

Self-adaptive potential-based stopping criteria for Particle Swarm Optimization with forced moves

Bernd Bassimir¹ · Manuel Schmitt¹ · Rolf Wanka¹

Received: 27 June 2019 / Accepted: 18 August 2020 / Published online: 10 September 2020 © The Author(s) 2020

Abstract

We study the variant of Particle Swarm Optimization that applies random velocities in a dimension instead of the regular velocity update equations as soon as the so-called *poten-tial* of the swarm falls below a certain small bound in this dimension, arbitrarily set by the user. In this case, the swarm performs a *forced move*. In this paper, we are interested in how, by counting the forced moves, the swarm can decide for itself to stop its movement because it is improbable to find better candidate solutions than the already-found best solution. We formally prove that when the swarm is close to a (local) optimum, it behaves like a blind-searching cloud and that the frequency of forced moves exceeds a certain, objective function-independent value. Based on this observation, we define stopping criteria and evaluate them experimentally showing that good candidate solutions can be found much faster than setting upper bounds on the iterations and better solutions compared to applying other solutions from the literature.

Keywords Particle Swarm Optimization · Stopping criterion · Heuristics

1 Introduction

Particle Swarm Optimization (PSO) is a meta-heuristic for the so-called continuous black box optimization problems, which means that the objective function is not explicitly known, e.g., in form of a closed formula. PSO produces good results in a variety of different real-world applications. The classical PSO as first introduced by [Kennedy and Eberhart (1995); Eberhart and Kennedy (1995)] is an iterative algorithm which tries to improve a fixed number of candidate solutions, each represented by a particle. The algorithm can be very easily implemented and adapted to the users' applications, which lead to increased

Bernd Bassimir bernd.bassimir@fau.de

> Manuel Schmitt manuel.schmitt@fau.de

Rolf Wanka rolf.wanka@fau.de

¹ Department of Computer Science, University of Erlangen-Nuremberg, Cauerstraße 11, 91058 Erlangen, Germany

attention not only among computer scientists. In order to further improve performance, many authors present changes to the original, "plain," or classical PSO scheme (for exact definitions, see Sect. 3) to improve the quality of the returned solution.

A serious problem of PSO is the phenomenon of premature stagnation, i.e., the convergence of the swarm to a non-optimal solution. To address the issue of premature stagnation, van den Bergh and Engelbrecht (2002) introduce Guaranteed Convergence PSO (GCPSO), where a random perturbation that scales with the number of global attractor updates is added to the best particle. They show on a set of benchmarks that the PSO finds a local minimum with an improved convergence speed and later prove convergence to local optima. Kennedy and Eberhart (1995) first realized that adding a noise term, called craziness, can prove beneficial for PSO performance. Adding a noise term to the movement of the particle has been theoretically addressed by Lehre and Witt (2013). To overcome such stagnation, these authors propose Noisy PSO that adds a "noise" term to the velocity at every move. They prove that for the Noisy PSO started on a certain simple 1-dimensional objective function the first hitting time of the δ -neighborhood of the *global* optimum is finite. However, as proved by Schmitt and Wanka (2013a), premature stagnation of classical PSO does not occur at all when the search space is 1-dimensional and the objective function is continuous. In the 1-dimensional case, PSO provably finds at least a local optimum, almost surely (in the well defined sense of continuous probability theory; see [Chow and Teicher 1978, p. 20) and (Chung 1974, p. 233)]. Furthermore, Schmitt and Wanka (2013a) show a similar result for a slightly modified PSO in the general D-dimensional case [for stagnation-related results regarding the unmodified PSO, see Raß et al. (2015)]. This slightly modified PSO assigns a small random velocity in solely one dimension only if the so-called *potential* of the swarm—a fundamental, measurable quantity of the swarm of all particles in this dimension falls below a certain (arbitrary and small) bound δ . In the following, we call such random velocity particle moves forced steps and the PSO variant f-PSO. The f-PSO provably finds a local optimum almost surely [Schmitt and Wanka (2013a)]. Monitoring the swarm's potential and increasing it from time to time by the forced steps is the key ingredient to mathematically proving successful convergence to a (local) optimum. In this paper, we will use the potential and the forced steps of f-PSO for overcoming another problem of heuristics, the question of when to stop the algorithm.

The goal of this paper is to characterize the behavior of the f-PSO when it is close to a (local) optimum. Our experiments and mathematical investigations show that the number of forced moves does not only increase significantly when the distance to the next local optimum falls below a certain bound, but that additionally the number of forced moves performed close to a local optimum is independent of the objective function. In particular, we prove mathematically by potential arguments that the swarm "pulsates" in a cloud around the best candidate solution found so far. First rudimentary experimental evidence for this potential-based behavior is reported by Bassimir et al. (2014). Note that a similar behavior has been described recently by Yi et al. (2018) for the classical PSO. Therefore, by measuring the frequency of occurrences of forced moves, this frequency can act as a stopping criterion, so the swarm may come to a self-determined halt. All findings are experimentally supported.

The paper is organized as follows. Section 2 provides an introduction and related work to stopping criteria. Section 3 presents a description of the classical and forced PSO (f-PSO) and the definition of the potential of a swarm. In Sect. 4, we experimentally and mathematically analyze the behavior of f-PSO when it comes close to a (local) optimum. In Sect. 5, based on the mathematical analysis in Sect. 4, we present the new termination criteria and show their practicability by experimental evaluations.

2 Stopping criteria

An important problem that arises in the context of PSO and other iterative optimization methods (see (Solis and Wets 1981, p. 23f) and further papers, referenced in this section) is when to stop the iterations and to return the best found admissible solution. Usually, the process is terminated (i) when an upper limit on the number of iterations is reached, or (ii) when an upper limit on the number of evaluations of the objective function is reached, or (iii) when the chance of achieving significant improvements in further iterations is extremely low. The choice of the mentioned two upper limits obviously depends on the concrete objective function which means that the user has to have detailed knowledge of the objective function and to "intervene" by hand. Furthermore, these limits often depend on the parameters used. For a more detailed description of the problems regarding limits as stopping criteria, see Engelbrecht (2014). In (iii), it is desired and advantageous that the algorithm adaptively decides when to stop, so external intervention is not necessary anymore. To achieve this for PSO, many criteria were introduced in the literature. A commonly used criterion is the swarm diameter, i.e., the maximum distance between two particles, as a measure for the expansion and thus the movement capability of the swarm. When PSO is extended with forced steps, this and also other criteria of this kind do not work because the expansion of the overall swarm is forcefully kept above a certain value and can no longer converge like it does for the classical PSO algorithm. Just the globally best position found by the swarm converges to an optimum.

Adaptive stopping criteria for PSO have been investigated by Zielinski et al. (2005) and Zielinski and Laur (2007). In these papers, a list of upper limit-based and adaptive termination criteria is presented and in experiments applied to (2-dimensional) established benchmark functions and a real-world problem, resp. Kwok et al. (2007) introduce a stopping criterion for PSO based on the rate of improvements found by the swarm in a given time interval. Quality of the termination is assured using the nonparametric sign-test enforcing a low false-positive rate. A further stopping approach due to Ong and Fukushima (2015) combines PSO with gene matrices.

Stopping criteria have also been studied in the context of other non-PSO optimization approaches. Ribeiro et al. (2011) introduce a stopping criterion that stops the GRASP (greedy randomized adaptive search procedures) algorithm when the probability of an improvement is below a threshold under an experimentally fitted normal distribution. Safe et al. (2004) present a study of various aspects associated with the specification of termination conditions for simple genetic algorithms. Aytug and Koehler (2000) introduce a stopping criterion that stops a genetic algorithm when the optimal solution is found with a specified confidence and thus no real further progress can be expected. Sharma and Rangaiah (2013) present a termination criterion for a multi-objective differential evolution algorithm based on a taboo list for similar solutions. In a very general, ground-breaking investigation, Solis and Wets (1981) consider random search in general and also address the question of stopping criteria. In this context, Dorea (1990) presents two adaptive stopping criteria. Stopping criteria in the context of multi-objective optimization is presented by Martí et al. (2016).

3 Definitions

Due to the wide variety of existing PSO variants, we first state the exact "classical" PSO algorithm on which our work is based.

Definition 1 (*Classical GBest PSO Algorithm*) A *swarm* \mathscr{S} of *N* particles moves through the *D*-dimensional search space \mathbb{R}^D with $\mathscr{D} = \{1, ..., D\}$ being the set of indices specifying

dimensions. Each particle $n \in \mathscr{S}$ consists of a *position* $X^n \in \mathbb{R}^D$, a *velocity* $V^n \in \mathbb{R}^D$ and a *local attractor* $L^n \in \mathbb{R}^D$, storing the best position particle *n* has visited so far. Additionally, the particles of the swarm share information via the *global attractor* $G \in \mathbb{R}^D$, describing the best point *any* particle has visited so far, i.e., as soon as a particle has performed its move¹, it possibly updates the global attractor immediately (sometimes called asynchronous PSO, see Carlisle and Dozier (2001) among others).

The actual movement of the swarm is governed by the following *movement equations* where χ , c_1 , $c_2 \in \mathbb{R}^+$ are some positive constants to be fixed later, and **r** and **s** are drawn u. a. r. from $[0, 1]^D$ every time the equation is applied.

$$V^{n} := \chi \cdot V^{n} + c_{1} \cdot \mathbf{r} \odot (L^{n} - X^{n}) + c_{2} \cdot \mathbf{s} \odot (G - X^{n})$$
$$X^{n} := X^{n} + V^{n}$$

Here, \odot denotes entrywise multiplication (Hadamard product). The application of the equations on particle *n* is called the *move* of *n*. When all particles have executed their moves, the swarm has executed one *iteration*.

Now we repeat the definition of a swarm's potential measuring how close it is to convergence, i.e., we describe a measure for its movement. A swarm with high potential should be more likely to reach search points far away from the current global attractor, while the potential of a converging swarm approaches 0. These considerations lead to the following definition by Schmitt and Wanka (2013b). The original definition by Schmitt and Wanka (2013a) is more complex just for technical reasons. We can omit the α and the square-root as these do not influence the required properties and the α and square-root are not used in the proposed algorithm.

Definition 2 (*Potential*) Fix a moment of the computation of swarm \mathscr{S} . For $d \in \mathscr{D}$, the current *potential* Φ_d of \mathscr{S} in dimension d is

$$\Phi_d := \sum_{n=1}^{N} (\underbrace{|V_d^n| + |G_d - X_d^n|}_{=: \phi_d^n}) = \sum_{n=1}^{N} \phi_d^n$$

 $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_D)$ is the total potential of \mathscr{S} , and ϕ_d^n (*partial potential*) is the *contribution* of particle *n* to the potential of \mathscr{S} in dimension *d*.

Note that the potential of the swarm has an entry in every dimension. The swarm comes to a halt if $\Phi \rightarrow (0, ..., 0)$. So, if the particles come close to *G*, the swarm may stop even if *G* is a non-optimal point, an incident that is called (premature) stagnation. The value of Φ can actually be computed at any given time and, hence, be used for decisions on the swarm. Between two different dimensions, the potential difference might be large, and "transferring" potential from one dimension to another is not possible due to the movement equations as each dimension is calculated independently. On the other hand, along the same dimension the particles influence each other and can transfer potential from one particle to the other. This is the reason why there is no potential of individual particles, but only their contribution to the potential.

To address the phenomenon of stagnation, Schmitt and Wanka (2015) slightly modified the PSO movement equations from Definition 1 by "recharging" potential and, hence, keeping the swarm moving.

¹ the particles' moves are executed sequentially, so there is some arbitrary order of the particles.

Definition 3 (*f-PSO*) The modified movement of the swarm is governed by the following movement equations where χ , c_1 , c_2 , $\delta \in \mathbb{R}^+$ are some positive constants to be fixed later, *r*, *s* and *t* are drawn u. a. r. from [0, 1] for every move and dimension of a particle.

$$V_d^n := \begin{cases} (2 \cdot t - 1) \cdot \delta, & \text{if } \forall n' \in \mathscr{S} : \phi_d^{n'} < \delta \quad \text{[forced velocity update]} \\ \chi \cdot V_d^n + c_1 \cdot r \cdot (L_d^n - X_d^n) + c_2 \cdot s \cdot (G_d - X_d^n), \text{ otherwise } \quad \text{[regular velocity update]} \\ X^n := X^n + V^n. \end{cases}$$

If the forced velocity update applies for a particle, we call its move and the corresponding dimension in the move *forced*. An iteration of the swarm is called *forced* if during this iteration at least one particle performs a forced move. The whole method is called *f-PSO*.

If it is necessary to identify the values at the beginning of iteration *i*, we write $X_d^{n,i}$, $V_d^{n,i}$ etc.

f-PSO Algorithm 1: **input** : Objective function $f : \mathbb{R}^D \to \mathbb{R}$, number of particles N output: $G \in \mathbb{R}^D$ 1 for $n = 1 \rightarrow N$ do Initialize X^n randomly; 2 Initialize V^n with $\vec{0}$; 3 Initialize $L^n := X^n$; 4 5 Initialize $G := \operatorname{argmin} f$; $\{L^n | n \in \mathscr{S}\}$ 6 repeat // Iterations; t, r, s drawn u.a.r. from [0,1] every time for $n = 1 \rightarrow N$ do 7 for $d = 1 \rightarrow D$ do 8 // We use $\delta = 10^{-7}$ in our experiments, see Sect. 4.2.1 and Sect. 5 Calculate new partial potentials $\phi_d^{n'}$ for all $n' \in \mathscr{S}$ 9 if $\forall n' \in \mathscr{S} : \phi_d^{n'} < \delta$ then 10 // Execute a forced velocity update $V_d^n := (2 \cdot t - 1) \cdot \delta$; // hence, $V_d^n \in [-\delta, \delta]$ 11 else 12 // Execute the usual, regular velocity update // We use $\chi = 0.72984$, $c_1 = 1.49617$, $c_2 = 1.49617$ in our experiments $V_d^n := \chi \cdot V_d^n + c_1 \cdot r \cdot (L_d^n - X_d^n) + c_2 \cdot s \cdot (G_d - X_d^n);$ 13 $X_d^n := X_d^n + V_d^n;$ 14 if $f(X^n) \leq f(L^n)$ then $L^n := X^n$; 15 if $f(X^n) \leq f(G)$ then $G := X^n$; 16 17 until termination criterion met // One can use our new criteria developed in Sec. 5;

18 **return** *G*;

Algorithm 1 provides a formal and detailed overview of f-PSO. The introduction of forced velocity updates guarantees that the swarm (or more precisely, the global attractor G) does not converge to a non-optimal point almost surely, but finds a local optimum (Schmitt and Wanka 2013a). In our analysis and for the experiments, we used the common parameter settings $\chi = 0.72984$, $c_1 = 1.49617$, $c_2 = 1.49617$ and N > 1 as suggested and used by Clerc and Kennedy (2002), which are parameter settings that are widely used in the literature.

4 Behavior of the f-PSO algorithm

The idea of the modification of the classical PSO algorithm is to help the swarm overcome (premature) stagnation. The modification is implemented in the calculation of the new velocity during a move of a particle (see Definition 3 and Lines 10 and 11 of Algorithm 1). The new velocity of a particle in the current dimension *d* is drawn u. a. r. from $[-\delta, \delta]$ when for all particles $n' \in \mathcal{S}$ their contributions $\phi_d^{n'}$ to the potential in dimension *d* is less than δ , which means that the range in which the swarm can optimize is small in dimension *d*.

The main topic of this section is to examine if at such a situation the particle swarm uses from now on only forced moves or, more desirable, the swarm will recover and continue using regular velocity updates. In Schmitt and Wanka (2015), it is shown that indeed the latter is the case, unless the global attractor G is already in the neighborhood of a local optimum. As in this case almost all moves are forced, we will see in Sect. 5 that this can be translated into a criterion to stop the PSO's execution.

4.1 Experiments

We first introduce the notion of the forcing frequency to quantify the number of applications of the forced velocity update (Line 11 of Algorithm 1).

Definition 4 (*Absolute and relative forcing frequency*) Let *I* denote a (time) interval of *II* iterations during a run of the *f-PSO* algorithm.

- (a) Let $\sigma(I, d)$ denote the number of times forced velocity updates (Line 11) have been executed in interval *I* for dimension $d, d \in \{1, ..., D\}$. $\sigma(I, d)$ is called the *absolute forcing frequency* in dimension *d* over interval *I*.
- forcing frequency in dimension d over interval I.
 (b) σ(I) = Σ^D_{d=1} σ(I, d) is the total number of forced velocity updates (Line 11) in interval I counted over all dimensions. σ(I) is called the *absolute forcing frequency* over interval I.

Analogously, the *relative forcing frequency* is $\sigma(I, d)/|I|$ and $\sigma(I)/|I|$, respectively.

To explore the behavior of the particle swarm with respect to the absolute forcing frequency $\sigma(I, d)$, we performed a series of experiments on the well-known benchmark functions Schwefel, ROSENBROCK, RASTRIGIN, H. C. ELLIPTIC, SPHERE, GRIEWANK and ACKLEY (for a comprehensive overview of these and many other benchmark functions, see (Helwig 2010, Sect. 4.2) and Suganthan et al. (2005)). SPHERE and H. C. ELLIPTIC will be analyzed in detail in Sect. 4.2. The tests were performed with Raß' HiPPSO (see Raß (2020)) a high precision implementation of PSO, in order to rule out the influence of insufficient computer systems' precision when applying a forced velocity update with δ close to the precision of a, e.g., long double variable in C++. The tests were performed with N = 3 particles, D = 30 dimensions, $\delta = 10^{-7}$ and the well-known swarm parameters already mentioned in Sect. 3. The number of particles was chosen to be small to increase the length (measured by the number of iterations) of the exploration phase. With a larger number of particles, more positions of the search space are explored per iteration as there are more function evaluations, and particles tend to come closer to an optimum in the same number of iterations compared to a particle swarm with fewer number of particles therefore ending the exploration phase quite early. Given this number of particles, we see a better distinction between the exploration phase and the exploitation phase and can observe the transition between these two phases. The overall results, however, are still valid for larger swarm sizes, e.g., see Appendix A. For a discussion on how to chose δ , see Sects. 4.2.1 and 5.

SCHWEFEL Fig. 1 presents the measurements obtained when optimizing the SCHWEFEL function with D = 30. The absolute forcing frequencies per dimension over the intervals $I_i = [0 \dots 50, 000 \cdot i]$ and the function values f(G) of the current global attractors G at the end of each interval are depicted. One can see that the absolute forcing frequency (the difference of two successive values of $\sigma(I_i, d)$) is relatively small if f(G) is far away from the (unique) optimum value and increases considerably when f(G) approaches the optimum. Additionally, one can see that the gradient of the absolute forcing frequency, i.e., the relative forcing frequency, tends to be constant for all $i \ge i_0$ after some value i_0 . This is a showcase and remains true for most of the other tested benchmark functions.

ROSENBROCK During the arbitrary precision tests on the standard benchmark functions, processing the ROSENBROCK function showed a slightly different behavior. Metaphorically speaking, there is a small "banana-shaped" valley in this function that leads



Fig. 1 Optimizing the SCHWEFEL function: development of the absolute forcing frequencies $\sigma(I_i, d)$, $d \in \{1, ..., 30\}$, over the intervals $I_i = [0 ... 50,000 \cdot i]$ (bundle of lines marked by triangles) compared to the development of the function value f(G) of the global best position *G* after 50,000 $\cdot i$ iterations (single red line marked by circles). N = 3 particles and $\delta = 10^{-7}$ were used



Fig. 2 Optimizing the ROSENBROCK function: development of the absolute forcing frequencies $\sigma(I_i, d)$, $d \in \{1, ..., 30\}$, over the intervals $I_i = [0 ... 50,000 \cdot i]$ compared to the development of the function value f(G) of the global best position G after 50,000 $\cdot i$ iterations (single red line marked by circles). N = 3 particles and $\delta = 10^{-7}$ were used. Noticeable is the constant behavior of $\sigma(I_i, d)$ with f(G) far away from the optimal value of 0, and the separation into forcing dimensions and non-forcing dimensions

from a local optimum to the global optimum. If the attractors are in this valley (which they are quite early), the chance to improve in some dimensions is quite small and improvements can easily be voided by setbacks in other dimensions. If this happens, forced velocity updates will be applied in some dimensions while in the other dimensions the particles use the usual, regular velocity updates, which can be seen clearly in Fig. 2. The chance to get moving into the optimum's direction again is not zero, but quite small which leads to a long phase of near-stagnation as can be seen in Fig. 2. Eventually the swarm will re-start moving again, but it may take a long time to do so.

AckLEY Another phenomenon may occur on functions like AckLEY, where there are many local optima. Here, it might happen that the global attractor G is already close to a (good) local optimum, but the local attractor L^n of some particle n is close to a slightly worse local optimum. In such a situation, the potential is governed by the distance between G and L^n , and the condition for executing a forced velocity update is not satisfied, even though a good local near-optimum value f(G) has already been found. But our experiments showed that this happens with only a small probability. However, with increased swarm sizes this behavior becomes more likely.

4.2 Experiments on the SPHERE Function, Phases, And Theoretical Analysis

We now analyze the swarm's final behavior with the help of experiments on SPHERE (the results for H. C. ELLIPTIC are similar). These experiments show that the final behavior can be classified into two main phases: the approaching phase and the pulsation phase that in

turn has three distinct sub-phases: forced phase, lockout phase, recovery phase. These findings hold for all objective functions, as the swarm is in this final phase similar to a blind searching cloud with no or at least very close to zero updates of the attractors.

So, when the global attractor reaches the δ -neighborhood of a local optimum, the number of global attractor updates decreases and the distance the global attractor moves becomes small. At that time, the whole swarm will start to approach the global attractor. At some point during this *approaching phase*, the local attractors will be close to the global attractor and the swarm will lose much of its potential. As described by Schmitt and Wanka (2015), the introduction of forced velocity updates in the movement equations of the PSO algorithm leads to the situation that the actual positions of the particles cannot approach each other closer than a value of about δ and the swarm enters a kind of *pulsation phase*: periodically contracting and expanding around the global attractor. However, the global attractor still converges (in the mathematical "infinite time" sense) to the local optimum. Also, the positions of the local attractors may further contract.

It can be observed that when the attractors are all almost at the same position and the particles converge to this position, the number of dimensions that are forced begins to increase. During the pulsation phase, almost all particles are forced and the relative forcing frequency $\sigma(I, d)/|I|$ begins to stagnate at a certain value $\partial \sigma_d$ (the gradient in the figures). Hence, also $\sigma(I)/|I|$ begins to stagnate at $\partial \sigma$. When this stagnation is reached, the f-PSO will become a periodic pulsating process and will behave almost like a blind search in a box with a side-length in the order of magnitude of δ around the global attractor. Hence, the objective function *f* will become irrelevant and the value $\partial \sigma$ is independent of *f*.

4.2.1 Experimental identification of parameters influencing $\partial \sigma$

In order to identify the parameters that influence $\partial \sigma$, a series of experiments were performed with the SPHERE function using a finite precision PSO algorithm. To study a pure pulsation phase, the global and local attractors were initially all set to **0**, which is the global optimum of this function. With the swarm parameters χ , c_1 and c_2 being fixed, the two crucial remaining parameters that might influence $\partial \sigma$ are

- the number D of dimensions and
- the number N of particles,

whereas the choice of δ has no influence at all. By changing δ the range of ϕ_d causing a forced step is influenced for a dimension $d \in \mathcal{D}$. However, the random velocity assigned by the forced step in this dimension d is influenced in the same magnitude. Therefore the change of δ has no influence on the absolute forcing frequency as these two effects cancel each other.

Figures 3 and 4 show the effect of changes in the dimension number *D* and swarm size *N*, resp., on the stagnation value $\partial \sigma$ for fixed *II*. The actual size of the fixed interval *II* has no real influence on the absolute forcing frequency, disregarding the higher standard deviations with a lower intervals lengths as shown in Fig. 5.

The use of a forced move in a given dimension depends only on the velocities and positions of all particles relative to the global attractor G in this dimension. The dimensions are independent of each other; therefore, the stagnation value (recall that |I| is fixed) for each number D of dimensions does not change with an increased number of dimensions as shown in Fig. 3. The same is not true for varying the number of particles, see Fig. 4. If the



Fig. 3 Optimizing the SPHERE function: Absolute forcing frequency $\sigma(I)$ over intervals *I* of length |I| = 50,000 relative to the number *D* of dimensions at the global optimum with N = 5 particles and $\delta = 10^{-7}$, varying the number *D* of dimensions. Shown are the average value and standard deviation of 100 trials taking the average values of 10 intervals per trial

length |I| of the interval and the number *D* of dimensions are fixed, increasing the number of particles increases the number of applications of the movement equations by $|I| \cdot D$ for each new particle. Therefore, more dimensions can be forced in a given interval, but there are also more particles that must have a partial potential ϕ less than δ .

4.2.2 Mathematical analysis of the sub-phases for arbitrary objective functions

We can identify three sub-phases during the pulsation phase which are repeated in a fixed order for a given dimension:

- (i) Starting with a *forced phase*, where each particle has its move in this dimension forced,
- (ii) followed by a *lockout phase* with the probability of a forced move being zero and
- (iii) finally a *recovery phase* in which the particles attempt to converge to the global attractor until one move of a particle gets forced again and the cycle repeats itself.

To fully understand why this leads to a smaller forced frequency $\sigma(I)$ per particle, see Fig. 4, we have to take a closer look at the three sub-phases of the pulsation phase. As described in Definition 3, a dimension $d \in \mathcal{D}$ is forced during the move of a particle when all particles n' including the particle itself have contribution $\phi_d^{n'}$ less than δ in this dimension d. For $i \in \mathbb{N}$, $n \in \mathcal{S}$, and $d \in \mathcal{D}$ let $F_{i,n,d}$ be the {0, 1}-indicator variable being 1 iff in iteration i and dimension d particle n has the velocity update forced. The global and the local attractors are constant, and in particular, it can be assumed that $G_d^i = G$ for fixed G. We get the following observations regarding a forced move in a given dimension.



Fig. 4 Optimizing the SPHERE function: Absolute forcing frequency $\sigma(I)$ over intervals *I* of length |I|=50,000 relative to the number *N* of particles at the global optimum with D = 15 dimensions and $\delta = 10^{-7}$, varying the number of particles. Shown are the average value and (the very small, almost invisible) standard deviation of 100 trials taking the average values of 10 intervals per trial

The first phase we analyze is the lockout phase. Recall that $\phi_d^{n,i} = |V_d^{n,i}| + |G - X_d^{n,i}|$ is the contribution of particle *n* to the potential of \mathscr{S} in dimension *d* and iteration *i*.

Lemma 1 If $\phi_d^{n,i+1} \ge \delta$ for particle n, then $\forall n' \in \mathcal{S}, n' > n : F_{i,n',d} = 0$ and $\forall n' \in \mathcal{S}, n' \le n : F_{i+1,n',d} = 0$.

Proof With the attractors being constant, the only way for particle *n* to have contribution ϕ_d^n less than δ is to make a swarm move that reduces the sum of the velocity and the distance to the global attractor to less than δ . The earliest time this is possible is in the next iteration.

By Lemma 1, we get the length of the lockout phase as N, the number of particles.

For the forced phase, we have to look at the probability that for a particle *n* with a forced move in dimension d, $\phi_d^n < \delta$ still applies in the next iteration.

Lemma 2 Let $n \in \mathcal{S}, d \in \mathcal{D}$. Then for n < N: $P[F_{i,n+1,d} = 1 | F_{i,n,d} = 1] = \frac{1}{2}$ and $P[F_{i+1,1,d} = 1 | F_{i,N,d} = 1] = \frac{1}{2}$

Proof As each dimension is independent and the same calculations are performed in each dimension, we may restrict our analysis to one dimension *d* and omit the index *d* in the following proofs.

W. l. o. g., let n = 1. Let $\Delta = G - X^{1,i}$ the (signed) distance of particle 1 to the fixed global attractor. Hence, $\phi^{1,i} = |V^{1,i}| + |\Delta|$.

$$\begin{split} & \mathbb{P}\big[F_{i,2} = 1 \mid F_{i,1} = 1\big] = \mathbb{P}\Big[\phi^{1,i-1} < \delta \mid F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \ge 0 \mid F_{i,1} = 1\big] \\ & + \mathbb{P}\Big[\phi^{1,i+1} < \delta \mid \Delta \ge 0 \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \ge 0 \mid F_{i,1} = 1\big] \\ & = \mathbb{P}\Big[\phi^{1,i+1} < \delta \mid \Delta \ge 0 \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \le 0 \mid F_{i,1} = 1\big] \\ & + \mathbb{P}\Big[\phi^{1,i+1} < \delta \mid (\Delta \ge V^{1,i+1} \ge \Delta \ge 0) \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \ge V^{1,i+1} \ge 0 \mid F_{i,1} = 1\big] \\ & + \mathbb{P}\Big[\phi^{1,i+1} < \delta \mid (\Delta \ge 0 \ge V^{1,i+1}) \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \ge 0 \ge V^{1,i+1} \mid F_{i,1} = 1\big] \\ & + \mathbb{P}\Big[\phi^{1,i+1} < \delta \mid (\Delta \le V^{1,i+1} < 0) \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \ge 0 \ge V^{1,i+1} \mid F_{i,1} = 1\big] \\ & + \mathbb{P}\Big[\phi^{1,i+1} < \delta \mid (\Delta \le V^{1,i+1} < 0) \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \le 0 \le V^{1,i+1} \mid F_{i,1} = 1\big] \\ & + \mathbb{P}\Big[\phi^{1,i+1} < \delta \mid (\Delta \le 0 \le V^{1,i+1}) \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \le 0 \le V^{1,i+1} \mid F_{i,1} = 1\big] \\ & + \mathbb{P}\Big[\phi^{1,i+1} < \delta \mid (\Delta \le 0 \le V^{1,i+1}) \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \ge 0 \lor V^{1,i+1} \mid F_{i,1} = 1\big] \\ & + \mathbb{P}\Big[\Delta < \delta \mid (\Delta \ge V^{1,i+1} \ge 0) \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \ge V^{1,i+1} \ge 0 \mid F_{i,1} = 1\big] \\ & + \mathbb{P}\Big[\Delta < \delta \mid (\Delta \ge V^{1,i+1} \ge 0) \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \ge V^{1,i+1} \ge 0 \mid F_{i,1} = 1\big] \\ & + \mathbb{P}\Big[\Delta - 2 \cdot V^{1,i+1} < \delta \mid (V^{1,i+1} \ge \Delta \ge 0) \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \ge V^{1,i+1} < 0 \mid F_{i,1} = 1\big] \\ & + \mathbb{P}\Big[\Delta - 2 \cdot V^{1,i+1} < \delta \mid (\Delta \ge 0 \ge V^{1,i+1}) \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \ge 0 \ge V^{1,i+1} \mid F_{i,1} = 1\big] \\ & + \mathbb{P}\Big[\Delta - 2 \cdot V^{1,i+1} < \delta \mid (\Delta \ge 0 \ge V^{1,i+1}) \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \ge 0 \ge V^{1,i+1} \mid F_{i,1} = 1\big] \\ & + \mathbb{P}\Big[\Delta \le V^{1,i+1} - \Delta < \delta \mid (\Delta \ge 0 \ge V^{1,i+1}) \land F_{i,1} = 1\big] & \cdot \mathbb{P}\Big[\Delta \ge 0 \ge V^{1,i+1} \mid F_{i,1} = 1\big] \\ & = \frac{1}{2} \cdot \left(\frac{1}{2} \cdot \frac{\delta - |\Delta|}{2 \cdot \delta} + 1 \cdot \frac{|\Delta|}{2 \cdot \delta} + 1 \cdot \frac{|\Delta|}{2 \cdot \delta} + \frac{1}{2} \cdot \frac{\delta - |\Delta|}{2 \cdot \delta} + \frac{\delta - |\Delta|}{2 \cdot \delta} \cdot \frac{1}{2} + \frac{\delta - |\Delta|}{2 \cdot \delta} \cdot \frac{1}{2} \right) \right] \\ & = \frac{1}{2} \end{aligned}$$

With this probability, we can now analyze the length of the forced phase.

Lemma 3 Given $F_{i,n,d} = 1$, define Y as the number of consecutive particle steps, during which dimension d of the respective particle is forced, starting with particle n at iteration i. Then for $k \ge 1$ it holds that $P[Y = k] = \frac{1}{2^{k-1}}$.

Proof Again, we omit the dimension d from the variables.

By Lemma 1, we know that if for one particle $|V^{n,i+1}| + |G - X^{n,i+1}| \ge \delta$ holds, the next particle cannot be forced and with Lemma 2, $P[F_{i,n+1} = 1|F_{i,n} = 1] = \frac{1}{2}$. By assumption, $P[F_{i,n} = 1] = 1$. By induction on *k*, we have:



Fig. 5 Absolute forcing frequency $\sigma(I)$ over intervals *I* relative to the length |I| at the global optimum of the SPHERE function with D = 15 dimensions and $\delta = 10^{-7}$ and N = 5 particles varying the length |I| of the measured intervals. Shown are the average value and standard deviation of 100 trials taking the average values of 10 intervals per trial

$$\begin{split} \mathsf{P}[Y=2] = \mathsf{P}\left[\bigwedge_{\nu=0}^{2-1} F_{i+\lfloor\frac{n+\nu}{N}\rfloor, n+\nu \bmod N} = 1\right] \\ = \mathsf{P}\Big[F_{i+\lfloor\frac{n+1}{N}\rfloor, n+1 \bmod N} = 1 \mid F_{i,n} = 1\Big] \cdot \mathsf{P}\big[F_{i,n} = 1\big] \\ = \mathsf{P}\Big[F_{i+\lfloor\frac{n+1}{N}\rfloor, n+1 \bmod N} = 1 \mid F_{i,n} = 1\Big] = \frac{1}{2} = \frac{1}{2^{2-1}} \\ \mathsf{P}[Y=k] = \mathsf{P}\left[\bigwedge_{\nu=0}^{k-1} F_{i+\lfloor\frac{n+\nu}{N}\rfloor, n+\nu \bmod N} = 1\Big] \\ = \mathsf{P}\bigg[F_{i+\lfloor\frac{n+(k-1)}{N}\rfloor, n+(k-1) \bmod N} = 1\Big| \bigwedge_{\nu=0}^{k-2} F_{i+\lfloor\frac{n+\nu}{N}\rfloor, n+\nu \bmod N} = 1\bigg] \\ \cdot \mathsf{P}\bigg[\bigwedge_{\nu=0}^{k-2} F_{i+\lfloor\frac{n+\nu}{N}\rfloor, n+\nu \bmod N} = 1\bigg] \\ = \frac{1}{2} \cdot \frac{1}{2^{k-2}} = \frac{1}{2^{k-1}} \end{split}$$

us at least an idea of the behavior of the particle during this phase.

Finally for the recovery phase, we cannot give an exact length, as it depends on the positions of the particles in the search space. The following Lemma 4 and Fig. 6, however, give

Springer



Fig. 6 Frequency of the distance to the global attractor of one particle in one dimension each iteration for 5,000,000 iterations in a run with 5 particles and 15 dimensions and $\delta = 10^{-7}$ initialized at the global optimum. The dashed lines indicate the interval of $[-\delta, \delta]$ with 94.4354% of the measurements in the indicated interval

Lemma 4 The probability for a particle n to get a partial potential ϕ_d^n greater than δ when not forced for the first time after a chain of forced moves is

$$\mathbb{P}[\phi^{n,i+1} \ge \delta \mid F_{i,n,d} = 0 \land F_{i-1,n,d} = 1] \ge \frac{1}{2} \cdot \left(1 - \frac{1}{2\chi}\right).$$

Proof By assumption, $L^n = G$. Again, *d* is omitted from the variables and let ρ be the sum of the weighted random variables; hence, $\rho = (c_1 \cdot r + c_2 \cdot s)$.

Applying the PSO movement equations leads to

$$\phi^{n,i+1} = |V^{n,i+1}| + |G - X^{n,i+1}| = |V^{n,i+1}| + |G - X^{n,i} - V^{n,i+1}|$$

= $|\chi \cdot V^{n,i} + \rho \cdot \Delta| + |\Delta - \chi \cdot V^{n,i} - \rho \cdot \Delta|$
= $|\chi \cdot V^{n,i} + \rho \cdot \Delta| + |-\chi \cdot V^{n,i} - (\rho - 1) \cdot \Delta|$
= $|\chi \cdot V^{n,i} + \rho \cdot \Delta| + |\chi \cdot V^{n,i} + (\rho - 1) \cdot \Delta|$
 $\ge |\chi \cdot V^{n,i} + \rho \cdot \Delta + \chi \cdot V^{n,i} + (\rho - 1) \cdot \Delta|$
= $|2 \cdot \chi \cdot V^{n,i} + (2 \cdot \rho - 1) \cdot \Delta|$

Hence,

$$\begin{split} & \mathbb{P}[\phi^{n,i+1} \geq \delta \mid F_{i,n} = 0 \land F_{i-1,n} = 1] \\ & \geq \mathbb{P}[|2\chi \cdot V^{n,i} + (2 \cdot \rho - 1) \cdot \Delta| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}\left[|2\chi \cdot V^{n,i} + (2 \cdot \rho - 1) \cdot \Delta| \geq \delta| \begin{vmatrix} V^{n,i} \cdot (2 \cdot \rho - 1) \cdot \Delta \geq 0 \\ \land F_{i-1,n} = 1 \end{vmatrix} \right] \\ & \cdot \mathbb{P}[V^{n,i} \cdot (2 \cdot \rho - 1) \cdot \Delta \geq 0 \mid F_{i-1,n} = 1] \\ & + \mathbb{P}\left[|2\chi \cdot V^{n,i} + (2 \cdot \rho - 1) \cdot \Delta| \geq \delta| \begin{vmatrix} V^{n,i} \cdot (2 \cdot \rho - 1) \cdot (G - x^{n,i}) < 0 \\ \land F_{i-1,n} = 1 \end{vmatrix} \right] \\ & \cdot \mathbb{P}[V^{n,i} \cdot (2 \cdot \rho - 1) \cdot \Delta < 0 \mid F_{i-1,n} = 1] \\ & \geq \mathbb{P}[|2\chi \cdot V^{n,i}| \geq \delta \mid V^{n,i} \cdot (2 \cdot \rho - 1) \cdot \Delta \geq 0 \land F_{i-1,n} = 1] \\ & \cdot \mathbb{P}[V^{n,i} \cdot (2 \cdot \rho - 1) \cdot \Delta \geq 0 \mid F_{i-1,n} = 1] \\ & \geq \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & \geq \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \geq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \leq \delta \mid F_{i-1,n} = 1] \\ & = \mathbb{P}[2\chi \cdot |V^{n,i}| \leq \delta \mid F_{i-1,n$$

Under the assumption that $F_{i-1,n} = 1$, we have $V^{n,i} \sim \mathcal{U}(-\delta, \delta)$ and $|V^{n,i}| \sim \mathcal{U}(0, \delta)$ (\mathcal{U} denotes the continuous uniform distribution). Therefore, we get $\alpha \ge 1 - \frac{1}{2\chi}$ and can derive $\beta = \frac{1}{2}$ by the following equations.

$$\begin{split} & \mathbf{P} \Big[V^{n,i} \cdot (2 \cdot \rho - 1) \cdot \Delta \geq 0 \Big] \\ &= \mathbf{P} \Big[V^{n,i} \cdot (2 \cdot \rho - 1) \cdot \Delta \geq 0 \mid V^{n,i} \geq 0 \Big] \cdot \mathbf{P} \Big[V^{n,i} \geq 0 \Big] \\ &+ \mathbf{P} \Big[V^{n,i} \cdot (2 \cdot \rho - 1) \cdot \Delta \geq 0 \mid V^{n,i} < 0 \Big] \cdot \mathbf{P} \Big[V^{n,i} < 0 \Big] \\ &= \frac{1}{2} \cdot (\mathbf{P} [(2 \cdot \rho - 1) \cdot \Delta \geq 0] + \mathbf{P} [(2 \cdot \rho - 1) \cdot \Delta < 0]) \\ &= \frac{1}{2} \end{split}$$

Note that the lower bound on the probability in Lemma 4 is negative for $\chi < \frac{1}{2}$. The actual probability depends on the position in the search space. With $\chi > \frac{1}{2}$, the probability for a particle *n* to get partial potential $\phi^n > \delta$ is positive independent of the position of the particle. Therefore, the length of the recovery phase will decrease drastically with smaller χ values.

We may conclude that for a given dimension we get the following behavior in the pulsation phase.

Theorem 1 Let \mathscr{S} be a swarm of particles at a (local) optimum of an arbitrary function f, i.e., $\forall n \in \mathscr{S} : L^n = G = \operatorname{Opt}(f)$ with $\operatorname{Opt}(f)$ being a (local) optimum of the function f. The swarm repeats three phases independently for each dimension:

- the forced phase of length $k \ge 1$ with probability $\frac{1}{2^{k-1}}$,
- the lockout phase of length N,
- and the recovery phase of length $\ell \geq 0$

leading to a stagnation of $\sigma(I)/|I| \rightarrow \partial \sigma$ at a (local) optimum depending on N and D.

Starting at one forced particle, the chance for the next k particles to be forced has probability 2^{-k+1} as stated in Lemma 3 which is independent of the number of particles. At some particle *u*, this sequence of forced particles ends, and in this moment the forced phase ends. With Lemma 1, at least the following N particles will not be forced which constitutes the lockout phase. The next particle that can be forced is $(u + 1) \mod N = v$, and for that all particles n that move between u and v need to keep partial potential $\phi_{1}^{n} < \delta$. All these particles use the regular swarm movement equations from Definition 3, and, with G being constant, they are independent of each other. As shown in Lemma 4, the chance for a particle *n* to get partial potential $\phi_d^n > \delta$ can be quite probable depending on the actual swarm parameters. Hence, with increasing N the chance for a particle n to get partial potential $\phi_d^n > \delta$ increases exponentially in the order of the probability of the complementary event. The exact probability for this event depends on the position of the particle relative to the global attractor. The same is true for the number of iterations a particle needs to fix its partial potential ϕ_d^n , but with normal swarm behavior the particle starts to converge to the global attractor and therefore decreases its partial potential. Figure 6 shows the distance of the particles to the global attractor at the global optimum. We can see that the particles tend to converge to the attractor with high probability thus ending the recovery phase. Therefore, the particles once again start to get forced and the same pattern repeats itself.

Influence of parameter selection To get a more diverse understanding of the influence of the parameter selection on the forcing frequency, we performed a series of experiments on the ROSENBROCK function with different swarm parameters and the swarm initialized once at and once near the global optimum. We tested the $\chi=0.1, c_1=c_2=0.1, \quad \chi=0.1, c_1=c_2=2.1, \quad \chi=0.6, c_1=c_2=1.7,$ parameters $\chi = 0.7, c_1 = c_2 = 0.3, \ \chi = 0.9, c_1 = c_2 = 0.1$ and $\chi = 0.9, c_1 = c_2 = 3$ taken from Trelea (2003) and $\chi = 0.72984$, $c_1 = 2.04$, $c_2 = 0.94$ taken from Carlisle and Dozier (2001). For all parameters except $\chi = 0.9, c_1 = c_2 = 3$ and $\chi = 0.1, c_1 = c_2 = 2.1$ we could observe the same behavior as with the parameters used in the rest of this work. For parameters like $\chi = 0.9, c_1 = c_2 = 3$ that lead to a diverging swarm as expected no forcing is done, as the swarm increases its distance from the global attractor. An interesting behavior can be observed for the parameters $\chi = 0.1, c_1 = c_2 = 2.1$. For these parameters, we get a larger forcing frequency when initialized only near the global optimum than the forcing frequency at an optimum, resulting from updates of the global attractor. With these parameters, it is unlikely for particles to increase their distance to the global attractor and far more likely to further reduce it. This behavior can be verified by the convergence of the swarm using the classical movement equations with these swarm parameters on the inclined plane function. After a particle starts the lockout phase, the other particles might move the global attractor in the direction of this particle which can cause the lockout phase to be stopped prematurely. However, the positions before and after the improvement of the global attractor need to be closer than δ , as otherwise other particles could have their partial potential increased above δ . We therefore get a running phase that is only driven by the forced moves and with updates in the range of delta. As a representative for negative χ , we tested the parameter set $\chi = -0.7$, $c_1 = c_2 = 0.5$ taken from Trelea (2003). At the optimum, we can see a similar behavior to parameter sets with positive χ values, as the forced and lockout phases are independent of the swarm parameters. However, in the approaching phase we could observe an increased number of forced steps due to the decreasing factor of the old velocity and therefore smaller overall velocities. We have not tested negative values for c_1 or c_2 . If either c_1 or c_2 are negative, a particle instead of moving towards the respective attractor will move directly away from it. If c_1 is negative, L is not an attractor in the classical sense,

but as long as c_2 is positive, G is still an attractor. Analogously, if c_2 is negative, G is not an attractor, whereas L is an attractor, as long as c_1 is positive. For such parameter sets, it is questionable if convergence appears at all.

We can therefore conclude that for parameters that are considered good for PSO, i.e., the algorithm shows reasonable convergence behavior (even if c_1 or c_2 are negative), we can expect the same behavior of f-PSO in respect to forced steps.

5 Termination criteria

In Sect. 4 above, we discussed the behavior of the particle swarm when the global and local attractors are directly at a local optimum. In Schmitt and Wanka (2015), it is shown that with f-PSO the global attractor converges (in the infinite-time sense) to an (at least local) optimum almost surely. However, this convergence can be very slow. Therefore, the particle swarm should be stopped at some time and either be re-started with, when indicated, different parameter settings, or, if the current global attractor does not satisfy the desired quality, to use a different algorithm. As the swarm does not know when it is close to a local optimum and due to the forced steps, the swarm will just converge up to a δ -neighborhood of the global attractor. When the distance of a particle to the global attractor is smaller than δ and the velocity is small then a forced move is performed, which sets the velocity of the particle to a value of up to δ . The particles therefore stay in a range of around δ around the global attractor, see Fig. 6. The global attractor in turn approaches a (local) optimum.

With our observations from Sect. 4, this means that the forcing frequency will exceed a fixed value which can be measured. In the following, we will define a new stopping criterion that stops the algorithm when the f-PSO algorithm reaches this frequency. Note that reaching this forcing frequency is just an indication that the swarm is close to an optimum. We cannot provide bounds on the probability for a swarm to be classified as stagnating while improvements are still possible. However, our experiments in Sect. 5.2 on a selection of the standard benchmark sets show that the stopping criteria is rarely triggered when no stagnation is present.

In case of plateaus, however, the swarm will keep moving inside the plateau and therefore does not converge to a single point. One would not expect that convergence is detected in this situation. Indeed, the stagnation frequency can only be attained if δ is larger than the size of this plateau.

Note that in an idealized setting with arbitrary precision the choice of δ has no influence on the stagnation frequency (see Sect. 4.2.1). However, calculations with actual machine numbers like, e.g., long double variables in C++, have only limited precision. Due to this imprecision, artificial plateaus can be created during the function evaluations, e.g., there might be an interval $U \subset \mathbb{R}$ with positive length such that for all $x, y \in U$ with $x \neq y$, $f(x) \neq f(y)$, but $f(x) \approx f(y)$. In such a case, x and y might yield the same objective function value due to the limited precision.

Choosing δ too small in the order of the size of such plateaus produces a low forcing frequency. Therefore, our termination criterion may not be fulfilled. In our test $\delta = 10^{-7}$ turned out to be a sufficient value to prevent such issues. This issue can be avoided by using a high precision PSO platform like HiPPSO [Raß (2020)].

5.1 Full stagnation criterion

By Theorem 1, we know that the absolute and the relative forcing frequency become fixed if the swarm would be at a (local) optimum, independent of the objective function. So one can use an arbitrary known function with known optimum to experimentally determine this frequency by starting the swarm in the optimum, as we did in Sect. 4.2 for SPHERE.

We will use this observation to describe the first simple termination criterion: we just stop the swarm when the measured absolute forcing frequency begins to stagnate and is close to the fixed forcing frequency around a (local) optimum. In the mathematical analysis of the behavior at an optimum in Sect. 4.2.2, we required the local and global attractors to be at the same point. This however does not happen with arbitrary precision and is quite unlikely otherwise. We can only expect the attractors to be close to an optimum and consequently also close to each other. Therefore, the measured forcing frequency can achieve values that are close to the forcing frequency at an optimum but do not necessarily match them. To compensate for this deviation, we introduce the new parameter γ as a tolerance. If better solutions would be required, one could reduce γ as this reduces the allowed tolerance. As a guideline, one can use a small percentage of the forcing frequency at an optimum.

Definition 5 Let $\sigma_{\text{stag}}(N, D, \mu)$ be the objective function-independent, fixed absolute forcing frequency around an optimum when intervals are considered with interval length μ , and let $\gamma \in \mathbb{N}$. The criterion " (γ, μ) -FULL STOP" is to terminate the f-PSO execution iff $\sigma_{\text{stag}}(N, D, \mu) - \sigma(I) \leq \gamma$ for some interval I with $|I| = \mu$.

We tested (γ , μ)-FULL STOP in Line 17 of Algorithm 1, i.e., f-PSO, on our benchmark functions with N = 5 particles, D = 15 dimensions, $\delta = 10^{-7}$ and interval length $\mu = 50,000$. The standard swarm parameters were used.

The stagnation frequency σ_{stag} (5,15,50,000) = 318,350 was measured on the SPHERE function with all particles initialized at the global optimum. We used $\gamma = 1350$, so we terminated the execution when $\sigma(I) \ge 317,000$ for intervals *I* of length μ =50,000.

We compared the solutions obtained with f-PSO that used (γ, μ) -FULL STOP with the results obtained by f-PSO that run for 15,000,000 iterations. Table 1 shows the median (not the average) and standard deviation on 500 runs of the norm of the gradient of the objective function at the global attractor with and without the use of criterion (γ , μ)-FULL STOP. Note that the measurements are done in the same run for both values, e.g., the swarm was not stopped after the termination was signaled by (γ, μ) -FULL STOP only the value of the gradient was recorded. In our test, we want to measure the quality of the returned solution in respect to the convergence to a (local) optimum. In the literature, the fitness value of the returned solution is used. However, when testing multi-modal functions we cannot expect the swarm to converge to the same optimum in every run and therefore the fitness value cannot represent a measure for the convergence to a local optimum. To account for multimodal functions, we use the norm of the gradient at the global attractor, i.e., the 2-norm of the partial derivative vector which can be calculated for all used benchmark functions, as a measure for the convergence. For continuous one-dimensional functions, the derivative is zero for local minima/maxima. The same holds for more dimensions. The less the norm of the gradient, the better the global attractor. A small value in this measure indicates that the improvements in the region around the global attractor will be small, as we can only guarantee a convergence to a local optimum, with no indication on the function value.

Function f	iter _{term}	$ \nabla f(G_{\text{iter}_{term}}) $	$ \nabla f(G_{15,000,000}) $
Sphere	$100,000 \pm 0$	$6.65 \times 10^{-8} \pm 6.93 \times 10^{-9}$	$3.60 \times 10^{-8} \pm 3.83 \times 10^{-9}$
H. C. Elliptic	$300,000 \pm 154,324$	$2.22 \times 10^{-5} \pm 1.93 \times 10^{-5}$	$1.17 \times 10^{-5} \pm 1.09 \times 10^{-5}$
Schwefel	$150,000 \pm 110,707$	$1.94 \times 10^{-7} \pm 1.20 \times 10^{-7}$	$1.07 \times 10^{-7} \pm 5.96 \times 10^{-8}$
RASTRIGIN	$100,000 \pm 5000$	$1.34 \times 10^{-5} \pm 1.61 \times 10^{-6}$	$7.82 \times 10^{-6} \pm 1.15 \times 10^{-6}$
ROSENBROCK	850,000 ± 490,538	$3.09 \times 10^{-5} \pm 6.45 \times 10^{-6}$	$2.04 \times 10^{-5} \pm 3.80 \times 10^{-6}$
GRIEWANK	$200,000 \pm 2,234,323$	$9.19 \times 10^{-9} \pm 2.03 \times 10^{-9}$	$8.41 \times 10^{-9} \pm 1.83 \times 10^{-9}$
Ackley	$200,\!000 \pm 3,\!600,\!092$	$2.37 \times 10^{-7} \pm 3.57 \times 10^{-2}$	$1.34 \times 10^{-7} \pm 3.57 \times 10^{-2}$

Table 1 Experimental results for (γ, μ) -FULL STOP

iter_{term} denotes the number of iterations (measured every μ th iteration) until (γ , μ)-FULL stop terminates the execution of f-PSO (median and with \pm the standard deviation are shown). G_t denotes the global attractor after *t* iterations. The final two columns present the median and (with \pm) the standard deviation of the gradient at the global attractor over 500 runs of the f-PSO with and without (γ , μ)-FULL stop. In the tests, N = 5, D = 15, $\mu = 50,000$, $\sigma_{\text{stag}}(N, D, \mu) = 318,350$, $\gamma = 1350$, and $\delta = 10^{-7}$ were used. The less the gradient, the better G_t

However, we are only interested if the particle swarm is near a local optimum. Furthermore, as the potential of the swarm is at around δ we do not expect the swarm to switch to another local optimum as this is quite unlikely. Therefore, the gradient provides us with a sufficient measure in regards to the convergence.

On simple benchmark functions like SPHERE criterion (γ , μ)-FULL STOP produces sufficiently good output. We see that the time of termination is significantly earlier than with pre-specified number of iterations and that the difference in the gradient at termination is in almost all cases in the same order of magnitude.

5.2 Partial stagnation criterion

On more complex functions like ROSENBROCK and ACKLEY, we can identify two scenarios that interfere with the FULL STOP termination criterion.

The first scenario can be observed on the ROSENBROCK function. It may occur that some dimensions are already highly optimized, i.e., close to an optimal coordinate, while others are still far away from optimum coordinates, a situation which leads to late termination. With ROSENBROCK when the particles and the attractors are in the valley between the local and the global attractor, the chance to improve on the not yet optimized dimensions is voided by the worsening in the already good dimensions. Therefore, the swarm will not change the local (and, thus, the global) attractors and the attractors are far away from each other. In this case, the optimized dimensions reached the average stagnation frequency $\sigma_{\text{stag}}(N, D, \mu)/D$, but the other dimensions will not get forced much or not at all as shown in Fig. 2.

The second scenario becomes visible on the ACKLEY function. This function has many local optima with a global optimum at **0**. Unlike ROSENBROCK, there is not much difference in the behavior of the particles in the different dimensions. Here, the problem originates from many local optima with only a small difference in function values. This can lead to the global attractor being in a better local optimum than some of the local attractors. Given this situation, the particles have a very small probability to actually improve their local

attractor. At this point, the algorithm should terminate, but given the different positions of the global and the local attractors these particles cannot decrease their velocity below the necessary bounds to reach the probability of a forced step at the stagnation frequency. The current σ will stay around some value, which is different from the stagnation frequency of each dimension and with a fluctuation larger than the fluctuation at the stagnation frequency. This phenomenon accounts for the large standard deviation, when optimizing the ACKLEY function.

To account for such situations, we formulate an extended termination criterion.

Definition 6 Let $\sigma_{\text{stag}}(N, D, \mu)$ be as defined in Definition 5, and let $\kappa \in [1, D]$. The criterion " (κ, γ, μ) -partial stop" is to terminate the f-PSO execution iff for some interval *I* with $|I| = \mu$ holds $\sigma(I) \ge \kappa \cdot \frac{\sigma_{\text{stag}}(N, D, \mu) - \gamma}{D}$

Definition 6 introduces the additional parameter κ . It can be interpreted as how many dimensions are required to reach their stagnation frequency such that f-PSO is stopped. Actually, if $\kappa = D$, we have (γ, μ) -FULL STOP. By this parameter, the user can choose the degree of convergence necessary before terminating. On complex objective functions that take a long time to evaluate, a small κ may be favorable leading to earlier termination and, hence, saving a lot of running time for applying further subsequent optimization methods to further improve the solution quality. On simple, easily to be evaluated functions such as Sphere, one may choose $\kappa = D$, which results in (γ, μ) -full stop, as already mentioned. With (κ, γ, μ) -PARTIAL STOP, both interfering scenarios are accounted for. There is no differentiation between dimensions, so it makes no difference that like in the first scenario some dimensions d are forced such that they have reached (or are close to) their stagnation value $\partial \sigma_d$, while in other dimensions still almost no forcing takes place. Similarly, it might happen, like in the second scenario, that in all dimensions the frequency is less than the stagnation value, but the overall sum has reached the stagnation value up to a value of $\frac{\kappa}{D}$. We executed a first series of experiments similar to the case with the criterion (γ, μ) -FULL STOP. Again, N = 5 particles, D = 15 dimensions, $\delta = 10^{-7}$, interval length $\mu = 50,000$, $\sigma_{\text{stag}}(N, D, \mu) = 318,350 \text{ and } \gamma = 1350.$ We tested (κ, γ, μ) -partial stop with $\kappa = 2$ and $\kappa = 8$, and for the fixed iteration limit, we used 15,000,000 iterations. We repeated the experiments 500 times.

Tables 2 and 3 show the results of the experiments. We can see that the median of the gradient of the terminated PSO is in the same order of magnitude as the gradient of the PSO after 15, 000, 000 iterations with almost the same standard deviation. We can conclude that our termination criterion can stop the PSO reliable when almost no improvement is to be expected. This is further emphasized by the geometric mean being in the same order of magnitude as shown in Table 3. Given only small deviations of the gradient between the different values of κ as shown in Table 2 we conclude that in most cases a small value of κ is sufficient without a significant loss in the quality of the returned solution. However, the ROSENBROCK measurement suggests that on complex objective functions, a small value of κ might lead to too early termination.

If there are many particles and/or a small number of dimensions, the neighborhood of a local optimum can be reached way sooner. Therefore, the question comes up whether the interval length can be reduced without losing the quality of (κ, γ, μ) -PARTIAL STOP. We repeated our experiments for an interval length of $\mu = 5000$. Here, N = 5 particles, D = 15 dimensions, $\delta = 10^{-7}$ and adjusted, $\sigma_{stag}(N, D, \mu) = 31,835$, and $\gamma = 135$ were applied. The results are presented in Tables 4 and 5.

As presumed for the simple functions, f-PSO with (κ, γ, μ) -PARTIAL STOP terminates significantly earlier with only a small decline in the quality of the returned solution. For ROSENBROCK, the difference in the quality of the returned solution with different choices of the parameter κ becomes more distinct. Furthermore, it should be noted that as shown in Fig. 5 the standard deviation increases drastically with shorter intervals. In practice, a small tolerance is used when testing if the stagnation value is reached. Given this standard deviation on small intervals, a tolerance that compensates this standard deviation can lead to a too early termination when the stagnation value is not reached and the swarm would recover from the stagnation. This leads to a trade-off between interval length and the risk of a termination when the swarm has not reached or is close to a local optimum.

We further compared our new stopping criterion against three stopping criteria introduced by Zielinski and Laur (2007).

The three stopping criteria used were:

- ImpBest: Termination is signaled if the improvement of the best function value falls below a specified bound.
- ImpAv: Termination is signaled if the arithmetic mean of the improvements for the whole population falls below a specified bound.
- MaxDistQuick: Termination is signaled if the maximal distance to the best individual falls below a certain bound for the best particles. The particles are sorted and only the best particles are considered as specified by a parameter.

We used the parameters as suggested by the authors. We only changed the threshold to 10^{-7} , the same range as our δ , for a better comparison with our stopping criteria.

Table 6 shows for the three stopping criteria the median and the standard deviation of the iterations when the PSO was terminated and the gradient of the global attractor at that time over 100 runs. Stopping criterion *ImpBest* shows the worst results of the three tested criteria, while the results for *ImpAvg* and *MaxDistQuick* are comparable on simple functions like SPHERE. However, on more complex functions *MaxDistQuick* outperformed *ImpAvg* with much less standard deviation.

If we compare the results with the results obtained using our introduced stopping criterion and the values after 15, 000, 000 iterations shown in Table 4, we can see that our stopping criterion stops the PSO algorithm later than *MaxDistQuick*; however, we achieve overall better results for the median and standard deviation and for nearly all functions our stopping criterion is in the same order of magnitude as the value after 15, 000, 000 iterations. This is not the case for the three stopping criteria *ImpBest*, *ImpAv* and *MaxDistQuick* that are on functions like ACKLEY way worse by several orders of magnitude.

We can conclude that our stopping criterion stops the PSO algorithm when the probability for an update of the global attractor is low, which is not the case for the tested stopping criteria from the literature. Furthermore, our stopping criterion only has to count the number of forced steps while more complex calculations are necessary for *ImpBest*, *ImpAv* and *MaxDistQuick*.

If we increase the swarm size, the observations still hold for almost all benchmark functions, see Table 7. For a more detailed discussion, see Appendix A. The situation as described in Sect. 5.1 for the ACKLEY function is amplified as more local attractor are present with more particles. Consequently, it occurs on the functions ROSENBROCK and ACKLEY.

Function f	iter _{<i>k</i>=2}	$ \nabla f(G_{\operatorname{iter}_{\kappa=2}}) $	$\text{iter}_{\kappa=8}$	$ \nabla f(G_{\operatorname{iter}_{\kappa=8}}) $	$ \nabla f(G_{15,000,000}) $
Sphere	50,000	7.24×10^{-8}	50,000	7.27×10^{-8}	3.60×10^{-8}
	±0	$\pm 7.50 \times 10^{-9}$	±0	$\pm 8.17 \times 10^{-9}$	$\pm 3.83 \times 10^{-9}$
H. C. Elliptic	50,000	2.52×10^{-5}	50,000	2.44×10^{-5}	1.17×10^{-5}
	±0	$\pm 2.22 \times 10^{-5}$	±0	$\pm 2.24 \times 10^{-5}$	±1.09 × 10 ⁻⁵
Schwefel	50,000 ±0	2.22×10^{-7} ±1.24 × 10 ⁻⁷	50,000 ±0	2.14×10^{-7} $\pm 1.31 \times 10^{-7}$	$1.07 \times 10^{-7} \pm 5.96 \times 10^{-8}$
RASTRIGIN	50,000 ±0	1.45×10^{-5} ±1.65 × 10 ⁻⁶	50,000 ±0	1.46×10^{-5} ±1.72 × 10 ⁻⁶	$7.82 \times 10^{-6} \pm 1.15 \times 10^{-6}$
ROSENBROCK	100,000	8.54×10^{-5}	150,000	7.02×10^{-5}	2.04×10^{-5}
	± 730,696	$\pm 3.78 \times 10^{-5}$	±469,614	±1.80 × 10^{-5}	$\pm 3.80 \times 10^{-6}$
Griewank	50,000	1.01×10^{-8}	50, 000	1.03×10^{-8}	8.41×10^{-9}
	±0	$\pm 2.43 \times 10^{-9}$	±0	$\pm 2.39 \times 10^{-9}$	$\pm 1.83 \times 10^{-9}$
Ackley	150,000	2.69×10^{-7}	200,000	2.51×10^{-7}	1.34×10^{-7}
	±3,094,564	$\pm 3.52 \times 10^{-2}$	± 3,469,939	$\pm 1.78 \times 10^{-2}$	$\pm 3.57 \times 10^{-2}$

Table 2 Experimental results for (κ, γ, μ) -PARTIAL STOP (for notions, see Table 1)

Median and (with ±) the standard deviation over 500 runs of the stopped f-PSO with N = 5, D = 15, $\mu = 50,000$, $\sigma_{stag}(N, D, \mu) = 318,350$, $\gamma = 1350$, and $\delta = 10^{-7}$, and $\kappa = 2$ and $\kappa = 8$, resp., against an "unstopped" f-PSO terminated after 15,000,000 iterations

Table 3 Experimental results for (κ, γ, μ) -PARTIAL STOP (for notions, see Table 1)

Function f	$iter_{\kappa=2}$	$ \nabla f(G_{\mathrm{iter}_{\kappa=2}}) $	$\text{iter}_{\kappa=8}$	$ \nabla f(G_{\mathrm{iter}_{\kappa=8}}) $	$ \nabla f(G_{15,000,000}) $
Sphere	50,000	7.15×10^{-8}	50,000	7.12×10^{-8}	3.54×10^{-8}
H. C. Elliptic	50,000	1.95×10^{-5}	50,000	1.95×10^{-5}	9.41×10^{-6}
Schwefel	50,000	1.04×10^{-7}	50,000	1.07×10^{-7}	1.04×10^{-7}
RASTRIGIN	50,000	1.43×10^{-5}	50,000	1.44×10^{-5}	7.81×10^{-6}
ROSENBROCK	153,672	8.80×10^{-5}	176,989	6.91×10^{-5}	2.03×10^{-5}
GRIEWANK	50,000	1.02×10^{-8}	50,000	1.04×10^{-8}	8.46×10^{-9}
Ackley	1,321,300	2.94×10^{-7}	1,609,300	2.90×10^{-7}	1.64×10^{-7}

Geometric Mean over 500 runs of the stopped f-PSO with N = 5, D = 15, $\mu = 50,000$, $\sigma_{stag}(N, D, \mu) = 318,350$, $\gamma = 1350$, and $\delta = 10^{-7}$, and $\kappa = 2$ and $\kappa = 8$, resp., against an 'unstopped' f-PSO terminated after 15,000,000 iterations

However, we can argue that in this situation the swarm is still exploring the search space, indicated by a potential larger than δ and might find better solutions therefore the algorithm intentionally should not be terminated.

6 Conclusions

This paper focused on the behavior of the f-PSO algorithm at a (local) optimum. We could identify three distinct phases that are repeated when the swarm reaches an optimum, which leads to a stagnation of the forcing frequency introduced in this work as a measure for the influence of the f-PSO modification. We showed that at an optimum the swarm behaves like a blind searching algorithm and that therefore the number of forced steps in a time

Function f	iter -	$ \nabla f(G, \cdot) $	iter -	$ \nabla f(G, \cdot) $	$ \nabla f(G_1, \dots, M_n) $
	πer _{κ=2}	$\prod \sqrt{O_{iter_{\kappa=2}}}$	ner _{K=8}	$ \mathbf{v} (\mathbf{O}_{iter_{\kappa=8}}) $	11 •) (015,000,000)11
Sphere	5000	1.10×10^{-7}	10,000	9.52×10^{-8}	3.63×10^{-8}
	±0	±1.37 × 10 ⁻⁸	±2200	±1.22 × 10 ⁻⁸	$\pm 3.73 \times 10^{-9}$
H. C. Elliptic	5000 ± 1310	1.11×10^{-4} ±1.54 × 10 ⁻⁴	10,000 ±0	3.93×10^{-5} $\pm 3.73 \times 10^{-5}$	1.28×10^{-5} ±1.00 × 10^{-5}
Schwefel	15,000	3.87×10^{-7}	15,000	3.12×10^{-7}	1.01×10^{-7}
	±2502	$\pm 2.18 \times 10^{-7}$	±1034	$\pm 1.82 \times 10^{-7}$	±6.02 × 10 ⁻⁸
RASTRIGIN	5000	2.12×10^{-5}	5000	1.97×10^{-5}	7.82×10^{-6}
	±0	$\pm 2.69 \times 10^{-6}$	<u>+</u> 2449	$\pm 2.64 \times 10^{-6}$	$\pm 1.13 \times 10^{-6}$
ROSENBROCK	65,000	1.40×10^{-3}	80,000	1.74×10^{-4}	2.04×10^{-5}
	±442,699	±1.04 × 10^{-3}	± 381,917	$\pm 5.34 \times 10^{-5}$	$\pm 4.03 \times 10^{-6}$
Griewank	5000	1.25×10^{-8}	5000	1.21×10^{-8}	8.41×10^{-9}
	±0	$\pm 3.41 \times 10^{-9}$	<u>+</u> 2466	$\pm 3.25 \times 10^{-9}$	$\pm 1.83 \times 10^{-9}$
Ackley	150,000	3.71×10^{-7}	170,000	3.40×10^{-7}	1.34×10^{-7}
	± 3,262,633	±4.20 × 10 ⁻²	± 3,311,552	±6.36 × 10^{-2}	$\pm 3.57 \times 10^{-2}$

Table 4 Experimental results for (κ, γ, μ) -PARTIAL STOP with reduced interval length (for notions, see Table 1)

Median and (with \pm) the standard deviation over 500 runs of the stopped f-PSO with N = 5, D = 15, $\mu = 5000$, $\sigma_{stag}(N, D, \mu) = 31,835$, $\gamma = 1350$, and $\delta = 10^{-7}$, and $\kappa = 2$ and $\kappa = 8$, resp., against an "unstopped" f-PSO terminated after 15,000,000 iterations

Table 5 Experimental results for (κ, γ, μ) -partial stop with reduced interval length (for notions, see Table 1)

Function f	$\text{iter}_{\kappa=2}$	$ \nabla f(G_{\mathrm{iter}_{\kappa=2}}) $	$\text{iter}_{\kappa=8}$	$ \nabla f(G_{\operatorname{iter}_{\kappa=8}}) $	$ \nabla f(G_{15,000,000}) $
Sphere	5000	1.09×10^{-7}	8337	9.41×10^{-8}	3.58×10^{-8}
H. C. Elliptic	5262	9.68×10^{-5}	10,000	3.15×10^{-5}	9.62×10^{-6}
Schwefel	12,276	3.84×10^{-7}	15,150	3.15×10^{-7}	1.04×10^{-7}
RASTRIGIN	5000	2.09×10^{-5}	6587	1.96×10^{-5}	7.75×10^{-6}
ROSENBROCK	92,560	1.41×10^{-3}	111,448	1.80×10^{-4}	2.01×10^{-5}
GRIEWANK	5000	1.27×10^{-8}	6669	1.22×10^{-8}	8.46×10^{-9}
Ackley	1,326,660	5.92×10^{-7}	1,344,320	3.73×10^{-7}	1.64×10^{-7}

Geometric Mean over 500 runs of the stopped f-PSO with N = 5, D = 15, $\mu = 5000$, $\sigma_{stag}(N, D, \mu) = 31,835$, $\gamma = 1350$, and $\delta = 10^{-7}$, and $\kappa = 2$ and $\kappa = 8$, resp., against an "unstopped" f-PSO terminated after 15,000,000 iterations

interval of specified size, the absolute forcing frequency, is larger than a fitness function independent number. Thus, when reaching this frequency it can be presumed that the swarm is close to an optimal solution, and it can be stopped. We introduced two new stopping criteria for the f-PSO algorithm and experimentally verified their performance on a set of standard benchmark functions. These stopping criteria need the f-PSO modification; however, as this modification is quite small it can be applied to most PSO variants in the literature.

-			-				
	ImpBest		ImpAv		MaxDistQuic	MaxDistQuick	
Function f	iter	$ \nabla f(G) $	iter	$ \nabla f(G) $	iter	$ \nabla f(G) $	
Sphere	1638 ±282	1.30×10^{-2} ±5.34 × 10^{-2}	2371 ±308	6.36×10^{-7} $\pm 3.32 \times 10^{-7}$	2307 ±289	2.90×10^{-6} $\pm 2.72 \times 10^{-6}$	
H. C. Elliptic	1385 ±955	$5.84 \pm 1.99 \times 10^{6}$	2653 ±395	1.56×10^{-3} ±1.77 × 10 ⁻³	3142 ±586	6.71×10^{-4} $\pm 1.17 \times 10^{-3}$	
Schwefel	2355 ±2335	$79.79 \pm 1.05 \times 10^3$	9210 ±1791	5.84×10^{-6} $\pm 2.87 \times 10^{2}$	8798 ±1168	1.61×10^{-5} $\pm 3.02 \times 10^{-5}$	
RASTRIGIN	423 ±581	$149.46 \pm 3.69 \times 10^2$	2045 ±538	1.32×10^{-4} ±4.38 × 10 ¹	1980 ±321	5.85×10^{-4} ±1.21 × 10 ⁻³	
Rosenbrock	3102 ±7029	$155.63 \pm 9.50 \times 10^9$	17, 961 ±16, 102	4.55×10^{-1} $\pm 4.54 \times 10^{4}$	42,824 ±234,029	3.85×10^{-3} ±1.81	
Griewank	1067 ±322	6.20×10^{-3} $\pm 5.56 \times 10^{-2}$	2102 ±304	6.19×10^{-8} $\pm 6.38 \times 10^{-3}$	2060 ±319	2.12×10^{-7} $\pm 2.62 \times 10^{-7}$	
Ackley	25	1.71	62	1.81	115,257	8.30×10^{-6}	

Table 6 Experimental results of the stopping criteria ImpBest, ImpAv and MaxDistQuick

Median and (with \pm) the standard deviation over 100 runs of the stopped f-PSO with N = 5, D = 15 and the standard swarm parameters

 ± 27

 $\pm 1.79 \times 10^{-1}$

 $\pm 2,201,190$

 $\pm 1.87 \times 10^{-1}$

 Table 7
 Stopping criterion for an increased number of particles

 ± 13

 $\pm 1.88 \times 10^{-1}$

Function f	$\text{iter}_{\kappa=8}$	$ \nabla f(G_{\mathrm{iter}_{\kappa=8}}) $	$ \nabla f(G_{15\times 10^6}) $	$\text{iter}_{\kappa=8}$	$ \nabla f(G_{\mathrm{iter}_{\kappa=8}}) $	$ \nabla f(G_{15\times 10^6}) $
Sphere	50,000	2.74×10^{-8}	1.02×10^{-8}	50,000	4.19×10^{-9}	7.37×10^{-10}
	±0	$\pm 3.32 \times 10^{-9}$	±1.21 × 10 ⁻⁹	±0	$\pm 6.80 \times 10^{-10}$	$\pm 1.16 \times 10^{-10}$
H. C. Ellip-	50,000	8.50×10^{-6}	2.66×10^{-6}	50, 000	$1.53 \times 10^{-6} \pm 1.14 \times 10^{-6}$	1.89×10^{-7}
tic	±0	$\pm 7.76 \times 10^{-6}$	$\pm 2.51 \times 10^{-6}$	±0		$\pm 2.03 \times 10^{-7}$
Schwefel	50,000	1.06×10^{-7}	3.39×10^{-8}	50,000	2.04×10^{-8}	3.07×10^{-9}
	±0	$\pm 5.44 \times 10^{-8}$	$\pm 2.05 \times 10^{-8}$	±0	$\pm 1.17 \times 10^{-8}$	$\pm 2.20 \times 10^{-9}$
RASTRIGIN	50,000 ±0	5.59×10^{-6} $\pm 9.61 \times 10^{-7}$	$2.78 \times 10^{-6} \pm 5.06 \times 10^{-7}$	50, 000 ±0	1.72×10^{-6} $\pm 2.81 \times 10^{-7}$	1.30×10^{-6} $\pm 2.06 \times 10^{-7}$
ROSENBROCK	15, 000, 000	3.48×10^{-17}	3.48×10^{-17}	15,000,000	3.47×10^{-17}	3.47×10^{-17}
	± 6,868,757	$\pm 3.21 \times 10^{-6}$	$\pm 3.21 \times 10^{-6}$	±5,658,620	$\pm 2.96 \times 10^{-7}$	$\pm 2.96 \times 10^{-7}$
Griewank	50,000 ±0	8.14×10^{-9} $\pm 2.21 \times 10^{-9}$	$7.53 \times 10^{-9} \pm 2.20 \times 10^{-9}$	50, 000 ±0	6.96×10^{-9} $\pm 2.03 \times 10^{-9}$	6.68×10^{-9} $\pm 2.22 \times 10^{-9}$
Ackley	$\begin{array}{c} 1,150,000\\ \pm4,745,144 \end{array}$	$7.68 \times 10^{-8} \pm 6.76 \times 10^{-9}$	7.53×10^{-8} $\pm 7.05 \times 10^{-9}$	$4,850,000 \pm 6,206,492$	7.47×10^{-8} $\pm 5.66 \times 10^{-9}$	7.62×10^{-8} $\pm 5.80 \times 10^{-9}$

Acknowledgements We would like to thank Alexander Raß for fruitful discussions and the anonymous referees for valuable remarks.

Funding Research funded in parts by the School of Engineering of the University of Erlangen-Nuremberg. Open Access funding provided by Projekt DEAL.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

A Increased swarm sizes

At an optimum, i.e., $\forall n : L^n = G$, increasing the swarm size does not change the fundamental behavior of the swarm in the pulsation phase, see Fig. 7.

As the fundamental behavior at an optimum is unchanged, it remains to explore how the increased swarm size impacts our stopping criterion. On most benchmark functions, the number of iterations after which the algorithm was stopped decreases as the swarm size increases. However, on complex multimodal benchmark functions there are runs where our stopping criterion did not stop the algorithm in the first 15,000,000 iterations. This issue was mostly encountered on the ROSENBROCK function. The reason that the stopping criterion was not triggered comes from the issue described in Sect. 5.1 for the ACKLEY function. Some local attractors get stuck in a local optimum and the probability for improvements that decrease the distance between the local and global attractors is close to zero. The swarm was not able to improve the local attractors after some point, and the distance is large enough that no forced moves are performed. The probability for such a situation gets higher with an increase in the number of particles. However, as the velocity of the particles in such a situation is still at around 10^{-1} or even higher in some runs, we can argue that the



Fig. 7 Cumulative number of forced steps at the global optimum of the SPHERE benchmark function for different swarm sizes

swarm is still in the exploration phase, and therefore, the algorithm intentionally should not be stopped even though the global attractor is coincidentally at the global optimum.

References

- Aytug, H., & Koehler, G. J. (2000). New Stopping Criterion for Genetic Algorithms. European Journal of Operational Research, 126(3), 662–674. https://doi.org/10.1016/S0377-2217(99)00321-5.
- Bassimir, B., Schmitt, M., & Wanka, R. (2014). How Much Forcing is Necessary to Let the Results of Particle Swarms Converge? In Proceedings of International Conference on Swarm Intelligence Based Optimization (ICSIBO), pp 98–105, https://doi.org/10.1007/978-3-319-12970-9_11
- Carlisle, A., & Dozier, G. (2001). An Off-the-Shelf PSO. In Proceedings of the Workshop on Particle Swarm Optimization
- Chow, Y. S., & Teicher, H. (1978). Probability Theory. Berlin: Springer. https://doi. org/10.1007/978-1-4612-1950-7.
- Chung, K. L. (1974). Elementary Probability theory with Stochastic Processes. Berlin: Springer. https:// doi.org/10.1007/978-0-387-21548-8.
- Clerc, M., & Kennedy, J. (2002). The Particle Swarm—Explosion, Stability, and Convergence in a Multidimensional Complex Space. *IEEE Transactions on Evolutionary Computation*, 6, 58–73. https:// doi.org/10.1109/4235.985692.
- Dorea, C. C. Y. (1990). Stopping rules for a Random Optimization Method. SIAM Journal on Control and Optimization, 28(4), 841–850. https://doi.org/10.1137/0328048.
- Eberhart RC, Kennedy J (1995) A New Optimizer using Particle Swarm Theory. In Proceedings of 6th International Symposium on Micro Machine and Human Science, pp 39–43, https://doi. org/10.1109/MHS.1995.494215
- Engelbrecht, A.P. (2014). Fitness function evaluations: A fair stopping condition? In 2014 IEEE Symposium on Swarm Intelligence (SIS), pp 181–188, https://doi.org/10.1109/SIS.2014.7011793
- Helwig, S. (2010). Particle swarms for constrained optimization. PhD thesis, Department of Computer Science, University of Erlangen–Nuremberg, Germany, urn:nbn:de:bvb:29-opus-19334
- Kennedy, J., Eberhart, R.C. (1995). Particle Swarm Optimization. In Proceedings of IEEE International Conference on Neural Networks, vol 4, pp 1942–1948, https://doi.org/10.1109/ICNN.1995.488968
- Kwok, N.M., Ha, Q.P., Liu, D.K., Fang, G., Tan, K.C. (2007). Efficient Particle Swarm Optimization: a termination condition based on the decision-making approach. In Proceedings of IEEE Congress on Evolutionary Computation (CEC), pp 3353–3360, https://doi.org/10.1109/CEC.2007.4424905
- Lehre, P.K., Witt, C. (2013). Finite first hitting time versus stochastic convergence in particle swarm optimisation. In: Advances in Metaheuristics, Operations Research/Computer Science Interfaces Series, vol 53, Springer, pp 1–20, https://doi.org/10.1007/978-1-4614-6322-1_1
- Martí, L., García, J., Berlanga, A., & Molina, J. (2016). A stopping criterion for multi-objective optimization evolutionary Algorithms. *Information Sciences*, 367–368, 700–718. https://doi. org/10.1016/j.ins.2016.07.025.
- Ong, B. T., & Fukushima, M. (2015). Automatically terminated Particle Swarm Optimization with principal component analysis. *International Journal of Information Technology & Decision Making*, 14(1), 171–194. https://doi.org/10.1142/S0219622014500837.
- Raß, A. (2020). High Precision Particle Swarm Optimization Algorithm (HiPPSO). Journal of Open Research Software, 8, 4. https://doi.org/10.5334/jors.282.
- Raß, A., Schmitt, M., Wanka, R. (2015). Explanation of stagnation at points that are not local optima in Particle Swarm Optimization by potential analysis. In Companion of Proceedings of 17th Genetic and Evolutionary Computation Conference (GECCO), pp 1463–1464, https://doi. org/10.1145/2739482.2764654
- Ribeiro, C.C., Rosseti, I., Souza, R.C. (2011). Effective probabilistic stopping rules for randomized metaheuristics: GRASP implementations. In Proceedings of 5th International Conference on Learning and Intelligent Optimization (LION), pp 146–160, https://doi.org/10.1007/978-3-642-25566-3_11
- Safe, M., Carballido, J., Ponzoni, I., Brignole, N. (2004). On stopping criteria for genetic Algorithms. In Proceedings of 17th Brazilian Symposium on Artificial Intelligence (SBIA), pp 405–413, https:// doi.org/10.1007/978-3-540-28645-5_41

- Schmitt, M., & Wanka, R. (2013a). Particle Swarm Optimization almost surely finds local optima. In Proceedings of 15th Genetic and Evolutionary Computation Conference (GECCO), pp 1629–1636, https://doi.org/10.1145/2463372.2463563
- Schmitt, M., & Wanka, R. (2013b). Particles prefer walking along the axes: Experimental insights into the behavior of a particle swarm. In Companion of Proceedings of 15th Genetic and Evolutionary Computation Conference (GECCO), pp 17–18, https://doi.org/10.1145/2464576.2464583
- Schmitt, M., & Wanka, R. (2015). Particle Swarm Optimization almost surely finds local optima. Theoretical Computer Science 561, Part A:57–72, https://doi.org/10.1016/j.tcs.2014.05.017, extended, full version of Schmitt and Wanka (2013a)
- Sharma, S., & Rangaiah, G. P. (2013). An improved multi-objective differential evolution with a termination criterion for optimizing chemical processes. *Computers & Chemical Engineering*, 56, 155– 173. https://doi.org/10.1016/j.compchemeng.2013.05.004.
- Solis, F. J., & Wets, R. J. B. (1981). Minimization by random search techniques. Mathematics of Operations Research, 6(1), 19–30. https://doi.org/10.1287/moor.6.1.19.
- Suganthan, P.N., Hansen, N., Liang, J.J., Deb, K., Chen, Y.P., Auger, A., & Tiwari, S. (2005). Problem definitions and evaluation criteria for the CEC 2005 special session on real-parameter optimization. Technical report, KanGAL Report Number 2005005 (Kanpur Genetic Algorithms Laboratory, IIT Kanpur)
- Trelea, I. C. (2003). The Particle Swarm Optimization Algorithm: Convergence analysis and parameter selection. *Information Process Letter*, 85, 317–325. https://doi.org/10.1016/S0020-0190(02)00447 -7.
- van den Bergh, F., & Engelbrecht, A.P. (2002). A new locally convergent particle swarm optimiser. In Proceedings of IEEE International Conference on Systems, Man and Cybernetics (SMC), vol 3, pp 94–99, https://doi.org/10.1109/ICSMC.2002.1176018
- Yi, D., & Seppi, K.D., & Goodrich, M.A. (2018). Understanding Particle Swarm Optimization: A component-decomposition perspective. In Proceedings of IEEE Congress on Evolutionary Computation (CEC), https://doi.org/10.1109/CEC.2018.8477832
- Zielinski, K., & Laur, R. (2007). Stopping criteria for a constrained single-objective Particle Swarm Optimization Algorithm. *Informatica*, *31*, 51–59.
- Zielinski, K., Peters, D., & Laur, R. (2005). Stopping criteria for single-objective optimization. In Proceedings of 3rd International Conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.