# New derivative-free nonmonotone line search methods for unconstrained minimization

Filip Nikolovski* Irena Stojkovska†

**Abstract**

Two new derivative-free nonmonotone line search methods for unconstrained optimization are proposed and analyzed. Convergence is established under standard conditions. Numerical results show good performance of the proposed methods.

**Keywords:** unconstrained optimization, nonmonotone line-search, derivative-free methods

**AMS subject classification:** 65K05, 90C56

## 1 Introduction

Let us consider the problem of unconstrained optimization:

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is bounded from below and has continuous partial derivatives that are not available.

A line search method for solving the problem (1), generates a sequence of iterates $\{x_k\} \in \mathbb{R}^n$ by the iterative formula $x_{k+1} = x_k + \alpha_k d_k$, where $d_k$ is a search direction at $x_k$ and $\alpha_k > 0$ is a step size which is usually chosen to minimize the objective function $f$ along the direction $d_k$. For more details about line search methods see [6]. The majority of line search methods require decrease in $f$ at each iteration, which means that the corresponding sequence of function values $\{f(x_k)\} = \{f_k\}$ monotonically decreases. This may sometimes

*Department of Mathematics, Faculty of Natural Sciences and Mathematics, Ss. Cyril & Methodius University, Gazi Baba bb, 1000 Skopje, Macedonia, e-mail: filipnikolovski@gmail.com

†Department of Mathematics, Faculty of Natural Sciences and Mathematics, Ss. Cyril & Methodius University, Gazi Baba bb, 1000 Skopje, Macedonia, e-mail: irenatra@iunona.pmf.ukim.edu.mk, irena.stojkovska@gmail.com

result in a slow convergence. On the other hand, the nonmonotone line search methods tend to converge faster and avoid converging to local minima, see [3, 4, 8, 9]. All of the above mentioned nonmonotone methods require the gradient of the objective function $f$ and are thus unsuitable for the problem (1), or for problems where the objective function $f$ is not smooth. In such situations, derivative-free methods based only on values of the objective function are more appropriate, see [6].

Diniz-Ehrhardt et al. [2] have proposed a derivative-free line search strategy which combines and extends the ideas form [3] and [4]. For given sequences $\{\eta_k\}$ and $\{\beta_k\}$, $k = 0, 1, 2, \ldots$ such that:

$$\eta_k > 0,\ \sum_{k=0}^{\infty} \eta_k = \eta < \infty \text{ and } \beta_k > 0,\ \lim_{k \in K} \beta_k = 0 \Rightarrow \lim_{k \in K} \nabla f(x_k) = 0, \tag{2}$$

where $\nabla f$ is the gradient of $f$ and $K \subseteq \mathbb{N}$ is an infinite set of indices, and a direction $d_k$, the step size $\alpha_k > 0$ is determined using the following nonmonotone line search rule:

$$f(x_k + \alpha_k d_k) \leq \bar{f}_k + \eta_k - \alpha_k^2 \beta_k, \tag{3}$$

with $\bar{f}_k = \max\{f(x_{k-j}) \mid 0 \leq j \leq m(k) - 1\}$, where $m(k) = \min\{k+1, M\}$ for $k = 0, 1, 2, \ldots$ and some $M \in \mathbb{N}$. We will refer to this method as *$M$-method.* In [2], the convergence with probability 1 is established for the $M$-method, when the search directions are randomly chosen and independent, bounded and descent with a fixed probability $p > 0$.

In this paper, we propose new derivative-free nonmonotone line search methods based on the line search rule (3) where the term $\bar{f}_k$ is chosen differently. We will establish the same convergence result as in [2], under same conditions. Numerical results show that the proposed new methods are competitive with the one from [2]. The rest of the paper is organized as follows: the new methods are formulated in section 2, while their convergence is established in section 3. The numerical results are presented in section 4.

## 2 New non-monotone line search rules

Assume that the sequences $\{\eta_k\}$ and $\{\beta_k\}$ satisfy the conditions (2). Let $\{r_k\}$, $r_k \in [0, 1]$ for all $k = 0, 1, \ldots$. The first new line search rule has the form of (3), where $\bar{f}_k$ is chosen to be $\bar{f}_k = C_k$, where $\{Q_k\}$ and $\{C_k\}$ are recursive sequences defined by:

$$Q_{k+1} = r_k Q_k + 1, \qquad C_{k+1} = \frac{r_k Q_k (C_k + \eta_k) + f_{k+1}}{Q_{k+1}}, \tag{4}$$

with $Q_0 = 1$, $C_0 = f_0$. We will refer to it as $C_k$-*line search rule* and to the corresponding method as $C_k$-*method*. The idea for a sequence $\{C_k\}$ comes from Zhang and Hager [9], and is used by many authors, e.g. [1].

The second line search rule also has the form of (3), where $\bar{f}_k$ is defined by $\bar{f}_k = \max\left\{f_k, \sum_{r=0}^{m(k)-1} \lambda_{k_r} f_{k-r}\right\}$, where $m(k) = \min\{k+1, M\}$ for $k = 0, 1, 2, \ldots$ and some $M \in \mathbb{N}$, and $\lambda_{k_r}$ are such that $\lambda_{k_r} \geq \lambda$, $r = 0, 1, ..., m(k)-1$ and $\sum_{r=0}^{m(k)-1} \lambda_{k_r} = 1$, for all $k = 0, 1, \ldots$ and some $\lambda \in (0, 1]$. We will refer to it as $\lambda$-*line search rule* and to the corresponding method as $\lambda$-*method*. The idea of a convex combination of last $M$ functional values has origin in Ulbrich [7], but it has not been given much attention until in [8].

Now we state the model algorithm for $C_k$- and $\lambda$-method.

**Algorithm 1 (Model algorithm ($C_k$-method / $\lambda$-method)).**
Choose $x_0 \in \mathbb{R}^n$ and sequences $\{\eta_k\}$ and $\{\beta_k\}$ that satisfy the conditions (2).

**Step 1.** Compute the search direction $d_k$ such that $\|d_k\| \leq \Delta$.

**Step 2.** Compute the term $\bar{f}_k$ (according to $C_k$-method / $\lambda$-method).

**Step 3.** Choose $0 < \alpha_k \leq 1$ such that $f(x_k + \alpha_k d_k) \leq \bar{f}_k + \eta_k - \alpha_k^2 \beta_k$.

**Step 4.** Set $x_{k+1} = x_k + \alpha_k d_k$, $k = k+1$ and go to Step 1. □

Let us note that $\eta_k > 0$ ensures that the line-search rule in Step 3 is satisfied for a sufficiently small step size $\alpha_k$, so the Algorithm 1 is well defined.

## 3 Convergence results

First we will prove some useful properties of the $C_k$- and $\lambda$-method.

**Lemma 1.** *Let $\{x_k\}$ be an iterative sequence generated by Algorithm 1 using the $C_k$-method. Then $f_k \leq C_k \leq C_{k-1} + \eta_{k-1}$ for all $k \in \mathbb{N}$.*

**Proof.** For all $k \in \mathbb{N}$, by Step 3 in Algorithm 1 we have $f_k \leq C_{k-1} + \eta_{k-1} - \alpha_{k-1}^2 \beta_{k-1} \leq C_{k-1} + \eta_{k-1}$. Then, the proof proceeds as the proof of Lemma 2.2 in [1]. □

**Lemma 2.** *Let $\{x_k\}$ be an iterative sequence generated by Algorithm 1 using the $\lambda$-method. Then:*

$$f_k \leq f_0 + \sum_{j=0}^{k-1} \eta_j - \lambda \sum_{j=0}^{k-2} \alpha_j^2 \beta_j - \alpha_{k-1}^2 \beta_{k-1} \leq f_0 + \sum_{j=0}^{k-1} \eta_j - \lambda \sum_{j=0}^{k-1} \alpha_j^2 \beta_j.$$

**Proof.** The proof is by induction. For $k = 1$, since $\lambda \leq 1$, and Step 3 from Algorithm 1 we have $f_1 \leq f_0 + \eta_0 - \alpha_0^2 \beta_0 \leq f_0 + \eta_0 - \lambda \alpha_0^2 \beta_0$. Assume that the assumption is true for all $j$, $1 \leq j \leq k$. We have two cases:

**Case 1.** $\max\left\{f_k, \sum_{r=0}^{m(k)-1} \lambda_{k_r} f_{k-r}\right\} = f_k$

**Case 2.** $\max\left\{f_k, \sum_{r=0}^{m(k)-1} \lambda_{k_r} f_{k-r}\right\} = \sum_{r=0}^{m(k)-1} \lambda_{k_r} f_{k-r}$

The proof can be completed by following the technique of the proof of Lemma 1 in [8], with additional consideration of the conditions (2). □

Let us note that Lemma 1 and Lemma 2 imply that a sequence $\{x_k\}$ generated by Algorithm 1 is such that $x_k \in \{x \in \mathbb{R}^n \mid f(x) \leq f_0 + \eta\}$ for all $k$. The next lemma gives another useful property shared by the $C_k$- and $\lambda$-method.

**Lemma 3.** *Let $\{x_k\}$ be an iterative sequences generated by Algorithm 1. Then, there exists an infinite subset of indices $K \subseteq \mathbb{N}$ such that $\lim_{k\in K} \alpha_k^2\beta_k = 0$.*

**Proof.** If $C_k$-method is used we have:

$$C_{k+1} = \tfrac{r_k Q_k(C_k+\eta_k)+f_{k+1}}{Q_{k+1}} \leq \tfrac{r_k Q_k(C_k+\eta_k)+C_k+\eta_k-\alpha_k^2\beta_k}{Q_{k+1}} = C_k + \eta_k - \tfrac{\alpha_k^2\beta_k}{Q_{k+1}}.$$

Now, summing up the first $k+1$ inequalities we have $C_{k+1} \leq C_0 + \sum_{j=0}^{k} \eta_j - \sum_{j=0}^{k} \frac{\alpha_j^2\beta_j}{Q_{j+1}}$. From Lemma 1, since $f$ is bounded from below with a constant $K > 0$, for all $k$ we have:

$$\sum_{j=0}^{k} \frac{\alpha_j^2\beta_j}{Q_{j+1}} \leq C_0 + \sum_{j=0}^{k} \eta_j - C_{k+1} \leq f_0 + \eta - f_{k+1} \leq f_0 + \eta - K,$$

which implies that $\sum_{j=0}^{\infty} \frac{\alpha_j^2\beta_j}{Q_{j+1}} < +\infty$. Since $Q_{j+1} \leq j+2$, we have that $\liminf_{j\to\infty} \alpha_j^2\beta_j = 0$.

If $\lambda$-method is used, by Lemma 2 we have:

$$f_k \leq f_0 + \sum_{j=0}^{k-1} \eta_j - \lambda \sum_{j=0}^{k-1} \alpha_j^2\beta_j \leq f_0 + \eta - \lambda \sum_{j=0}^{k-1} \alpha_j^2\beta_j.$$

Now, since $f$ is bounded from below with a constant $K > 0$ and $\lambda > 0$, for all $k$ we have:

$$\sum_{j=0}^{k-1} \alpha_j^2\beta_j \leq \frac{f_0 + \eta - f_{k+1}}{\lambda} \leq \frac{f_0 + \eta - K}{\lambda},$$

which implies that $\sum_{j=0}^{\infty} \alpha_j^2\beta_j < +\infty$ and $\liminf_{j\to\infty} \alpha_j^2\beta_j = 0$, which completes the proof. □

Let us note that, if $0 < r_k < 1$ for all $k$ in $C_k$-method, then the stronger result $\lim_{j\to\infty} \alpha_j^2\beta_j = 0$ can be proven. Now we state one of the main theorems for the proposed derivative-free nonmonotone line search methods. Since $\bar{f}_k \geq f_k$ is valid for both $C_k$- and $\lambda$-method, the proof follows from Lemma 3, mimicing the proof of Theorem 1 in [2].

**Theorem 1.** *Let $\{x_k\}$ be an iterative sequence generated by Algorithm 1. Assume that $(x^*, d)$ is a limit point of the subsequence $\{(x_k, d_k)\}_{k\in K}$ where $K \subseteq \mathbb{N}$ is an infinite subset of indices. Then $\nabla f(x^*)^T d \geq 0$.* □

Using Lemma 3, under additional assumptions for the search directions $d_k$, such as: $\|d_k\| \in [\Delta_{\min}, \Delta_{\max}]$ and $\nabla f(x_k)^T d_k \leq -\theta\|\nabla f(x_k)\|\|d_k\|$ for infinitely many $k$, where $0 < \Delta_{\min} < \Delta_{\max} < \infty$ and $0 < \theta < 1$, it can be proven that Algorithm 1 finds stationary points up to any arbitrary precision (see Corollary 1 in [2]). In order to have truly derivative-free methods, we can relax the above additional conditions in such a way that the search directions are randomly chosen, independent, bounded and descent with a fixed probability $p > 0$. This way, convergence with probability 1 to a stationary point can be established (see Theorem 2 in [2]).

## 4 Numerical results

In this section we present some of the results from testing $M$-, $C_k$-, $\lambda$- and the “monotone” line-search method (the last one by using $M = 1$ in (3)). All test problems are from Moré et al. [5] and all of them have optimal function value $f^* = 0$. The sequences in (2) have been chosen such that $\beta_k \equiv 1$ and $\eta_k = f_0/k^{1.1}$. Search direction $d_k = -\hat{g}_k/\sigma_k$ is used, where $\hat{g}_k$ is a discrete gradient approximation of $\nabla f(x_k)$, and $\sigma_k$ is the spectral coefficient (see [2]). The rest of parameters for $M$-, $C_k$-, $\lambda$-method are $M = 5$, $r_k = 0.85$ for all $k$ and $\lambda_{k_r} = 1/m(k)$ for all $r$. The results are reported in Table 1, where *It* denotes the number of iterations, *Evalf* denotes number of function evaluations, $f$ is the function value at the last iterate, *normg* is the norm of the gradient approximation at the last iterate. All methods tested on the first problem (MGH26) stopped after the maximum of 5000 iterations was reached, all methods except $C_k$-method tested on the second (MGH22) and third (MGH21) problem stopped after the maximum of 500000 function evaluations was exceeded, and the $C_k$-method tested on the second and third problem stopped when $|f_k| \leq 10^{-9}$. Additional numerical results are also available at `http://www.institutzamatematika.com/index.php/Irena_Stojkovska_Curriculum_Vitae`. All results show good performance of the proposed new derivative-free nonmonotone line search methods compared with the $M$-method. There are many problems for which nonmonotone methods converge faster to the solution then the “monotone” one, which is as expected.

## 5 Conclusions

In this paper two new derivative-free nonmonotone line search methods were proposed and their convergence established under same assumptions as in [2]. Numerical results show good performance of the proposed methods. The fact that nonmonotone methods are able to avoid local minima, leads to the idea to test these nonmonotone methods on noisy functions.

| Prb | $n$ | It | Evalf | $f$ | normg |
|---|---|---|---|---|---|
| | | "Monotone" line search / $M$-method, [2] | | | |
| MGH26 | 10 | 5000 / 5000 | 422555 / 422885 | $2.80E-05$ / $2.80E-05$ | $2.36E-05$ / $7.40E-06$ |
| MGH22 | 100 | 2428 / 2427 | 500182 / 500048 | $4.21E-06$ / $1.22E-04$ | $3.81E-03$ / $2.19E-01$ |
| MGH21 | 100 | 2424 / 2425 | 500073 / 500137 | $1.02E-05$ / $5.52E-07$ | $2.88E-03$ / $6.65E-04$ |
| | | $C_k$-method / $\lambda$-method | | | |
| MGH26 | 10 | 5000 / 5000 | 382384 / 342569 | $2.80E-05$ / $2.80E-05$ | $6.41E-05$ / $6.21E-06$ |
| MGH22 | 100 | 533 / 2429 | 109040 / 500165 | $4.25E-11$ / $1.22E-04$ | $1.43E-06$ / $5.92E-03$ |
| MGH21 | 100 | 1907 / 2444 | 390414 / 500193 | $9.97E-10$ / $2.26E-08$ | $3.03E-05$ / $1.47E-04$ |

Table 1: Results for tested line search methods with spectral gradient direction

## References


[1] Cheng, W., Li, D. H. (2009) *A derivative-free nonmonotone line search and its application to the spectral residual method*, IMA J. of Num. Anal., 29(3), 814–825.

[2] Diniz-Ehrhardt, M. A., Martíez, J. M., Raydán, M. (2008) *A derivative-free nonmonotone line-search technique for unconstrained optimization*, J. of Comp. App. Math., 219(2), 383–397.

[3] Grippo, L., Lampariello, F., Lucidi, S. (1986) *A nonmonotone line search technique for Newton's method*, SIAM J. on Num. Anal., 23(4), 707–716.

[4] Li, D. H., Fukushima, M. (2000) *A derivative-free line search and global convergence of Broyden-like method for nonlinear equations*, Opt. Meth. and Soft., 13(3), 181–201.

[5] Moré, J. J., Garbow, B. S., Hillstrom, K. E. (1981) *Testing unconstrained optimization software*, ACM Transactions on Mathematical Software (TOMS), 7(1), 17–41.

[6] Nocedal, J., Wright, S. J. (2006) *Numerical optimization. Second edition*, New York, Springer-Verlag.

[7] Ulbrich, M. (2001) *Nonmonotone trust-region methods for bound-constrained semismooth equations with applications to nonlinear mixed complementarity problems*, SIAM J. on Optimiz., 11(4), 889–917.

[8] Yu, Z., Pu, D. (2008) *A new nonmonotone line search technique for unconstrained optimization*, J. of Comp. App. Math., 219(1), 134–144.

[9] Zhang, H., Hager, W. W. (2004). *A nonmonotone line search technique and its application to unconstrained optimization*, SIAM J. on Optimiz., 14(4), 1043–1056.