Modifying SQP Method for Constrained Optimization

Elaf Sulaiman Khaleel [1] and Eman T. Hamed [2]

[1] dept. of Mathematics, college of Computers Sciences and Mathematics, University of Mosul

[2] dept. of Operation Researches and Intelligent Techniques, college of Computers Sciences and Mathematics, University of Mosulelaf.sulaiman@uomosul.edu.iq, dr.emantarik@uomosul.edu.iq

Abstract

We study a variant of SQP technology to find the optimal solution of nonlinear programing with the equality and inequality constrained optimization , Two techniques have been developed, the first is the derivation of a new approximation for the Hessian matrix, which has proven to be positively defined, and the second is was designed to take advantage of the structure present in the Hessian of the augmented Lagrangian function for problem and showed local superliner convergence and that at the minimizer x*.

Keywords: ***augmented Lagrangian, SQP technology, constrained optimization .***

DOINumber:10.14704/nq.2022.20.8.NQ44918 **NeuroQuantology 2022; 20(8): 8985-8996**

I. INTRODUCTION

The Constrained Nonlinear Programming Suite (1) was a fundamental challenge to various gradient optimization techniques. The Sequential Quadratic Programming (SQP) technique sophisticated by Wilson[1], han [2,3] and powell[4,5] that utilises the active group strategy in solving sub-problems of Quadratic Programming (QP) has been shown to be effective in determining local optima points.

min $Q(x)$ ... (1)

s.t

$e_i \leq 0$ for i=1,2,……,p

$c_j = 0$ for j=1,2,……,m

$Q : R^n \rightarrow R$ Smooth function e_i is inequality constrained and c_j equality constrained.

One of the most important developments and successes in continuous improvement is the quasi-Newton effective Methods for solving problems of miniaturization. The main reasons for this success are: Positive definite quasi-Newton updates is compatible with line search rules that maintain global convergence. The Hessian matrix can be approximated at every iteration and positive definiteness still preserved Newton's method (or quasi-Newton methods) is used to directly solve the original problem's KKT conditions using SQP technique, also known as sequential or recursive quadratic programming approaches. [6].

The SQP is a kind of iterative method, at iterate k it needs to solve a QP subproblem

Min $g_k^T d + \frac{1}{2} d^T \bar{d}$ (2)

s.t. $e_i + \nabla e_i\, d \leq 0$

$c_i + \nabla c_i\, d = 0$

The iteration has the form

$$x_{k+1} = x_k + \alpha_k d_k \qquad (3)$$

Where d_k solves (2) and α_k is step length chose to reduces the value of merit function form (1). where $\mathbb{B}$ is an n×n matrix. Often
$\mathbb{B}$ is required to be positive definite, it updated by BFGS at every step by means of a quasi-Newton update formula, and g is the
gradient of Q at x. In particular, the BFGS update formula is given by

$$\mathbb{B}_{k+1} = \mathbb{B}_k - \frac{\mathbb{B}_k s_k s_k^T \mathbb{B}_k}{s_k^T \mathbb{B}_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} \qquad (4)$$

$$d_{k+1} = -g_{k+1} + \left(\frac{g_{k+1}^T y_k - t_k g_{k+1}^T s_k}{y_k^T s_k}\right) s_k + \frac{g_{k+1}^T s_k}{y_k^T s_k} y_k \quad (5)$$

so the three new terms, according to Yao et al. [12] In 2020. which had an entirely unique t.

as well planned a statement stating that if tk is dfdclose to ($t_k > \frac{\| y_k\|^2}{y_k^T s_k}$), then the path will be zigzag direction of search results [13]. hence, the chosen the anther choice to t_k :

$$t_k = 1 + 2\frac{\| y_k\|^2}{y_k^T s_k} \quad (6)$$

$$d_{k+1} = -\left[I - \left(\frac{y_k s_k^T + s_k y_k^T}{y_k^T s_k}\right) + t_k \frac{s_k s_k^T}{y_k^T s_k}\right] g_{k+1} \quad (7)$$

Then the new approximation of BFGS is

$$d_{k+1} = -\mathbb{H}_k g_{k+1} \quad (8)$$

$$\mathbb{H} = I - \left(\frac{y_k s_k^T + s_k y_k^T}{y_k^T s_k}\right) + t_k \frac{s_k s_k^T}{y_k^T s_k} \quad (9)$$

Where

$$t_k = 1 + 2\frac{\| y_k\|^2}{y_k^T s_k} \quad (10)$$

Second: New Augmented Lagrange BFGS in SQP technological

Let the Lagrange function

$$(x, p, r) = Q(x) + p^T c(x) + r\, e(x) \quad (11)$$

where $\lambda \in R^m$ is Both the vector of Lagrange multipliers and only the Lagrange multiplier itself are commonly referred to.

These methods make use of an augmented Lagrangian, which is an altered version of the Lagrangian used in (5).

The iteration has the form

$$x_{k+1} = x_k + \alpha_k d_k$$

Where d_k is

$$d_{k+1} = \begin{cases} -\nabla(x, p, r) & for\ k = 0 \\ -\nabla(x, p, r) + \beta_k d_k & for\ k \geq 1 \end{cases} \quad (12)$$

and α_k is step length chose to reduces the value of merit function form (1).where $g_k = \nabla P(x, p, r)$, and

$$y_k = \nabla_x(x_{k+1}, p_{k+1}, r) - \nabla_x P(x_k, p_{k+1}, r) \quad (13)$$

Researchers have struggled for years to come up with alternative formulations of the Lagrangian SQP approach to address the lack of positive clarity for $\nabla^2 P(x, p,)$ As an initial alternative, the augmentative Lagrangian associated with Problem (1) is substituted for the Lagrangian in the following equation:

( ) ( ) 1 ( ) 2 ( )

$$\bar{P}(x,\wp,r) = P(x,\wp,r) + \wp\,(c(x)) - r\ln(e(x)) \tag{14}$$

Take note of the augmented Lagrangian's Hessian for a local solution to the problem (1).

$$\mathcal{H}(x,\wp,r) = \nabla^2 P(x,\wp,r) = \nabla^2 \bar{P}(x,\wp,r) + \wp\,\mathbb{C}(x)\mathbb{C}(x) + r\,\frac{\mathbb{D}(x)\mathbb{D}^T(x)}{e^2(x)} \tag{15}$$

Where $\mathbb{C}(x) = \nabla c(x)$ and $\mathbb{D}(x) = \nabla e(x)$

Now a new correlation has been established between New approximation of Hessian and New Augmented Lagrange by replacing the new $\mathbb{H}$ of eq (9) with the $\nabla^2(x, \wp, r)$ in augmented Lagrangian's Hessian so it is positive defined and symmetric

Then we obtained

$$(\quad) ^2 (\quad)^2 ( ) ( ) \quad \mathbb{D}(x)\mathbb{D}^T(x) \quad (16)$$

$$\mathcal{H}\ x,\wp,r = \nabla\ P\ x,\wp,r = \bar{\mathbb{H}} + \mathbb{C}\ x\ \mathbb{C}\ x\ \overline{+2\ e3}(x)$$
$\wp$

The methodologies for changing penalty parameters and Lagrange multipliers are crucial in the algorithm's development. Algorithms are affected by updates in different ways. In order to preserve convergence, the algorithm needs to construct a nondecreasing sequence k. The penalty settings in the algorithm will be different for each component of constrained.

We minimize the augmented Lagrangian function and then updating the multiplier vector in the outer iteration by some formula depending on the current knowledge to be more precisely kept as a constant.

$$\wp_i^{k+1} = \wp_i^k + \frac{2}{\wp_i^k} c_i(x) \quad (17)$$

$$r_i^{k+1} = r_i^k - \frac{i}{i}\, e_i(x) \quad (18)$$

ALGORITHM : SQP LAGRANGIAN BFGS METHOD.

8989

Algorithm A .(Outer iteration)

Step 1 let $^0_i > 0, i = 1, \dots, m$; tolerances $\epsilon_1 \geq 0$ and $\epsilon_2 \geq 0$; $\beta > 1$; starting points $x_0 \in \mathbb{R}^n$ and $\lambda^0 \in \mathbb{R}^m$. Set k = 0.Step 2 circulator an approximate minimizer x_{k+1} of (3) by using Algorithm B

Step 3 If $\|\nabla(x, \wp, r)\| \leq \epsilon_1$ and $\|g_k(x_{k+1})\| \leq \epsilon_2$, stop x_{k+1} is optimal solution of the constrained optimization problem (1).

Step 4 Update penalty parameters and the Lagrange multiplier vector from eq (17) &(18)Step 5 Set $k = k + 1$ and go to step2.

The internal iteration iterative procedure used in step 2 of algorithm A can be take from

Algorithm B (Inner iteration)

Step 1 Start with an initial point $x_k \in \mathbb{R}^n$ which is obtained from Algorithm A , a $n \times n$ positive definite symmetric matrix $\mathbb{H}_k$ to approximate the inverse of the Hessian matrix of $P(x, \wp, r)$ and a tolerances $\epsilon^k \geq 0$.

Step 2 Compute the gradient of the augmented Lagrangian function, at point x_k, and set d_k FROM EQ (12)Step 3 compute an acceptable step size a_k in the direction d_k from (3)

Step 4 Test the new point x_{k+1} for optimality. If $\|\nabla(x, \wp, r)\| \leq \epsilon^k$, terminate the iterative process and return to step 3 in Algorithm B . Otherwise, go to step 5.

Step 5 Update the Hessian matrix from eq (16). Set $x_k = x_{k+1}$, $\mathbb{H}_k = \mathbb{H}_{k+1}$ and go to step 2 .

The Stability of modified SQP technique

It is common practice for descent algorithms to be stable since one assures that the function to be minimized is reduced at each stage of the process. In this part of the article, it will be demonstrated that the direction of search $-\mathbb{H}g_k$is downward, which means that k can always be selected to have a positive value. If, and only if, the conditions are met, the direction will be downward.

Theorem 1

The matrix $\mathbb{H}_{k+1}$ the new QN- BFGS method, which is defined by eq (9), achieved the following QN- condition

$\mathbb{H}y = ps$

$$\mathbb{H}y_k = y_k^T - \frac{\|y_k\|^2}{y_k^T s_k} s_k^T - \frac{y_k^T s_k}{y_k^T s_k} y_k^T + t \frac{y_k^T s_k}{y_k^T s_k} s_k^T \quad (19)$$

$$\mathbb{H}y_k = [t_k + \frac{\|y_k\|^2}{y_k^T s_k}] s_k^T \quad (20)$$

$$p = t_k + \frac{\|y_k\|^2}{y_k^T s_k} \quad (21)$$

It satisfied the QN condition $\mathbb{H}y = ps$

Theorem 2

If $\mathbb{H}_1$ is a positive definite matrix, then all matrix $\mathbb{H}_{k+1}$ which is generated by (9) is also positive definite, i.e. $z^T\mathbb{H}_k z > 0$,for any vector $z \neq 0$.

poof:

$$z^T \left[I - \left(\frac{y_k s_k^T + s_k y_k^T}{y_k^T s_k}\right) + t_k \frac{s_k s_k^T}{y_k^T s_k}\right] z \qquad (22)$$

$$\Rightarrow z^T z - \left(\frac{z^T y_k s_k^T z + z^T s_k y_k^T z}{y_k^T s_k}\right) + t_k \frac{z^T s_k s_k^T z}{y_k^T s_k}$$

$$\Rightarrow z^T\mathbb{H}_k z = z^T z - \frac{2(z^T y_k)(s_k^T z)}{y_k^T s_k} + t_k \frac{(z^T s_k)^2}{y_k^T s_k}$$

Now there are tow case

If $(z^T y_k)(s_k^T z)$ have different signal

$$\Rightarrow z^T\mathbb{H}_k z = z^T z - \frac{2(z^T y_k)(s_k^T z)}{y_k^T s_k} + t_k \frac{(z^T s_k)^2}{y_k^T s_k} > 0 \qquad (23)$$

8991

If $(z^T y_k)(s_k^T z)$ has same signal

$$\Rightarrow z^T\mathbb{H}_k z \geq z^T z - \frac{2(z^T y_k)(s_k^T z)}{y_k^T s_k} \qquad (24)$$

Let $p = \frac{2(z^T y_k)(s_k^T z)}{y_k^T s_k}$ where $p < z^T z$

Then $z^T\mathbb{H}_k z \geq z^T z > 0$ That mean $\mathbb{H}_k$ is positive defind

Lemma (1)

When Q is twice continuously differentiable and positive constants , v exist, Byrd & Nocedal [14] demonstrated that theBFGS method is globally converging.

$$u\|z\|^2 \leq z^T\mathbb{H}_k z \leq v\|z\|^2 \qquad (25)$$

for all $z \in R^n$ and all $x \in donen$. It is important to note that (25) means that x* is the only solution to f . Assuming (25), they showed local superliner convergence and that at the minimizer x*, $\mathbb{H}_k$ is Lipschitz continuous.[15].

Theorem:3

Let $\mathbb{H}_k$is a symmetric positive definite matrix for BFGS update in SQP if Q is twice continuously differentiable and thereexist positive constants , v such that

$$u\|s\|^2 \leq s^T\mathbb{H}_k s \leq v\|s\|^2 \qquad (26)$$

for all $z \in R^n$ and all $x \in donen$. Note that (25) implies that f has a unique minimizer x* . They proved the local superlinear convergence under assumption (25) and that $\mathbb{H}_k$ is Lipschitz continuous at the minimizer x*.

$$s^T \left[I - \left(\frac{y_k s_k^T + s_k y_k^T}{y_k^T s_k}\right) + t_k \frac{s_k s_k^T}{y_k^T s_k}\right] s \qquad (27)$$

$$\Rightarrow s^T s - \quad y^T s_k \qquad (s^T y_k \qquad)(s^T s)$$

$k\ y^T s_k$
$(s^T s)(y\quad s^T s_k s^T_k s$ (28)
$^T s)$
$-\quad + t_k$
$y^T s_k\quad {}_k\ y^T s_k\quad k\quad y^T s_k$

$\Rightarrow s^T \mathbb{H}\quad s = \|s\|^2 - \|s\|^2 (s^T y_k) - \|s\|^2 + t\ \|s\|^2 \|s\| 2$ (29)

$k\quad y^T_k s_k\quad k\ {}_k y^T s_k$
$T\quad [\ \|s\|^2\ (s^T y_k)$
$\|\ \|^2$

$\Rightarrow s\quad \mathbb{H}_k\ s \leq t_k\ y^T_k\ ]\ s$ (30)
$s_k - y^T s_k$

Let $\mathcal{v} = t\quad \|s\|^2\quad -$ (31)
$(s^T y_k)$

$k\ y^T s_k\quad {}_k\ y^T s_k$
$s^T \mathbb{H}_k\ s \leq \mathcal{v}\|s\|^2$ (32)

Theorem:4

Consider problem to minimize $Q(x)$subject $e_i \leq 0$,for i=1,2,......,p $c_j = 0$, for j=1,2,......,m . For a local minimum, let KKT be satisfied by the second order sufficiency condition for KKT. Described $J = \{j: e_j(\bar{x}) = 0\}$, $I = \{i: e_j(\bar{x}) \leq 0\}$and the cone $C = \{d \neq 0 , e_j(\bar{x})d = 0 \ for \ j \in J \ and \ e_j(\bar{x})d \geq 0 \ for \ i \in I\}$ Then, if there exists a u such that $u_k > u_{k+1}$ therefore $\mathcal{H}(x,u) = \nabla^2 P(x,p,r)$ is positive definite $\forall d \in C$ then $\bar{x}$ is strict local minimum for eq.(1) for all $u \geq 0$.

Proof:

A local minimum in C cone KKT must fulfill what $(\bar{x},\bar{p},\bar{r})$ is called the second-order sufficiency requirement for a solution. And Hessian $\nabla^2(x,p,r)$ of the Lagrange function for eq.(1). Let's say there exists a d_k with $\|d_k\| = 1$, such that

$$d_k^T\nabla^2(x,p,r)d_k = d_k^T\nabla^2(x,p,r)d_k + \frac{2}{p} d_k^T \mathbb{C}(x)\mathbb{C}^T(x)d_k + r d_k^T \frac{\mathbb{D}(x)\mathbb{D}^T(x)}{e^2(x)} d_k \quad (33)$$

Since $\nabla^2(x,p,r)$ is positive incisive on the C, then we have $d_k^T\nabla^2 P(x,p,r)d_k \geq 0$ for all $i \in I$ or $j \in J$ consequently $\bar{x}$ a strict local minimum. Now, on the contrary, if there does not exist a μ_k such that $\nabla^2(x,u)$ is positive definite for $u_k > u_{k+1}$; then it must be the case that, given any u_k = , k = 1,2,3, … .. h. There exists a a d_k with $\|d_k\| = 1$, such that

$$d_k^T\nabla^2 P(\bar{x},\bar{p},r)d_k = d_k^T\nabla^2 P(\bar{x},\bar{p},r)d_k + \frac{2}{p} d_k^T\mathbb{C}(\bar{x})\mathbb{C}^T(\bar{x})d_k + r d_k^T\frac{\mathbb{D}(\bar{x})\mathbb{D}^T(\bar{x})}{e^2(\bar{x})}d_k < 0 \quad (34)$$

Since $\|\bar{d}_k\| = 1$ for all k, $\{d_k\}$ with limit point a d_k with $\|\bar{d}_k\| = 1$ we show that $\mathbb{C}^T(\bar{x})d_k = 0 \ \forall j \in J$ $\mathbb{C}(\bar{x})d_k = 0$ $\forall i \in I$ and $\mathbb{D}(\bar{x})d_k = 0$. If $d_k^T\nabla^2(\bar{x},\bar{p},\bar{r})d_k$ is constant. For eq.(34) then we have $d_k^T\nabla^2 P(\bar{x},\bar{p},\bar{r})d_k \leq 0$,this is contradiction the second –order condition. Therefore we have $d_k^T\nabla^2(\bar{x},\bar{p},\bar{r})d_k$ for all $i \in I$ or $j \in J$ consequently $\bar{x}$ a strict local minimum .

8993

Experimental Results and Comparisons

To demonstrate the performance efficiency of all newly proposed algorithms, two comparison experiments were performed. We carried out a variety of numerical experiments, the results of which may be condensed into three distinct types of comparison activities, in order to validate the efficacy of the suggested algorithms. The first thing that needs to be done is to evaluate SQP technological utilizing the new derivative of approximation of Hessian in light of the standard SQP algorithms. With regard to the second efficiency, a comparison of

several optimized SQP technological with those that arealready New Augmented Lagrange BFGS.
We consider the conditions below the discontinuation criterion
$\|\nabla(x, p, r)\| \leq \epsilon_1$ and $\|g_k(x_{k+1})\| \leq \epsilon_2$
we employed the Dolan and more'[16] technique.

The following figures (1-3) is illustrate the results using the Dolan and more'. Displays the Dolan-More performance profile for these methods, which are susceptible to the frequency of suitable performance when compared to the basic methods.
Performance profile:

1

0.9

0.8

0.7

0.6

0.5

0.4

0.3

0.2

0.1

0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1
Tau

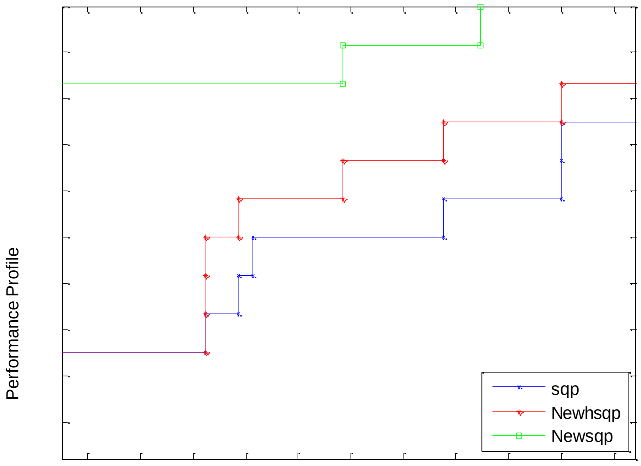

FIGURE 1 . data relating to how well the aforementioned methods perform in terms of function evaluations

Performance profile:
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.5 1 1.5 2 2.5 3
Tau

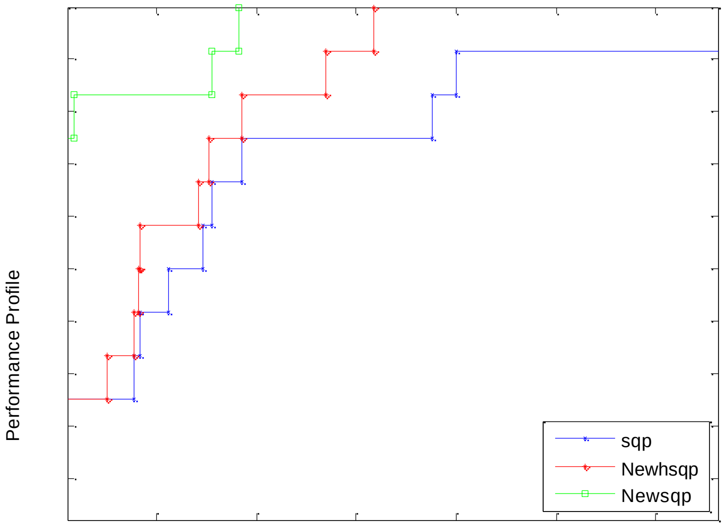

FIGURE 2 . data relating to how well the aforementioned methods perform in terms of iteration evaluations

Performance profile:
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.5 1 1.5 2 2.5 3
Tau

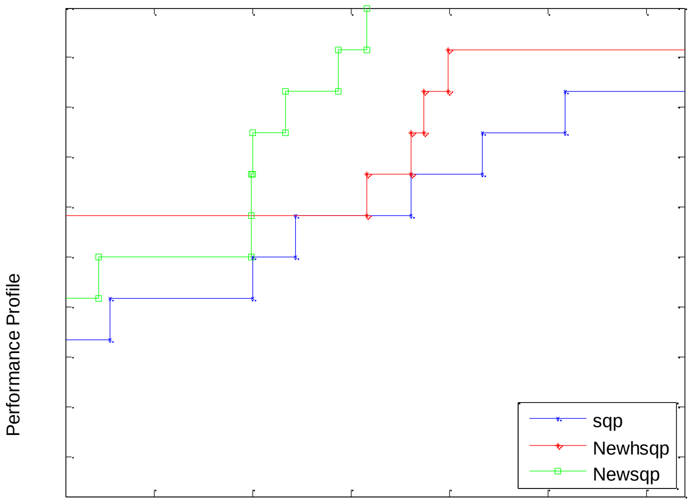

FIGURE 3 . data relating to how well the aforementioned methods perform in terms of gradients evaluations

By examining the Dolan-More performance profile, which is measured in CPU time, we may conclude from the three forms shown that the new method is particularly suitable for tackling problems with numerous dimensions

ACKNOWLEDGMENT

The College of Computer Sciences and Mathematics at the University of Mosul in the Republic of Iraq is funding theresearch.

REFERENCES

[1] Wilson, Robert B. "A simplicial algorithm for concave programming." Ph. D. Dissertation, Graduate School of Bussiness Administration (1963).
[2] Han, Shih-Ping. "A globally convergent method for nonlinear programming." Journal of optimization theory and applications 22.3 (1977): 297-309.
[3] Han, Shih-Ping. "Superlinearly convergent variable metric algorithms for general nonlinear programming problems." Mathematical Programming 11.1 (1976): 263-282.
[4] Powell, Michael JD. "Algorithms for nonlinear constraints that use Lagrangian functions." Mathematical programming 14.1 (1978): 224-248.
[5] Powell, Michael JD. "A fast algorithm for nonlinearly constrained optimization calculations." Numerical analysis. Springer, Berlin, Heidelberg, 1978. 144-157.
[6] Burke, James V., and Shih-Ping Han. "A robust sequential quadratic programming method." Mathematical Programming 43.1 (1989): 277-303.
[7] Boggs, Paul T., Anthony J. Kearsley, and Jon W. Tolle. "A global convergence analysis of an algorithm for large-scale nonlinear optimization problems." SIAM Journal on Optimization 9.4 (1999): 833-862.
[8] Byrd, Richard H., Richard A. Tapia, and Yin Zhang. "An SQP augmented Lagrangian BFGS algorithm for constrained optimization." SIAM Journal on Optimization 2.2 (1992): 210-241.
[9] Gill, Philip E., Walter Murray, and Michael A. Saunders. "SNOPT: An SQP algorithm for large-scale constrained optimization." SIAM review 47.1 (2005): 99-131.
[10] Ameli, Kazem, Alireza Alfi, and Mohammadreza Aghaebrahimi. "A fuzzy discrete harmony search algorithm applied to annual cost reduction in radial distribution systems." Engineering Optimization 48.9 (2016): 1529-1549.
[11] Mousavi, Yashar, and Alireza Alfi. "A memetic algorithm applied to trajectory control by tuning of fractional order proportional-integral-derivative controllers." Applied Soft Computing 36 (2015): 599-617.
[12] Yao, Shengwei, et al. "A one-parameter class of three-term conjugate gradient methods with an adaptive parameter choice." Optimization Methods and Software 35.6 (2020): 1051-1064.
[13] Alhawarat, Ahmad, et al. "A descent four-term conjugate gradient method with global convergence properties for large-scale unconstrained optimisation problems." Mathematical Problems in Engineering 2021 (2021).
[14] Byrd, Richard H., and Jorge Nocedal. "A tool for the analysis of quasi-Newton methods with application to unconstrained minimization." SIAM Journal on Numerical Analysis 26.3 (1989): 727-739.
[15] Chen, Xiaojun. "Convergence of the BFGS method for LC^1 convex constrained optimization." SIAM Journal on Control and Optimization 34.6 (1996): 2051-2063.
[16] Dolan, Elizabeth D., and Jorge J. Moré. "Benchmarking optimization software with performance profiles." Mathematical programming 91.2 (2002): 201- 213.