On the Global Convergence of Derivative Free Methods for Unconstrained Optimization.

SIAM J. Optim.
Vol. 13, No. 1, pp. 97–116

ON THE GLOBAL CONVERGENCE OF DERIVATIVE-FREE METHODS FOR UNCONSTRAINED OPTIMIZATION*

STEFANO LUCIDI† AND MARCO SCIANDRONE‡

Abstract. In this paper, starting from the study of the common elements that some globally convergent direct search methods share, a general convergence theory is established for unconstrained minimization methods employing only function values. The introduced convergence conditions are useful for developing and analyzing new derivative-free algorithms with guaranteed global convergence. As examples, we describe three new algorithms which combine pattern and line search approaches.

Key words. unconstrained minimization, derivative-free methods

AMS subject classifications. 90C56, 65K05

PII. S1052623497330392

1. Introduction.

In this paper, we consider the problem of the form

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

where $f : R^n \to R$ is a continuously differentiable function and where the first order derivatives of f can be neither calculated nor approximated explicitly.

The interest in studying minimization algorithms for solving these optimization problems derives from the increasing demand from industrial and scientific applications for such tools. Many derivative-free methods have been proposed in the literature; descriptions of these methods can be found, for instance, in [13] and [19].

An important class of such methods is formed by the so-called *direct search methods*, which base the minimization procedure on the comparison of objective function values computed on suitable trial points. Two particular subclasses of globally convergent direct search methods are the following:

- *pattern search methods* (see, e.g., [2], [6], [16], [19]), which present the distinguishing feature of evaluating the objective function on specified geometric patterns;
- *line search methods* (see, e.g., [1], [4], [5], [8], [10], [11], [12], [17], [20]), which draw their inspiration from the gradient-based minimization methods and perform one dimensional minimizations along suitable directions.

These two classes of methods present different interesting features. In fact, the pattern search methods can accurately sample the objective function in a neighborhood of a point and, hence, can identify a "good" direction, namely, a direction along which the objective function decreases significantly. The line search algorithms can perform large steps along the search directions and, hence, can exploit to a large extent the possible goodness of the directions. Therefore it could be worthwhile to combine

*Received by the editors November 18, 1997; accepted for publication (in revised form) November 28, 2001; published electronically June 5, 2002. This work was supported by Agenzia Spaziale Italiana, Rome, Italy.
http://www.siam.org/journals/siopt/13-1/33039.html

†Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza," Via Buonarroti 12 00185 Roma, Italy (lucidi@dis.uniroma1.it).

‡Istituto di Analisi dei Sistemi ed Informatica, CNR, Viale Manzoni 30 00185 Roma, Italy (sciandro@iasi.rm.cnr.it).

these approaches in order to define new classes of derivative-free algorithms that could exploit as much as possible their different features, namely, algorithms which are able to determine "good" directions and to perform "significant" steplengths along such directions. Some examples of methods combining different direct search approaches have already been proposed in [3], [10], [14], [15], [17], [18]. In this paper, on the basis of the convergence analyses reported in [5], [7], and [16] for pattern and line search methods, respectively, we give general sufficient conditions for ensuring the global convergence of a sequence of points. These conditions, which do not require any information on first order derivatives, can be used as the basis for developing new globally convergent derivative-free algorithms and, in particular, algorithms which can follow a mixed pattern-line search approach.

More specifically, in section 2, we start by identifying the common key features of the pattern and line search methods which are behind their global convergence properties. This analysis indicates that the global convergence of a derivative-free algorithm can be guaranteed by satisfying some minimal and quite natural requirements on the search directions used and on the sampling of the objective function along these directions. Then, in section 3, we analyze theoretical requirements regarding the search directions. In section 4, we define general conditions sufficient to ensure global convergence without gradient information. Finally, in section 5, we propose new globally convergent algorithms which combine pattern and line search approaches. The appendix contains the proofs of two technical results.

Notation. We indicate by $\|\cdot\|$ the Euclidean norm (on the appropriate space). A subsequence of $\{x_k\}$ corresponding to an infinite subset K will be denoted by $\{x_k\}_K$. Given two sequences of scalars $\{u_k\}$ and $\{v_k\}$ such that

$$\lim_{k\to\infty} u_k = 0 \qquad \text{and} \qquad \lim_{k\to\infty} v_k = 0,$$

we say that $u_k = o(v_k)$ if

$$\lim_{k\to\infty} \frac{u_k}{v_k} = 0.$$

As usual we say that a set of directions $\{p^1, p^2, \dots, p^r\}$ positively span R^n if for every $x \in R^n$ there exist $\lambda_i \geq 0$, for $i = 1, \dots, p$, such that

$$x = \sum_{i=1}^{r} \lambda_i p^i.$$

Finally, we denote by e^i, with $i = 1, \dots, n$, the orthonormal set of the coordinate directions.

2. Preliminary remarks. It is well known that, when the gradient is available, to define a globally convergent algorithm for unconstrained problems is not a difficult task. In fact, at each iteration, the gradient allows us

- to compute and select a "good" descent direction, namely, a direction along which the objective function decreases with a suitable rate;
- to determine a "sufficiently" large steplength along a descent search direction, namely, a steplength which is able to exploit the descent property of the search direction by enforcing a significant decrease in the value of the objective function relative to the norm of the gradient.

When the gradient is not available, we lose information about the local behavior of the objective function. In fact, the ith component $\nabla_i f$ of the gradient is the directional derivative of the objective function along the vector e^i, and $-\nabla_i f$ is the directional derivative along the vector $-e^i$. Therefore, the whole gradient vector provides the rate of change of the objective function along the $2n$ directions $[e^1, e^2, \dots, e^n, -e^1, -e^2, \dots, -e^n]$. This fact guarantees that the gradient information characterizes quite accurately the local behavior of the objective function in a neighborhood of the point at which the derivatives are computed.

Most of the algorithms belonging to the class of direct search methods follow, more or less visibly, the same strategy to overcome the lack of first order information contained in the gradient. Their common approach is based on the idea of investigating the behavior of the objective function in a neighborhood of the generic point by sampling the objective function along a set of directions. Clearly each of these algorithms presents properties and features which depend on the particular choice of the sets of directions and on the particular way in which the samplings of the objective function are performed.

The directions to be used in a derivative-free algorithm should be such that the local behavior of the objective function along them is sufficiently indicative of the local behavior of the function in a neighborhood of a point. Roughly speaking, these directions should have the property that, performing finer and finer samplings of the objective function along them, it is possible either

(i) to realize that the current point is a good approximation of a stationary point of the objective function, or

(ii) to find a specific direction along which the objective function decreases.

The important point is to identify larger and larger classes of sets of search directions which can be used to define globally convergent derivative-free algorithms. To this end, in the next section, we propose a general condition which formally characterizes classes of sets of directions complying with the properties (i) and (ii).

In addition to contributing to the previous points (i) and (ii), the method of sampling has the task of guiding the choice of the new point so as to ensure that the sequence of points produced by the algorithm is globally convergent towards stationary points of the objective function. On the basis of the common features of the sampling techniques of the direct search methods proposed in [5], [7], and [16], in section 4 we define sufficient conditions on the samplings of the objective function along suitable directions for the global convergence of a derivative-free method. Similar conditions were given in [18]; however, the ones proposed in this work are more general.

3. Search directions. Before describing our analysis, we recall the following basic assumption on the objective function.

Assumption A1. The function $f : R^n \to R$ is continuously differentiable.

As said before, the first step in defining a direct search method is to associate a suitable set of search directions p_k^i, $i = 1, \dots, r$, with each point x_k produced by the algorithm. This set of directions should have the property that the local behavior of the objective function along them provides sufficient information to overcome the lack of the gradient.

Here, we introduce a new condition which characterizes the sets of directions p_k^i, $i = 1, \dots, r$, that satisfy this property. This condition requires that the distance between the points generated by an algorithm and the set of stationary points of the objective function tends to zero if and only if the directional derivatives of the objective function along the directions p_k^i, $i = 1, \dots, r$, tend to assume nonnegative values. Formally we have the following condition.

Condition C1. Given a sequence of points $\{x_k\}$, the sequence directions $\{p_k^i\}$, $i = 1, \dots, r$, are bounded and such that

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0 \quad \text{if and only if} \quad \lim_{k\to\infty} \sum_{i=1}^{r} \min\{0, \nabla f(x_k)^T p_k^i\} = 0.$$

By drawing our inspiration from some results established in [7] and [16], we state the following proposition, which points out a possible interest in the sets of directions satisfying Condition C1.

PROPOSITION 3.1. *Let $\{x_k\}$ be a bounded sequence of points and let $\{p_k^i\}$, $i = 1, \dots, r$, be sequences of directions which satisfy Condition* C1. *For every $\eta > 0$, there exist $\gamma > 0$ and $\delta > 0$ such that, for all but finitely many k, if x_k satisfies $\|\nabla f(x_k)\| \ge \eta$, then there exists a direction $p_k^{i_k}$, with $i_k \in \{1, \dots, r\}$, for which*

$$f(x_k + \alpha p_k^{i_k}) \le f(x_k) - \gamma\alpha \|\nabla f(x_k)\| \|p_k^{i_k}\| \tag{3.1}$$

for all $\alpha \in (0, \delta]$.

Proof. For the proof, see the appendix. □

The previous proposition guarantees that, whenever the current point is not a stationary point, it is possible to enforce sufficient decrease of the objective function by using sets of directions satisfying Condition C1. In other words, this ensures that Condition C1 implies that the sets of directions are able to comply with the requirement (ii) discussed in section 2.

From a theoretical point of view, Proposition 4.1, given in the next section, shows that Condition C1 is a sufficient requirement for the search directions to ensure the global convergence of the sequence of iterates (or at least one subsequence) to a stationary point of f. Roughly speaking, the role of Condition C1 in the field of derivative-free methods can be similar to that of the *gradient-related condition* used in the field of gradient-based algorithms. In fact, Condition C1 can be considered a mild technical condition on the sets of search directions which can be either naturally satisfied or easily enforced in a derivative-free algorithm (see Algorithm 3 in section 5).

In order to show that Condition C1 is a viable requirement on the search directions, we report two classes of sets of directions satisfying Condition C1 and some examples of these classes. The classes introduced here generalize the ones proposed in [7].

Classes of sets of search directions.

(a) The sequences $\{p_k^i\}$, with $i = 1, \dots, r$, are bounded, and every limit point $(\bar{p}^1, \dots, \bar{p}^r)$ of the sequence $\{p_k^1, \dots, p_k^r\}$ is such that the vectors $\bar{p}^i$, with $i = 1, \dots, r$, positively span R^n.

(b) The sequences $\{p_k^i\}$, with $i = 1, \dots, r$, are bounded; the vectors p_k^i, $i = 1, \dots, n$, are uniformly linearly independent; and, for all k, there exists a direction p_k^{n+j}, with $j \geq 1$, given by

$$p_k^{n+j} = \sum_{\ell=1}^{2n} \rho_k^\ell \frac{(v_k^1 - v_k^\ell)}{\tilde{\xi}_k^\ell}, \tag{3.2}$$

where

- the sequences $\{\rho_k^\ell\}$, $\ell = 1, \dots, 2n$, are bounded and such that $\rho_k^\ell \geq 0$ with $\rho_k^{2n} \geq \bar{\rho} > 0$ for all k;
- $\{v_k^1, v_k^2, \dots, v_k^{2n}\} = \{z_k^1, z_k^1 + \xi_k^1 p_k^1, z_k^2, z_k^2 + \xi_k^2 p_k^2, \dots, z_k^n, z_k^n + \xi_k^n p_k^n\}$, with the points v_k^ℓ, for $\ell = 1, \dots, 2n$, ordered (and possibly relabeled) so that

  $$f(v_k^1) \leq f(v_k^2) \leq \cdots \leq f(v_k^{n-1}) \leq \cdots \leq f(v_k^{2n}), \tag{3.3}$$

  and the sequences $\{\xi_k^i\}$ and $\{z_k^i\}$, for $i = 1, \dots, n$, are such that, for all k,

  $$\xi_k^i > 0, \tag{3.4}$$

  $$\frac{\max\limits_{i=1,\dots,n} \{\xi_k^i\}}{\min\limits_{i=1,\dots,n} \{\xi_k^i\}} \leq c_1, \tag{3.5}$$

  $$\|z_k^i - x_k\| \leq c_2 \xi_k^i, \tag{3.6}$$

  where c_1,$c_2 > 0$, and such that

  $$\lim_{k \to \infty} \xi_k^i = 0; \tag{3.7}$$

- the sequences $\{\tilde{\xi}_k^\ell\}$, $\ell = 1, \dots, 2n$, are such that $\min_{i=1,\dots,n}\{\xi_k^i\} \leq \tilde{\xi}_k^\ell \leq \max_{i=1,\dots,n}\{\xi_k^i\}$.

For the classes of sets of search directions we can state the following proposition.

PROPOSITION 3.2. *Let $\{x_k\}$ be a bounded sequence of points, and let $\{p_k^i\}$, $i = 1, \dots, r$, be sequences of directions belonging to class* (a) *or class* (b)*. Then, Condition* C1 *is satisfied.*

Proof. For the proof, see the appendix. ∎

Two examples of sets of directions belonging to the classes (a) and (b) are described in [7]. These classes are defined starting from a set of n uniformly linearly independent search directions, for example,

$$p_k^1 = e^1, \qquad p_k^2 = e^2, \qquad \dots, \qquad p_k^n = e^n. \tag{3.8}$$

Then, to obtain a set of class (a), it is sufficient to consider also the directions

$$p_k^{n+1} = -e^1, \qquad p_k^{n+2} = -e^2, \qquad \dots, \qquad p_k^{2n} = -e^n$$

or just the direction

$$p_k^{n+1} = -\sum_{i=1}^{n} e^i.$$

A set of class (b) can be obtained by adding to (3.8) the direction

$$p_k^{n+1} = \frac{x_k - x_k^{max}}{\xi_k},$$

where $x_k^{max} = \arg\max_{i=1,\dots,n}\{f(x_k+\xi_k p_k^i)\}$ and $\xi_k \to 0$ for $k \to \infty$. This corresponds to setting

$$z_k^i = x_k, \qquad \xi_k^i = \xi_k \qquad \text{for } i = 1,\dots,n,$$

$$\xi_k^{2n} = \xi_k,$$

$$\rho_k^1 = \rho_k^2 = \cdots = \rho_k^{2n-1} = 0, \qquad \rho_k^{2n} = 1.$$

A new class of sets of search directions satisfying Condition C1 will be defined within Algorithm 3 proposed in section 5. In particular, this class is constructed during the minimization procedure so as to exploit as much as possible the information on the objective function obtained in the preceding iterations.

4. Global convergence conditions. In this section we show that the global convergence of an algorithm can be guaranteed by means of the existence of suitable sequences of points along search directions p_k^i, $i = 1,\dots,r$, satisfying Condition C1. In particular, by using Condition C1 we can characterize a stationary point of f with the fact that the objective function does not decrease locally along the directions p_k^i, $i = 1,\dots,r$, in points sufficiently close to the current point x_k. This leads to the possibility of defining new general conditions for the global convergence of derivative-free algorithms by means of the existence of sequences of points showing that the objective function does not decrease along the directions p_k^i, $i = 1,\dots,r$. These conditions, even if very simple and intuitive, allow us to identify some minimal requirements on acceptable samplings of the objective function along the directions p_k^i, $i = 1,\dots,r$, that guarantee the global convergence of the method.

In the remainder of the paper we suppose that the following standard assumption holds.

Assumption A2. The level set

$$\mathcal{L}_0 = \{x \in R^n : f(x) \le f(x_0)\}$$

is compact.

The following proposition describes a set of global convergence conditions.

PROPOSITION 4.1. *Let $\{x_k\}$ be a sequence of points; let $\{p_k^i\}$, $i = 1,\dots,r$, be sequences of directions; and suppose that the following conditions hold:*

(a) *$f(x_{k+1}) \le f(x_k)$;*

(b) *$\{p_k^i\}$, $i = 1,\dots,r$, satisfy Condition* C1;

(c) *there exist sequences of points $\{y_k^i\}$ and sequences of positive scalars $\{\xi_k^i\}$, for $i = 1,\dots,r$, such that*

$$f(y_k^i + \xi_k^i p_k^i) \ge f(y_k^i) - o(\xi_k^i), \tag{4.1}$$

$$\lim_{k\to\infty} \xi_k^i = 0, \tag{4.2}$$

$$\lim_{k\to\infty} \|x_k - y_k^i\| = 0. \tag{4.3}$$

Then,

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{4.4}$$

Proof. From (a) it follows that $\{f(x_k)\}$ is a nonincreasing sequence, so that $\{x_k\}$ belongs to the compact set $\mathcal{L}_0$ and admits at least one limit point. Let $\bar{x}$ be any limit point of $\{x_k\}$. Then, there exists a subset $K_1 \subseteq \{0, 1, \ldots\}$ such that

$$\lim_{k\to\infty, k\in K_1} x_k = \bar{x},$$

$$\lim_{k\to\infty, k\in K_1} p_k^i = \bar{p}^i, \quad i = 1, \ldots, r.$$

Using (4.3), it follows that

$$\lim_{k\to\infty, k\in K_1} y_k^i = \bar{x}, \qquad i = 1, \ldots, r.$$

Now, recalling (4.1) for all $k \geq 0$, we have

$$f(y_k^i + \xi_k^i p_k^i) - f(y_k^i) \geq -o(\xi_k^i), \quad i = 1, \ldots, r. \tag{4.5}$$

By the mean-value theorem, we can write

$$f(y_k^i + \xi_k^i p_k^i) - f(y_k^i) = \xi_k^i \nabla f(u_k^i)^T p_k^i, \quad i = 1, \ldots, r, \tag{4.6}$$

where $u_k^i = y_k^i + \lambda_k^i \xi_k^i p_k^i$, with $\lambda_k^i \in (0, 1)$. By substituting (4.6) into (4.5), we obtain

$$\nabla f(u_k^i)^T p_k^i \geq -o(\xi_k^i), \quad i = 1, \ldots, r. \tag{4.7}$$

Now, it is easily seen from (4.2), taking into account the boundedness of p_k^i, that $u_k^i \to \bar{x}$ as $k \to \infty$ and $k \in K_1$. Hence, by the continuity of ∇f, from (4.7) and recalling (4.2), we get

$$\lim_{k\to\infty, k\in K_1} \nabla f(u_k^i)^T p_k^i = \nabla f(\bar{x})^T \bar{p}_i \geq 0, \quad i = 1, \ldots, r.$$

Then, recalling (b) and Condition C1, we have that

$$\nabla f(\bar{x}) = 0.$$

As $\bar{x}$ is any limit point of $\{x_k\}$, we conclude that

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0. \qquad \square$$

Roughly speaking, according to (c), for each search direction p_k^i, the existence of suitable points y_k^i and $y_k^i + \xi_k^i p_k^i$ related to the "current" point x_k is assumed (see (4.2) and (4.3)) whenever a "failure" of a (sufficient) strict decrease of f occurs (see (4.1)).

Then, also considering (4.2), we have that at the point y_k^i the directional derivative of f along p_k^i can be approximated by a quantity which tends to be nonnegative. Therefore, due to the property of the search directions expressed by Condition C1, the global convergence of the sequence $\{x_k\}$ can be ensured by requiring that the failure points "cluster" more and more around x_k (see (4.3)).

Similar conditions were given in [18]; however, those of Proposition 4.1 are more general in the requirements placed on both the search directions p_k^i and the trial steps ξ_k^i, $i = 1, \ldots, r$.

The use of directions satisfying Condition C1 and the result of producing sequences (or subsequences) of points that satisfy the hypothesis of Proposition 4.1 are the common elements of the globally convergent derivative-free algorithms proposed in [5] and [16], which consider the pattern and line search approaches, respectively. This point is discussed in more detail in [9], where the known global convergence results of different algorithms are reobtained by using Condition C1 and Proposition 4.1. With regard to (4.1) and (4.2) of Proposition 4.1(c), we note only that

- in the pattern search algorithms, the failures (4.1) (with $o(\xi_k^i) = 0$) occur "naturally" by requiring only a simple decrease of f, while (4.2) follows by imposing further restrictions on the search directions and on the steplengths;
- in the line search algorithms, (4.1) and (4.2) are satisfied by enforcing a "sufficient" decrease of f depending on ξ_k^i and without imposing further restrictions on the search directions.

5. New globally convergent algorithms. In this section we try to motivate further the possible practical interest of the analysis performed in sections 3 and 4, by showing that Condition C1 and Proposition 4.1 can play the role of guidelines for defining new derivative-free algorithms and for analyzing their convergence properties.

Since the conditions given in Proposition 4.1 capture some common theoretical features of pattern and line search approaches, they are suitable for defining algorithms which combine these two approaches. In particular, our aim is to propose algorithms which are able to

- get sufficient information on the local behavior of the objective function f, like in a pattern strategy;
- exploit the possible knowledge of a "good" direction, like in a line search strategy.

In this section, as examples, we describe three new algorithms (Algorithm 1, Algorithm 2, and Algorithm 3). The basic idea of these algorithms is to sample, at each iteration k, the objective function f along a set $\{p_k^i\}_{i=1}^r$ of search directions. This is performed with the aim of detecting a "promising" direction (like in a pattern strategy), that is, a direction along which the objective function decreases "sufficiently." Then, once such a direction has been detected, a "sufficiently" large step is performed along it. Both the "sufficient" decrease of the objective function and the "sufficient" steplength are evaluated by means of criteria derived from the line search approach. These criteria, requiring sufficient decrease of the objective function, are stronger than the ones used in the pattern search algorithms (where the simple reduction of f is allowed). However, as we said before, they allow us more freedom in the choice of search directions and in the steplengths used to sample the objective function.

In particular, in Algorithm 1 and Algorithm 2, we assume that the sets of search directions satisfying Condition C1 are given. Algorithm 1 is very simple, and its scheme is similar to that of a pattern search algorithm. For this algorithm we can prove that at least one accumulation point of the sequence produced is a stationary point of f. In Algorithm 2 a line search technique is introduced to exploit as much as possible a promising direction identified by the algorithm. For this algorithm we prove that any convergent subsequence generated by the algorithm tends to a stationary point of f. The approach of Algorithm 3 is the same as that of Algorithm 2; the distinguishing feature of Algorithm 3 is that of using sets of $n+1$ directions, in which the first n are given and the last one is computed on the basis of the information iteratively obtained with the aim of identifying a "good" direction. For this algorithm we prove the same convergence result stated for Algorithm 2.

The first algorithm is the following.

ALGORITHM 1.

Data. $x_0 \in R^n$, $\tilde{\alpha}_0 > 0$, $\gamma > 0$, $\theta \in (0,1)$.

Step 0. Set $k = 0$.

Step 1. If there exists $y_k \in R^n$ such that

$$f(y_k) \leq f(x_k) - \gamma \tilde{\alpha}_k,$$

then go to Step 4.

Step 2. If there exists $i \in \{1, \dots r\}$ and an $\alpha_k \geq \tilde{\alpha}_k$ such that

$$f(x_k + \alpha_k p_k^i) \leq f(x_k) - \gamma (\alpha_k)^2,$$

then set $y_k = x_k + \alpha_k p_k^i$, $\tilde{\alpha}_{k+1} = \alpha_k$ and go to Step 4.

Step 3. Set $\tilde{\alpha}_{k+1} = \theta \tilde{\alpha}_k$ and $y_k = x_k$.

Step 4. Find x_{k+1} such that $f(x_{k+1}) \leq f(y_k)$, set $k = k+1$, and go to Step 1.

Algorithm 1 follows an approach similar to that of a pattern search algorithm. In particular, at each iteration it is possible to accept any single point for which sufficient decrease of the objective function is realized (Step 1). The stepsize α_k is reduced only when it is not possible to locally enforce the sufficient reduction of f along the search directions p_k^i, for $i = 1, \dots, r$ (Steps 2–3). At Step 4 the algorithm can accept any point which produces an improvement of the objective function with respect to the selected point y_k.

We note that, at Step 2, any extrapolation technique can be attempted to determine a good stepsize α_k whenever a suitable direction has been detected. However, the use of an extrapolation technique is not necessary to guarantee global convergence. (In particular, it is enough to use $\alpha_k = \tilde{\alpha}_k$.) Furthermore, we point out that, even if a set of r search directions p_k^i, $i = 1, \dots, r$, is associated to the current point x_k, so long as a sufficient decrease condition has been satisfied along a direction $p_k^{\bar{i}}$, the remaining directions can be ignored. This is a feature that Algorithm 1 has in common with the weak form of pattern search algorithms (see [16]).

Finally, we observe also that Step 1 and Step 4 allow the possibility of using any approximation scheme for the objective function to produce a new better point.

The convergence properties of the algorithm are reported in the following proposition.

PROPOSITION 5.1. *Let $\{x_k\}$ be the sequence produced by Algorithm 1. Suppose* *that the sequences of directions* $\{p_k^i\}_{i=1}^r$ *satisfy Condition* C1*. Then we have*

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{5.1}$$

Proof. We prove (5.1) by showing that conditions (a), (b), and (c) of Proposition 4.1 are satisfied (at least) by a subsequence of $\{x_k\}$.

Condition (a) follows from the instructions of the algorithm. Condition (b) is obviously true. Therefore we concentrate on Condition (c).

We can split the iteration sequence $\{k\}$ into three parts, K_1, K_2, and K_3, namely, those iterations where the test at Step 1 is satisfied, those where the test at Step 2 is satisfied, and those where Step 3 is performed. In particular, if $k \in K_1$, we have

$$f(x_{k+1}) \le f(x_k) - \gamma\tilde{\alpha}_k; \tag{5.2}$$

if $k \in K_2$, we have

$$f(x_{k+1}) \le f(x_k) - \gamma(\alpha_k)^2 \le f(x_k) - \gamma(\tilde{\alpha}_k)^2; \tag{5.3}$$

and if $k \in K_3$, we have

$$f(x_k + \tilde{\alpha}_k p_k^i) > f(x_k) - \gamma(\tilde{\alpha}_k)^2 \qquad \text{for } i = 1, \dots, r. \tag{5.4}$$

If K_1 is an infinite subset, then (5.2), the compactness of the level set $\mathcal{L}_0$, the continuity assumption on f, and Condition (a) imply

$$\lim_{k\to\infty, k\in K_1} \tilde{\alpha}_k = 0. \tag{5.5}$$

Now, let us assume that K_2 is an infinite subset. From (5.3), by repeating the same reasoning, we obtain

$$\lim_{k\to\infty, k\in K_2} \tilde{\alpha}_k = 0. \tag{5.6}$$

Now for each $k \in K_3$ let m_k be the biggest index such that $m_k < k$ and $m_k \in K_1 \cup K_2$. Then we have

$$\tilde{\alpha}_{k+1} = \theta^{k-m_k}\tilde{\alpha}_{m_k}. \tag{5.7}$$

(We can assume that $m_k = 0$ if the index m_k does not exist; that is, K_1 and K_2 are empty.)

As $k \to \infty$ and $k \in K_3$, either $m_k \to \infty$ (if $K_1 \cup K_2$ is an infinite subset) or $(k - m_k) \to \infty$ (if $K_1 \cup K_2$ is finite). Therefore, (5.7) together with (5.5) and (5.6) or the fact that $\theta \in (0, 1)$ yields

$$\lim_{k\to\infty, k\in K_3} \tilde{\alpha}_k = 0. \tag{5.8}$$

Thus, by using (5.5), (5.6), and (5.8), we can write

$$\lim_{k\to\infty} \tilde{\alpha}_k = 0. \tag{5.9}$$

From (5.9) it follows that there exists an infinite subset $K \subseteq \{0, 1, \dots\}$ such that $\tilde{\alpha}_{k+1} < \tilde{\alpha}_k$ for all $k \in K$; namely, Step 3 is performed for all $k \in K$. Therefore, we have $K \subseteq K_3$, and hence (5.4) holds for all $k \in K$. Now, with reference to condition (c) of Proposition 4.1, for each $k \in K$ we set

$$\xi_k^i = \tilde{\alpha}_k, \quad y_k^i = x_k, \qquad i = 1, \dots, r. \tag{5.10}$$

Then we have

$$f(y_k^i + \xi_k^i p_k^i) \geq f(y_k^i) - \gamma \left(\xi_k^i\right)^2;$$

moreover, recalling (5.9), it follows that

$$\lim_{k\to\infty, k\in K} \xi_k^i = 0,$$

so that (4.1) and (4.2) hold. Finally, (4.3) follows directly from (5.10), and this concludes the proof. ∎

Now we define pattern-line search algorithms producing sequences of points with the stronger property that every limit point is a stationary point of f. This additional property can be obtained by investigating in more detail the behavior of the objective function along the search directions p_k^i, $i = 1, \dots, r$, and by using a derivative-free line search technique to ensure sufficiently large movements along any "good" direction identified by the algorithm. The first of these algorithms is the following.

ALGORITHM 2.

Data. $x_0 \in R^n$, $\tilde{\alpha}_0^i > 0$, $i = 1, \dots, r$, $\gamma > 0$, $\delta, \theta \in (0, 1)$.

Step 0. Set $k = 0$.

Step 1. Set $i = 1$ and $y_k^1 = x_k$.

Step 2. If $f(y_k^i + \tilde{\alpha}_k^i p_k^i) \leq f(y_k^i) - \gamma(\tilde{\alpha}_k^i)^2$, then
compute α_k^i by *LS Procedure*($\tilde{\alpha}_k^i, y_k^i, p_k^i, \gamma, \delta$)
and set $\tilde{\alpha}_{k+1}^i = \alpha_k^i$;
else set $\alpha_k^i = 0$ and $\tilde{\alpha}_{k+1}^i = \theta\tilde{\alpha}_k^i$.
Set $y_k^{i+1} = y_k^i + \alpha_k^i p_k^i$.

Step 3. If $i < r$, set $i = i + 1$ and go to Step 2.

Step 4. Find x_{k+1} such that

$$f(x_{k+1}) \leq f(y_k^{r+1}),$$

set $k = k + 1$, and go to Step 1.

LS PROCEDURE($\tilde{\alpha}_k^i, y_k^i, p_k^i, \gamma, \delta$).

Compute $\alpha_k^i = \min\{\delta^{-j}\tilde{\alpha}_k : j = 0, 1, \dots\}$ such that

$$f(y_k^i + \alpha_k^i p_k^i) \leq f(x_k) - \gamma\big(\alpha_k^i\big)^2, \tag{5.11}$$

$$f\left(y_k^i + \frac{\alpha_k^i}{\delta} p_k^i\right) \geq \max\left[f(y_k^i + \alpha_k^i p_k^i), f(y_k^i) - \gamma\bigg(\frac{\alpha_k^i}{\delta}\bigg)^2\right]. \tag{5.12}$$

At each iteration k the algorithm examines the behavior of the objective function along all the search directions p_k^i, $i = 1, \dots, r$ (Steps 1–3). However, whenever it detects a direction p_k^i where the function is sufficiently decreased, the algorithm produces a new point by performing a "sufficiently" large movement along this direction. This point is determined by means of a suitable stepsize α_k^i computed by a line search technique (LS Procedure). At Step 4, similarly to Algorithm 1, the new point x_{k+1} can be the point y_k^{r+1} produced by Steps 1–3 or any point where the objective function is improved with respect to $f(y_k^{r+1})$. This fact, as said before, allows us to adopt any approximation scheme for the objective function to produce a new better point and hence to improve the efficiency of the algorithm without affecting its convergence properties.

Comparing Algorithms 1 and 2, it is easy to observe that Algorithm 2 requires stronger conditions to produce the new point. In fact, all the directions must be investigated at each iteration, and the use of a line search technique is necessary. However, in Algorithm 2 it is possible to associate to each direction p_k^i a different initial stepsize $\tilde{\alpha}_k^i$, which is updated on the basis of the behavior of the objective function along p_k^i observed in the current iteration. This feature can be useful when the search directions are the same for all iterations ($p_k^i = \bar{p}^i$, $i = 1, \dots, r$, for all k). In fact, in this case, the instructions of the algorithm should guarantee that the initial stepsizes $\tilde{\alpha}_k^i$, $i = 1, \dots, r$, take into account the different behavior of f along different search directions.

Finally, we note that Algorithm 2, similarly to the strong form of pattern search algorithms, is required to examine, at each iteration, the local behavior of f along all the r directions p_k^i, $i = 1, \dots, r$. However, at each iteration the current point x_k is updated by means of intermediate points y_k^{i+1} whenever sufficient decrease of f is obtained along any of the search directions p_k^i, $i \in \{1, \dots, r\}$.

From a theoretical point of view, it is possible to state the following convergence result, which is stronger than the one obtained for Algorithm 1.

PROPOSITION 5.2. *Let $\{x_k\}$ be the sequence produced by Algorithm 2. Suppose that the sequences of directions $\{p_k^i\}_{i=1}^r$ satisfy Condition C1. Then, Algorithm 2 is well defined and we have*

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{5.13}$$

Proof. In order to prove that Algorithm 2 is well defined, we must show that, given an integer $i \le r$ such that the test of Step 2 is satisfied, there exists a finite integer j for which (5.11) and (5.12) hold with $\alpha_k^i = \delta^{-j}\tilde{\alpha}_k^i$. With this goal, we give a proof by contradiction. We assume that either

$$f(y_k^i + \delta^{-j}\tilde{\alpha}_k^i p_k^i) < f(y_k^i) - \gamma\big(\delta^{-j}\tilde{\alpha}_k^i\big)^2 \qquad \text{for all } j$$

or

$$f(y_k^i + \delta^{-j-1}\tilde{\alpha}_k^i p_k^i) < f(y_k^i + \delta^{-j}\tilde{\alpha}_k^i p_k^i) \le f(y_k^i) - \gamma\big(\delta^{-j}\tilde{\alpha}_k^i\big)^2 \qquad \text{for all } j.$$

Then, taking the limits for $j \to \infty$, we obtain in both cases that f is unbounded below, which contradicts Assumption A2.

Now we prove (5.13) by showing that conditions (a), (b), and (c) of Proposition 4.1 are satisfied.

Condition (a) follows from the instructions of the algorithm. Condition (b) is obviously true. Then we must show that condition (c) holds.

We first prove that for $i = 1, \dots, r$ we have

$$\lim_{k\to\infty} \alpha_k^i = 0 \tag{5.14}$$

and

$$\lim_{k\to\infty} \tilde{\alpha}_k^i = 0. \tag{5.15}$$

From the instructions of the algorithm we have

$$f(x_{k+1}) \le f(y_k^{r+1}) \le f(x_k) - \gamma \sum_{i=1}^{r} \left(\alpha_k^i\right)^2,$$

so that, since $\{x_k\}$ belongs to the compact set $\mathcal{L}_0$, $\{f(x_k)\} \to \bar{f}$ and hence $\alpha_k^i \to 0$, for $i = 1, \dots r$. Given $i \in \{1, \dots r\}$, we split the iteration sequence $\{k\}$ into two parts, K and $\bar{K}$, namely, those iterations where $\alpha_k^i > 0$ and those where $\alpha_k^i = 0$. For all $k \in K$ we have $\alpha_k^i \ge \tilde{\alpha}_k^i$, so that, if K is an infinite subset, it follows that

$$\lim_{k\to\infty, k\in K} \tilde{\alpha}_k^i = 0. \tag{5.16}$$

For each $k \in \bar{K}$, let m_k be the biggest index such that $m_k < k$ and $m_k \in K$. (We can assume $m_k = 0$ if the index m_k does not exist, that is, K is empty.) Then we have

$$\tilde{\alpha}_k^i = (\theta)^{k-m_k} \tilde{\alpha}_{m_k}^i.$$

As $k \to \infty$ and $k \in \bar{K}$, either $m_k \to \infty$ (if K is an infinite subset) or $(k - m_k) \to \infty$ (if K is finite). Therefore, (5.16) and the fact that $\theta \in (0, 1)$ imply $\lim_{k\to\infty, k\in\bar{K}} \tilde{\alpha}_k^i = 0$. Now, with reference to condition (c) of Proposition 4.1, we set

$$\xi_k^i = \begin{cases} \dfrac{\alpha_k^i}{\delta} & \text{if } k \in K, \\ \tilde{\alpha}_k^i & \text{if } k \in \bar{K}. \end{cases} \tag{5.17}$$

Then, we have $f(y_k^i + \xi_k^i p_k^i) \ge f(y_k^i) - (\xi_k^i)^2$; moreover, recalling (5.14) and (5.15), it follows that $\lim_{k\to\infty} \xi_k^i = 0$, so that (4.1) and (4.2) hold. Finally, since we have

$$\|y_k^i - x_k\| \le \sum_{j=1}^{i-1} \alpha_k^j \|p_k^j\|,$$

by again using (5.14) it follows that

$$\lim_{k\to\infty} \|x_k - y_k^i\| = 0, \tag{5.18}$$

so that even (4.3) is satisfied and this concludes the proof. ∎

Remark. By the proof of Proposition 5.2, in particular by (5.18), we note also that

$$\lim_{k\to\infty} \|\nabla f(y_k^i)\| = 0 \qquad \text{for } i = 1, \dots, r+1. \qquad \square$$

We conclude this section by describing Algorithm 3. This algorithm and Algorithm 2 differ only in their search directions. In particular, we recall that in Algorithm

2 the sets of directions $\{p_k^i\}_{i=1}^r$ satisfying Condition C1 are given. In Algorithm 3 we instead assume that, at each iteration, only n linearly independent directions are given. Then the algorithm, on the basis of the behavior of the objective function along these directions, determines a further direction that should have a good descent property and that is able (with the other directions) to ensure the global convergence of the sequence produced.

ALGORITHM 3.

Data. $x_0 \in R^n$, $c > 0$, $\tilde{\alpha}_0^i > 0$, $i = 1, \dots, n+1$, $\gamma > 0$, $\delta, \theta \in (0, 1)$.

Step 0. Set $k = 0$.

Step 1. Set $i = 1$, $y_k^1 = x_k$, $V_k = \{y_k^1\}$, $S_k = \{\emptyset\}$.

Step 2. If $f(y_k^i + \tilde{\alpha}_k^i p_k^i) \le f(y_k^i) - \gamma(\tilde{\alpha}_k^i)^2$, then
compute α_k^i by *LS Procedure*$(\tilde{\alpha}_k^i, y_k^i, p_k^i, \gamma, \delta)$ and
set $\tilde{\alpha}_{k+1}^i = \alpha_k^i$, $V_k = V_k \cup \{y_k^i + \alpha_k^i p_k^i\}$,
$S_k = S_k \cup \{\alpha_k^i\}$;
else set $\alpha_k^i = 0$, $\tilde{\alpha}_{k+1}^i = \theta\tilde{\alpha}_k^i$, $V_k = V_k \cup \{y_k^i + \tilde{\alpha}_k^i p_k^i\}$,
$S_k = S_k \cup \{\tilde{\alpha}_k^i\}$.
Set $y_k^{i+1} = y_k^i + \alpha_k^i p_k^i$.

Step 3. If $i < n$, set $i = i + 1$ and go to Step 2.

Step 4. Compute $\alpha_k^{min} = \min_{\alpha \in S_k}\{\alpha\}$ and $\alpha_k^{max} = \max_{\alpha \in S_k}\{\alpha\}$.
If $\frac{\alpha_k^{max}}{\alpha_k^{min}} \le c$, then compute p_k^{n+1} such that

$$p_k^{n+1} = \frac{v_k^{max} - v_k^{min}}{\xi_k},$$

where $v_k^{max} = \arg\max_{v \in V_k}\{f(v)\}$,
$v_k^{min} = \arg\min_{v \in V_k}\{f(v)\}$, and
$\xi_k \in [\alpha_k^{min}, \alpha_k^{max}]$;
else set $p_k^{n+1} = -\sum_{i=1}^n p_k^i$.

Step 5. If $f(y_k^n + \tilde{\alpha}_k^{n+1} p_k^{n+1}) \le f(y_k^n) - \gamma(\tilde{\alpha}_k^{n+1})^2$, then
compute α_k^{n+1} by *LS Procedure*$(\tilde{\alpha}_k^i, y_k^i, p_k^{n+1}, \gamma, \delta)$
and set $\tilde{\alpha}_{k+1}^{n+1} = \alpha_k^{n+1}$;
else set $\alpha_k^{n+1} = 0$ and $\tilde{\alpha}_{k+1}^{n+1} = \theta\tilde{\alpha}_k^{n+1}$.
Set $y_k^{n+1} = y_k^n + \alpha_k^{n+1} p_k^{n+1}$.

Step 6. Find x_{k+1} such that

$$f(x_{k+1}) \le f(y_k^{n+1}),$$

set $k = k + 1$, and go to Step 1.

Steps 1–3 are essentially the same as those of Algorithm 2. In these steps the algorithm produces the points y_k^i, with $i = 1, \dots, n$, by examining the behavior of the objective function along the linearly independent directions p_k^i, with $i = 1, \dots, n$. At Step 4 we check whether the steplengths used to sample the objective function along the n directions have been "sufficiently regular," namely, whether the ratio between the biggest steplength and the smallest one is not too high. In this case, the objective function values corresponding to points generated along the n linearly independent directions are sufficiently representative of the local behavior of f. Hence, the direction p_k^{n+1} is computed taking these values into account, and it is given by the direction (suitably scaled) from the point with the highest objective value to the point with the lowest objective value. The aim is to approximate the direction of steepest descent. Whenever the test on the ratio between the biggest steplength and the smallest one is not satisfied, the direction p_k^{n+1} is chosen in such a way that the set $\{p_k^1, \dots, p_k^{n+1}\}$ is a positive basis for R^n. Roughly speaking, the test at Step 4 can be viewed as a derivative-free angle condition which, as for the usual angle condition adopted in gradient-based algorithms, allows us to define sets of search directions satisfying Condition C1 and hence to ensure the global convergence of the algorithm.

At Step 5, the point y_k^{n+1} is produced by essentially repeating the instructions of Step 2 for the computed direction p_k^{n+1}. Finally, according to Step 6, the algorithm can update the current point by any point which produces an improvement of the objective function value with respect to $f(y_k^{n+1})$. Now we prove the following convergence result.

PROPOSITION 5.3. *Let $\{x_k\}$ be the sequence produced by Algorithm 3. Suppose that the vectors $\{p_k^i\}$, with $i = 1, \dots, n$, are bounded and uniformly linearly independent. Then Algorithm 3 is well defined and we have*

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{5.19}$$

Proof. In order to prove the thesis, since Algorithm 3 is an instance of Algorithm 2, we need only show that the sets of directions $\{p_k^i\}_{i=1}^{n+1}$ satisfy Condition C1. First let us suppose that there exists an index $\bar{k}$ such that for all $k \ge \bar{k}$ we have

$$p_k^{n+1} = -\sum_{i=1}^{n} p_k^i.$$

Then, the sets p_k^i, with $i = 1, \dots, n+1$, belong to the class (a) of sets of search directions defined in section 3, and hence Condition C1 is satisfied.

Now, let us consider any subset $K \subseteq \{0, 1, \dots\}$ such that, for all $k \in K$, p_k^{n+1} is given by

$$p_k^{n+1} = \frac{v_k^{max} - v_k^{min}}{\xi_k},$$

according to Step 4. The instructions of this step imply

$$\frac{\alpha_k^{max}}{\alpha_k^{min}} \le c \qquad \text{for all } k \in K. \tag{5.20}$$

In this case, we can prove that the sets p_k^i, with $k \in K$ and $i = 1, \dots, n+1$, belong to the class (b) of sets of search directions defined in section 3. In fact, we can define

$$z_k^i = y_k^i, \qquad \xi_k^i = \begin{cases} \alpha_k^i & \text{if } \alpha_k^i > 0, \\ \tilde{\alpha}_k^i & \text{otherwise}, \end{cases} \qquad \text{for } i = 1, \dots, n,$$

and we can set

$$\rho_k^1 = \rho_k^2 = \cdots = \rho_k^{2n-1} = 0, \quad \rho_k^{2n} = 1,$$

$$\tilde{\xi}_k^{2n} = \xi_k,$$

so that (3.2) becomes

$$p_k^{n+1} = \frac{(v_k^1 - v_k^{2n})}{\tilde{\xi}_k^{2n}} = \frac{v_k^{max} - v_k^{min}}{\xi_k}.$$

The conditions on ρ_k^l, with $l = 1, \ldots, 2n$, are obviously satisfied. Recalling the definitions of ξ_k^i for $i = 1, \ldots, n$, we have that (3.4) holds; moreover, the test at Step 4 implies that (3.5) is satisfied with $c_1 = c$ (see (5.20)). Regarding (3.6), recalling the boundedness of $\{p_k^j\}$ with $j = 1, \ldots, n$, we can write for all $i \in \{1, \ldots, n\}$

$$\|z_k^i - x_k\| \le \sum_{j=0}^{i-1} \xi_k^j \|p_k^j\| \le \max_{l=1,\ldots,n} \xi_k^l \sum_{j=0}^{i-1} \|p_k^j\| \le c\xi_k^i \sum_{j=0}^{i-1} \|p_k^j\| \le \tilde{c}\xi_k^i,$$

so that (3.6) holds with $c_2 = \tilde{c}$. Finally, by repeating the same reasoning used in the proof of Proposition 5.2, we can prove (5.14), (5.15), and (5.18), so that (3.7) is satisfied. $\square$

6. Conclusions. In this work we have tried to establish a general convergence theory for unconstrained optimization without derivatives. Toward that aim, we have stated a set of conditions by satisfying which a pattern search or a line search algorithm is guaranteed to enjoy global convergence. On the basis of the theoretical analysis, we have defined new derivative-free algorithms which combine pattern and line search approaches. Future work will be devoted to designing an efficient code and to performing computational experiments in order to thoroughly investigate the practical interest of the proposed approach.

7. Appendix.

Proof of Proposition 3.1. Assume, by contradiction, that the assertion of the proposition is false. Therefore, there exists a value $\eta > 0$ such that, for every pair γ_t, δ_t, we can find an index $k(t)$ and scalars $\alpha_{k(t)}^i$, with $i = 1, \ldots, r$, for which we have

$$\|\nabla f(x_{k(t)})\| \ge \eta,$$

$$f(x_{k(t)} + \alpha_{k(t)}^i p_{k(t)}^i) > f(x_{k(t)}) - \gamma_t \alpha_{k(t)}^i \|\nabla f(x_{k(t)})\| \|p_{k(t)}^i\|,$$

and

$$0 < \alpha_{k(t)}^i \le \delta_t$$

for all $i \in \{1, \ldots, r\}$. Now, taking into account the boundedness of $\{x_k\}$, we have that there exist (by relabeling if necessary) sequences $\{x_k\}$, $\{\gamma_k\}$, $\{\delta_k\}$, $\{\alpha_k^i\}$, $\{p_k^i\}$, with $i = 1, \ldots, r$, such that

$$x_k \to \bar{x}, \tag{7.1}$$

$$\gamma_k \to 0, \tag{7.2}$$

$$\delta_k \to 0, \tag{7.3}$$

$$\alpha_k^i \le \delta_k, \tag{7.4}$$

$$f(x_k + \alpha_k^i p_k^i) > f(x_k) - \gamma_k \alpha_k^i \|\nabla f(x_k)\| \|p_k^i\|. \tag{7.5}$$

By the continuity assumption, we have that $\|\nabla f(\bar{x})\| \ge \eta$; then, by using Condition C1, for k sufficiently large there exists an index $i \in \{1, \ldots, r\}$ such that

$$\nabla f(x_k)^T p_k^i \le \rho < 0. \tag{7.6}$$

Now, by (7.3), (7.4), and the boundedness of $\{p_k^i\}$ for $i = 1, \ldots, r$, we have that

$$\lim_{k\to\infty} \alpha_k^i \|p_k^i\| = 0 \tag{7.7}$$

for all $i \in \{1, \ldots, r\}$. By (7.5) and the mean-value theorem, we can write

$$\nabla f(x_k)^T p_k^i + (\nabla f(x_k + \theta_k^i \alpha_k^i p_k^i) - \nabla f(x_k))^T p_k^i \geq -\gamma_k \|\nabla f(x_k)\| \|p_k^i\|, \tag{7.8}$$

where $\theta_k^i \in (0, 1)$. From (7.7), (7.8), (7.2) and recalling again the boundedness of $\{p_k^i\}$, we get a contradiction with (7.6) for k sufficiently large. $\square$

Proof of Proposition 3.2. If $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$, then the boundedness of $\{p_k^i\}$ for $i = 1, \ldots, r$ implies that $\lim_{k\to\infty} \min\{0, \nabla f(x_k)^T p_k^i\} = 0$, $i = 1, \ldots, r$.

In order to prove that

$$\lim_{k\to\infty} \sum_{i=1}^{r} \min\{0, \nabla f(x_k)^T p_k^i\} = 0 \tag{7.9}$$

implies

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0, \tag{7.10}$$

we assume, by contradiction, that the assertion is false. Therefore, taking into account the boundedness of $\{x_k\}$, there exist a subset $K_1 \subseteq \{0, 1, \ldots\}$ and a positive number η such that

$$\lim_{k\to\infty, k\in K_1} x_k = \bar{x}, \tag{7.11}$$

$$\|\nabla f(\bar{x})\| \geq \eta > 0. \tag{7.12}$$

Now we distinguish the two classes of sets of search directions.

Class (a). By recalling the assumptions on the sets of search directions of this class, we have that we can find a subset $K_2 \subseteq K_1$ such that we have

$$\lim_{k\to\infty, k\in K_2} p_k^i = \bar{p}^i, \qquad i = 1, \ldots, r,$$

where $\bar{p}^1, \ldots, \bar{p}^r$ positively span R^n. Therefore, we can write

$$-\nabla f(\bar{x}) = \sum_{i=1}^{r} \beta^i \bar{p}^i, \tag{7.13}$$

with $\beta^i \geq 0$ for $i = 1, \ldots, r$. Then, recalling (7.12), we obtain

$$-\eta^2 \geq \sum_{i=1}^{r} \beta^i \nabla f(\bar{x})^T \bar{p}^i. \tag{7.14}$$

From (7.14), recalling the continuity assumption on ∇f, it follows that

$$\lim_{k\to\infty, k\in K_2} \sum_{i=1}^{r} \min\{0, \nabla f(x_k)^T p_k^i\} = \sum_{i=1}^{r} \min\{0, \nabla f(\bar{x})^T \bar{p}^i\} < 0,$$

which contradicts (7.9).

Class (b). By the boundedness assumptions on the sequences $\{p_k^i\}$, with $i = 1,\dots,r$, and $\{\rho_k^l\}$, with $l = 1,\dots,2n$, we have that there exists a subset $K_2 \subseteq K_1$ such that we have

$$\lim_{k\to\infty, k\in K_2} p_k^i = \bar{p}^i, \qquad i = 1,\dots,r, \tag{7.15}$$

$$\lim_{k\to\infty, k\in K_2} \rho_k^l = \bar{\rho}^i, \qquad l = 1,\dots,2n, \tag{7.16}$$

where $\bar{\rho}^{2n} \geq \bar{\rho} > 0$.

From the definitions of $\tilde{\xi}_k^l$ and v_k^l with $l = 1,\dots,2n$, the boundedness of $\{p_k^i\}$ with $i = 1,\dots,r$ (for the sake of simplicity, we assume $\|p_k^i\| = 1$), and (3.5), (3.7), (3.6), it follows that the vectors $(v_k^l - v_k^1)/\tilde{\xi}_k^l$ are bounded. In fact, from (3.6), for k sufficiently large and for each $l \in \{1,\dots,2n\}$ we can write

$$\|v_k^l - v_k^1\| \leq \|v_k^l - x_k\| + \|x_k - v_k^1\| \leq \sigma_1^l \xi_k^l + \sigma_2^l \xi_k^1$$

with $\sigma_1^l, \sigma_2^l > 0$. From the assumptions on $\tilde{\xi}_k^l$, by (3.5) we have

$$\frac{1}{c_1} \leq \frac{\xi_k^i}{\tilde{\xi}_k^l} \leq c_1 \tag{7.17}$$

for each $i \in \{1,\dots,n\}$ and for each $l \in \{1,\dots,2n\}$. Then, the boundedness of $\{p_k^i\}$, with $i = 1,\dots,r$, implies the boundedness of $(v_k^l - v_k^1)/\tilde{\xi}_k^l$ for $l = 1,\dots,2n$. Hence we have

$$\lim_{k\to\infty, k\in K_2} \frac{v_k^l - v_k^1}{\tilde{\xi}_k^l} = \bar{y}^l, \qquad l = 1,\dots,2n. \tag{7.18}$$

Furthermore, (3.7) and (7.15) imply

$$\lim_{k\to\infty, k\in K_2} v_k^l = \bar{x}, \qquad l = 1,\dots,2n. \tag{7.19}$$

From (3.3), for all $k \geq 0$ and for $l = 1,\dots,2n$, we can write

$$f(v_k^l) - f(v_k^1) \geq 0,$$

from which, by using the mean-value theorem, it follows that

$$\tilde{\xi}_k^l \nabla f\left(v_k^l + \theta_k^l \tilde{\xi}_k^l \frac{v_k^l - v_k^1}{\tilde{\xi}_k^l}\right)^T \left(\frac{v_k^l - v_k^1}{\tilde{\xi}_k^l}\right) \geq 0, \tag{7.20}$$

with $\theta_k^l \in (0,1)$. Then, recalling (7.11) and (7.18), taking into account (3.7), and by using the continuity assumption on ∇f for $l = 1,\dots,2n$, we have

$$\lim_{k\to\infty, k\in K_2} \nabla f\left(v_k^l + \theta_k^l \tilde{\xi}_k^l \frac{v_k^l - v_k^1}{\tilde{\xi}_k^l}\right)^T \left(\frac{v_k^l - v_k^1}{\tilde{\xi}_k^l}\right) = \nabla f(\bar{x})^T \bar{y}^l \geq 0. \tag{7.21}$$

Now, from (3.2), we get

$$\nabla f(x_k)^T p_k^{n+j} = \sum_{l=1}^{2n} \rho_k^l \nabla f(x_k)^T \frac{(v_k^1 - v_k^l)}{\tilde{\xi}_k^l}. \tag{7.22}$$

On the other hand, from (7.9) and recalling the continuity assumption on ∇f, it follows that

$$\lim_{k\to\infty, k\in K_2} \nabla f(x_k)^T p_k^i = \nabla f(\bar{x})^T \bar{p}^i = l^i \geq 0, \quad i = 1,\dots,r. \tag{7.23}$$

Therefore, from (7.22), taking the limits for $k \to \infty$ and $k \in K_2$, we obtain

$$\nabla f(\bar{x})^T \bar{p}^{n+j} = -\sum_{l=1}^{2n} \bar{\rho}_l \nabla f(\bar{x})^T \bar{y}^l \geq 0,$$

where $\bar{\rho}^l \geq 0$ and $\bar{\rho}^{2n} > 0$. Hence, recalling (7.21), it follows that

$$\nabla f(\bar{x})^T \bar{y}^{2n} = 0. \tag{7.24}$$

Now, from (3.3), we get

$$\frac{f(v_k^{2n}) - f(v_k^1)}{\tilde{\xi}_k^{2n}} \geq \frac{f(z_k^i + \xi_k^i p_k^i) - f(z_k^i)}{\tilde{\xi}_k^{2n}}, \qquad i = 1,\dots,n. \tag{7.25}$$

By using the mean-value theorem, we have

$$\frac{f(v_k^{2n}) - f(v_k^1)}{\tilde{\xi}_k^{2n}} = \frac{\nabla f\left(v_k^1 + \theta_k \alpha_k \frac{v_k^{2n} - v_k^1}{\tilde{\xi}_k^{2n}}\right)^T (v_k^{2n} - v_k^1)}{\tilde{\xi}_k^{2n}}, \tag{7.26}$$

$$\frac{f(z_k^i + \xi_k^i p_k^i) - f(z_k^i)}{\tilde{\xi}_k^{2n}} = \nabla f(z_k^i + u_k^i \xi_k^i p_k^i)^T p_k^i \frac{\xi_k^i}{\tilde{\xi}_k^{2n}}, \tag{7.27}$$

with $\theta_k \in (0,1)$, $u_k^i \in (0,1)$, $i = 1,\dots,n$.

By substituting (7.26) and (7.27) into (7.25), taking the limits for $k \to \infty$ and $k \in K_2$, and recalling (7.17) and the continuity assumption on ∇f, we obtain

$$\nabla f(\bar{x})^T \bar{y}^{2n} \geq \nabla f(\bar{x})^T \bar{p}^i \frac{1}{c_1}, \qquad i = 1,\dots,n.$$

Then, from (7.23) and (7.24), it follows that

$$\nabla f(\bar{x})^T \bar{p}^i = 0, \qquad i = 1,\dots,n.$$

The linear independence of $\bar{p}^i$, with $i = 1,\dots,n$, implies

$$\nabla f(\bar{x}) = 0,$$

which contradicts (7.12). $\square$

Acknowledgments. The authors are grateful to the referees for their helpful comments and suggestions, which led to significant improvements in the paper.

REFERENCES

[1] R. De Leone, M. Gaudioso, and L. Grippo, *Stopping criteria for linesearch methods without derivatives*, Math. Programming, 30 (1984), pp. 285–300.

[2] J. E. Dennis, Jr., and V. Torczon, *Direct search methods on parallel machines*, SIAM J. Optim., 1 (1991), pp. 448–474.
[3] T. Glad and A. Goldstein, *Optimization of functions whose values are subject to small errors*, BIT, 17 (1977), pp. 160–169.
[4] L. Grippo, *A class of unconstrained minimization methods for neural network training*, Optim. Methods Softw., 4 (1994), pp. 135–150.
[5] L. Grippo, F. Lampariello, and S. Lucidi, *Global convergence and stabilization of unconstrained minimization methods without derivatives*, J. Optim. Theory Appl., 56 (1988), pp. 385–406.
[6] R. Hooke and T. A. Jeeves, *Direct search solution of numerical and statistical problems*, J. ACM, 8 (1961), pp. 212–229.
[7] R. M. Lewis and V. Torczon, *Rank Ordering and Positive Bases in Pattern Search Algorithms*, Technical report TR 96-71, ICASE, NASA Langley Research Center, Hampton, VA, 1996.
[8] S. Lucidi and M. Sciandrone, *Numerical results for unconstrained optimization without derivatives*, in Nonlinear Optimization and Applications, G. Di Pillo and F. Giannessi, eds., Plenum Publishing, New York, 1995, pp. 261–270.
[9] S. Lucidi and M. Sciandrone, *On the Global Convergence of Derivative Free Methods for Unconstrained Optimization without Derivatives*, Technical report R. 18-96, DIS, Università di Roma "La Sapienza," Rome, 1996.
[10] R. Mifflin, *A superlinearly convergent algorithm for minimization without evaluating derivatives*, Math. Programming, 9 (1975), pp. 100–117.
[11] M. J. D. Powell, *An efficient method for finding the minimum of a function of several variables without calculating derivatives*, Comput. J., 7 (1964), pp. 155–163.
[12] M. J. D. Powell, *Unconstrained minimization algorithms without computation derivatives*, Boll. Unione Mat. Ital., 9 (1974), pp. 60–69.
[13] M. J. D. Powell, *Direct search algorithms for optimization calculations*, Acta Numer., 7 (1998), pp. 287–336.
[14] A. S. Rykov, *Simplex direct search algorithms*, Automat. Remote Control, 41 (1980), pp. 784–793.
[15] A. S. Rykov, *Simplex methods of direct search*, Engineering Cybernetics, 18 (1980), pp. 12–18.
[16] V. Torczon, *On the convergence of pattern search algorithms*, SIAM J. Optim., 7 (1997), pp. 1–25.
[17] P. Tseng, *Fortified-descent simplicial search method: A general approach*, SIAM J. Optim., 10 (1999), pp. 269–288.
[18] Yu Wen-ci, *Positive basis and a class of direct search techniques*, Scientia Sinica, Special Issue of Mathematics, 1 (1979), pp. 53–67.
[19] M. H. Wright, *Direct search methods: Once scorned, now respectable*, in Proceedings of the 1995 Dundee Biennial Conference in Numerical Analysis, D. F. Griffiths and G. A. Watson, eds., Addison-Wesley Longman, Harlow, United Kingdom, 1996, pp. 191–208.
[20] W. J. Zangwill, *Minimizing a function without calculating derivatives*, Comput. J., 10 (1967), pp. 293–296.