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

Konstantin Krechetov

Academic group: j4132c

Accepted by

Dr Petr Chunaev

St. Petersburg

2020

Goal

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

Formulation of the problem

Noisy linear function should be approximated with linear and rational functions using Gradient Descent, Conjugate Gradient Descent, Newton's method and Levenberg-Marquardt algorithm. Those results and number of iterations must be compared and demonstrated.

Brief theoretical part

For some problems it is possible to calculate first- and second-order function derivatives. This information gives a significant benefit for optimization problem.

The simplest method that uses such information is Gradient Descent. Each iteration it calculates the function gradient and makes a step against its direction (In direction of antigradient). With adjusting the step size, this algorithm may be accelerated to find extremum faster.

The problem is local extremum. For such object functions there are modified Gradient methods such as "Gradient descent with Momentum" (It has simulation term of physical momentum to give opportunity to step out of local minimum).

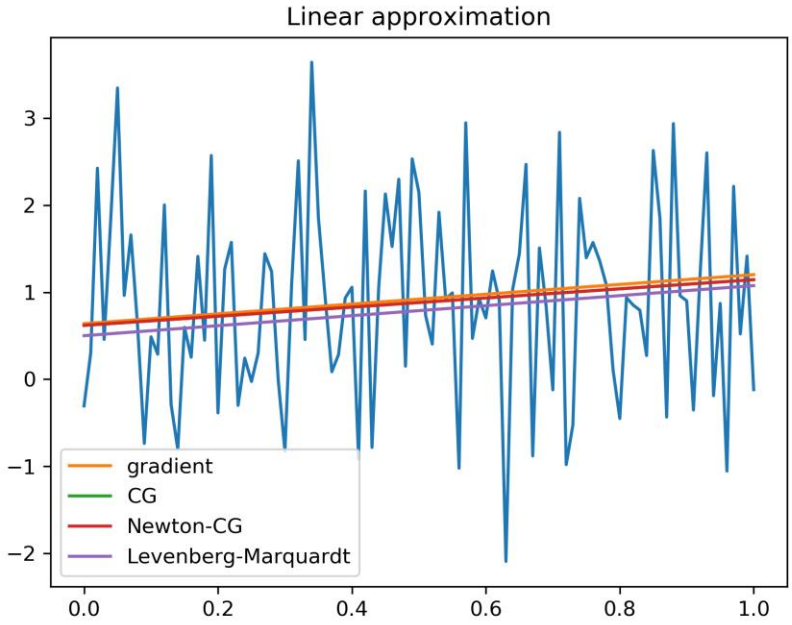
Besides, second-order info may be useful. For example, Newton's Method uses Hessian and its inverse to find optimum. This method requires considered function to be convex and have invertible Hessian. Otherwise, it ***does not guarantee finding global*** *extremum or can't be calculated.*

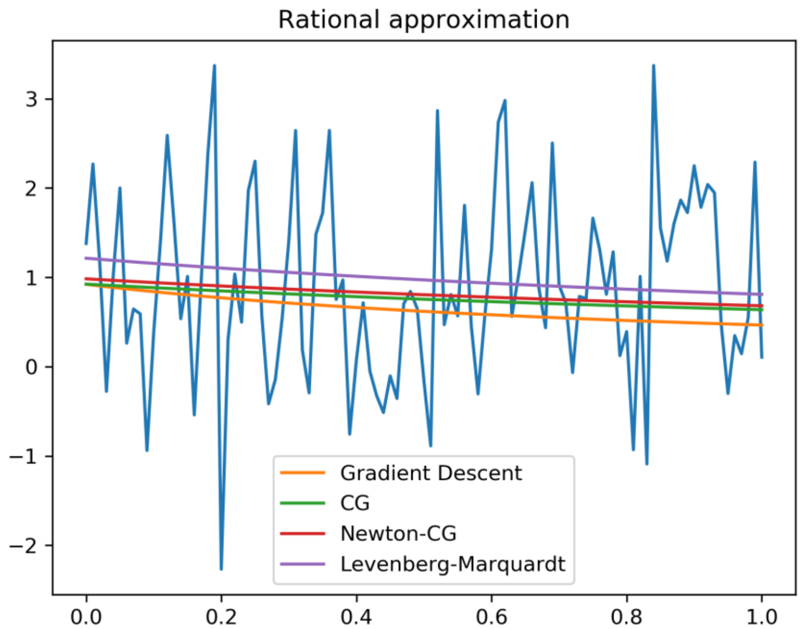
Usually it is difficult to compute the Hessian. Other methods called Quasi-Newton propose to find an approximation to the Hessian to simplify the scheme. In this work Quasi-Newton Hessian approximation is used for Newton's method.

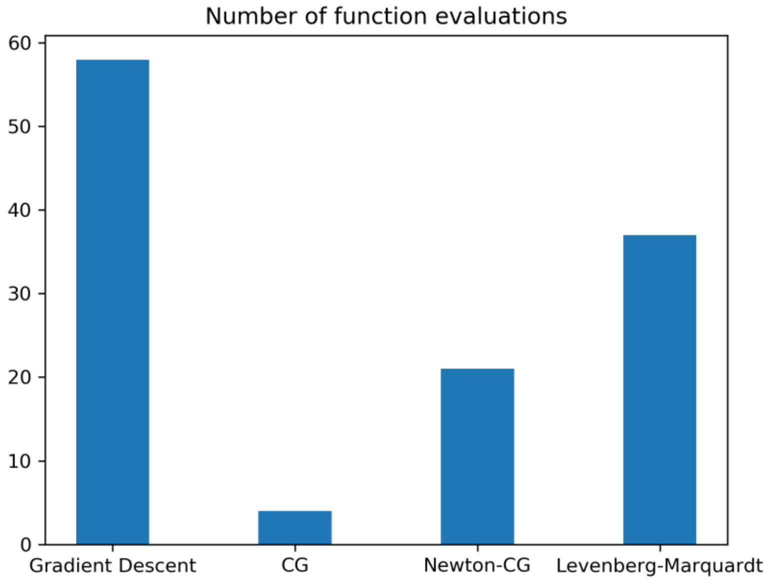
Levenberg–Marquardt algorithm (LMA) also uses Jacobian and Hessian. In case of least-squares problem these matrices J and H have special form. As with many fitting algorithms, the LMA finds only a local minimum. The LMA interpolates between the Gauss–Newton algorithm (GNA) and the method of gradient descent.

Results

The figures of functions and solutions by different methods are demonstrated below. The bar plots show number of function evaluations for each optimization method.







Conclusions

Thus, we can see that all methods find ***local*** *minimum and* ***depends on its initial estimate****. In this work Conjugate Gradient Descent and Newton's methods seemed most reliable because two others often found further local minimum.*

What is more, they took less function evaluations. However, it also depends on hyper-parameters.

Appendix

Source code is available on

https://github.com/KostyaKrechetov/ITMO-Analysis-and-development-of-algorithms/tree/master/Task3