A derivative-free filter method with local variations for systems of nonlinear equations

Irena Stojkovska*

Abstract

In this paper a new derivative-free line-search filter method for system of nonlinear equations will be presented. The line-search rule will be based on a new concept of a forcing function that helps to establish the convergence results under weaker conditions and on the other hand gives more general insight to the line-search procedure. The multidimensional filter that will be used within the line-search procedure will help to avoid small step sizes from backtracking or interpolation. Truly derivative-free character of the method is obtained by occasional choice of a random search directions. The convergence in a probabilistic sense will be established. The proposed method can be applied also for unconstrained optimization problems, nonlinear complementarity problems, variational inequality problems.

Key words. systems of nonlinear equations, derivative-free method, line-search method, filter method

AMS subject classification. 65H10, 90C56

1 Introduction

Consider the following system of nonlinear equations

$$F(x) = 0, \tag{1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable function. If we denote with $f(x) = \frac{1}{2}\|F(x)\|^2 \geq 0$, then problem (1) is equivalent to the following unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x), \tag{2}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable function. We are interested in case when derivatives of F are not available.

*Department of Mathematics, Faculty of Natural Sciences and Mathematics, St. Cyril and Methodius University, Gazi Baba b.b., 1000 Skopje, Macedonia, e-mail: irenatra@iunona.pmf.ukim.edu.mk

One of the most known iterative methods for system of nonlinear equations are line-search methods that for a given iterate x_k search along the search direction d_k for a step size $\alpha_k > 0$ that satisfies some line-search rule, and define the new iterate with $x_{k+1} = x_k + \alpha_k d_k$. Common requirement is d_k to be a descent direction which will ensure reduction on f at each iteration. Different choices of the search direction d_k define different line-search methods, among which the most popular are quasi-Newton methods that instead of true Jacobian ∇F they use some derivative-free approximation (see [9]). That is why the line-searches that require the calculation of derivatives are not suitable for quasi-Newton methods. Therefore, it is desirable to develop a derivative-free line-search. Some results on this subject are given in [1, 8]. Dealing with random search directions is also suitable for developing truly derivative-free method if the line-search rule is without using of derivatives. Papers like [2] give contributions in this field.

Filter methods are one of the latest developments in global optimization algorithms and were firstly proposed by Fletcher in 1996, [3]. First filter methods are a kind of alternative to penalty functions used in constrained nonlinear programming optimization algorithms [11]. The main purpose of filters is allowing the full Newton step to be taken more often and thus inducing global convergence of the method. Filter methods are extended to solve nonlinear equations [1, 5], unconstrained optimization [7].

Using a new concept of forcing function we are going to propose a new derivative-free line-search rule. The forcing function acts like a generalization, and helps to establish convergence results under weaker assumptions. The multidimensional filter will be combined with the line-search procedure which will help to avoid the small step sizes form backtracking or interpolation. And at the end, by occasionally choosing a random search directions, the convergence in probabilistic sense will be established.

The paper is organized as follows. In Section 2, a new definition of forcing function is given, a new derivative-free line-search rule is presented, and the filter algorithm is stated. Convergence of the proposed filter method is established in Section 3.

2 Filter algorithm

First, we will give a new definition of a *forcing function*, and then we will present our new line-search rule with local variations using the concept of forcing functions. Then the multidimensional filter will be combined with this line-search rule, and the new filter algorithm will be presented.

Definition 1. The function $\sigma : (0, +\infty) \to (0, +\infty)$ is a *forcing function*, if for any sequence $\{t_i\} \subset (0, +\infty)$, the following conditions are satisfied

(i) $\lim_{i\to\infty} \sigma(t_i) = 0 \ \Rightarrow\ \lim_{i\to\infty} t_i = 0$,

(ii) $\lim_{i\to\infty} t_i = 0 \ \Rightarrow\ \lim_{i\to\infty} \frac{\sigma(t_i)}{t_i} = 0$.

An example of a forcing function according to the above Definition 1 is $\sigma(t) = ct^p$ where $c > 0$ and $p > 1$. Some authors use the forcing function that satisfies the first condition (i) [10], and some of them use forcing function that satisfies the second condition (ii) [6]. Incorporating these two conditions into one definition of a forcing function helps us by strengthening the conditions for forcing function to weaker the assumptions for establishing the global convergence of the method that we are going to propose.

Given an iterate x_k we search along the direction d_k for a step size $0 < \alpha_k \leq 1$ that will satisfy the following line-search rule

$$f(x_k + \alpha_k d_k) \leq (1 + \eta_k) f(x_k) - \sigma(t_k), \tag{3}$$

where σ is a forcing function, $t_k = \alpha_k \|F(x_k)\|$ and the sequence $\{\eta_k\}$ is such that

$$\eta_k > 0,\ k = 0, 1, 2, ...,\ \sum_{k=0}^{\infty} \eta_k = \eta < \infty. \tag{4}$$

The presence of term $\eta_k > 0$ in the line-search rule (3) guaranties that (3) holds for sufficiently small α_k i.e. the iteration that is based on the line-search rule (3) is well-defined. Another feature of the line-search rule (3) is that the nonmonotonicity may occur i.e. the case when $f(x_{k+1}) > f(x_k)$ can happened. Actually when iterates are far from solution the method can be more nonmonotone, and as the iterates approach the solution the method becomes more monotone. This strategy is desirable, at first to avoid local optima when far from the solution and than to speed up the convergence when near the solution. Similar reasoning for better numerical performance is presented in [12].

We will also note that the following line-search rule

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \eta_k - \sigma(t_k), \tag{5}$$

where σ is a forcing function, $t_k = \alpha_k \|F(x_k)\|$ and the sequence $\{\eta_k\}$ is defined by (4), will act in a similar way, and the convergence results for it can be obtained similarly.

Not every trial step size that is unsatisfactory for the line-search rule (3) will be rejected by our new method. Here is the definition of the filter and the rule for acceptance of a point with a help of this filter. We will use the multidimensional filter defined in [7], which is similar as in [1, 5]. The equations (1) are partitioned into m sets $\{F_i(x)\}_{i \in I_j}$, $j = 1, ..., m$, with the property $\{1, ..., n\} = I_1 \cup ... \cup I_m$, $I_j \cap I_k = \varnothing$, $j \neq k$ and the *filter functions* are defined as

$$\phi_j(x) \stackrel{def}{=} ||F_{I_j}(x)|| \quad \text{for} \quad j = 1, ..., m \tag{6}$$

where $|| \cdot ||$ is the Euclidean norm and F_{I_j} is the vector whose components are the components of F indexed by I_j. With this notation x^* is the solution of (1) if and only if $\phi_j(x^*) = 0$ for all $j = 1, ..., m$. The following abbreviations will be used

$$\phi(x) \stackrel{def}{=} (\phi_1(x), ..., \phi_m(x))^T,\ \phi_k \stackrel{def}{=} \phi(x_k) \text{ and } \phi_{j,k} \stackrel{def}{=} \phi_j(x_k).$$

A *filter* is a list $\mathcal{F}$ of m-tuples of the form $(\phi_{1,k}(x), \phi_{2,k}(x), ..., \phi_{m,k}(x))$ such that

$$\phi_{j,k} < \phi_{j,l} \text{ for at least one } j \in \{1, ..., m\} \text{ and for all } k \neq l.$$

A point x_1 *dominates* a point x_2 whenever

$$\phi_j(x_1) \leq \phi_j(x_2) \text{ for all } j = 1, ..., m.$$

Therefore we say that the filter keeps all iterates that are not dominated by other iterates in the filter. We will use a filter to construct an additional acceptability condition for a new trial iterate $x_k^+ = x_k + \alpha_k d_k$ when the step size α_k does not satisfy the line search rule (3).

We say that a new trial point x_k^+ is *acceptable for the filter* $\mathcal{F}$ if

$$\forall \phi_l \in \mathcal{F} \quad \exists j \in \{1, ..., m\} \quad \phi_j(x_k^+) < \phi_{j,l} - \gamma_\phi \delta_1(||\phi_l||, ||\phi_k^+||), \tag{7}$$

where $\gamma_\phi \in (0, 1)$ is a small positive constant and

$$\delta_1(||\phi_l||, ||\phi_k^+||) = \max\{||\phi_l||, ||\phi_k^+||\}.$$

When a trial point is acceptable for the filter, we *add the trial point to the filter* immediately i.e. we add the m-tuple $\phi_k^+ = \phi(x_k^+) = (\phi_1(x_k^+), ..., \phi_m(x_k^+))^T$ to the filter $\mathcal{F}$.

When we add a new trial point x_k^+ to the filter, we remove all points from the filter that are dominated by the trial point x_k^+, which means that we remove m-tuples $(\phi_{1,k_i}, ..., \phi_{m,k_i})^T \in \mathcal{F} \setminus \{\phi_k^+\}$ from the filter if

$$\phi_j(x_k^+) \leq \phi_{j,k_i} \quad j = 1, ..., m.$$

Every trial point which is acceptable for the filter is taken as a new iterate.

If a trial point x_k^+ is not acceptable to the filter we reduce the step size α_k and check again if the line-search rule (3) is satisfied. The proposed filter algorithm gives its best performance in the case when $m = n$ and $I_j = \{j\}$ - similarly as it is discussed in [4]. In this case we will have $\phi_j(x) = |F_j(x)|$, $\phi(x) = F(x)$, $\phi_k = F(x_k)$ and $\phi_{j,k} = |F_j(x_k)|$, $j = 1, 2, ..., n$. So, when we use filter in our optimization procedure, we accept those points in which at least one of the equations from problem (1) reduces in norm. With this we avoid to reduce the step size when line-search rule is not satisfied, so the optimization procedure might be fasten.

Now we can state our filter algorithm.

Algorithm 1. *Filter algorithm*

Step 0. Choose $x_0 \in \mathbb{R}^n$ and set $k = 0$.

Step 1. If $F(x_k) = 0$ then Stop. Otherwise, compute the search direction d_k. Take $\alpha_k = 1$ and go to Step 2.

Step 2. Let the trial point be $x_k^+ = x_k + \alpha_k d_k$. If the line-search rule (3) is satisfied then go to Step 3, else if x_k^+ is acceptable to the filter, then add x_k^+ to the filter, remove all points from the filter that are dominated by x_k^+ and go to Step 3. Otherwise, reduce α_k and repeat Step 2.

Step 3. Take the next iterate $x_{k+1} = x_k^+$, put $k = k + 1$ and go to Step 1.

The search direction d_k can be chosen randomly in Step 1, which means that it might happen that d_k is not a descent direction. In the next section we are going to impose an assumption on the search direction d_k in order to establish the convergence result. The requirement imposed in this assumption will be easy to satisfy, in some probabilistic sense, as we will see later.

The reduction of α_k in Step 2 can be done by the backtracking procedure, or by interpolation on the interval $[\tau_1\alpha_k, \tau_2\alpha_k]$, where $0 < \tau_1 < \tau_2 < 1$ are given constants.

At the beginning of Algorithm 1, the filter is initialized to be empty $\mathcal{F} = \varnothing$ or to be $\mathcal{F} = \{\phi(x) : \phi_j(x) \geq \phi_{j\,\max}, j = 1, ..., m\}$ for any $\phi_{j\,\max} > \phi_j(x_0)$, $j = 1, ..., m$ (similarly as in [11]).

Note that if there are finitely many values of ϕ_k that are added to the filter, then for all k large enough (for all $k \geq k_0$ where ϕ_{k_0} is the last m-tuple added to the filter), Algorithm 1 will be the same algorithm without filter. This discussion leads us to differ two cases, the first one when finitely many points are added to the filter, and the second one when there are infinitely many points that are added to the filter. Convergence results, presented in the next section, are established by considering these two cases.

3 Convergence results

Dealing with filter makes us to impose the following assumption for the values ϕ_k that are added to the filter, as it is done in [7].

Assumption 1. There exists a constant $C > 0$ such that for all values ϕ_k that are added to the filter by Algorithm 1 the following stands

$$\|\phi_k\| \leq C.$$

Let $\mathfrak{L}$ be the level set defined by

$$\mathfrak{L} = \{x : ||F(x)|| \leq e^{\frac{\eta}{2}} \max\{C, ||F(x_0)||\}\}, \tag{8}$$

where $\eta > 0$ is the value defined with (4).

Than we have the following theorem.

Theorem 1. *Let the Assumption 1 holds and let $\{x_k\}$ be a sequence generated by Algorithm 1. Then $\{x_k\} \subset \mathfrak{L}$.*

Proof. Similar as in [7]. For each two consecutive iterates x_{k-1} and x_k which were not added to the filter the line-search rule (3) is satisfied which means that

$$||F(x_k)|| \leq (1+\eta_{k-1})^{\frac{1}{2}}||F(x_{k-1})||, \tag{9}$$

because of $\sigma(t_k) > 0$ and $f(x) = \frac{1}{2}\|F(x)\|^2$.

If for the iterate x_k there are no values $\phi_{k'}, k' \leq k$ that are added to the filter, then we have

$$\begin{aligned} ||F(x_k)|| &\leq (1+\eta_{k-1})^{\frac{1}{2}}||F(x_{k-1})|| \\ &\leq (1+\eta_{k-1})^{\frac{1}{2}}(1+\eta_{k-2})^{\frac{1}{2}}||F(x_{k-2})|| \\ &\vdots \\ &\leq \Big(\prod_{j=0}^{k-1}(1+\eta_j)^{\frac{1}{2}}\Big)||F(x_0)|| \\ &\leq ||F(x_0)||\Big(\frac{1}{k}\sum_{j=0}^{k-1}(1+\eta_j)\Big)^{\frac{k}{2}} \\ &= ||F(x_0)||\Big(1+\frac{1}{k}\sum_{j=0}^{k-1}\eta_j\Big)^{\frac{k}{2}} \qquad (10) \\ &\leq ||F(x_0)||\Big(1+\frac{\eta}{k}\Big)^{\frac{k}{2}} \\ &\leq e^{\frac{\eta}{2}}||F(x_0)||, \qquad (11) \end{aligned}$$

where $\eta > 0$ is defined with (4).

Now let $x_{i(k)}$ be the last iterative point before x_k such that $\phi_{i(k)}$ was added to the filter. Then transforming in a similar way as in (10) we have

$$\begin{aligned} ||F(x_k)|| &\leq ||F(x_{i(k)})||\Big(1+\frac{1}{k-i(k)}\sum_{j=i(k)}^{k-1}\eta_j\Big)^{\frac{k-i(k)}{2}} \\ &\leq ||F(x_{i(k)})||\Big(1+\frac{\eta}{k-i(k)}\Big)^{\frac{k-i(k)}{2}} \\ &\leq e^{\frac{\eta}{2}}||F(x_{i(k)})|| \\ &\leq e^{\frac{\eta}{2}}C, \qquad (12) \end{aligned}$$

where $\eta > 0$ is defined with (4) and $C > 0$ is a constant from Assumption 1. From (11) and (12) it implies that $\{x_k\} \subset \mathfrak{L}$. ■

The following assumption on the objective function $f(x)$ is imposed throughout the paper.

Assumption 2. The function $f(x)$ is bounded from below on the level set $\mathfrak{L}$.

When there are finitely many points that are added to the filter by Algorithm 1, then we can obtain the following result.

Lemma 3.1. *Let Assumptions 1-2 hold and let $\{x_k\}$ be a sequence generated by Algorithm 1. If there are finitely many points that are added to the filter by Algorithm 1, then*

$$\lim_{k\to\infty} \alpha_k \|F(x_k)\| = 0.$$

Proof. Let ϕ_{k_0} is the last m-tuple that was added to the filter by Algorithm 1. So, for all $k \geq k_0$, the line-search rule (3) is satisfied by some trial point $x_k^+ = x_k + \alpha_k d_k$ which is taken as a new iterate x_{k+1} i.e. we have

$$f(x_{k+1}) \leq (1+\eta_k) f(x_k) - \sigma(t_k) \text{ for all } k \geq k_0.$$

Let $k \geq k_0$ be arbitrary. Then, because $\eta_k > 0$ implies $1 + \eta_k > 1$ for all k, we have

$$\begin{aligned} f(x_{k+1}) &\leq (1+\eta_k) f(x_k) - \sigma(t_k) \\ &\leq (1+\eta_k)(1+\eta_{k-1}) f(x_{k-1}) - (1+\eta_k)\sigma(t_{k-1}) - \sigma(t_k) \\ &\leq (1+\eta_k)(1+\eta_{k-1}) f(x_{k-1}) - \sigma(t_{k-1}) - \sigma(t_k) \\ &\;\;\vdots \\ &\leq \Big(\prod_{j=k_0}^{k} (1+\eta_j)\Big) f(x_{k_0}) - \sum_{j=k_0}^{k} \sigma(t_j). \end{aligned}$$

From the last inequality we have

$$\sum_{j=k_0}^{k} \sigma(t_j) \leq \Big(\prod_{j=k_0}^{k} (1+\eta_j)\Big) f(x_{k_0}) - f(x_{k+1}).$$

Considering that $f(x_{k_0}) = \frac{1}{2}\|F(x_{k_0})\|^2 = \frac{1}{2}\|\phi_{k_0}\|^2 \leq \frac{1}{2}C^2$ by Assumption 1, and that $f(x_{k+1}) \geq M$ for some positive constant $M > 0$ by Assumption 2, and transforming the product in a similar way as in (10) i.e.

$$\begin{aligned} \prod_{j=k_0}^{k} (1+\eta_j) &\leq \Big(\frac{1}{k-k_0+1} \sum_{j=k_0}^{k} (1+\eta_j)\Big)^{k-k_0+1} \\ &\leq \Big(1 + \frac{1}{k-k_0+1} \sum_{j=k_0}^{k} \eta_j\Big)^{k-k_0+1} \\ &\leq \Big(1 + \frac{\eta}{k-k_0+1}\Big)^{k-k_0+1} \leq e^{\eta}, \end{aligned}$$

we obtain

$$\sum_{j=k_0}^{k} \sigma(t_j) \leq \Big(\prod_{j=k_0}^{k} (1+\eta_j)\Big) f(x_{k_0}) - f(x_{k+1}) \leq \frac{1}{2} e^{\eta} C^2 - M, \tag{13}$$

where without loss of generality we can assume that $\frac{1}{2}e^{\eta}C^2 - M > 0$, since $C > 0$ is the upper bound, and $M > 0$ is the lower bound. Now, let $k \to \infty$ in (13), and we have

$$\sum_{j=k_0}^{\infty} \sigma(t_j) < \infty,$$

which implies that

$$\lim_{j\to\infty} \sigma(t_j) = 0.$$

From Definition 1 we have that

$$\lim_{j\to\infty} t_j = 0, \tag{14}$$

and $t_j = \alpha_j \|F(x_j)\|$ gives

$$\lim_{j\to\infty} \alpha_j \|F(x_j)\| = 0,$$

which completes the proof. ■

For the case when finitely many points are added to the filter we make the following assumption on the search direction.

Assumption 3. There exists $0 < \Delta_{\min} < \Delta_{\max} < \infty$ and $0 < \theta < 1$ such that for all $k \in K$, where K is an infinite subset of $\mathbb{N}$, the search direction d_k satisfies

$$\Delta_{\min} \leq d_k \leq \Delta_{\max} \text{ and } \langle d_k, g(x_k)\rangle \leq -\theta \, \|g(x_k)\|\|d_k\|,$$

where g is the gradient of f.

We also make the following assumption on the level set $\mathfrak{L}$.

Assumption 4. The level set $\mathfrak{L}$ is bounded.

Theorem 2. *Let Assumptions 1-4 hold and let $\{x_k\}$ be a sequence generated by Algorithm 1. If there are finitely many points that are added to the filter by Algorithm 1, then*

$$\liminf_{k\to\infty} \|F(x_k)\| = 0.$$

Proof. From Lemma 3.1 we have that

$$\lim_{k\to\infty} \alpha_k \|F(x_k)\| = 0.$$

If $\lim_{k\to\infty} \|F(x_k)\| = 0$ the proof is done. Otherwise, we have that

$$\lim_{k\to\infty} \alpha_k = 0. \tag{15}$$

Therefore for k large enough we have that $\alpha_k < 1$. Without loss of generality assume that $\alpha_k < 1$ for all $k \geq k_0$, where k_0 is the index of the last m-tuple ϕ_{k_0} that was added to the filter by Algorithm 1. This means that for $k \geq k_0$ there exists $\alpha'_k > \alpha_k$ for which the line-search rule (3) is violated. For those α'_k we have

$$\lim_{k\to\infty} \alpha'_k = 0 \tag{16}$$

and

$$f(x_k + \alpha'_k d_k) > (1 + \eta_k)f(x_k) - \sigma(t_k).$$

Therefore, since $\eta_k > 0$ we have

$$\frac{f(x_k + \alpha'_k d_k) - f(x_k)}{\alpha'_k} > -\frac{\sigma(t_k)}{\alpha'_k},$$

for all $k \geq k_0$. By the Mean Value Theorem, for all $k \geq k_0$ there exists $\xi_k \in [0, 1]$ such that

$$\langle g(x_k + \xi_k \alpha'_k d_k), d_k\rangle > -\frac{\sigma(t_k)}{\alpha'_k},$$

where g is the gradient of f. Therefore, for all $k \geq k_0$,

$$\langle g(x_k + \xi_k \alpha'_k d_k) - g(x_k), d_k\rangle + \langle g(x_k), d_k\rangle > -\frac{\sigma(t_k)}{\alpha'_k}.$$

Having in mind that $\alpha'_k > \alpha_k$ we have for all $k \geq k_0$,

$$\begin{aligned} \langle g(x_k), d_k\rangle &> -\frac{\sigma(t_k)}{\alpha'_k} - \langle g(x_k + \xi_k \alpha'_k d_k) - g(x_k), d_k\rangle \\ &> -\frac{\sigma(t_k)}{\alpha_k} - \|g(x_k + \xi_k \alpha'_k d_k) - g(x_k)\|\|d_k\|. \end{aligned} \tag{17}$$

Again from Definition 1 and (14) we have that

$$\lim_{k\to\infty} \frac{\sigma(t_k)}{t_k} = 0.$$

This limit implies that

$$\lim_{k\to\infty} \frac{\sigma(t_k)}{\alpha_k} = \lim_{k\to\infty} \frac{\sigma(t_k)}{\alpha_k \|F(x_k)\|}\|F(x_k)\| = \lim_{k\to\infty} \frac{\sigma(t_k)}{t_k}\|F(x_k)\| = 0, \tag{18}$$

since $\|F(x_k)\|$ is bounded above (Theorem 1).

Now, since d_k, for all $k \in K$, is bounded (Assumption 3), g is continuous and from (16) and (18) we have that

$$\lim_{k\in K} \left(\frac{\sigma(t_k)}{\alpha_k} + \|g(x_k + \xi_k \alpha'_k d_k) - g(x_k)\|\|d_k\|\right) = 0.$$

The last limit and (17) imply

$$\lim_{k\in K} \langle g(x_k), d_k\rangle \geq 0. \tag{19}$$

From Theorem 1 we have that $\{x_k\} \subset \mathfrak{L}$, so Assumption 4 implies that $\{x_k\}$ is bounded. The boundness of $\{x_k\}$, continuity of g and Assumption 3 imply that there exists the infinite subset $K_1 \subset K$ such that

$$\lim_{k \in K_1} \langle g(x_k), d_k \rangle \leq 0. \tag{20}$$

So, from (19) and (20) we have that

$$\lim_{k \in K_1} \langle g(x_k), d_k \rangle = 0. \tag{21}$$

Then, by Assumption 3 we have

$$\lim_{k \in K_1} \|g(x_k)\| \|d_k\| = 0. \tag{22}$$

Since, $\|d_k\| \geq \Delta_{\min} > 0$ for all k by Assumption 3, implies that

$$\lim_{k \in K_1} \|g(x_k)\| = 0. \tag{23}$$

Having in mind that g is the gradient of f and $f(x) = \frac{1}{2}\|F(x)\|^2$, the last limit (23) means that

$$\lim_{k \in K_1} \|F(x_k)\| = 0,$$

which completes the proof. ■

Now we will consider the case when infinitely many points are added to the filter by Algorithm 1. We can establish stronger result than in Theorem 2.

Theorem 3. *Let Assumption 1 holds and let $\{x_k\}$ be a sequence generated by Algorithm 1. If there are infinitely many points that are added to the filter by Algorithm 1, then*

$$\lim_{k \to \infty} \|F(x_k)\| = 0.$$

Proof. The proof is the same as in [7], where g is substituted with F and the sequence $\{\varepsilon_k\}$ is substituted with $\{\eta_k\}$. ■

From last two theorems, the main convergence result follows as a consequence.

Corollary 1. *Let Assumptions 1-4 hold and let $\{x_k\}$ be a sequence generated by Algorithm 1. Then*

$$\liminf_{k \to \infty} \|F(x_k)\| = 0, \tag{24}$$

and every accumulation point of $\{x_k\}$ is the solution of (1).

Proof. With Theorem 2 and Theorem 3 the two possible cases were considered, when there are finitely many points added to the filter and when there are infinitely many points added to the filter. And in both of the cases the limit (24) is valid.

From Theorem 1 we have that $\{x_k\} \subset \mathfrak{L}$, so Assumption 4 implies that $\{x_k\}$ is bounded. Let x^* be an accumulation point of $\{x_k\}$. Continuity of F and the limit (24) implies that $F(x^*) = 0$ i.e. x^* is the solution of problem (1). ■

Since we want to have truly derivative-free method, the second part of the requirement in Assumption 3 is impossible to check because it depends on knowing $g(x_k)$. As it is discussed in [2], we may overcome this difficulty with occasional choice of a random search direction. So, due to its geometrical meaning, the requirement in Assumption 3 is easy to satisfy with the probability greater then some fixed $0 < p < 1$. We can prove the following result.

Corollary 2. *Let Assumptions 1-2 and Assumption 4 hold. Let for all k, a search direction d_k is chosen randomly such that*

(i) $d_0, d_1, d_2, \ldots$ *are independent n-dimensional random variables,*

(ii) there exists $0 < \Delta_{\min} < \Delta_{\max} < \infty$ *and* $0 < \theta < 1$ *such that for all* $k \in \mathbb{N}$, *search directions d_k satisfy the requirement from Assumption 3 with probability greater then some fixed* $0 < p < 1$.

Let $\{x_k\}$ be a sequence generated by Algorithm 1. Then, with probability 1, $\liminf_{k\to\infty} \|F(x_k)\| = 0$, *and every accumulation point of $\{x_k\}$ is the solution of (1) with probability 1.*

Proof. Similar as in [2]. Let $k \in \mathbb{N}$. There exists $l(k) \geq k$ such that $l(k) \in K$, and the probability of the event

$$\Delta_{\min} \leq d_{l(k)} \leq \Delta_{\max} \text{ and } \langle d_{l(k)}, g(x_{l(k)})\rangle \leq -\theta \, \|g(x_{l(k)})\| \|d_{l(k)}\| \tag{25}$$

is greater than $p > 0$. Therefore, the probability of the existence of an infinite subset $K_1 \subset \mathbb{N}$ such that (25) holds for all $k \in K_1$ is equal to 1. This implies that the probability of the existence of an infinite subset $K_1 \subset K$ such that the requirement from Assumption 3 holds for all $k \in K_1$ is equal to 1. Therefore, by Corollary 1, $\liminf_{k\to\infty} \|F(x_k)\| = 0$ with probability 1, and every accumulation point of $\{x_k\}$ is the solution of (1) with probability 1, as we wanted to prove. ■

4 Conclusions

In this paper a new derivative-free line-search filter method for system of nonlinear equations is presented and the convergence results for the same method are established. The line-search rule is based on a new concept of a forcing function that globalizes the method and helps to establish the convergence results under weaker conditions. The multidimensional filter that is combined with the line-search procedure helps to avoid small step sizes by accepting those points in which at least one of the equations reduces in norm. And it is shown that with occasional choice of a random search direction, we can ensure the truly derivative-free method and we can establish its convergence in a probability sense.

This method can be applied for nonlinear complementarity problems, variational inequality problems, and also for unconstrained optimization problems, as it is discussed at the beginning. It might be interesting to spread the idea on nonmonotone line-search procedures, and to adopt the method for optimization in noisy environment.

References

[1] Y. Deng, Zh. Liu, *Two derivative-free algorithms for nonlinear equations,* Optimization Methods and Software, Vol. 23, No. 3 (2008), pp.395-410

[2] M. A. Diniz-Ehrhardt, J. M. Martinez, M. Raydan, *A derivative free nonmonotone line-search technique for unconstrained optimization,* J. of Comp. and App. Math., Vol. 219 (2008), No. 2, pp.383-397

[3] R. Fletcher, S. Leyffer, *Nonlinear programming without a penalty function,* Math. Program., 91 (2002), pp.239-270

[4] R. Fletcher, S. Leyffer, Ph. Toint, *A brief history of filter methods,* SIAG/OPT Views-and-News, A Forum for the SIAM Activity Group on Optimization, Vol. 18, No. 1 (2007), pp.2-12

[5] N. I. M. Gould, S. Leyffer, Ph. L. Toint, *A multidimensional filter algorithm for nonlinear equations and nonlinear least-squares,* SIAM J. Optim., Vol. 15, No. 1 (2004), pp.17-38

[6] T. G. Kolda, R. M. Lewis, V. Torezon, *Optimization by direct search: New perspectives on some classical and modern methods*, SIAM Review, Vol.45, No.3 (2003), pp.385-482

[7] N. Krejic, Z. Luzanin, I. Stojkovska, *Gauss-Newton-based BFGS method with filter for unconstrained minimization*, Applied Mathematics and Computation, Vol.211, No.2 (2009), pp.354-362

[8] D. H. Li, M. Fukushima, *A derivative-free line search and global convergence of Broyden-Like method for nonlinear equations,* Optim. Methods and Software, No.13 (2000), pp.181-201

[9] J. Nocedal, S. J. Wright, *Numerical optimization,* Springer-Verlag, New York, (1999)

[10] W. Y. Sun, J. Y. Han, J. Sun, *Global convergence of non-monotone descent methods for unconstrained optimization problems,* J. Comput. Appl. Math., No.146 (2002), pp.89-98

[11] A. Wächter, L. T. Biegler, *Line search filter methods for nonlinear programming: motivation and global convergence,* SIAM J. Optim., Vol. 16, No. 1 (2005), pp.1-31

[12] H. Zhang, W. W. Hager, *A nonmonotone line search technique and its application to unconstrained optimization,* SIAM J. Optim., Vol 14 (2004), No. 4, pp.1043-1056

ФИЛТЕР МЕТОД СО ЛОКАЛНИ ВАРИЈАЦИИ БЕЗ ПРЕСМЕТУВАЊЕ НА ИЗВОДИ ЗА РЕШАВАЊЕ НА СИСТЕМИ ОД НЕЛИНЕАРНИ РАВЕНКИ

Ирена Стојковска

Апстракт. Во овој труд изложен е нов филтер метод со линиско пребарување без пресметување на изводи за решавање на системи од нелинеарни равенки. Правилото на линиско пребарување е базирано на нов концепт на принудувачка функција (forcing function) при што воспоставена е конвергенција под послаби услови. Повеќедимензионалниот филтер кој се користи при линиското пребарување помага во избегнувањето на малите чекори добиени од процедурата на константно намалување на чекорот (backtracking) или од интерполацијата. Повремено е дозволено случајно бирање на правецот на пребарување со што се избегнува пресметувањето на изводите или нивна апроксимација. Воспоставена е конвергенција во веројатносна смисла. Предложениот метод исто така може да се примени и при решавање на задачи на безусловна оптимизација, нелинеарни комплементарни проблеми и варијационо неравенствени проблеми.