Mathematical Analysis
& Convex Optimization

Vol. 3 (2022), No. 2, 69–76
https:\\ maco.lu.ac.ir
DOI: 10.52547/maco.3.2.7

Research Paper

A NEW MODIFIED LINE SEARCH ALGORITHM TO SOLVE LARGE-SCALE NON-SMOOTH NON-CONVEX OPTIMIZATION PROBLEM

SAEED BANIMEHRI AND HAMID ESMAEILI*

Abstract. In this paper, a new modified line search Armijo is used in the diagonal discrete gradient bundle method to solve large-scale non-smooth optimization problems. The new principle causes the step in each iteration to be longer, which reduces the number of iterations, evaluations, and the computational time. In other words, the efficiency and performance of the method are improved. We prove that the diagonal discrete gradient bundle method converges with the proposed monotone line search principle for semi-smooth functions, which are not necessarily differentiable or convex. In addition, the numerical results confirm the efficiency of the proposed correction.

MSC(2010): 90C26; 49J52; 90C56.
Keywords: Non-smooth optimization, Derivative-free optimization, Diagonal discrete gradient bundle method, line search.

1. Introduction and Background

In this paper, we are considering the non-smooth optimization problem of the form

$$\begin{aligned} &\min \quad && f(x) \\ &s.t. \quad && x \in R^n, \end{aligned} \tag{1.1}$$

where the objective function $f : R^n \to R$ is supposed to be semi-smooth and the number of variables n is supposed to be large. Note that no differentiability or convexity assumptions for problem (1.1)are made. Non-smooth optimization problems are encountered in many usage areas; for instance, they are used in economics mechanics, engineering, control theory, optimal shape design, machine learning and data mining, for cluster analysis and classification (see [2])these problems are applied in a large scale.

The diagonal discrete gradient bundle method [11] is a derivative-free method for the large-scale non-smooth optimization. The diagonal discrete gradient bundle method combines discrete gradient method [1] and the diagonal bundle method [12]. The discrete gradient method is a derivative-free method for the small-scale non-smooth optimization, while the diagonal bundle method is a successor of the limited memory bundle method utilizing subgradient information to solve large-scale non-smooth optimization problems.

To solve optimization problems, we use the line search technique to find the step length. Exact line search methods to calculate step length can be expensive and time consuming.

Date: Received: May 25, 2022, Accepted: September 24, 2022.

*Corresponding author.

Therefore, some inexact line search techniques [4] namely the Armijo technique, the Wolfe technique and the Goldstein technique have been proposed to determine an acceptable step length t_k. Monotone line search method is a new approach to determine the step length in optimization problems. This method reduces the line search range to find the largest step length in each iteration and avoids being confined to a narrow valley as much as possible [4, 7].

Of the problems with the diagonal discrete gradient bundle method, one is that it may not work for some problems. Our purpose in this article is to modify the above method. In the diagonal discrete gradient bundle method, Armijo line search method is used to calculate the step length. In this paper, we present a monotone line search algorithm to solve non-smooth optimization problems. In this algorithm, we combine a monotone strategy with the modified Armijo rule and design a new algorithm that will probably choose a larger step length at each iteration. This strategy may reduce the number of iterates, time and function evaluations and can improve the efficiency of the new approach. Numerical results show that the new approach to solving non-smooth optimization problems is robust and efficient.

The paper is organized as follows. In Section 2, we introduce the diagonal discrete gradient bundle method. In Section 3, we describe a new monotone Armijo line search algorithm and present its convergence properties. Section 4 shows numerical results of the algorithm. Finally, Section 5 concludes the paper.

2. Diagonal discrete gradient bundle method

In this section, we introduce the diagonal discrete gradient bundle method, which uses the ideas of the variable metric bundle method [8] to calculate the null step, simple aggregate, and the subgradient locality measure. The diagonal discrete gradient bundle method uses discrete gradients instead of subgradients in our calculations and the search direction d_k is calculated using the diagonal variable metric update as

$$d_k = -D_k v_k$$

where v_k is discrete gradient and D_k is the diagonal variable metric update.

In order to specify a new step into the search direction d_k, the diagonal discrete gradient bundle method uses the so-called armijo line search (see [9, 17]) for: a new iteration point x_{k+1} and a new auxiliary point y_{k+1} created such that

$$x_{k+1} = x_k + t_L^k d_k \quad and \quad y_{k+1} = x_k + t_R^k d_k \quad \forall\, k \geq 1,$$

with $y_1 = x_1$, where $t_R^k \in (0, t_{\max}]$ and $t_L^k \in \left[0, t_R^k\right]$ are step sizes, and $t_{\max} > 1$ is the upper bound for the step size. A necessary condition for a serious step is to have

$$t_L^k = t_R^k > 0 \quad and \quad f(x_{k+1}) \leq f(x_k) - \varepsilon_L^k t_R^k w_k \tag{2.1}$$

where $\varepsilon_L^k \in \left(0, \frac{1}{2}\right)$ is a line search parameter and $w_k > 0$ represents the acceptable amount of descent of f at x_k. If (2.1) is satisfied, we set $x_{k+1} = y_{k+1}$ and a serious step is taken. On the other hand, a null step is taken if

$$t_R^k > t_L^k = 0 \quad and \quad -\beta_{k+1} + d_k^T v_{k+1} \geq -\varepsilon_R^k w_k,$$

where $\varepsilon_R^k \in \left(\varepsilon_L^k, \frac{1}{2}\right)$ is a line search parameter and $v_{k+1} \in V_0\left(y_{k+1}, \zeta_k\right)$ (the closed convex set of discrete gradients $V_0\left(y_{k+1}, \zeta_k\right)$ is an approximation to the subdifferential $\partial f\left(y_{k+1}\right)$ for sufficiently small $\zeta_k > 0$). Moreover, β_{k+1} is analogous to the subgradient locality measure [13, 14] used in standard bundle methods, that is

$$\beta_{k+1} = \max\left\{\left|f(x_k) - f(y_{k+1}) + (y_{k+1} - x_k)^T v_{k+1}\right|, \gamma\|y_{k+1} - x_k\|^2\right\}.$$

Here, $\gamma > 0$ is a distance measure parameter supplied by the user. In the case of a null step, we set $x_{k+1} = x_k$ but information about the objective function is increased because we store the auxiliary point y_{k+1} and the corresponding auxiliary discrete gradient $v_{k+1} \in V_0(y_{k+1}, \zeta_k)$. The diagonal discrete gradient bundle method uses the original discrete gradient v_k after the serious step and the aggregate subgradient $\tilde{v}_k$ after the null step for direction finding (i.e. we set $v_k = \tilde{v}_k$ if the previous step is a serious step). The aggregation procedure is carried out by determining multipliers λ_i^k satisfying $\lambda_i^k > 0$ for all $i = \{1, 2, 3\}$ and $\sum_{i=1}^{3} \lambda_i^k = 1$ that minimize a simple quadratic function

$$\varphi(\lambda_1, \lambda_2, \lambda_3) = [\lambda_1 v_m + \lambda_2 v_{k+1} + \lambda_3 \tilde{v}_k]^T D^k [\lambda_1 v_m + \lambda_2 v_{k+1} + \lambda_3 \tilde{v}_k] + 2\left(\lambda_2 \beta_{k+1} + \lambda_3 \tilde{\beta}_k\right).$$

Here, $v_m \in V_0(x_k, \zeta_k)$ is the current discrete gradient, $v_{k+1} \in V_0(y_{k+1}, \zeta_k)$ is the auxiliary discrete gradient, and $\tilde{v}_k$ is the current aggregate discrete gradient from the previous iteration ($\tilde{v}_1 = v_1$). In addition, β_{k+1} is the current subgradient locality measure and $\tilde{\beta}_k$ is the current aggregate subgradient locality measure ($\tilde{\beta}_1 = 0$) (see [17]).

The resulting aggregate discrete gradient $\tilde{v}_{k+1}$ and aggregate subgradient locality measure $\tilde{\beta}_{k+1}$ are computed by

$$\tilde{v}_{k+1} = \lambda_1^k v_m + \lambda_2^k v_{k+1} + \lambda_3^k \tilde{v}_k \quad \textit{and} \quad \tilde{\beta}_{k+1} = \lambda_2^k \beta_{k+1} + \lambda_3^k \tilde{\beta}_k$$

Due to this simple aggregation procedure, only one trial point y_{k+1} and the corresponding discrete gradient $v_{k+1} \in V_0(y_{k+1}, \zeta_k)$ need to be stored.

We need to consider how to update the matrix D_k and, at the same time, to find the search direction d_k The basic idea in direction finding is the same as that with the limited memory bundle method. However, due to the usage of null steps some modifications similar to the variable metric bundle methods has to be made: If the previous step is a null step, the matrix D_k is formed by using the limited memory $SR1$ update (see [12]). This update formula gives us a possibility to preserve the boundedness and some other properties of generated matrices that are required in the proof of global convergence.

The stopping parameter w_k at iteration k is defined by

$$w_k = -\tilde{v}_k^T d_k + 2\tilde{\beta}_k$$

and the algorithm stops if $w_k < \epsilon$ for some user specified $\epsilon > 0$. The parameter w_k is also used during the line search procedure to represent the desirable amount of descent. (See [11] for more details on the diagonal discrete gradient bundle method)

3. **Modified Armijo line search conditions**

More practical strategies perform an inexact line search to identify a steplength that achieves adequate reductions in f at minimal cost. These strategies choose the steplength t_k guaranteeing a sufficient reduction in function values while this might induce the overall algorithm to converge. Some conditions proposed for acceptance of steplength t_k, namely the Armijo, Wolfe and Goldstein conditions [12, 4]. A practical and common criterion for terminating linear search is the Armijo condition. This condition is to control a sufficient decrease in the objective function and the basic idea of the Armijo condition is to guarantee that the chosen steplength value is not too large.

We use the condition

$$f\left(x_{k+1}\right) \leq f\left(x_k\right) + \varepsilon_1 t_R^k w_k + \theta_1 h\left(t_R^k, d_k\right) \tag{3.1}$$

to produce a serious step, where t_R^k is the largest number in $\left\{s, \rho s, \rho^2 s, ...\right\}$, with $\rho \in (0,1)$ and

$$s_k = -\frac{v_k^T d_k}{d_k^T B_k d_k}.$$

In condition(3.1), $h\left(t_R^k, d_k\right)$ is obtained from the following equation:

$$h\left(t_R, d\right) = \begin{cases} 0 \quad if \quad t_R = 0 \\ \exp^{-\left(\frac{t_R^2\|d\|^2}{2}\right)} \quad if \quad t_R > 0 \end{cases} \tag{3.2}$$

is a function and

$$0 < \varepsilon_2 \leq \varepsilon_1 < \frac{1}{2} \quad and \quad \varepsilon_2 + \varepsilon_1 < \frac{1}{2} \quad and \quad 0 < \theta_1 < 1 \tag{3.3}$$

are auxiliary parameters.

We can see the right-hand side of the approach is greater than the right-hand side of the standard Armijo rule, so a larger steplength is possible for the algorithm to gain. These changes may reduce the number of iterations and function evaluations for attaining the same optimum.

3.1. **modified Armijo line search Procedure. Initial.** Consider positive parameters $\varepsilon_A, \varepsilon_L, \varepsilon_R, \varepsilon_T$ satisfying $\varepsilon_T + \varepsilon_A < \varepsilon_R < \frac{1}{2}$ and $\varepsilon_L < \varepsilon_T$, distance measure parameter $\gamma > 0$, an interpolation parameter $\kappa \in \left(0, \frac{1}{2}\right)$ and $\theta_1 \in (0,1)$. All of these parameters are constant.

Step i . Set $t_A = 0$ and $t = t_U$.

Step ii . Calculate $f\left(x_k + td_k\right)$, $v_{k+1} \in V_0\left(x_{k+1}, \zeta\right)$ and
$\beta_{k+1} = \max\left\{\left|f\left(x_k\right) - f\left(y_{k+1}\right) + \left(y_{k+1} - x_k\right)^T v_{k+1}\right|, \gamma\|y_{k+1} - x_k\|^2\right\}$. If $f\left(x_k + td_k\right) \leq f\left(x_k\right) + \varepsilon_1 t_R^k w_k + \theta_1 h\left(t_R^k, d_k\right)$ set $t_A = t$; otherwise, set $t_U = t$.

Step iii (**serious step**). If $f\left(x_k + td_k\right) \leq f\left(x_k\right) + \varepsilon_1 t_R^k w_k + \theta_1 h\left(t_R^k, d_k\right)$ set $t_R = t_L = t$ and return.

Step iv (**Null step**). If $-\beta_{k+1} + d_k^T v_{k+1} \geq -\varepsilon_R^k w_k$, set $t_R = t$, $t_L = 0$ and return.

Step v. Choose $t \in \left[t_L + \kappa\left(t_U - t_L\right), t_U - \kappa\left(t_U - t_L\right)\right]$ by some interpolation procedure, and go to **Step ii**.

Lemma 3.1. *Suppose that the sequence $\{x_k\}$ is generated by modified Armijo line search. If $\tilde{t}$ and t are steplengths which satisfy in the standard Armijo rule and modified Armijo rule, respectively, then $\tilde{t} \leq t$ and new Armijo rule is well-defined.*

Proof. If $\tilde{t}$ and t are the steplengths which satisfy in the standard Armijo rule and the new Armijo-type line search method, respectively, then we have

$$f\left(x_k + \widetilde{t}_k d_k\right) - f\left(x_k\right) \leq \varepsilon_1 \widetilde{t}_R^k w_k \leq \varepsilon_1 \widetilde{t}_R^k w_k + \theta_1 \widetilde{t}_R^k h\left(t_R, d\right). \tag{3.4}$$

This implies that $\tilde{t} \leq t$. Now, by Taylor's theorem

$$\begin{aligned}
&\lim_{t\to 0^+} \frac{f(x_k) - f(x_k + td_k) + t(-\varepsilon_1 w_k + \theta_1 h(t_R, d))}{t} \\
&= \lim_{t\to 0^+} \frac{f(x_k) - \left(f(x_k) + tv_k^T d_k + o(t\|d_k\|)\right) + t(-\varepsilon_1 w_k + \theta_1 h(t_R, d))}{t} \\
&= -v_k^T d_k - \varepsilon_1 w_k + \theta_1 h(t_R, d) > 0.
\end{aligned} \tag{3.5}$$

So, there exists a $\widehat{t}_k > 0$ such that

$$f(x_k + td_k) \leq f(x_k) + t(-\varepsilon_1 w_k + \theta_1 h(t_R, d)) \quad \forall t \in \left[0, \widehat{t}_k\right] \tag{3.6}$$

Therefore, the new Armijo line search is well-defined. □

3.2. **Convergence analysis.** In this section, we show the global convergence of the diagonal discrete gradient bundle algorithm. The convergence of diagonal discrete gradient bundle algorithm is described in [11]. In [**?**], it is shown that the monotone line search method is well-defined. We will continue to show that the monotone line search procedure terminates in a finite number of iterations. First, the monotone line search procedure has been proved to be finite under the assumption of upper semi-smoothness when subgradients are used.

Lemma 3.2. *Let f satisfy the following semi-smoothness hypothesis. For any $x \in R^n$, $d \in R^n$ and sequences $\left\{\hat{t}_i\right\} \subset R_+$ and $\{\hat{g}_i\} \subset R^n$ satisfying $\hat{t}_i \downarrow 0$ and $\hat{g}_i \in \partial f\left(x_k + \hat{t}_i d\right)$, one has*

$$\lim_{i\to\infty} \sup\ \hat{g}_i^T d \geq \lim_{i\to\infty} \inf\ \frac{f\left(x_k + \hat{t}_i d\right) - f(x_k)}{\hat{t}_i} \tag{3.7}$$

Then, the monotone line search procedure terminates in a finite number of iterations.

Proof. The proof becomes identical to the proof of Theorem 3.6 in [17]. □

The set of discrete gradients is an approximation to the subdifferential if the function is semi-smooth. Since the class of semi-smooth functions includes the class of upper semi-smooth functions, we here assume that the objective function f is semi-smooth. Now, due to assumption of semi-smoothness and subgradient, the monotone Armijo line search procedure is also finite when subgradients are replaced with discrete gradients.

4. Numerical experiments

As already said, the test set used in our experiments consists of extensions of classical academic non-smooth minimization problems from the literature. That is, problems $1-8$ were first introduced in [10]. These problems can be formulated with any number of variables. Note that in the computation of both the Armijo line search and monotone Armijo line search, more than n function evaluations are needed for each iteration. Here, we examine problems with dimensions of 50, 200 and 1000 variables.We perform our experiments in MATLAB 8.1 programming environment.

We say that a solver finds the solution with respect to a tolerance $\epsilon > 0$ if

$$\left|\frac{f_{k+1} - f_k}{1 + f_k}\right| \leq \varepsilon$$

and

$$\left| \frac{\|x_{k+1} - x_k\|}{1 + \|x_k\|} \right| \leq \varepsilon$$

Where f_{k+1} and x_{k+1} are the values of the function and the optimal point in the current iteration, f_k and x_k are the values of the function and the optimal point in the previous iteration. We have accepted the results with respect to the tolerance $\epsilon = 10^{-3}$. For the diagonal discrete gradient bundle method, these are

$$\varepsilon_L = 10^{-4} \quad \varepsilon_R = 0.25 \quad t_{\min} = 10^{-12} \quad t_{\max} = 1000 \quad \gamma = 10^{-4} \quad \theta_1 = 0.85.$$

We put $\tau_k = \frac{F_{k-1}}{F_k}$ in the modified line search condition, where F_k is the sum of the first k sentences of the Fibonacci sequence(see [6]).Also, in this paper, multipliers λ_i^k for $i = \{1, 2, 3\}$ are calculated by the default optimization method in MATLAB.

The results are summarized in Tables $1 - 3$ where we have compared the efficiency of the conditions both in terms of the computational time and the number of function evaluations (nf , evaluations for short). In details, these results suggest that the proposed algorithm has promising behaviour encountering with medium-scale and large-scale unconstrained optimization problems and it is superior to the considered algorithm in the all cases.

Problem	**Armijo line search nf/time**	**Modified Armijo line search nf/time**
1	3201/0.077	3128/0.053
2	*Fail*	11,783/0.92
3	8132/0.513	7856/0.398
4	14,189/0.302	13,889/0.258
5	5,400/0.268	4,693/0.213
6	3,345/0.09	3,122/0.05
7	11,471/0.18	10,782/0.10
8	13,356/0.11	12,997/0.09

Table 1. Summary of the results with 50 variables

Problem	**Armijo line search nf/time**	**Modified Armijo line search nf/time**
1	42,815/1.168	41,782/0.997
2	*Fail*	52,364/2.690
3	44,654/1.638	43,541/1.489
4	113,918/10.291	112,273/5.05
5	41,104/1.663	40,963/1.528
6	279,301/5.36	252,177/4.95
7	58,965/7.235	55,236/6.589
8	51,297/1.34	49,484/1.01

Table 2. Summary of the results with 200 variables.

Problem	**Armijo line search nf/time**	**Modified Armijo line search nf/time**
1	$1,821,133/47.864$	$1,633,963/43.873$
2	$Fail$	$832,896/42.426$
3	$1,506,899/81.411$	$1,463,998/75.878$
4	$2,932,844/39.886$	$2,634,231/38.123$
5	$129,091/47.276$	$124,827/45.321$
6	$176,458/55.45$	$175,892/40.35$
7	$203,372/61.32$	$201,309/52.37$
8	$5,390,563/59.45$	$5,242,908/53.931$

TABLE 3. Summary of the results with 1000 variables

The results are summarized in table $1-3$ where we have compared the efficiency of the conditions both in terms of the computational time and the number of function evaluations (nf , evaluations for short).The phrase **Fail** indicates that the method in question is not able to solve the problem.In problem 2, the old method is not able to solve the problem, but the modified method is able to solve the problem. In details, these results suggest that the proposed algorithm has promising behaviour encountering with medium-scale and large-scale unconstrained optimization problems and it is superior to the considered algorithm in all cases. Summarizing the results of tables 1, 2 and 3 implies that modified diagonal discrete gradient bundle method is superior to the presented algorithm with respect to the number of iterations and function evaluations.

5. Conclusion

In this paper, we present a correction for the diagonal discrete gradient bundle method. In this modification, we focus on a new approach to a new monotone line search. This rule produces a larger step size, especially when the repetition is far from optimal. We proved the global convergence of this method for semi-smooth functions that are not necessarily differentiable and convex. The numerical experiments confirm the efficiency of the proposed correction compared to the diagonal discrete gradient bundle method to solve large-scale non-smooth optimization problems.

5.1. References.

References

[1] A.M. Bagirov and B. Karasözen and M. Sezer, Discrete gradient method: derivative-free method for nonsmooth optimization. *Journal of Optimization Theory and Applications*, (2): 317-334, 2008.

[2] A.M. Bagirov and N. Karmitsa and M.M. Mäkelä, *Introduction to Nonsmooth Optimization*, volume 12 of *Theory, Practice and Software*. Springer International Publishing, Switzerland, 2014.

[3] J. Barzilai and J.M. Borwein, Two-point step size gradient methods. *IMA journal of numerical analysis*, (1): 141-148, 1988.

[4] S. Bojari and M.R. Eslahchi, Global convergence of a family ofmodified BFGSmethods under a modified weak-Wolfe–Powell line search for nonconvex functions. *Scientific Reports*, (18): 219-244, 2020.

[5] Y.H. Dai and L.Z. Liao, R-linear convergence of the Barzilai and Borwein gradient method. *IMA Journal of Numerical Analysis*, (1): 1-10, 2002.

[6] R. Grimaldi, *Fibonacci and Catalan Numbers: an introduction*, John Wiley and Sons, 2012.
[7] I. Griva and S.G. Nash and A. Sofar, *Linear and nonlinear optimization.*, volume 108 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009.
[8] M. Haarala, *The Large-scale nonsmooth optimization: variable metric bundle method with limited memory*, University of Jyväskylä, 2004.
[9] N. Haarala and K. Miettinen and M.M. Mäkelä, Globally convergent limited memory bundle method for large-scale nonsmooth optimization. *Mathematical Programming*, (1): 181-205, 2007.
[10] M. Haarala and K. Miettinen and M.M. Mäkelä, New limited memory bundle method for large-scale nonsmooth optimization. *Optimization Methods and Software*, (6): 673-692, 2004.
[11] N. Karmitsa, Diagonal discrete gradient bundle method for derivative free nonsmooth optimization. *Optimization*, (8): 1599-1614, 2016.
[12] N. Karmitsa, Diagonal bundle method for nonsmooth sparse optimization. *Journal of Optimization Theory and Applications*, (3): 886-905, 2015.
[13] C. Lemarechal and J.J. Strodiot and A. Bihain, On a bundle algorithm for nonsmooth optimization. *Nonlinear programming 4*, 245-282, 1981.
[14] R. Mifflin, A modification and an extension of Lemaréchal's algorithm for nonsmooth minimization,. *Nondifferential and variational techniques in optimization*, 77-90, 1980.
[15] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM Journal on Optimization*, (1): 26-33, 1997.
[16] M. Raydan, On the Barzilai and Borwein choice of steplength for the gradient method. *IMA Journal of Numerical Analysis*, (3): 321-326, 1993.
[17] J. Vlˇcek and L. Lukšan, Globally convergent variable metric method for nonconvex nondifferentiable unconstrained minimization. *Journal of Optimization Theory and Applications*, (2): 407-430, 2001.

(Saeed Banimehri) Department of Mathematics, Bu-Ali Sina University, Hamedan, Iran.
Email address: `s.banimehri@sci.basu.ac.ir`

(Hamid Esmaeili) Department of Mathematics, Bu-Ali Sina University, Hamedan, Iran.
Email address: `esmaeili@basu.ac.ir`