Optim Lett
DOI 10.1007/s11590-014-0816-9

ORIGINAL PAPER

(Non) convergence results for the differential evolution method

Marco Locatelli · Massimiliano Vasile

Received: 23 April 2014 / Accepted: 6 October 2014

Abstract In this paper we deal with the convergence properties of the differential evolution (DE) algorithm, a rather popular stochastic method for solving global optimization problems. We are going to show there exist instances for which the basic version of DE has a positive probability not to converge (stagnation might occur), or converges to a single point which is not a local minimizer of the objective function, even when the objective function is convex. Next, some minimal corrections of the basic DE scheme are suggested in order to recover convergence with probability one to a local minimizer at least in the case of strictly convex functions.

Keywords Differential evolution · Stagnation · Convergence

1 Introduction

Differential evolution (DE in what follows) is one of the most popular and effective stochastic algorithms for the solution of global optimization (GO in what follows) problems

$$\min_{\mathbf{x}\in D\subset\mathbb{R}^d} f(\mathbf{x}),$$

M. Locatelli (✉)
Dipartimento di Ingegneria dell'Informazione, Università di Parma,
Via G.P. Usberti, 181/A, 43124 Parma, Italy
e-mail: locatell@ce.unipr.it

M. Vasile
Department of Mechanical and Aerospace Engineering, University of Strathclyde,
75 Montrose Street, Glasgow, Scotland G1 1XJ, UK
e-mail: massimiliano.vasile@strath.ac.uk

Published online: 18 October 2014

where D is a compact domain, assumed to be a box in what follows. Since its first appearance (see [9,10]), a large number of papers about DE have been published in the literature. A search in SCOPUS for the publications containing DE in their title returned more than 4,000 documents. These include books, like, e.g. [7], collection of papers [3], surveys [4]. The effectiveness of DE is illustrated by the fact that DE, or its variants, has always ranked among the best approaches at the competitions held during the IEEE Congress on Evolutionary Computation (CEC) conferences.

Algorithm 1 is the most basic version of DE.

```
Set the parameter values F > 0, C_R ∈ (0, 1) and the integer value n_pop ≥ 4 ;
Initialize the iteration counter k := 0;
Generate and evaluate an initial population P_0 = {x_{1,0}, ..., x_{n_pop,0}} made up by n_pop random
points drawn from the uniform distribution over D ;
while stopping criterion is false do
    for i = 1, ..., n_pop do
        Select at random distinct indices i_1, i_2, i_3 from {1, ..., n_pop} \ {i};
        for j = 1, ..., d do
            Set e(j) = 1 with probability C_R, otherwise set e(j) = 0;
            if e(j) = 1 then
                Set u^j_{i,k} = (x^j_{i_3,k} − x^j_{i,k}) + F(x^j_{i_2,k} − x^j_{i_1,k})
            end
            else
                u^j_{i,k} = 0
            end
        end
        if f(x_{i,k} + u_{i,k}) < f(x_{i,k}) then
            Set x_{i,k+1} = x_{i,k} + u_{i,k}
        end
        else
            x_{i,k+1} = x_{i,k}
        end
        Set k := k + 1;
    end
end
```

Algorithm 1: Basic version of DE

There are different variants of the basic version, like, e.g., the version where $\mathbf{x}_{i_3,k}$ is replaced by the best member of the population (the one with the lowest function value). However, we do not discuss such variants here and we refer to the existing literature.

In this paper we do not discuss the practical performance of DE but we consider its theoretical properties. In particular, we would like to study the convergence properties of DE. In fact, we are going to prove that DE is *not* able to guarantee convergence to global minimizers even over simple optimization problems (like, e.g., problems with a convex objective function f, for which each local minimizer is also a global minimizer). However, we will also suggest some possible minimal modifications of the basic DE scheme in order to recover, at least in some cases, convergence with probability one (see, e.g., [14]) to the set of global minimizers. We warn the reader that the results about the non convergence of DE should not be used to diminish its relevance. Indeed, they should be taken with some care since they are based on situations created ad hoc “against” DE, while in practice, as already pointed out, the behavior of DE is highly satisfactory. However, when DE performs poorly, the results may be helpful to identify the reasons for such poor performance and, consequently, also to identify more effective variants of DE.

The paper is structured as follows. In Sect. 2 we will discuss some instances that have a positive probability that DE does not converge not only with general continuous objective functions, but also when the objective function is strictly convex. In Sect. 3 we suggest several minimal modifications of the basic DE scheme which allow to recover convergence with probability one.

2 Non convergence results for DE

It is easily seen that if the whole population in DE collapses to a single point, then no further progress is possible. However, it has been observed (see [6]) that situations where no further progress is possible might occur even when the population is not collapsed into a single point. This phenomenon is called *stagnation* and is basically related to the finite number of possible moves which DE might perform at each iteration (see also [1]). In particular, if we denote by P the current population (made up by at least four distinct members) and by Q the (finite) set of all points which *may* be generated at the current iteration, then if $Q \cap P = \emptyset$, there always exists at least one function f (in fact, infinitely many functions) such that stagnation will occur. This is easily seen by choosing f so that

$$\min_{\mathbf{x} \in Q} f(\mathbf{x}) > \max_{\mathbf{x} \in P} f(\mathbf{x}). \tag{1}$$

If the above condition holds, no single candidate point is able to replace the current members of the population, which, thus, will not change. The probability of such an event tends to decrease with the increase of the population size because the cardinality of Q is also increased, but the occurrence of the event can not be excluded. A further way to reduce the probability of stagnation is by considering, as done in [13], a new version of DE where the constant factor F is multiplied by a continuous random variable. In this way, set Q has infinite cardinality. However, set Q consists of a collection of lines and if $Q \cap P = \emptyset$, then there always exists at least some continuous function f for which (1) holds. This means that stagnation occurs also in this case. As pointed out in [8] in the more general context of evolutionary strategies, stagnation can be avoided and convergence with probability one can be guaranteed if the probability associated to each set of positive measure is bounded away from 0. In the basic version of DE this can be attained by adding to $\mathbf{x}_{i,k} + \mathbf{u}_{i,k}$ a random perturbation vector $\mathbf{h}$ with a density bounded away from 0. This has been done in [2]. In that paper a different selection mechanism is employed. Given a target distribution π, the new candidate point $\mathbf{x}_{i,k} + \mathbf{u}_{i,k} + \mathbf{h}$ is accepted with the following probability

$$\min\left\{1, \frac{\pi(\mathbf{x}_{i,k} + \mathbf{u}_{i,k} + \mathbf{h})}{\pi(\mathbf{x}_{i,k})}\right\},$$

so that non improving moves may be accepted, i.e., candidate points with a function value worse than the current one are accepted with a positive probability, depending on π. In [2] convergence to the target distribution is proved. However, although these modifications guarantee convergence of DE, such convergence does not depend at all on the specific steps of DE. Any stochastic algorithm where at each iteration there is a positive probability of generating a point within any set with positive measure, converges in probability to the set of global minimizers. This is true even for the simplest stochastic GO algorithm, Pure Random Search, where at each iteration a uniform random point over the feasible region is sampled. In the above scheme this is guaranteed by the fact that $\mathbf{h}$ has density bounded away from 0, but the specific steps of the DE approach do not come into play in the convergence proof. Thus, let us go back to the original version of DE with a finite set of possible candidates. Given that for general continuous functions it is simple to establish non convergence for DE, we might wonder whether convergence results can be obtained if we restrict the class of functions. In [12] it was proved that for strictly quasi-convex functions DE always converges to a single point $\bar{\mathbf{x}}$ or, equivalently,

$$\max_{i,j=1,\ldots,n_{pop}} \|\mathbf{x}_{i,k} - \mathbf{x}_{j,k}\| \to 0 \quad k \to \infty,$$

if the condition that the indices i, i_1, i_2, i_3 are all different from each other, imposed in the classical DE scheme, is relaxed. In the proof in [12] only $i_1 \neq i_2$ was imposed. We report here the statement of the result proved in [12].

Theorem 2.1 *Let us consider the most basic version of DE where only $i_1 \neq i_2$ is imposed. Given a function f that is strictly quasi-convex over the compact and convex set D, and a population $P_k \in D$, then if $F \in (0, 1)$ the population P_k converges to a single point in D for $k \to \infty$ with probability one.*

Note that for general multimodal functions the above result can not be applied. However, if a local minimizer of a GO problem satisfies some regularity assumption (e.g., the Hessian at the local minimizer is positive definite), then a neighborhood can always be defined such that: (i) DE will be unable to accept points outside the neighborhood; (ii) the function is strictly convex within the neighborhood. Therefore, if at some iteration k the population P_k belongs to such a neighborhood, where the conditions of Theorem 2.1 are fulfilled, the same theorem guarantees convergence to a single point. It is interesting to remark that, if we keep the original assumption that the indices i, i_1, i_2, i_3 are all different, stagnation might occur even when f is a strictly convex function. The following example shows that there exist strictly convex functions for which the probability of stagnation after the generation of the random initial population P_0 is positive.

Example 2.1 Let $d = 2$ and $n_{pop} = 4$. Let the current population be

$$P = \{(1, 1),\ (1, -1),\ (-1, 1),\ (10, 10)\}, \tag{2}$$

and $F = 0.9$. Imposing the condition that all indices must be different, the set of candidate points Q (whose cardinality is 72) turns out to lie outside the convex hull of the four points, namely the triangle with vertices

$$(1, -1), \quad (-1, 1), \quad (10, 10),$$

while $(1, 1)$ is in the interior of the convex hull. Now one can easily build a strictly convex function with a level set which contains all these four points but does not contain any of the candidate points, which would thus all be rejected, causing stagnation of the population. This is the case, e.g., if the function has the following form

$$\begin{aligned} f(x, y) = &(x - 4.95)^2 + (y - 4.95)^2 \\ &+M \times \max\{0, -x - y, -20 - 9x + 11y, -20 + 11x - 9y\}, \end{aligned}$$

where $M > 0$. Note that f is the sum of a strictly convex function and a convex polyhedral one, and is, thus, strictly convex. The convex polyhedral function is equal to 0 within the convex hull of the four points and is positive outside it. The local and global minimizer of this function is the point $(4.95, 4.95)$, at which the value of f is 0. The value of f at the three points $(1, -1)$, $(-1, 1)$, $(10, 10)$ is 51.005, while it is equal to 31.205 at the fourth point $(1, 1)$. For M large enough, all 72 candidate points are outside the level set

$$\{(x, y) \,:\, f(x, y) \leq 51.005\},$$

so that stagnation occurs. We remark that, by continuity of the objective function, the same holds also for small perturbations of the population P. Thus, for the given function there is a positive probability that stagnation occurs even for a randomly generated initial population. In particular, the following experiment has been made with the same function f. We ran DE with $CR = 0.5$ and an initial population where each member is a slight random perturbation (within the interval $[0, 0.01]$) of the above population. DE has been stopped either if the maximal distance between members of the population fell below 10^{-4} (convergence to a point), or if a limit of 1,000 iterations was reached. We also kept track of the number of iterations at which at least one member of the population has been changed. It turned out that, over 100 runs of DE with initial populations generated as previously commented, the initial population never changed. Stagnation occurred in all these cases.

While this instance leads us to the conclusion that the basic version of DE might stagnate from the very beginning with a positive probability even for strictly convex functions, the same example shows that, as already pointed out at the end of the Introduction, these conclusions should be taken with some care, since they refer to situations created ad hoc, whose probability is, in fact, very small. Indeed, as soon as the generation of the initial population has been modified and its four members have been randomly generated within the box $[-1, 10] \times [-1, 10]$, stagnation never occurred and all 100 runs converged to the global minimizer.

In the above example, stagnation is avoided by imposing only $i_1 \neq i_2$ because in that case a positive probability exists that the algorithm generates candidate points which are convex combinations of two members of the population. Indeed, assume that:

- $i_3 = i_1$ and $i = i_2$;

- $f(\mathbf{x}_{i_2,k}) \geq f(\mathbf{x}_{i_1,k})$;
- $e(j) = 1$, $j = 1, \ldots, d$.

Then, the candidate point $\mathbf{x}_{i,k} + \mathbf{u}_{i,k}$ is equal to

$$(1 - F)\mathbf{x}_{i_1,k} + F\mathbf{x}_{i_2,k},$$

and is generated with a positive probability. Moreover, in view of the strict quasi-convexity assumption

$$f((1 - F)\mathbf{x}_{i_1,k} + F\mathbf{x}_{i_2,k}) < \max\{f(\mathbf{x}_{i_2,k}),\ f(\mathbf{x}_{i_1,k})\} = f(\mathbf{x}_{i_2,k}),$$

so that the candidate point will replace $\mathbf{x}_{i_2,k}$. This shows that stagnation can not occur. It is worthwhile to remark that strict quasi-convexity can not be relaxed into simple quasi-convexity, as will be shown in what follows.

Now, we might wonder whether the convergence point $\bar{\mathbf{x}}$ is always a local (in fact, global) minimizer of f over D. This is not guaranteed. Convergence to a point which is not a minimizer might occur in the following cases:

A1 at some iteration k all points of the population lie over an hyperplane

$$H = \{\mathbf{x} \in \mathbb{R}^d \mid \exists i \in \{1, \ldots, d\},\ b \in \mathbb{R},\ x_i = b\},$$

not containing the minimizer: in this case all candidate points generated by DE at the next iterations will also belong to H (DE is unable to escape from H) (note that such phenomenon has been already observed in [13]);

A2 all points in the population are very close to each other: then, the population might keep on changing but the infinitesimal progress does not allow to reach a minimizer.

While we currently do not have an example where convergence to a local minimizer is prevented by A2, the following example shows that case A1 may hold with a positive probability.

Example 2.2 Let

$$f(x, y) = x^2 + y^2, \quad D = [-1, 1] \times [-1, 1].$$

Let us also assume that at iteration k the following holds for each member of the population $\mathbf{x}_{i,k} = (x_{i,k}, y_{i,k})$, $i \in \{1, \ldots, n_{pop}\}$,

$$x_{i,k} < -\frac{1}{2}. \tag{3}$$

Note that there is a positive probability, namely $\left(\frac{1}{4}\right)^{n_{pop}}$, that a uniformly randomly generated initial population satisfies this condition. Let

$$s_1 \in \arg \max_{i=1,\ldots,n_{pop}} x_{i,k},$$

and s_2, s_3 be such that

$$0 < x_{s_2,k} - x_{s_3,k} < \frac{1}{2}, \tag{4}$$

and

$$x_{s_2,k}, x_{s_3,k} > \min_{i=1,\dots,n_{pop}} x_{i,k}. \tag{5}$$

With positive probability, for each $i \in \{1, \dots, n_{pop}\} \setminus \{s_1, s_2, s_3\}$ the displacement $\mathbf{u}_{i,k}$ is obtained as follows

$$\begin{aligned} i_1 = s_1 \quad i_2 = s_2 \quad i_3 = s_3, \\ e(x) = 1 \quad e(y) = 0. \end{aligned} \tag{6}$$

Note that the probability of this event can be computed. Indeed, we observe that: (i) the probability that $i_1 = s_1$ and $i_3 = s_3$ is $\frac{1}{n_{pop}-1}$, since i_1 and i_3 are randomly drawn from $\{1, \dots, n_{pop}\} \setminus \{i\}$; ii) the probability that $i_2 = s_2$ is $\frac{1}{n_{pop}-2}$, since i_2 is randomly drawn from $\{1, \dots, n_{pop}\} \setminus \{i, i_1\}$; (iii) the probability that $e(x) = 1$ is C_R and that of $e(y) = 0$ is $1 - C_R$. Then, since (6) must occur for each $i \in \{1, \dots, n_{pop}\} \setminus \{s_1, s_2, s_3\}$, the probability is

$$\left[C_R(1 - C_R) \left(\frac{1}{n_{pop} - 1} \right)^2 \left(\frac{1}{n_{pop} - 2} \right) \right]^{n_{pop}-3}.$$

Then

$$u^x_{i,k} = (x_{s_1,k} - x_{i,k}) + F(x_{s_2,k} - x_{s_3,k}), \quad u^y_{i,k} = 0, \tag{7}$$

and in view of (3) and (4) for $F < 1$ it holds that

$$f(\mathbf{x}_{i,k} + \mathbf{u}_{i,k}) < f(\mathbf{x}_{i,k}),$$

so that for each i

$$\mathbf{x}_{i,k+1} = \mathbf{x}_{i,k} + \mathbf{u}_{i,k}.$$

Then, in view of (7)

$$x_{i,k+1} = x_{s_1,k} + F(x_{s_2,k} - x_{s_3,k}).$$

We are left with members s_1, s_2, s_3 of the population. In this case at iteration k, a displacement can be generated such that these members do not change at iteration $k+1$ since the function values at $\mathbf{x}_{s_j,k+1}$ are not better than those at $\mathbf{x}_{s_j,k}$ for $j = 1, 2, 3$. This holds, e.g., by taking

- $i_1 \in \arg\min_{i=1,\dots,n_{pop}} x_{i,k} \neq s_1, s_2, s_3$ in view of (4)–(5);
- $i_2, i_3 \neq s_1, s_2, s_3$ and such that $x_{i_2,k} \leq x_{i_3,k}$;
- $e(x) = 1, \; e(y) = 0.$

Once again, one could compute the value of the (positive) probability by an analysis similar to that performed for event (6). Next, at iteration $k+2$ for $i \in \{s_1, s_2, s_3\}$ one can select

$$\begin{array}{l} i_1, i_2, i_3 \notin \{s_1, s_2, s_3\} \\ e(x) = 1 \quad e(y) = 0, \end{array}$$

so that

$$u^x_{s_j,k+1} = (x_{i_1,k+1} - x_{s_j,k+1}), \quad u^y_{s_j,k+1} = 0, \quad j = 1, 2, 3,$$

and

$$x_{s_j,k+2} = x_{i_1,k+1} \quad j = 1, 2, 3,$$

while for $i \notin \{s_1, s_2, s_3\}$ one can select

$$\begin{array}{l} i_1, i_2, i_3 \notin \{s_1, s_2, s_3, i\} \\ e(x) = 1 \quad e(y) = 0, \end{array}$$

so that

$$u^x_{i,k+1} = (x_{i_1,k+1} - x_{i,k+1}), \quad u^y_{i,k+1} = 0,$$

and

$$x_{i,k+2} = x_{i_1,k+1}.$$

Therefore, at the end of iteration $k+2$ all members have the same x-coordinate.

It is worthwhile to point out here that the same example with the convex (but not strictly convex) function $f(x, y) = x^2$ also shows that strict quasi-convexity is necessary to prove convergence to a single point: indeed, when all points have the same value for the coordinate x, no more progress is possible within the population. Moreover, we need to remark again that the above instance is created ad hoc and is not representative of the average behavior of DE.

3 Convergence recovery strategies

Throughout this section we will assume that the functions are continuously differentiable and that their stationary points lie in the interior of the feasible set. We construct several *minimal* modifications of the DE algorithm in order to guarantee that convergence with probability one to a local minimizer holds for strictly convex functions, and convergence with probability one to the set of stationary points holds for functions which are not necessarily convex ones. We first recall the convergence conditions of line search methods for unconstrained optimization, which include the following.

B1 the search direction should be "far enough" from orthogonality with respect to the gradient at the current point;

B2 the step should guarantee a "sufficient decrease" (see, e.g., the Wolfe or Armijo–Goldstein conditions).

Something similar to condition B1 should allow to avoid case A1, while something similar to condition B2 should prevent A2. Therefore, we would like to enforce B1 and B2 also in DE. In order to do that we will impose that at each iteration k of DE there exist two members of the population $\mathbf{x}_{p_k,k}, \mathbf{x}_{q_k,k}$ such that the following conditions hold:

$$\left|\cos(\theta_{h,k})\right| = \frac{\left|(\mathbf{x}_{q_k,k} - \mathbf{x}_{p_k,k})^T \mathbf{e}_h\right|}{\|\mathbf{x}_{q_k,k} - \mathbf{x}_{p_k,k}\|} \geq \eta_k \quad \forall\, h \in \{1, \ldots, d\}, \tag{8}$$

where $\mathbf{e}_h$ is h-th unit vector, and $\theta_{h,k}$ is the angle between $\mathbf{e}_h$ and $\mathbf{x}_{q_k,k} - \mathbf{x}_{p_k,k}$;

$$\|\mathbf{x}_{q_k,k} - \mathbf{x}_{p_k,k}\| \geq \rho_k \quad \text{and} \quad \rho_k \to 0; \tag{9}$$

$$\sum_{k=1}^{\infty} \eta_k \rho_k = +\infty. \tag{10}$$

Condition (10) requires that η_k and ρ_k do not converge to 0 too quickly; condition (8) requires that there exist two members of the population whose difference vector forms an angle with each coordinate direction "sufficiently far" from $\frac{\pi}{2}$ (since the cosine is limited from below by η_k), while condition (9) requires that the distance between these two members is at least ρ_k. Basically, (8) and (9), combined with (10), guarantee that B1 and B2 are satisfied.

Of course, in the standard DE approach we can not guarantee that these conditions are satisfied. However, the algorithm can be modified in such a way that, if at some iteration k the conditions, which are easy to check, are not satisfied, one can restore them by displacing and/or rotating the population or even a single point of the population. Note that this might imply that the maximum distance between points in the population is increased to ρ_k at iteration k. Since $\rho_k \to 0$ asymptotically, an increase at iteration k does not affect the overall convergence of the whole population to a single point $\bar{\mathbf{x}}$ but might change its speed. In any case, we do not want to discuss here how to restore the conditions, but just to prove that, under them, the convergence point $\bar{\mathbf{x}}$ is a minimizer of f.

Theorem 3.1 *Let f be a continuously differentiable strictly convex function, whose minimizer $\mathbf{x}^*$ over D lies in the interior of D. Then, if conditions (8)–(10) are satisfied, then*

$$\mathbf{x}_{i,k} \to \mathbf{x}^* \quad k \to +\infty, \quad i = 1, \ldots, n_{pop}.$$

with probability one.

Proof Note that, under the given assumptions, $\mathbf{x}^*$ is the unique local minimizer of f over D and is also the unique stationary point of f. Let us assume by contradiction that the population converges with a positive probability to a point $\bar{\mathbf{x}} \neq \mathbf{x}^*$. Then, $\bar{\mathbf{x}}$ is not a minimizer and a stationary point of f and, for k large enough, there exists some index h such that for each $i \in \{1, \ldots, n_{pop}\}$

$$\left|\left[\nabla f(\mathbf{x}_{i,k})\right]_h\right| \geq \delta > 0. \tag{11}$$

Now, let p_k, q_k be the indices of the two members satisfying (8)–(10) and let $i = i_3$. Now, with probability $\xi > 0$ DE will generate the following displacement for $\mathbf{x}_{i,k}$

$$[\mathbf{u}_{i,k}]_j = \begin{cases} 0 & j \neq h \\ \pm F[(\mathbf{x}_{p_k,k} - \mathbf{x}_{q_k,k})]_h & j = h. \end{cases} \tag{12}$$

In particular that holds if

$$e(j) = 0 \quad \forall\, j \neq h, \quad e(h) = 1.$$

The indices i_1 and i_2, and, consequently, the sign in (12), are chosen as follows:

- $i_1 = q_k, i_2 = p_k$, so that $[\mathbf{u}_{i,k}]_h = +F[\mathbf{x}_{p_k,k} - \mathbf{x}_{q_k,k}]_h$, if

$$\left[\nabla f(\mathbf{x}_{i,k})\right]_h [\mathbf{x}_{p_k,k} - \mathbf{x}_{q_k,k}]_h < 0;$$

- $i_1 = p_k, i_2 = q_k$, so that $[\mathbf{u}_{i,k}]_h = -F[\mathbf{x}_{p_k,k} - \mathbf{x}_{q_k,k}]_h$, if

$$\left[\nabla f(\mathbf{x}_{i,k})\right]_h [\mathbf{x}_{p_k,k} - \mathbf{x}_{q_k,k}]_h > 0.$$

In particular, with the above choices of i_1 and i_2 it is guaranteed that

$$\left[\nabla f(\mathbf{x}_{i,k})\right]_h [\mathbf{u}_{i,k}]_h < 0. \tag{13}$$

Now, using Taylor expansion, for large k, as a consequence of (8)–(12), with positive probability, we have that

$$\begin{aligned} f(\mathbf{x}_{i,k} + \mathbf{u}_{i,k}) - f(\mathbf{x}_{i,k}) &\approx \nabla f(\mathbf{x}_{i,k})^T \mathbf{u}_{i,k} = \left[\nabla f(\mathbf{x}_{i,k})\right]_h [\mathbf{u}_{i,k}]_h \\ &= -F\left|\left[\nabla f(\mathbf{x}_{i,k})\right]_h\right|\left|(\mathbf{x}_{q_k,k} - \mathbf{x}_{p_k,k})^T \mathbf{e}_h\right| \\ &\leq -F\eta_k \|\mathbf{x}_{p_k,k} - \mathbf{x}_{q_k,k}\|\delta \leq -\eta_k \rho_k F\delta < 0, \end{aligned} \tag{14}$$

so that, for k large enough, $\mathbf{x}_{i,k+1}$ will be updated with respect to $\mathbf{x}_{i,k}$. Therefore, at each (large enough) iteration there is a positive probability (independent from the iteration counter k) that at least one member of the population will improve its value by at least $-\eta_k \rho_k F\delta$. Since (10) holds, this means that the overall expected improvement Δ_k of the population at iteration k with respect to the initial population, defined as follows

$$\Delta_k = E\left[\sum_{r=1}^{n_{pop}} [f(\mathbf{x}_{0,r}) - f(\mathbf{x}_{k,r})]\right],$$

will diverge to $+\infty$ as $k \to \infty$. On the other hand, one should have that

$$\Delta_k \leq n_{pop}\left[\max_{\mathbf{x}\in D} f(\mathbf{x}) - \min_{\mathbf{x}\in D} f(\mathbf{x})\right] < \infty \quad \forall\, k.$$

Thus, we are lead to a contradiction.

As an alternative to imposing (9) we might consider a modification of the basic DE scheme, where the assumption that the parameter F is a fixed constant is relaxed. In that case, we are able to prove an even stronger result. Let us assume that F depends on k, $\mathbf{x}_{i_1,k}$, and $\mathbf{x}_{i_2,k}$, through the following formula

$$F = F_{k,i_1,i_2} = \frac{\rho_k}{\|\mathbf{x}_{i_1,k} - \mathbf{x}_{i_2,k}\|}, \tag{15}$$

where ρ_k is the same as in (9). Notice that with this choice for F the vector $\mathbf{x}_{i_2,k} - \mathbf{x}_{i_1,k}$ contributes to the definition of the direction of the displacement $\mathbf{u}_{i,k}$, but its norm does not contribute to the definition of the step length along this direction, which is instead controlled by ρ_k. Then, one can prove the following result, stating that, under suitable assumptions, the gradient of all members of the population converges to 0 (note that strict convexity of f is not required).

Theorem 3.2 *Let f be a continuously differentiable function with the stationary points that are in D being located in the interior of D. Then, if conditions* (8) *and* (10) *are satisfied and if F is defined as in* (15)*, then*

$$\|\nabla f(\mathbf{x}_{i,k})\| \to 0 \quad k \to +\infty, \quad i = 1, \ldots, n_{pop},$$

with probability one.

Proof Let us assume by contradiction that there exists a positive probability such that for some $r \in \{1, \ldots, n_{pop}\}$, and for some infinite subsequence K,

$$\|\nabla f(\mathbf{x}_{r,k})\| \geq \gamma > 0 \quad \forall\, k \in K.$$

Thus, there exists some index h and a further infinite subsequence K' of K such that

$$\left|\left[\nabla f(\mathbf{x}_{r,k})\right]_h\right| \geq \delta > 0 \quad \forall\, k \in K'. \tag{16}$$

Now, the same choices as in the proof of Theorem 3.1 are made. In particular : (a) p_k, q_k are the indices of the two members satisfying (8); (b) $i = i_3$ still holds; (c) i_1, i_2 are chosen in the same way as in Theorem 3.1, so that (13) holds. It is further required that $i = i_3 = r$, where $r \in \{1, \ldots, n_{pop}\}$ is the same as in (16). Again, Taylor expansion is used and for large k, as a consequence of (8), of (12) (with $i = r$), of (16), and of the definition (15) of $F = F_{k,p_k,q_k}$, with positive probability, we have that

$$
\begin{aligned}
f(\mathbf{x}_{r,k}+\mathbf{u}_{r,k})-f(\mathbf{x}_{r,k}) &\approx \nabla f(\mathbf{x}_{r,k})^T\mathbf{u}_{r,k}=\left[\nabla f(\mathbf{x}_{r,k})\right]_h[\mathbf{u}_{r,k}]_h\\
&=-F_{k,p_k,q_k}\left|\left[\nabla f(\mathbf{x}_{r,k})\right]_h\right|\left|(\mathbf{x}_{q_k,k}-\mathbf{x}_{p_k,k})^T\mathbf{e}_h\right|\\
&=-\frac{\rho_k\left|\left[\nabla f(\mathbf{x}_{r,k})\right]_h\right|\left|(\mathbf{x}_{q_k,k}-\mathbf{x}_{p_k,k})^T\mathbf{e}_h\right|}{\|\mathbf{x}_{q_k,k}-\mathbf{x}_{p_k,k}\|}\\
&\le -\eta_k\rho_k\delta<0.
\end{aligned}
$$

From this point the proof proceeds as in Theorem 3.1.

Note that under the assumptions of Theorem 3.1, function f has a unique local (global) minimizer, which is also the unique stationary point. Thus, in this case Theorem 3.2 proves convergence of the population to the global minimizer. For general functions one can not guarantee convergence to a single point. To see this, it is sufficient to think about a constant function: in that case for each population stagnation occurs, but all points are stationary ones.

We conclude the section with a remark.

Remark 3.1 Conditions (8)–(10) are strictly related to those imposed in pattern search methods for derivative-free optimization (see, e.g., [5,11]). More precisely, condition (8) is related to the coordinate search method based on the directions $\pm\mathbf{e}_i$, $i=1,\ldots,n$, of the coordinate axes. As seen in the proof of Theorem 3.1, DE would need to generate a displacement $\mathbf{u}_{q_k,k}$ along each of these $2n$ directions with a positive probability. In fact, what one would like to guarantee at each iteration is that DE generates a displacement $\mathbf{u}_{q_k,k}$ for $\mathbf{x}_{q_k,k}$ such that (14) holds. Therefore, any set of directions, generated by the DE heuristic, that satisfies (14), with positive probability (independent of the iteration counter), would also do the job.

4 Conclusions

In this paper the convergence properties of DE are discussed. It has been shown that there exist instances for which the basic version of DE stagnates or converges to a point which is not a local minimizer, even when the objective function is convex. These instances are created ad hoc. In fact, the practical performance of DE is much better than what suggested by these instances. However, their identification leads to insight into some minimal changes which can be made to the basic DE scheme, in order to recover convergence with probability one to a local minimizer at least in some special cases, e.g., when the objective function is strictly convex. These modifications allow DE to satisfy conditions which are strictly related to those imposed in pattern search methods for derivative-free optimization.

References

1. Ali, M.: Differential evolution with preferential crossover. Eur. J. Oper. Res. **181**, 1137–1147 (2007)
2. Braak, C.T.: A Markov Chain Monte Carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces. Stat. Comput. **16**, 239–249 (2006)
3. Chakraborty, U.K. (ed.): Advances in Differential Evolution. Springer, Berlin (2008)

4. Das, S., Suganthan, P.: Differential evolution: a survey of the state-of-the-art. IEEE Trans. Evol. Comput. **15**, 4–31 (2011)
5. Kolda, T., Lewis, R., Torczon, V.: Optimization by direct search: new perspectives on some classical and modern methods. SIAM Rev. **45**, 385–482 (2003)
6. Lampinen, J., Zelinka, I.: On stagnation of the differential evolution algorithm. In: Proceedings of the 6th international conference on soft computing MENDEL, pp. 76–83. Brno, Czech Republic (2000)
7. Price, K., Storn, R., Lampinen, J.: Differential Evolution. A Practical Approach to Global Optimization. Springer, Berlin (2005)
8. Rudolph, G.: Convergence of evolutionary algorithms in general search space. In: Proceedings of the IEEE International Conference On Evolutionary Computation. Nagoya, Japan (1996)
9. Storn, R., Price, K.: Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Optim. **11**, 341–359 (1997)
10. Storn, R., Price, K.V.: Differential evolution: a simple and efficient adaptive scheme for global optimization over continuous spaces (TR-95-012). ICSI, USA (1995)
11. Torczon, V.: On the convergence of pattern search algorithms. SIAM J. Optim. **7**, 1–25 (1997)
12. Vasile, M., Minisci, E., Locatelli, M.: An inflationary differential evolution algorithm for space trajectory optimization. IEEE Trans. Evol. Comput. **15**, 267–281 (2011)
13. Zaharie, D.: Critical values for control parameters of differential evolution algorithm. In: Proceedings of 8th international conference on soft computing, mendel 2002, pp. 62–67 (2002)
14. Zhigljavsky, A., Žilinskas, A.: Stochastic Global Optimization. Springer, Berlin (2008)