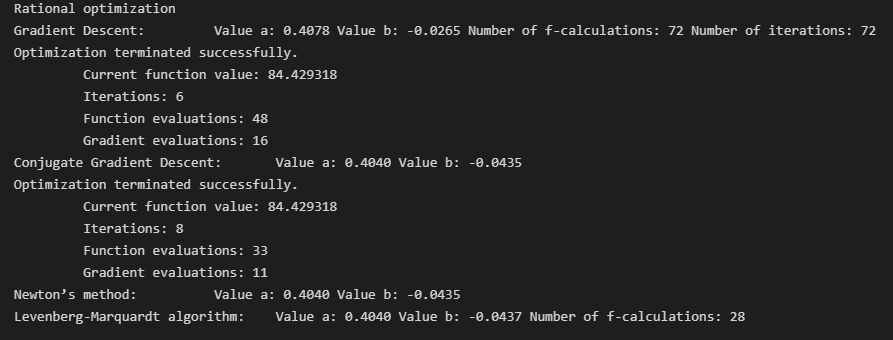


Picture 2 – rational optimization

As can be seen in the pictures 3 – 4 all algorithms converged to the parameter values that are almost identical with the slight difference, which cannot be seen on the plots.



Picture 3 – values obtained with linear optimization



Picture 3 – values obtained with rational optimization

Further, obtained results were compared with the results obtained by direct methods from Task 2. Number of function calculations and iterations for each method are presented in Table 3.

As expected, the gradient descent method required the largest number of iterations and the calculation of the function, since at each iteration it needs to

calculate the direction of motion and the values of the function. In turn, Newton's method requires

fewer iterations than gradient descent, however, it requires additional information about the second

derivatives of the original function. As a compromise for these two functions, can be used the

conjugate gradient method, which proved to be better both in the number of iterations and in

function calculations among the three mentioned methods, and does not require information about

the second derivatives of the original function. And, finally, the Levenberg-Marquardt method

became the most efficient in terms of iterations and calculations of the function. For almost every

dataset, it took 3 iterations to achieve the specified accuracy. This may be partly since this method

was taken from the scipy library, so it is optimized.

In both linear and rational cases of approximates, all gradient-based method required less function calls than direct methods, except for vanilla Gradient Descent with static learning rate which required less function calls than Gauss method, but more than Nelder-Mead algorithm.

The most effective in terms of function calls is Levenberg-Marquardt algorithm which required only 2 calls in the case of linear approximate and 12 calls in the case of rational approximate. The other two methods – Conjugate Gradient Descent and Newton’s method – required 5 function calls each for the linear approximate and 26 and 16 calls in the case of rational approximate. Best gradient-based method - Levenberg-Marquardt required 10-50 times less function calls than the best direct method for linear type of approximate.

**Conclusion**

Gradient-based methods of optimization were used to solve two-dimensional optimization problem introduced in the previous assignment. All methods converged to the same values of parameters as direct methods. The most effective among gradient-based methods in terms of function calls was Levenberg-Marquardt algorithm. Although this algorithm can only be used to solve least-squares problem with quadratic loss.

**Appendix**

Source code is available on <https://github.com/sophia-vdovkina/Analysis-and-development-of-algorithms/blob/main/Task%202/task.ipynb>