Hindawi Publishing Corporation
Abstract and Applied Analysis
Volume 2014, Article ID 507102, 6 pages
http://dx.doi.org/10.1155/2014/507102

Research Article

The Hybrid BFGS-CG Method in Solving Unconstrained Optimization Problems

Mohd Asrul Hery Ibrahim,[1] Mustafa Mamat,[2,3] and Wah June Leong[4]

[1] *School of Applied Sciences and Foundation, Infrastructure University Kuala Lumpur, 43000 Kajang, Malaysia*
[2] *Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Tembila Campus, 22200 Besut, Malaysia*
[3] *Department of Mathematics, Faculty of Science and Technology, Universiti Malaysia Terengganu (UMT), 21030 Kuala Terengganu, Malaysia*
[4] *Department of Mathematics, Faculty of Science, Universiti Putra Malaysia (UPM), 43400 Serdang, Malaysia*

Correspondence should be addressed to Mohd Asrul Hery Ibrahim; asrulhery@iukl.edu.my

Received 22 April 2013; Accepted 23 January 2014; Published 4 March 2014

Academic Editor: Lucas Jodar

In solving large scale problems, the quasi-Newton method is known as the most efficient method in solving unconstrained optimization problems. Hence, a new hybrid method, known as the BFGS-CG method, has been created based on these properties, combining the search direction between conjugate gradient methods and quasi-Newton methods. In comparison to standard BFGS methods and conjugate gradient methods, the BFGS-CG method shows significant improvement in the total number of iterations and CPU time required to solve large scale unconstrained optimization problems. We also prove that the hybrid method is globally convergent.

1. Introduction

The unconstrained optimization problem only requires the objective function as

$$\min_{x\in R^n} f(x), \tag{1}$$

where R^n is an n-dimensional Euclidean space and $f : R^n \to R$ is continuously differentiable. The iterative methods are used to solve (1). On the ith iteration, an approximation point x_i and the $(i+1)$th iteration are given by

$$x_{i+1} = x_i + \alpha_i d_i, \tag{2}$$

where d_i denotes the search direction and α_i denotes the step size. The search direction must satisfy the relation $g_i^T d_i < 0$, which guarantees that d_i is a descent direction of $f(x)$ at x_i. The different choices of d_i and α_i yield the different convergence properties. Generally the first order condition $\nabla f(x_*) = 0$ is used to check for local convergence to stationary point x_*. There are many ways to calculate the search direction depending on the method used, such as the steepest descent method, conjugate gradient (CG) method, Newton-Raphson method, and quasi-Newton method.

The different choices of the step size ensure that the sequence of iterates x_i defined by (2) is globally convergent with some rates of convergence. There are two ways to determine the values of the step size, the exact line search, and the inexact line search. For the exact line search, α_i is calculated by using the formula $\alpha_i = \operatorname{argmin}_{a>0}(f(x_i + \alpha_i d_i))$. However, it is difficult and often impossible to find the value of step size in practical computation using the exact line search. Hence, the inexact line search is proposed by previous researchers like Armijo [1], Wolfe [2, 3], and Goldstein [4] to overcome the problem. Recently Shi proposed a new inexact line search rule similar to the Armijo line search and analysed the global converge [5]. Shi also claimed that among several well-known inexact line search procedures published by previous researchers, the Armijo line search rule is one of the most useful and the easiest to be implemented in computational calculations. The Armijo line search rule can be described as follows:

$$\begin{gathered}\text{Given } s > 0, \quad \beta \in (0,1), \quad \sigma \in (0,1), \\ \alpha_i = \max\left\{s, s\beta, s\beta^2, \ldots\right\} \quad \text{such that} \\ f(x_i) - f(x_i + \alpha_i d_i) \geq -\sigma\alpha_i g_i^T d_i,\end{gathered} \tag{3}$$

$i = 0, 1, 2, \ldots$. Then, the sequence of $\{x_i\}_{i=0}^{\infty}$ is converged to the optimal point, x^*, which minimises f [6]. Hence, we will use the Armijo line search in this research associated with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method and the new hybrid method.

This paper is organised as follows. In Section 2, we elaborate the step size and search direction that are used in this research. Here, the BFGS method and CG method also will be presented. Then, the new hybrid method and convergence analysis will be discussed in Section 3. An explanation about the numerical results is provided in Section 4 and the paper ends with a short conclusion in Section 5.

2. The Search Direction

The different methods in solving unconstrained optimization problems depend on the calculation of search direction, d_i in (2). In this paper, we will focus on the CG method and quasi-Newton methods. The CG method is useful for finding the minimum value of functions or unconstrained optimization problems, which are introduced by [7]. The search direction of the CG method is

$$d_i = \begin{cases} -g_i, & i = 0, \\ -g_i + \beta_i d_{i-1}, & i \geq 2, \end{cases} \tag{4}$$

where $g_i = \nabla f(x_i)$ and β_i is known as the CG coefficient. There are many ways to calculate β_i and some well-known formulas are

$$\begin{gathered}\beta_i^{\mathrm{FR}} = \frac{g_i^T g_i}{\|g_{i-1}\|^2}, \\ \beta_i^{\mathrm{PR}} = \frac{g_i^T (g_i - g_{i-1})}{\|g_{i-1}\|^2}, \\ \beta_i^{\mathrm{HS}} = \frac{g_i^T (g_i - g_{i-1})}{(g_i - g_{i-1})^T d_{i-1}},\end{gathered} \tag{5}$$

where g_i and g_{i-1} are gradients of $f(x)$ at points x_i and x_{i-1}, respectively, while $\|\cdot\|$ is a norm of vectors and d_{i-1} is a search direction for the previous iteration. The above corresponding coefficients are known as Fletcher-Reeves (CG-FR) [7], Polak-Ribière (CG-PR) [8–11], and Hestenes-Stiefel (CG-HS) [12].

In quasi-Newton methods, the search direction is the solution of linear system

$$d_i = -H_i g_i, \tag{6}$$

where H_i is an approximation of Hessian. Initial matrix H_0 is chosen by the identity matrix, which subsequently updates by an update formula. There are a few update formulas that are widely used like Davidon-Fletcher-Powell (DFP), BFGS, and Broyden family formula. This research uses a BFGS formula in a classical algorithm and the new hybrid method. The update formula for BFGS is

$$H_{i+1} = H_i - \frac{H_i s_i s_i^T H_i}{s_i^T H_i s_i} + \frac{y_i y_i^T}{s_i^T y_i}, \tag{7}$$

with $s_i = x_i - x_{i-1}$ and $y_i = g_i - g_{i-1}$. The approximation that the Hessian must fulfil is

$$H_{i+1} s_i = y_i. \tag{8}$$

This condition is required to hold for the updated matrix H_{i+1}. Note that it is only possible to fulfil the secant equation if

$$s_i^T y_i > 0, \tag{9}$$

which is known as the curvature condition.

3. The New Hybrid Method

The modification of the quasi-Newton method based on a hybrid method has already been introduced by previous researchers. One of the studies is a hybridization of the quasi-Newton and Gauss-Seidel methods, aimed at solving the system of linear equations in [13]. Luo et al. [14] suggest the new hybrid method, which can solve the system of nonlinear equations by combining the quasi-Newton method with chaos optimization. Han and Neumann [6] combine the quasi-Newton methods and Cauchy descent method to solve unconstrained optimization problems, which is recognised as the quasi-Newton-SD method.

Hence, the modification of the quasi-Newton method by previous researchers spawned the new idea of hybridizing the classical method to yield the new hybrid method. Hence, this study proposes a new hybrid search direction that combines the concept of search direction of the quasi-Newton and CG methods. It yields a new search direction of the hybrid method which is known as the BFGS-CG method. The search direction for the BFGS-CG method is

$$d_i = \begin{cases} -H_i g_i, & i = 0, \\ -H_i g_i + \eta\left(-g_i + \beta_i d_{i-1}\right), & i \geq 1, \end{cases} \tag{10}$$

where $\eta > 0$ and $\beta_i = (g_i^T g_{i-1} / g_i^T d_{i-1})$.

Hence, the complete algorithms for the BFGS method, CG-HS, CG-PR, and CG-FR methods, and the BFGS-CG method will be arranged in Algorithms 1, 2, and 3, respectively.

Algorithm 1 (*BFGS method*). States the following.

Step 0. Given a starting point x_0 and $H_0 = I_n$, choose values for s, β, and, σ and set $i = 1$.

Step 1. Terminate if $\|g(x_{i+1})\| < 10^{-6}$ or $i \geq 10000$.

Step 2. Calculate the search direction by (6).

Step 3. Calculate the step size α_i by (3).

Step 4. Compute the difference between $s_i = x_i - x_{i-1}$ and $y_i = g_i - g_{i-1}$.

Step 5. Update H_{i-1} by (7) to obtain H_i.

Step 6. Set $i = i + 1$ and go to Step 1.

Algorithm 2 (*CG-HS, CG-PR, and CG-FR*). States the following.

Step 0. Given a starting point x_0, choose values for s, β, and σ and set $i = 1$.

Step 1. Terminate if $\|g(x_{k+1})\| < 10^{-6}$ or $i \geq 10000$.

Step 2. Calculate the search direction by (4) with respect to the coefficient of CG.

Step 3. Calculate the step size α_i by (3).

Step 4. Compute the difference between $s_i = x_i - x_{i-1}$ and $y_i = g_i - g_{i-1}$.

Step 5. Set $i = i + 1$ and go to Step 1.

Algorithm 3 (*BFGS-CG method*). States the following.

Step 0. Given a starting point x_0 and $H_0 = I_n$, choose values for s, β, and σ and set $i = 1$.

Step 1. Terminate if $\|g(x_{i+1})\| < 10^{-6}$ or $i \geq 10000$.

Step 2. Calculate the search direction by (10).

Step 3. Calculate the step size α_i by (3).

Step 4. Compute the difference between $s_i = x_i - x_{i-1}$ and $y_i = g_i - g_{i-1}$.

Step 5. Update H_{i-1} by (7) to obtain H_i.

Step 6. Set $i = i + 1$ and go to Step 1.

Based on Algorithms 1, 2, and 3 we assume that every search direction d_i satisfied the descent condition

$$g_i^T d_i < 0, \tag{11}$$

for all $i \geq 0$. If there exists a constant $c_1 > 0$ such that

$$g_i^T d_i \leq c_1 \|g_i\|^2 \tag{12}$$

for all $i \geq 0$, then the search directions satisfy the sufficient descent condition which can be proved in Theorem 6. Hence, we need to make a few assumptions based on the objective function.

Assumption 4. Consider the following.

H1: the objective function f is twice continuously differentiable.

H2: the level set L is convex. Moreover, positive constants c_1 and c_2 exist, satisfying

$$c_1 \|z\|^2 \leq z^T F(x) z \leq c_2 \|z\|^2, \tag{13}$$

for all $z \in R^n$ and $x \in L$, where $F(x)$ is the Hessian matrix for f.

H3: the Hessian matrix is Lipschitz continuous at the point x^*; that is, there exists the positive constant c_3 satisfying

$$\|g(x) - g(x^*)\| \leq c_3 \|x - x^*\| \tag{14}$$

for all x in a neighbourhood of x^*.

Theorem 5 (see [15, 16]). *Let $\{B_i\}$ be generated by the BFGS formula* (8)*, where B_1 is symmetric and positive definite, and $y_i^T s_i > 0$ for all i. Furthermore, assume that $\{s_i\}$ and $\{y_i\}$ are such that*

$$\frac{\|(y_i - G_*) s_i\|}{\|s_i\|} \leq \varepsilon_i \tag{15}$$

for some symmetric and positive definite matrix $G(x^*)$ and for some sequence $\{\varepsilon_i\}$ with the property $\sum_{i=1}^{\infty} \varepsilon_i < \infty$. Then

$$\lim_{i \to \infty} \frac{\|(B_i - G_*) d_i\|}{\|d_i\|} = 0 \tag{16}$$

and the sequence $\{\|B_i\|\}$, $\{\|B_i^{-1}\|\}$ are bound.

Theorem 6. *Suppose that Assumption 4 and Theorem 5 hold. Then condition* (12) *holds for all $i \geq 0$.*

Proof. From (12), we see that

$$\begin{aligned} g_i^T d_i &= -g_i^T B_i^{-1} g_i + \eta g_i^T \left(-g_i + \left(\frac{g_i^T g_{i-1}}{g_i^T d_{i-1}} \right) d_{i-1} \right) \\ &= -g_i^T B_i^{-1} g_i + \eta \left(-g_i^T g_i + \left(\frac{g_i^T g_{i-1}}{g_i^T d_{i-1}} \right) g_i^T d_{i-1} \right) \\ &= -g_i^T B_i^{-1} g_i + \eta \left(-g_i^T g_i + g_i^T g_{i-1} \right). \end{aligned} \tag{17}$$

Based on Powell [17], $g_i^T g_{i-1} \geq \varepsilon \|g_i\|^2$ with $\varepsilon = (0, 1]$, and

$$\begin{aligned} g_i^T d_i &= -g_i^T B_i^{-1} g_i + \eta \left(-\|g_i\|^2 + \varepsilon \|g_i\|^2 \right) \\ &\leq -\lambda_i \|g_i\|^2 + (-\eta + \eta\varepsilon) \|g_i\|^2 \\ &\leq c_1 \|g_i\|^2, \end{aligned} \tag{18}$$

where $c_1 = -(\lambda_i + \eta - \eta\varepsilon)$ which is bound away from zero. Hence, $g_i^T d_i \leq c_1 \|g_i\|^2$ holds. The proof is completed. □

Lemma 7. *Under Assumption 4, positive constants ϖ_1 and ϖ_2 exist, such that for any x_i and any d_i with $g_i^T d_i < 0$, the step size a_i produced by Algorithm 2 will satisfy either*

$$f(x_i + \alpha_i d_i) - f_i \leq -\varpi_1 \frac{(g_i^T d_i)^2}{\|d_i\|^2} \tag{19}$$

or

$$f(x_i + \alpha_i d_i) - f_i \leq \varpi_1 g_i^T d_i. \tag{20}$$

Proof. Suppose that $a_i < 1$, which means that (3) failed for a step size $a' \leq a_i/\tau$:

$$f(x_i + \alpha_i' d_i) - f(x_i) \leq \varpi a' g_i^T d_i. \tag{21}$$

Then, by using the mean value theorem, we obtain

$$f(x_{i+1}) - f(x_i) = \overline{g}^T (x_{i+1} - x_i), \tag{22}$$

where $\overline{g} = \nabla f(\overline{x})$, for some $\overline{x} \in (x_i, x_{i+1})$. Now, by the Cauchy-Schwartz inequality, we get

$$\begin{aligned}
\overline{g}^T (x_{i+1} - x_i) &= g^T (x_{i+1} - x_i) + (\overline{g} - g_i)^T (x_{i+1} - x_i) \\
&= g^T (x_{i+1} - x_i) + \|\overline{g} - g_i\| (x_{i+1} - x_i) \\
&\leq g^T (x_{i+1} - x_i) + \mu \|x_{i+1} - x_i\|^2 \\
&\leq g^T (a' d_i) + \mu \|a' d\|^2 \\
&\leq g^T (a' d_i) + \mu (a' \|d\|)^2.
\end{aligned} \tag{23}$$

Thus, from H3

$$(\varpi - 1) a' g_i^T d_i < a' (\overline{g} - g_i)^T d_i \leq M (a' \|d_i\|)^2, \tag{24}$$

which implies that

$$a_i \geq \tau a' > \tau (1 - \varpi) \frac{-g_i^T d_i}{M (a' \|d_i\|)^2}. \tag{25}$$

Substituting this into (21), we have

$$f(x_i + \alpha_i' d_i) - f(x_i) \leq c_2 \frac{-g_i^T d_i}{(a' \|d_i\|)^2}, \tag{26}$$

where $c_2 = \tau(1 - \varpi)/M$, which gives (19). □

Theorem 8 (global convergence). *Suppose that Assumption 4 and Theorem 5 hold. Then*

$$\lim_{i \to \infty} \|g_i\|^2 = 0. \tag{27}$$

Proof. Combining descent property (12) and Lemma 7 gives

$$\sum_{i=0}^{\infty} \frac{\|g_i\|^4}{\|d_i\|^2} < \infty. \tag{28}$$

Hence, from Theorem 6, we can define that $\|d_i\| \leq -c\|g_i\|$. Then, (28) will be simplified as $\sum_{i=0}^{\infty} \|g_i\|^2 < \infty$. Therefore, the proof is completed. □

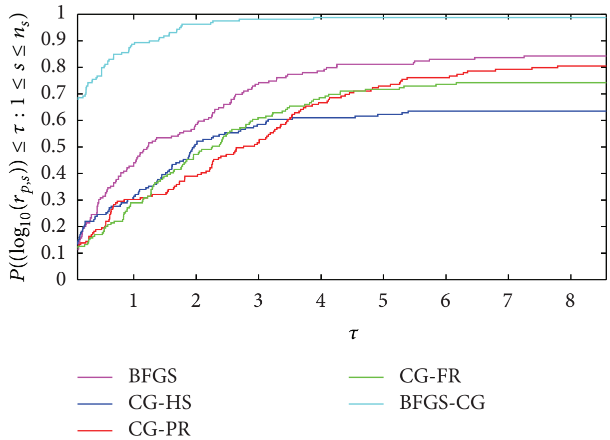

Figure 1: Performance profile in a $\log_{10}$ scale based on iteration.

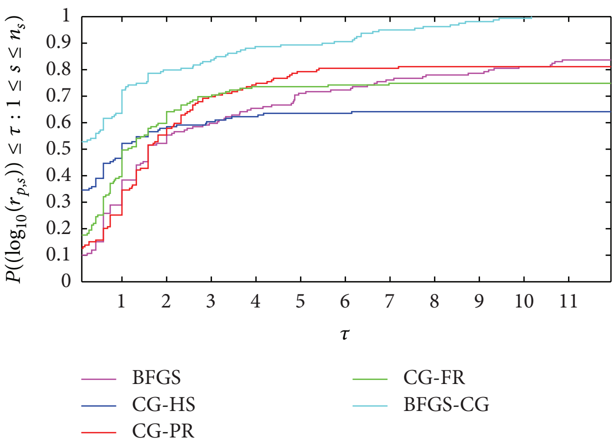

Figure 2: Performance profile in a $\log_{10}$ scale based on CPU time.

4. Numerical Result

In this section, we use the test problem considered by Andrei [18], Michalewicz [19], and Moré et al. [20] in Table 1 to analyse the improvement of the BFGS-CG method compared with the BFGS method and CG method. Each of the test problems is tested with dimensions varying from 2 to 1,000 variables. This represents a total of 159 test problems. As suggested by [20], for each of the test problems, the initial point x_0 will further subtract from the minimum point. In doing so, this leads us to test the global convergence properties and the robustness of our method. For the Armijo line search, we use $s = 1$, $\beta = 0.5$, and $\sigma = 0.1$. The stopping criteria we use are $\|g_i\| \leq 10^{-6}$ and the number of iterations exceeds its limit, which is set to be 10,000. In our implementation, the numerical tests were performed on an Acer Aspire with a Windows 7 operating system and using Matlab 2012 languages.

The performance results will be shown in Figures 1 and 2, respectively, using the performance profile introduced by Dolan and Moré [21]. The performance profile seeks to find how well the solvers perform relative to the other solvers on

Table 1: A list of problem functions.

Test problem	n-dimensional	Sources
Powell badly scaled	2	Moré et al. [20]
Beale	2	Moré et al. [20]
Biggs Exp 6	6	Moré et al. [20]
Chebyquad	4, 6	Moré et al. [20]
Colville polynomial	4	Michalewicz [19]
Variably dimensioned	4, 8	Moré et al. [20]
Freudenstein and Roth	2	Moré et al. [20]
Goldstein price polynomial	2	Michalewicz [19]
Himmelblau	2	Andrei [18]
Penalty 1	2, 4	Moré et al. [20]
Extended Powell singular	4, 8	Moré et al. [20]
Extended Rosenbrock	2, 10, 100, 200, 500, 1000	Andrei [18]
Trigonometric	6	Andrei [18]
Watson	4, 8	Moré et al. [20]
Six-hump camel back polynomial	2	Michalewicz [19]
Extended shallow	2, 4, 10, 100, 200, 500, 1000	Andrei [18]
Extended strait	2, 4, 10, 100, 200, 500, 1000	Andrei [18]
Scale	2	Michalewicz [19]
Raydan 1	2, 4	Andrei [18]
Raydan 2	2, 4	Andrei [18]
Diagonal 3	2	Andrei [18]
Cube	2, 10, 100, 200	Moré et al. [20]

a set of problems. In general, $P(\tau)$ is the fraction of problems with performance ratio τ; thus, a solver with high values of $P(\tau)$ or one that is located at the top right of the figure is preferable.

Figures 1 and 2 show that the BFGS-CG method has the best performance since it can solve 99% of the test problems compared with the BFGS (84%), CG-HS (65%), CG-PR (80%), and CG-FR (75%) methods. Moreover, we can also say that the BFGS-CG is the fastest solver on approximately 68% of the test problems for iteration and 52% of CPU time.

5. Conclusion

We have presented a new hybrid method for solving unconstrained optimization problems. The numerical results for a broad class of test problems show that the BFGS-CG method is efficient and robust in solving the unconstrained optimization problem. We also note that, as the size and complexity of the problem increase, greater improvements could be realised by our BFGS-CG method. Our future research will be to try the BFGS-CG method with coefficients of CG like Fletcher-Reeves, Hestenes-Stiefel, and Polak-Ribiére.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research was supported by Fundamental Research Grant Scheme (FRGS Vote no. 59256).

References

[1] L. Armijo, "Minimization of functions having Lipschitz continuous first partial derivatives," *Pacific Journal of Mathematics*, vol. 16, pp. 1–3, 1966.

[2] P. Wolfe, "Convergence conditions for ascent methods," *SIAM Review*, vol. 11, no. 2, pp. 226–235, 1969.

[3] P. Wolfe, "Convergence conditions for ascent methods. II: some corrections," *SIAM Review*, vol. 13, no. 2, pp. 185–188, 1971.

[4] A. A. Goldstein, "On steepest descent," *Journal of the Society for Industrial and Applied Mathematics A*, vol. 3, no. 1, pp. 147–151, 1965.

[5] Z.-J. Shi, "Convergence of quasi-Newton method with new inexact line search," *Journal of Mathematical Analysis and Applications*, vol. 315, no. 1, pp. 120–131, 2006.

[6] L. Han and M. Neumann, "Combining quasi-Newton and Cauchy directions," *International Journal of Applied Mathematics*, vol. 12, no. 2, pp. 167–191, 2003.

[7] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *The Computer Journal*, vol. 7, no. 2, pp. 149–154, 1964.

[8] N. Andrei, "Accelerated scaled memoryless BFGS preconditioned conjugate gradient algorithm for unconstrained optimization," *European Journal of Operational Research*, vol. 204, no. 3, pp. 410–420, 2010.

[9] E. Polak and G. Ribière, "Note on the convergence of methods of conjugate directions," *Revue Française d'Informatique et de Recherche Opérationnelle*, vol. 3, pp. 35–43, 1969.

[10] Z.-J. Shi and J. Shen, "Convergence of the Polak-Ribiére-Polyak conjugate gradient method," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 66, no. 6, pp. 1428–1441, 2007.

[11] G. Yu, L. Guan, and Z. Wei, "Globally convergent Polak-Ribière-Polyak conjugate gradient methods under a modified Wolfe line search," *Applied Mathematics and Computation*, vol. 215, no. 8, pp. 3082–3090, 2009.

[12] M. R. Hestenes and E. Stiefel, "Method of conjugate gradient for solving linear equations," *Journal of Research of the National Bureau of Standards*, vol. 49, no. 6, pp. 409–436, 1952.

[13] A. Ludwig, "The Gauss-Seidel-quasi-Newton method: a hybrid algorithm for solving dynamic economic models," *Journal of Economic Dynamics and Control*, vol. 31, no. 5, pp. 1610–1632, 2007.

[14] Y.-Z. Luo, G.-J. Tang, and L.-N. Zhou, "Hybrid approach for solving systems of nonlinear equations using chaos optimization and quasi-Newton method," *Applied Soft Computing*, vol. 8, no. 2, pp. 1068–1073, 2008.

[15] R. H. Byrd and J. Nocedal, "A tool for the analysis of quasi-Newton methods with application to unconstrained minimization," *SIAM Journal on Numerical Analysis*, vol. 26, no. 3, pp. 727–739, 1989.

[16] R. H. Byrd, J. Nocedal, and Y.-X. Yuan, "Global convergence of a class of quasi-Newton methods on convex problems," *SIAM Journal on Numerical Analysis*, vol. 24, no. 5, pp. 1171–1191, 1987.

[17] M. J. D. Powell, "Restart procedures for the conjugate gradient method," *Mathematical Programming*, vol. 12, no. 1, pp. 241–254, 1977.

[18] N. Andrei, "An unconstrained optimization test functions collection," *Advanced Modeling and Optimization*, vol. 10, no. 1, pp. 147–161, 2008.

[19] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, Berlin, Germany, 1996.

[20] J. J. Moré, B. S. Garbow, and K. E. Hillstrom, "Testing unconstrained optimization software," *ACM Transactions on Mathematical Software*, vol. 7, no. 1, pp. 17–41, 1981.

[21] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," *Mathematical Programming*, vol. 91, no. 2, pp. 201–213, 2002.