



HAL
open science

Weak convergence of particle swarm optimization

Vianney Bruned, André Mas, Sylvain Wlodarczyk

► **To cite this version:**

Vianney Bruned, André Mas, Sylvain Wlodarczyk. Weak convergence of particle swarm optimization. 2024. hal-01918943v3

HAL Id: hal-01918943

<https://hal.science/hal-01918943v3>

Preprint submitted on 18 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Weak convergence of particle swarm optimization

Vianney Bruned¹, André Mas², and Sylvain Wlodarczyck¹

¹Schlumberger Petroleum Services, Montpellier, 34000, France.

²IMAG, Univ Montpellier, CNRS, Montpellier, France.

Abstract

Particle swarm optimization algorithm is a stochastic meta-heuristic solving global optimization problems appreciated for its efficiency and simplicity. It consists in a swarm of interacting particles exploring a domain and searching a global optimum. The trajectory of the particles has been well-studied in a deterministic case. More recently authors shed some light on the stochastic approach to PSO, considering particles as random variables and studying them with probabilistic and statistical tools. These works paved the way to the present article. We focus here on weak convergence, the kind of stochastic convergence that appears in the Central Limit Theorem. We obtain three main results related to three different frameworks. These three settings depend either on the regime of the particles (oscillation or fast convergence) or on the sampling strategy (along the trajectory or in the swarm). The main application of these results is the construction of confidence intervals around the local optimum found by the swarm. The theorems are illustrated by a simulation study.

Keywords : Particle swarm optimization;Convergence;Central limit theorem

1 Introduction

intro

The particle swarm optimization algorithm (PSO), based on social interactions (behaviors of birds) was introduced in Eberhart and Kennedy (1995). Since then PSO has known a great popularity in many domains and gave birth to many variants of the original algorithm (see Zhang et al. (2015) for a survey of variants and applications). PSO is a stochastic meta heuristic solving an optimization problem without any evaluation of the gradient. The algorithm explores the search space in an intelligent way thanks to a population of particles interacting with each other and updating at each step their position and their velocity. The dynamic of the particles relies on two attractors: their personal best position (historical best position of the particle denoted p_n^s below), and the neighborhood best position (corresponding to the social component of the particles, denoted g_n^s). In the dynamic equation, the attractors are linked with a stochastic process in order to explore the search space. Algorithm 1 refers to the classical version of PSO with S particles and N iterations.

Algorithm 1 Classical PSO

Initialize the swarm of S particles with random positions x_0^s and velocities v_0^s over the search space.

for $n = 1$ to N **do**

 Evaluate the optimization fitness function for each particle.

 Update p_n^s (personal best position) and g_n^s (neighborhood best position).

 Change velocity (v_n^s) and position (x_n^s) according to the dynamic equation.

end for

algo:1

The convergence and stability analysis of PSO are important matters. In the literature, there are two kinds of convergence:

- the convergence of the particles towards a local or global optimum. This convergence is not obtained with the classical version of PSO. Van den Bergh and Engelbrecht (2010) and Schmitt and Wanka (2015) proposed a modified version of PSO to obtain the convergence.
- the convergence of each particle to a point (e.g. (Poli (2009))).

If we focus on the convergence of each particle to a point, a prerequisite is the stability of the trajectory of the particles. In a deterministic case, Clerc and Kennedy (2002) dealt with the stability of the particles with some conditions on the parametrization of PSO. Later, Kadiramanathan et al. (2006) used the Lyapunov stability theorem to study the stability. About the convergence of PSO, Van Den Bergh and Engelbrecht (2006) looked at the trajectories of the particles

and proved that each particle converges to a stable point (deterministic analysis). Under stagnation hypotheses (no improvement of the personal and neighborhood best positions), Poli (2009) gives the exact formula of the second moment. More recently, Bonyadi and Michalewicz (2016) or Cleghorn and Engelbrecht (2018) provided results for the order-1 and order-2 stabilities with respectively stagnant and non-stagnant distribution assumptions (both weaker than the stagnation hypotheses). Since our main results rely on martingale convergence theorems the other recent work Xu and Yu (2018) has to be mentioned as well.

Let us introduce some notations. We consider here a cost function $f : \mathbb{R}^d \rightarrow \mathbb{R}^+$ that should be minimized on a compact set Ω . Consequently the particles evolve in $\Omega \subset \mathbb{R}^d$.

Let $x_n^s \in \mathbb{R}^d$ $1 \leq s \leq S$ denote the position of particle number s in the swarm at step n . Let $(r_{j,n})_{j=1,\dots,n \geq 1}$ be sequences of independent random vectors in \mathbb{R}^d whose margins are uniformly distributed over $[0, 1]$ and denote by ω , c_1 and c_2 three positive constants which will be discussed later. Then the PSO algorithm considered in the sequel is defined by the two following equations (or dynamic equations, Poli (2009)):

$$\begin{cases} v_{n+1}^s = \omega \cdot v_n^s + c_1 r_{1,n} \odot (p_n^s - x_n^s) + c_2 r_{2,n} \odot (g_n^s - x_n^s), \\ x_{n+1}^s = x_n^s + v_{n+1}^s \end{cases} \quad \text{pso-def (1)}$$

where \odot stands for the Hadamard product:

$$u \odot v = (u_1 v_1, \dots, u_d v_d)$$

and p_n^s (resp. g_n^s) is the best personal position (resp. the best neighborhood position of the particle s):

$$\begin{aligned} p_n^s &= \underset{t \in \{x_0^s, \dots, x_n^s\}}{\operatorname{argmin}} f(t), \\ g_n^s &= \underset{t \in \{p_n^{s'} : s' \in \mathcal{V}(s)\}}{\operatorname{argmin}} f(t) \end{aligned}$$

with $\mathcal{V}(s)$ the neighborhood of particle s . This neighborhood depends on the swarm's topology: if the topology is called global (all the particles communicate between each other) then $g_n^s = g_n = \underset{t \in \{p_n^1, \dots, p_n^S\}}{\operatorname{argmin}} f(t)$ (see Lane et al. (2008)).

Our main objective is to provide (asymptotic) confidence sets for the global or local optimum of the cost function f . If $g = \underset{t \in \Omega}{\operatorname{argmin}} f(t)$ for some domain Ω , a confidence region for g is a random set $\Lambda \subset \mathbb{R}^d$ such that :

$$\mathbb{P}(g \in \Lambda) \geq 1 - \alpha$$

for some small $\alpha \in (0, 1)$. The set Λ depends on the swarm and is consequently random due to the random evolution of the particles. The probability symbol above, \mathbb{P} , depends on the distribution of the particles in the swarm. Let us illustrate the use of confidence interval for a real-valued PSO. Typically the kind of results we expect is: $g \in [m, M]$ with a probability larger than, say 99 %. This does not aim at yielding a precise estimate for g but defines a ‘‘control area’’ for g , as well as a measure of the imprecision and variability of PSO.

Convergence of the swarm will not be an issue here. In fact we assume that the personal and global best converge : see assumptions **A₂** and **B₂** below. We are interested in the ‘‘step after’’ : localizing the limit of the particles with high probability, whatever their initialization and trajectories.

Formally, confidence set estimation forces us to inspect order two terms (i.e. the rate of convergence), typically convergence of the empirical variance. The word *asymptotic* just means that the sample size increases to infinity.

The outline of the paper is the following. In the next section the three main results are introduced. They are all related to weak convergence of random variables and vectors (see Billingsley (1999) for a classical monograph) and obtained under three different sets of assumptions.

The two first consider the trajectory of single particles. The sample consists in the path of a fixed particle. We show that two different regimes should be investigated depending on the limiting behavior of p_n and g_n . Briefly speaking : if the limits of p_n and g_n are distinct, the particles oscillate between them (which is a well-known characteristics of PSO), if the limits of p_n and g_n coincide, then particles converge at a fast, exponential, rate.

In the oscillating case a classical Central Limit Theorem is obtained relying essentially on martingale difference techniques. In the non-oscillating situation, the particle converges quickly and we have to use random matrices products to obtain a non-standard CLT. As by-products of these two subsections we will retrieve confidence sets of the form $\Lambda(x_1^s \dots x_n^s)$, depending on the n positions of each particle x^s .

The third result states another classical CLT. The sample consists here in the whole swarm. This time the confidence set is of the form $\Lambda(x_n^1 \dots x_n^S)$, depending on the S particles of the swarm when the iteration step is fixed at n . A numerical study and simulations are performed in a Python environment. A discussion follows. The derivations of the main theorems are collected in the last section and in the Appendix.

2 Main results

The usual euclidean norm and associated inner product for vectors in \mathbb{R}^d are denoted respectively $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$. If X is a random vector with null expectation then $\mathbb{E}(X \otimes X) = \mathbb{E}(XX^t)$ is the covariance matrix of X . The covariance matrix is crucial since it determines the limiting Gaussian distribution in the Central Limit Theorem. We will need two kinds of stochastic convergence in the sequel. convergence in probability of X_n to X is denoted $X_n \xrightarrow{\mathbb{P}} X$. The arrow \hookrightarrow stands for convergence in distribution (weak convergence).

Except in section 2.3 we consider a single particle in order to alleviate notations. We drop the particle index so that $x_n^s = x_n$, $p_n^s = p_n$ and $g_n^s = g_n$.

At last we take for granted that particles are warm, reached an area of the domain were they fluctuate without exiting (condition **A**₁ below).

2.1 First case: oscillatory ($p \neq g$)

Denote $\xi_n = \max\{|p_n - p|, |g_n - g|\}$ and :

$$\theta = \frac{c_1 p + c_2 g}{c_1 + c_2}, \quad c = \frac{c_1 + c_2}{2}.$$

The following assumptions are required and discussed after the statement of Theorem 1.

A₁ : For all n , $x_n \in \Omega$ where Ω is a compact subset of \mathbb{R}^d .

A₂ : $\sqrt{N}\xi_N \xrightarrow{\mathbb{P}} 0$ when $N \rightarrow +\infty$.

A₃ : The inequality below connects c_1 , c_2 and ω :

$$2c \frac{1 - \omega}{1 + \omega} \left(1 + \omega - \frac{c}{2}\right) > \frac{c_1^2 + c_2^2}{12}$$

Before stating the Theorem we need a last notation. Let $\delta = (\delta_1, \dots, \delta_d) \in \mathbb{R}^d$. The notation $\text{diag}(\delta)$ stands for the diagonal $d \times d$ matrix with entries $\delta_1, \dots, \delta_d$ and $\delta^{\odot 2}$ is the vector in \mathbb{R}^d defined by $\delta^{\odot 2} = (\delta_1^2, \dots, \delta_d^2)$.

TH1

Theorem 1. Set $\mathfrak{L} = c \left(\frac{1-\omega}{1+\omega}\right) \left(1 + \omega - \frac{c}{2}\right)$ and

$$\mathfrak{C} = \frac{1}{24} \frac{c_1 c_2}{c^2} \frac{\mathfrak{L}}{\mathfrak{L} - \frac{c_1^2 + c_2^2}{24}}.$$

Denote finally $\Gamma = \mathfrak{C} \cdot \text{diag}(p - g)^{\odot 2}$ then :

$$\sqrt{N} \left(\frac{1}{N} \sum_{n=1}^N x_n - \theta \right) \hookrightarrow \mathcal{N}(0, \Gamma)$$

where $\mathcal{N}(0, \Gamma)$ denotes the Gaussian centered random vector of \mathbb{R}^d with covariance matrix Γ .

Discussion of the Assumptions:

We avoid here the assumption of stagnation: the personal and local best are not supposed to be constant but they oscillate around their expectation. The convergence occurs at a rate ensuring that neither g_n nor p_n are involved in the weak convergence of the particles x_n . Condition **A**₂ is specific of what we intend by a convergent PSO. It ensures that p_n or g_n have no impact on the weak convergence behavior of the particles. With other words Assumption **A**₂ requires that the oscillations of p_n and g_n around their expectations are negligible. We tried here to model the stagnation phenomenon which consists in sequence of iterations during which g_n (resp. p_n) remain constant for n in $[\underline{N}, \bar{N}]$ supposedly.

Note that assumption **A**₃ is exactly the condition found by Poli in Poli (2009) (see the last paragraph of section III) for defining order 2 stability. This condition may be extended to the case when $c_1 \neq c_2$, see Cleghorn and Engelbrecht (2018) and references therein. At last **A**₃ holds for the classical calibration appearing in Clerc and Kennedy (2002) (constriction constraints) with $c = 1.496172$ and $\omega = 0.72984$.

The next Corollary provides finally two kinds of by-products : asymptotic confidence intervals, in \mathbb{R} , for the coordinates of θ and confidence regions for θ in \mathbb{R}^d . Let $\alpha \in [0, 1]$, q_α be the $1 - \alpha$ quantile of the standard Gaussian distribution and $\chi_{1-\alpha}^{(d)}$ be the $1 - \alpha$ quantile of the Chi-square distribution with d degrees of freedom. Set also $\theta = (\theta_1, \dots, \theta_d)$ and $\bar{x}_N = (\bar{x}_{N,1}, \dots, \bar{x}_{N,d})$.

Corollary 2. Pick $\ell \in \{1, \dots, d\}$. An asymptotic confidence interval at level $1 - \alpha$ for θ_ℓ is directly derived from Corollary ?? :

$$\mathcal{I}_{1-\alpha}(\theta_\ell) = [\bar{x}_{N,\ell} - s_\ell(N, \alpha); \bar{x}_{N,\ell} + s_\ell(N, \alpha)]$$

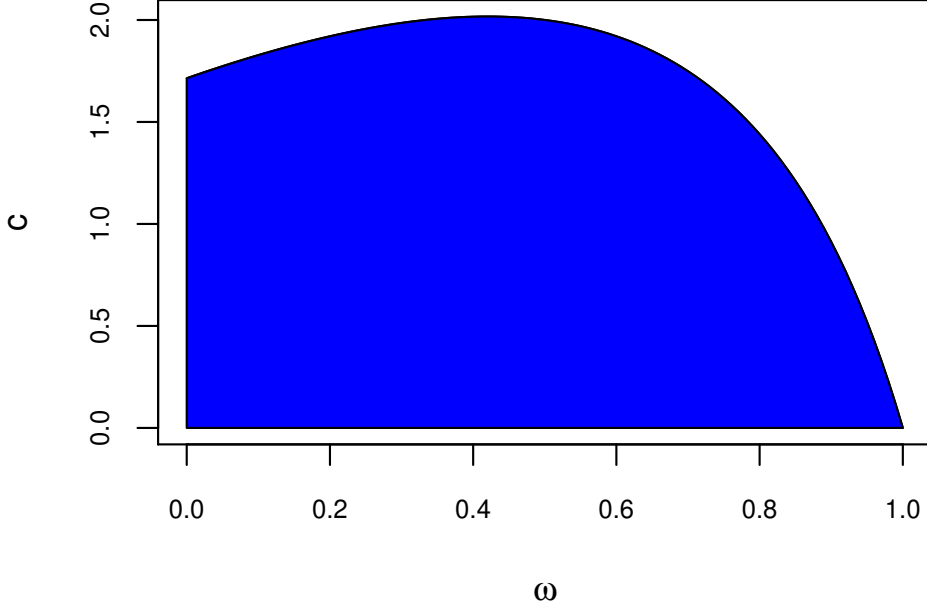


Figure 1: Display of constraint \mathbf{A}_3 in the plane (ω, c) .

fig:assumptionA3

with $s_\ell(N, \alpha) = |p_\ell - g_\ell| \sqrt{\frac{c}{N}} q_{1-\frac{\alpha}{2}}$.

An asymptotic confidence region at level $1 - \alpha$ for the vector $\theta = (p + g) / 2$ is :

$$\begin{aligned} \Lambda_{1-\alpha}(\theta) &= \left\{ t \in \mathbb{R}^d : N \left\| \Gamma^{-1/2} (t - \bar{x}_N) \right\|^2 \leq \chi_{1-\alpha}^{(d)} \right\} \\ &= \left\{ t = (t_1, \dots, t_n) : \sum_{\ell=1}^d \left(\frac{t_\ell - \bar{x}_{N,\ell}}{p_\ell - g_\ell} \right)^2 \leq \frac{c}{N} \chi_{1-\alpha}^{(d)} \right\}. \end{aligned}$$

We note however that the vector θ may not be of crucial interest for the initial optimization problem conversely to g . This point will be discussed in the last section.

2.2 Second case: non-oscillatory and stagnant ($p = g$)

In this section we study again a single particle and suppose once and for all that $x_n \in \mathbb{R}$. We assume throughout this subsection that the particle is under stagnation that is $p_n = p$ for n sufficiently large (see assumption \mathbf{B}_2 below). This assumption is strong but a more general framework leads to theoretical developments out of our scope. Starting from Equation (1), the PSO equation becomes this time:

$$x_{n+1} = (1 + \omega)x_n - \omega x_{n-1} + c(r_{1,n} + r_{2,n})(p - x_n).$$

Change the centering and consider $x_n - p = y_n$. The previous equation becomes :

$$y_{n+1} = (1 + \omega - c + c\varepsilon_n)y_n - \omega y_{n-1}, \tag{2}$$

where ε_n is the sum of two independent random variables with $\mathcal{U}[-1/2; 1/2]$ distribution.

Assuming that for all n $y_n \neq 0$, we have then :

$$\frac{y_{n+1}}{y_n} = (1 + \omega - c + c\varepsilon_n) - \omega \frac{y_{n-1}}{y_n} \tag{3}$$

It is plain that y_{n+1}/y_n defines a Markov chain (more precisely : a non-linear auto-regressive process) which will play a crucial role in the forthcoming results. It is shown in the proof section that y_{n+1}/y_n is Harris recurrent and has

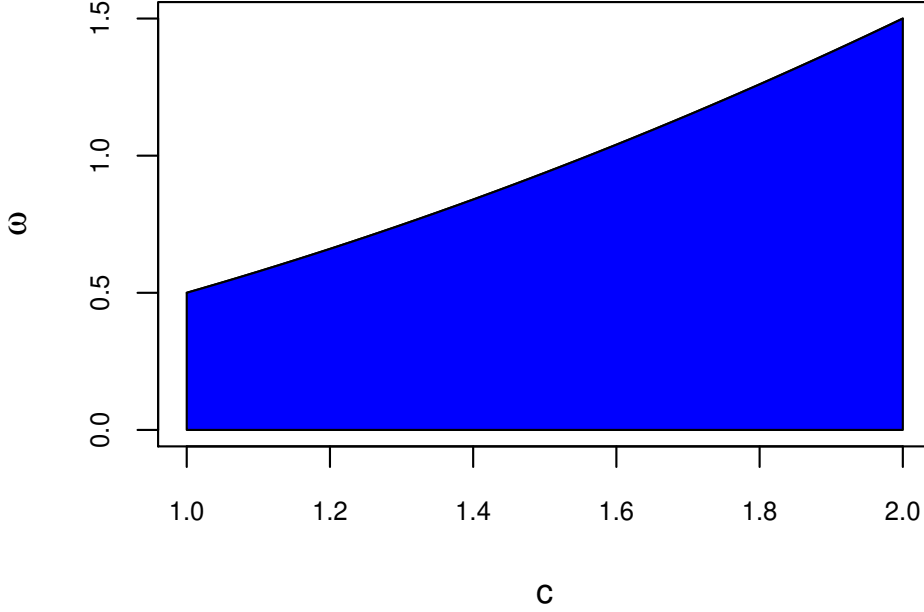


Figure 2: Display of constraint \mathbf{B}_1 in the plane (c, ω) .

fig:assumptionB2

consequently a stationary distribution denoted π . The definition of Harris recurrence needed is for instance Meyn and Tweedie (2012), beginning of Chapter 9. Take Z_n a copy of y_n/y_{n-1} with Z_0 a realization of π . Then define :

$$\begin{aligned}\mu_x &= \mathbb{E}_\pi \log |Z_0|, \\ \sigma_x^2 &= \text{Var}_\pi(\log |Z_0|) + 2 \sum_{k=1}^{+\infty} \text{Cov}_\pi(\log |Z_0|, \log |Z_k|).\end{aligned}$$

We are ready to introduce a new set of assumptions.

B₁ : $1 + \omega - c < \omega/c < (1 + c)/4$.

B₂ : For sufficiently large n $g_n = p_n = p = g$ is constant.

Before stating next Theorem notice that $\sum_{n=1}^N \log \left| \frac{y_n}{y_{n-1}} \right| = \log |y_N| - \log |y_0|$.

Theorem 3. Let $\omega \in (0, 1)$, $c > 1$, when **B₁₋₂** hold, then :

$$\frac{1}{\sqrt{N}} (\log |x_N - g_N| - N\mu_x) \hookrightarrow \mathcal{N}_1(0, \sigma_x^2)$$

when N tends to infinity. ^{theo:equal}

Remark 1. The theorem above is not a Central Limit Theorem for x_N . It is derived thanks to a CLT but it shows that the asymptotic distribution of $|x_N - g_N|$ is asymptotically log-normal with approximate parameters $N\mu_x$ and $N\sigma_x^2$.

Remark 2. The mean and variance μ_x and σ_x^2 are usually unknown but may be approximated numerically. We refer to the simulation section for more details.

Corollary 4. If $p_n = p$ an asymptotic confidence (non convex) region for g at level $1 - \alpha$ denoted $\Lambda_{1-\alpha}$ below may be derived from the preceding Theorem:

$$\begin{aligned}\Lambda_{1-\alpha}(g) &= \Lambda_{1-\alpha}^+ \cup \Lambda_{1-\alpha}^- \\ \Lambda_{1-\alpha}^+ &= \left[x_n + \exp\left(\mu_x + \frac{\sigma_x}{\sqrt{n}} q_{\alpha/2}\right), x_n + \exp\left(\mu_x - \frac{\sigma_x}{\sqrt{n}} q_{1-\alpha/2}\right) \right] \\ \Lambda_{1-\alpha}^- &= \left[x_n - \exp\left(\mu_x + \frac{\sigma_x}{\sqrt{n}} q_{1-\alpha/2}\right), x_n - \exp\left(\mu_x - \frac{\sigma_x}{\sqrt{n}} q_{\alpha/2}\right) \right]\end{aligned}$$

Remark 3. Under matrix form the equation (2) is purely linear but driven by a random matrix:

$$\begin{pmatrix} y_{n+1} \\ y_n \end{pmatrix} = \mathbf{y}_{n+1} = \begin{bmatrix} 1 + \omega - c(1 + \varepsilon_n) & -\omega \\ 1 & 0 \end{bmatrix} \begin{pmatrix} y_n \\ y_{n-1} \end{pmatrix} \quad \text{non-osc-PSO} \quad (4)$$

$$\mathbf{y}_{n+1} = \mathbf{S}_{n+1} \mathbf{y}_n, \quad \mathbf{y}_n = \mathbf{S}_n \mathbf{S}_{n-1} \dots \mathbf{S}_2 \mathbf{y}_1 = \mathbf{T}_n \mathbf{y}_1,$$

with $\mathbf{T}_n = \mathbf{S}_n \mathbf{S}_{n-1} \dots \mathbf{S}_2$. It is plain here that a classical Central Limit Theorem cannot hold for the sequence $(y_n)_{n \in \mathbb{N}}$. In the proofs we turn to asymptotic theory for the product of random matrices. We refer to the historical references: Furstenberg and Kesten (1960) and Berger (1984) who proved Central Limit Theorems for the regularity index of the product of i.i.d random matrices. Later Hennion (1997) generalized their results. But the assumptions of (almost surely) positive entries is common to all these papers. Other authors obtain similar results under different sets of assumptions (see Le Page (1982), Benoist and Quint (2016), and references therein), typically revolving around characterization of the semi-group spanned by the distribution of \mathbf{S}_n . These assumptions are uneasy to check here and we carried out to a direct approach with Markov chain fundamental tools.

2.3 The swarm at a fixed step

In this section we change our viewpoint. Instead of considering a single particle and sampling along its trajectory we will take advantage of the whole swarm but at a fixed and common iteration step. Our aim here is to localize the minimum of the cost function based on (x_n^1, \dots, x_n^S) . This time the particle index s varies up to S the swarm size, whereas the index n is fixed. In this subsection we assume that $S \uparrow +\infty$ and asymptotic is with respect to S . We do not drop n in order to see how the iteration steps influence the results. We still address only the case $x_n^s \in \mathbb{R}$ even if our results may be straightforwardly generalized to $x_n^s \in \mathbb{R}^d$. We provide below a Central Limit Theorem suited to the case when the number of particles in the swarm becomes large. In order to clarify the method, we assume that for all particles x_n^i in the swarm $p_n^i = g_n = p$. In other words, no local minimum stands in the domain Ω , which implies additional smoothness or convexity assumptions on the cost function f . This may be possible by a preliminary screening of the search space. Indeed a first (or several) run(s) of preliminary PSO(s) on the whole domain identifies an area where a single optimum lies. Then a new PSO is launched with initial values close to this optimum and with parameters ensuring that most of the particles will stay in the identified area.

So we are given (x_n^1, \dots, x_n^S) where S is the sample size. Basically, the framework is the same as in the non oscillatory case studied above for a single particle. From (4) we get with $\mathbf{y}_n^i = x_n^i - p$:

$$\mathbf{y}_n^i = \mathbf{T}_n^i \mathbf{y}_1^i,$$

$$\mathbf{T}_n^i = \prod_{j=2}^n \mathbf{S}_j, \quad \mathbf{S}_j = \begin{bmatrix} 1 + \omega - c(1 + \varepsilon_j^i) & -\omega \\ 1 & 0 \end{bmatrix}.$$

Assume that the domain Ω contains 0 and that for all s $(x_0^i, x_1^i)_{i \leq S}$ are independent, identically distributed and centered then from the decomposition above, for all n and s , $\mathbb{E} \mathbf{y}_n^i = 0$ and the $(\mathbf{y}_n^i)_{1 \leq i \leq S}$ are i.i.d too.

The assumptions we need to derive Theorem 5 below are:

C₁: The operational domain Ω contains 0 (and is ideally a symmetric set).

C₂: The couples $(x_0^i, x_1^i)_{i \leq S}$ are i.i.d. and centered.

C₃: For all i in $\{1, \dots, S\}$ $p_n^i = g_n = p$.

When S is large the following Theorem may be of interest and is a simple consequence of the i.i.d. CLT.

Theorem 5. Under assumptions **C₁₋₃** a Central Limit Theorem holds when S the number of particles in the swarm tends to $+\infty$:

$$\frac{1}{\sqrt{S}} \sum_{i=1}^S (x_n^i - g_n) \xrightarrow{S \rightarrow +\infty} \mathcal{N}(0, \sigma_n^2),$$

where $\sigma_n^2 = \mathbb{E} (x_n^1 - g_n)^2$ is estimated consistently by:

$$\hat{\sigma}_n^2 = \frac{1}{S} \sum_{i=1}^S (x_n^i - g_n)^2.$$

prop3

Remark 4. The convergence of $\hat{\sigma}_n^2$ to σ_n^2 is a straightforward consequence of the weak and strong laws of large numbers.

Denote $\bar{x}^S = (1/S) \sum_{i=1}^S x_n^i$. The Theorem above paves the way towards an asymptotic confidence interval.

Corollary 6. An asymptotic confidence interval at level $1 - \alpha$ for g is:

$$\Lambda_n(g) = \left[\bar{x}^S - \frac{\hat{\sigma}_n}{\sqrt{S}} q_{1-\alpha/2}, \bar{x}^S + \frac{\hat{\sigma}_n}{\sqrt{S}} q_{1-\alpha/2} \right].$$

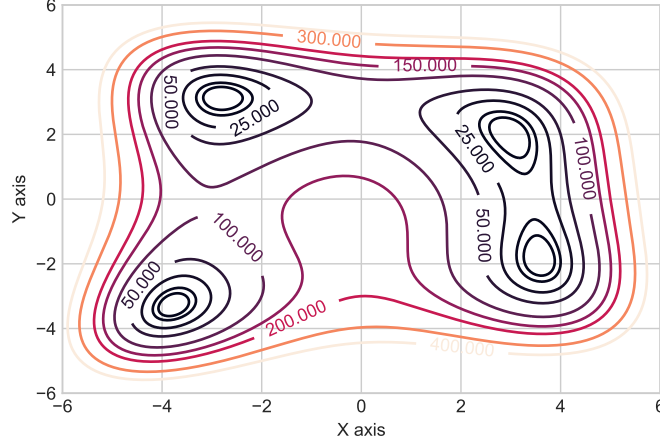


Figure 3: Contour of the Himmelblau's function in the space $[-6, 6]^2$. `fig:himmelblau_contour`

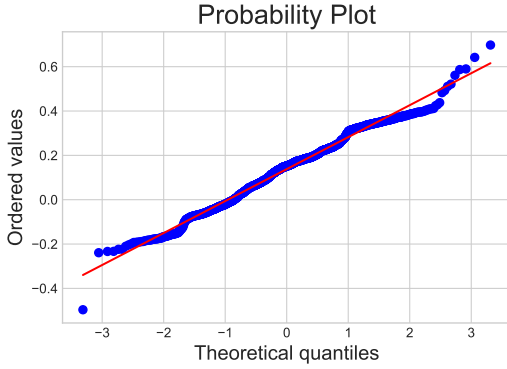


Figure 4: Normal probability plot of $H_1^s(N)$ on the first coordinate. `fig:droite_henry_x`

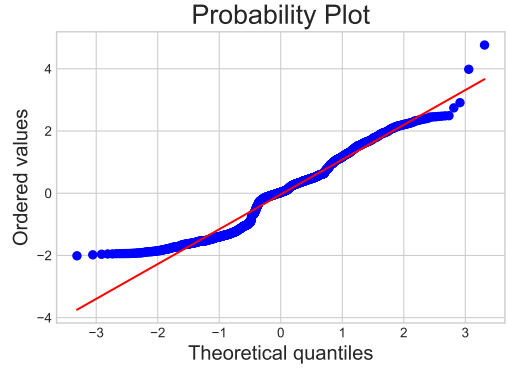


Figure 5: Normal probability plot of $H_1^s(N)$ on the second coordinate. `fig:droite_henry_y`

3 Simulation and numerical results

All along this section we take $c_1 = c_2 = c$.

The Himmelblau's function is chosen as example for our experiments. It is a 2 dimensional function with four local optima in $[-10, 10]^2$ defined by : $f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$. Figure 3 illustrates the contour of this function.

With the Himmelblau's function, we can observe the two different behaviors of the particles: oscillatory and non-oscillatory. The Himmelblau's function is positive and has four global minima in: $(3, 2)$, $(-2.81, 3.13)$, $(-3.77, -3.28)$, $(3.58, -1.84)$ where $f(x, y) = 0$. We use a ring topology (for a quick review of the different topologies of PSO see Lane et al. (2008)) for the algorithm in order to have both oscillating and non-oscillating particles. The latter converge quickly. The former go on running between two groups of particles converging to two distinct local optima.

3.1 Oscillatory case

We select particles oscillating between $(3.58, -1.84)$ and $(3, 2)$, these values could be both their personal best position or their neighborhood best position. In this case, the convergence of the p_n and g_n to $(3.58, -1.84)$ or $(3, 2)$ satisfies the conditions of Corollary ?? . We have to verify that the Gaussian asymptotic behavior of $H_1^s(N) = \sqrt{N} \left(\frac{1}{N} \sum_{n=1}^N x_n^s - \frac{p+s}{2} \right)$ for each s oscillating particle.

We launch PSO with a population of 200 particles and with 2000 iterations, $\omega = 0.72984$ and $c = 1.496172$. A ring topology was used to ensure the presence of oscillating particles. A particle is said oscillating if between the 500th and the 2000th iteration, Assumptions \mathbf{A}_{1-3} holds.

A visual tool to verify the normality of $H_1^s(N)$ for a particle is a normal probability plot. Figures 4 and 5 displays the normal probability plot of $H_1^s(N)$ respectively for the x axis and y axis. For each axis, the normality is confirmed: $H_1^s(N)$ fits well the theoretical quantiles.

To check the formula of the covariance matrix Γ , the confidence ellipsoid is also a good indicator (see Figure 6). For

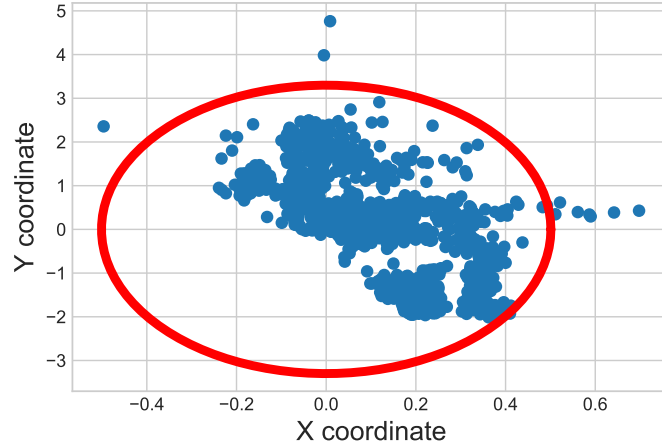


Figure 6: Trajectory of $H_1^s(N)$ for an oscillating particle in $[-10, 10]^2$. The confidence ellipsoid at a level of 85% is displayed in red. Around 99% of the trajectory of $H_1^s(N)$ is inside the ellipse. fig:ellipsoid

a single particle, $H_1^s(N)$ is not necessarily always inside the confidence ellipsoid and does not respect the percentage of the defined confidence level. Figure 7 shows the trajectory of x_n^s and $H_1^s(N)$ on the y axis, $H_1^s(N)$ remains bounded.

With 200 Monte-Carlo simulations of PSO (200 particles, and 2000 iterations), we select all the particles oscillating between $(3.58, -1.84)$ and $(3, 2)$, and for each of them we compute $H_1^s(2000)$. Figure 8 displayed the density of $H_1^s(2000)$ using 1150 oscillating particles. Almost 95% of the particles are inside the confidence ellipsoid of level 95% (represented in red).

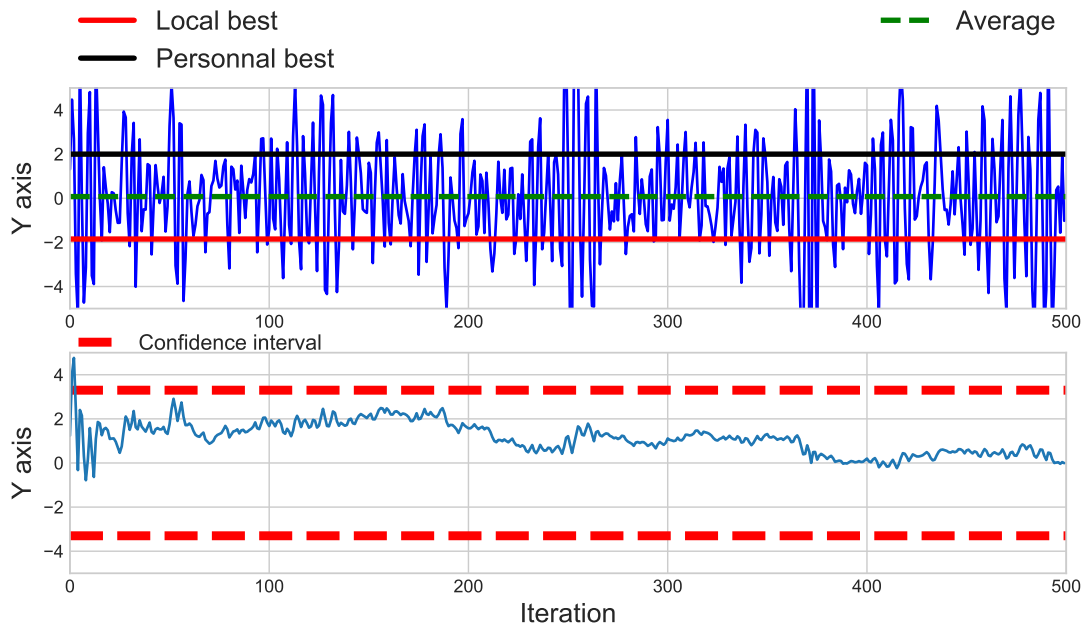


Figure 7: Top track: Trajectory of a oscillating particle on the y axis. The particle is oscillating between 2 and -1.84. Bottom track: corresponding trajectory of $H_1^s(N)$ on the y axis, the red dot are corresponding to the 95% confidence interval. The trajectory of $H_1^s(N)$ is well bounded. fig:trajectory_particle

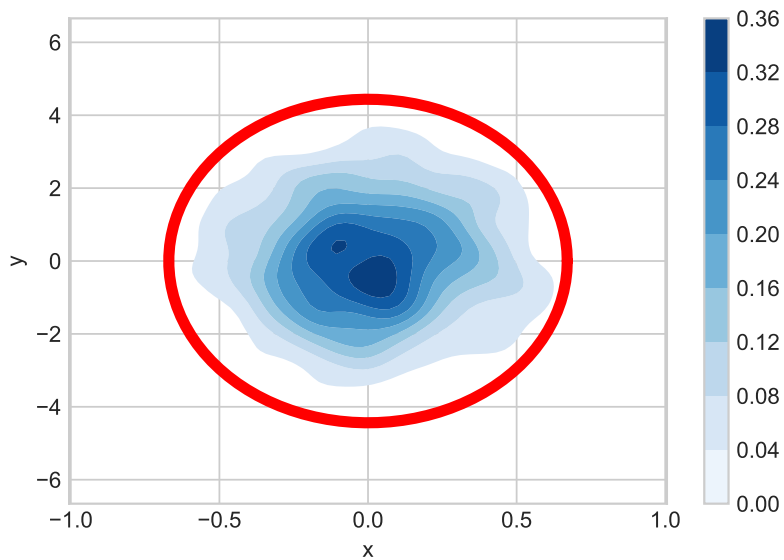


Figure 8: Density of $H_1(2000)$ with 1150 particles issued from Monte-Carlo simulations. The red ellipsoid is the 95% confidence ellipsoid. fig:density_mc_oscillating

3.2 Non-oscillatory case

We study now the behaviors of non-oscillating particles on the Himmelblau's function. We launch PSO with a population of 1000 particles and with 2000 iterations, $\omega = 0.72984$ and $c = 1.496172$. A ring topology was used to ensure the presence of enough particle converging to each local optimum. We select particles converging to $(3, 2)$, meaning that $p_n = g_n = p$ for a sufficiently large n . For the weak convergence of the particle, we consider:

$$H_2^s(N) = \frac{1}{\sqrt{N}} (\log |x_N^s - g_N| - N\mu_x).$$

First, it is easy to check the linear dependency of $\log |x_N^s - g_N|$ with a single display of the trajectory. Figure 9 illustrates this phenomenon for a single particle. We observed numerical issues when we reach the machine precision, but a numerical approximation of μ_x can be performed thanks to a linear regression.

Using many converging particles, a Monte Carlo approximation of μ_x is done. For the approximation of σ_x , a possibility is:

$$\bar{\sigma}_x^2 = \text{Var}_\pi(\log |X_0|) + 2 \sum_{k=1}^T \text{Cov}_\pi(\log |X_0|, \log |X_k|),$$

where $T = 20$. With near 240 converging particles to $(3, 2)$, we found that for the first coordinate:

$$\begin{aligned} \bar{\mu}_x &= -0.032, \\ \bar{\sigma}_x &= 0.156. \end{aligned}$$

We verify the asymptotic normality of $H_2(N)$ with a normal probability plot using the approximation of μ_x . Figure 10 displays the normal probability plot of $H_2(N)$ on the first coordinate, the theoretical quantiles are well fitted by $H_2(N)$. Figure 11 illustrates different trajectories of $H_2(N)$ on the first coordinate which are bounded by the 95% confidence interval deduced from $\bar{\sigma}_x$.

3.3 Swarm at a fixed step

To check the Theorem 5, we study:

$$H_3^n(S) = \frac{1}{\sqrt{S\hat{\sigma}_n}} \sum_{i=1}^S (x_n^i - g_n).$$

In practice, we encountered some difficulties to verify Theorem 5 because of the convergence rate of the particles. Indeed, when $p_n^s = g_n^s = p^s$, the particle s converges exponentially to g_n^s but the spread of the rate of convergence is large. As a consequence, at a fixed step of PSO, some particles could be considered as outliers because of a lower rate of convergence. Because of these particles qualified as belated, the asymptotic Gaussian behavior of $H_3^n(S)$ is not verified. A solution is to filter the particles and remove the belated particles. Figures 12 and 13 illustrate this phenomenon for the Himmelblau's function in 2D and with near 1500 converging particles to $(3, 2)$ over 500 iterations. In Figure 12, we compute without any filtering $H_3^n(S)$ and we notice that the Gaussian behavior is not verified and some jumps appeared. The presence of these "jumps" is due to belated particles which have a lower rate of convergence in comparison to the swarm. When we remove these particles with a classical outliers detection algorithm in Figure 13, Theorem 5 seems to be verified.

4 Conclusion

Our main theoretical contribution revolves around the three CLTs and the confidence regions derived from sampling either a particle's path or the whole swarm. Practically the confidence set $\Lambda_{1-\alpha}$ localizes, say g , with a high probability. High probability means larger than $1 - \alpha$. The simulations carried out in Python language tend to foster the theory.

This work was initiated in order to solve a practical issue in connection with oil industry and petrophysics. Yet in the previous section we confined ourselves to simulated data for several reasons. It appears that our method should be applied on real-data for a clear validation.

A second limitation of our work is the asymptotic set-up. The results are stated for samples large enough. This may not have an obvious meaning for people not familiar with statistics or probability. Practitioners usually claim that the Central Limit Theorem machinery works well for a sample size larger than thirty/forty whenever data are stable, say stationary (path with no jumps...). As a consequence the first iterations of the algorithm should be avoided to ensure a warming. When studying PSO the behavior of p_n and g_n turns out to be crucial too in order to ensure this stability, hence the validity of the theoretical results. However the control of p_n and g_n is a difficult matter, which explains our current assumptions on stagnation or "almost stagnation". However, in order to fix the question of the asymptotic versus non-asymptotic approach a work is on progress. We expect finite-sample concentration inequalities : the non-asymptotic counterparts of CLT.

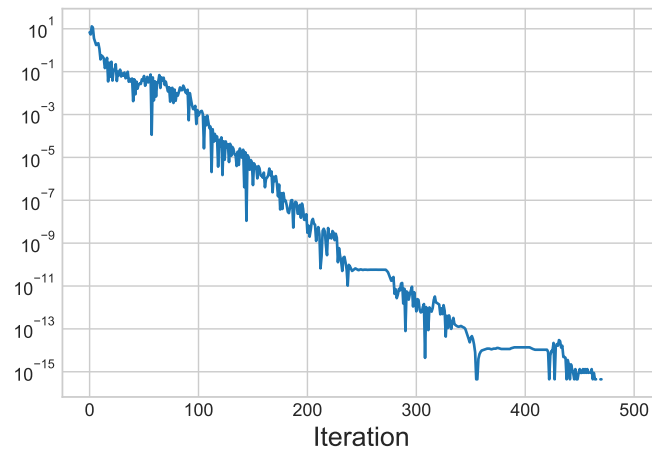


Figure 9: $|x_n - g_n|$ over 500 iterations in a logarithmic scale. After 300 iterations, we reach the computer precision. We take advantage of the linear behavior of $\log(|x_n - g_n|)$ on the 200 first iterations to perform a linear regression estimating μ_x . fig:himmelblau_converging_linear

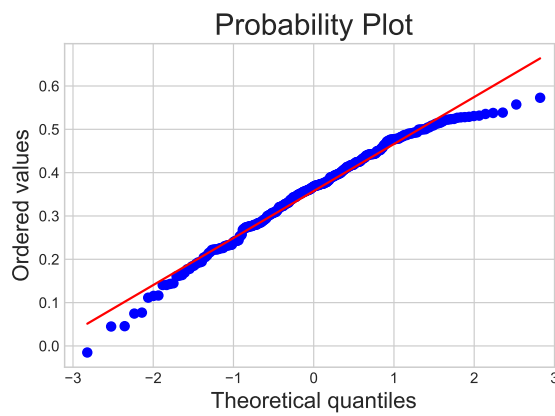


Figure 10: Normal probability plot of $H_2(N)$ on the first coordinate. fig:henry_x_conv

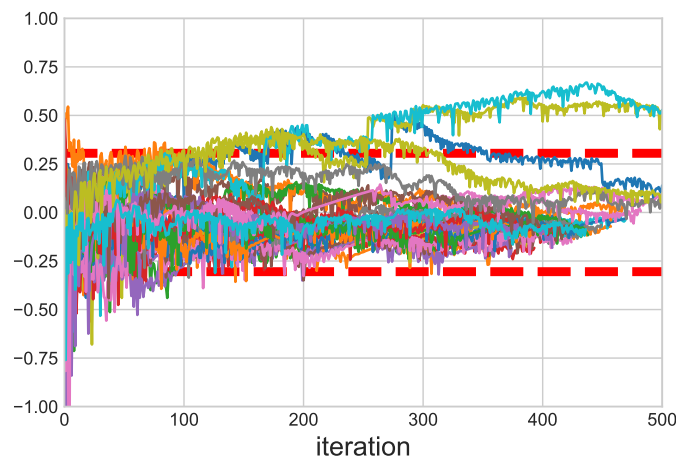


Figure 11: Trajectories of $H_2(N)$ on the first coordinate for five particles. The red dot represents the 95% confidence interval deduced from $\bar{\sigma}_x$. Trajectories stop around the 400 iterations (after an heating phase) due to numerical precision. fig:mu11519_02

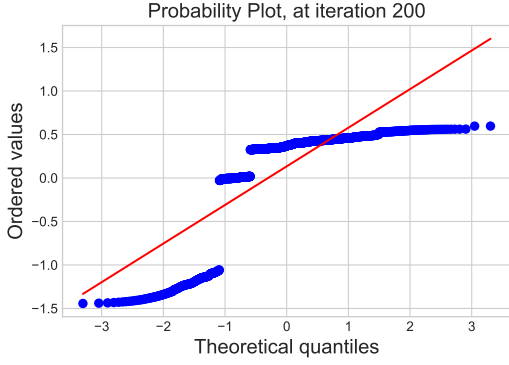


Figure 12: Normal probability plot of $H_3^{n=200}(S)$ at the 200th iteration on the first coordinate, using all the particles. We observe a discontinuity of the probability plot due to belated particles. `fig:H3_all`

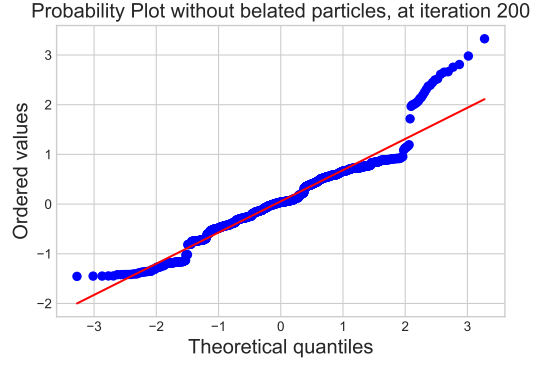


Figure 13: Normal probability plot of $H_3^{n=200}(N)$ at the 200th iteration on the first coordinate, computed without outlying or belated particles. `fig:H3_sub`

In the oscillating framework (see Corollary ??) our results involve the parameter denoted θ , the center of the segment $[p, g]$. Here we miss the target g for instance. We claim that our method may be adapted to tackle this issue but we will not elaborate on it. Briefly speaking : running two independent PSOs on the same domain with two distinct pairs (c_1, c_2) and (c'_1, c'_2) will provide, under additional assumptions on the local and global minima of the cost function, two CLTs involving :

$$\theta = \frac{c_1 g + c_2 p}{c_1 + c_2} \quad \text{and} \quad \theta' = \frac{c'_1 g + c'_2 p}{c'_1 + c'_2}.$$

A simple matrix inversion will then recover a joint CLT for the couple (g, p) .

5 Derivations of the results

We start with some notations. First we recall that the sup-norm for square matrices of size d defined by:

$$\|M\|_\infty = \sup_{x \neq 0} \frac{\|Mx\|}{\|x\|}.$$

The tensor product notation is appropriate when dealing with special kind of matrices for instance covariance matrices. Let u and v be two vectors of \mathbb{R}^d then $u \otimes v = uv^t$ (where v^t is the transpose of vector v) stands for the rank one matrix defined for all vector x by $(u \otimes v)(x) = \langle v, x \rangle u$. Besides $\|u \otimes v\|_\infty = \|u\| \|v\|$. The Hadamard product between vectors was mentioned earlier. Its matrix version may be defined a similar way. Let M and S be two matrices with same size then $M \odot S$ is the matrix whose (i, j) cell is $(M \odot S)_{i,j} = m_{i,j} s_{i,j}$. We recall without proof the following computation rule mixing Hadamard and tensor product. Let η, ε, u and v be four vectors in \mathbb{R}^d . Then:

$$(\eta \odot u) \otimes (\varepsilon \odot v) = (\eta \otimes \varepsilon) \odot (u \otimes v), \quad \text{Hadam (5)}$$

and the reader must be aware that the Hadamard product on the left-hand side operates between vectors whereas on the right-hand side it operates on matrices.

We will need \mathcal{F}_n^s the filtration generated by the path of particle number s up to step n : $\{x_0^s, \dots, x_n^s\}$ and $\mathcal{F}_n^{\mathcal{S}}$ the filtration generated by the swarm up to step n : $\{x_0^s, \dots, x_n^s : s = 1, \dots, S\}$.

We will also denote later $g_n = \mathbb{E}(g_n) + \xi_n$ and $p_n = \mathbb{E}(p_n) + v_n$ the expectation-variance decomposition of g_n and p_n where ξ_n and v_n are centered random vectors and support all the variability of g_n and p_n respectively.

5.1 First case: oscillatory

In this subsection we prove Theorem 1. We start with two Lemmas and a Proposition who will be invoked later.

Lemma 7. Let $\varepsilon_n^{(i)}$ and $\eta_n^{(j)}$ be any coordinate of the random vectors ε_n and η_n appearing in ?. Clearly, $\varepsilon_n^{(i)}$ and $\eta_n^{(j)}$ are not independent but not correlated and follow the same type of distribution. Besides:

$$\begin{aligned} \mathbb{E}\varepsilon^{(i)2} &= \mathbb{E}\eta^{(j)2} = 1/6, & \mathbb{E}\varepsilon^{(i)}\eta^{(j)} &= 0, \\ \mathbb{E}\varepsilon^{(i)3}\eta^{(j)} &= \mathbb{E}\varepsilon^{(i)}\eta^{(j)3} = 0. \end{aligned}$$

The proof is very simple hence omitted.

The next Lemma is also straightforward but will be frequently invoked later.

LLN

Lemma 8. Let E_n be a sequence of i.i.d centered random matrices with finite moment of order 4, let u_n and v_n two sequence of random vectors almost surely bounded and such that (u_n, v_n) is for all n independent from E_n then for the $\|\cdot\|_\infty$ norm:

$$\frac{1}{N} \sum_{n=1}^N E_n \odot (u_n \otimes v_n) \xrightarrow{\mathbb{P}} 0.$$

The proof is a simple application of Cauchy-Schwartz inequality.

Proof of Theorem 1:

First step : We show that the problem of convergence in distribution for $\frac{1}{N} \sum_{n=1}^N (x_n - \theta)$ may be shifted to proving a CLT for a martingale difference array.

Starting from the initial PSO equation :

$$x_{n+1} - x_n = \omega (x_n - x_{n-1}) + c_1 r_{1,n} \odot (p_n - x_n) + c_2 r_{2,n} \odot (g_n - x_n)$$

it is simple to derive :

$$x_{n+1} - x_n = \omega (x_n - x_{n-1}) + c_1 r_{1,n} \odot (p - x_n) + c_2 r_{2,n} \odot (g - x_n) + R_n$$

with

$$R_n = c_1 r_{1,n} \odot (p_n - p) + c_2 r_{2,n} \odot (g_n - g).$$

Denote $z_n = (x_n - \theta)$ then after some additional calculations we get :

$$z_{n+1} - (1 + \omega - c) z_n + \omega z_{n-1} = c_1 \varepsilon_{1,n} \odot \left(\frac{c_2}{c_1 + c_2} (p - g) - z_n \right) - c_2 \varepsilon_{2,n} \odot \left(\frac{c_1}{c_1 + c_2} (p - g) + z_n \right) + R_n \quad \text{decomp-base (6)}$$

with, for $i = 1, 2$: $r_{i,n} = \mathbb{E} r_{i,n} + \varepsilon_{i,n} = 1/2 + \varepsilon_{i,n}$ and $\varepsilon_{i,n}$ is the centered version of $r_{i,n}$.

Summing the above equation from $n = 1$ to $n = N - 1$:

$$\sum_{n=1}^N z_n = \frac{1}{c} \sum_{n=1}^N u_n + \mathcal{R}_N \quad \text{decomp-zn-un (7)}$$

with :

$$u_n = c_1 \varepsilon_{1,n} \odot \left(\frac{c_2}{c_1 + c_2} (p - g) - z_n \right) - c_2 \varepsilon_{2,n} \odot \left(\frac{c_1}{c_1 + c_2} (p - g) + z_n \right),$$

$$\mathcal{R}_N = \frac{1}{c} \left[\sum_{n=1}^N R_n + z_1 + (c - 1) z_N + \omega (z_{N-1} - z_0) \right]$$

Notice that u_n just above is a martingale difference with respect to the nested filtration $\mathcal{F}_n : \mathbb{E}(u_n | \mathcal{F}_{n-1}) = 0$ because $\varepsilon_{1,n}$ and $\varepsilon_{2,n}$ are centered and independent from z_n .

In view of conditions \mathbf{A}_1 and \mathbf{A}_2 :

$$\frac{\mathcal{R}_N}{\sqrt{N}} \xrightarrow{\mathbb{P}} 0,$$

then from (7) $\frac{1}{\sqrt{N}} \sum_{n=1}^N z_n$ converges in distribution if and only if $\frac{1}{\sqrt{N}} \sum_{n=1}^N u_n$ does.

Second step : it is shown below that weak convergence for $\frac{1}{\sqrt{N}} \sum_{n=1}^N u_n$ holds under the assumptions of the Theorem.

We aim at proving a Levy-Lindeberg version of the CLT for the series of u_n in two steps (Theorem 2.1.9 p. 46 and its corollary 2.1.10 in Duflo (1997)): first checking the Lyapunov condition holds (hence the Lindeberg's uniform integrability that ensures uniform tightness of the sequence) then ensuring convergence in probability of the conditional covariance structure of $(1/\sqrt{N}) \sum_{n=1}^N u_n$.

Here the Lyapunov condition holds trivially because we are faced with bounded martingale difference sequences. Indeed by assumption \mathbf{A}_1 and since the random variables $\varepsilon_{1,n}$ and $\varepsilon_{2,n}$ are almost surely bounded u_n is itself almost surely bounded.

We turn to the conditional covariance sequence of u_n :

$$\Gamma_N = \frac{1}{N} \sum_{n=1}^N \mathbb{E}(u_n \otimes u_n | \mathcal{F}_{n-1}).$$

Starting from the definition of u_n just below (7) leads after simple calculations to :

$$\begin{aligned}\mathbb{E}(u_n \otimes u_n | \mathcal{F}_{n-1}) &= \frac{c_1^2 c_2^2}{(c_1 + c_2)^2} \frac{1}{6} \mathbf{I} \odot [(p-g) \otimes (p-g)] + \frac{(c_1^2 + c_2^2)}{12} \mathbf{I} \odot (z_n \otimes z_n) \\ &\quad - 2 \frac{c_1 - c_2}{12} \frac{c_1 c_2}{c_1 + c_2} \mathbf{I} \odot [(p-g) \otimes z_n]\end{aligned}$$

Notice that, due to the operation $\mathbf{I} \odot (\cdot)$ $\mathbb{E}(u_n \otimes u_n | \mathcal{F}_{n-1})$ is a diagonal matrix with :

$$\begin{aligned}\mathbb{E}(u_n \otimes u_n | \mathcal{F}_{n-1}) &= \text{diag}(\tau_{i,n})_{1 \leq i \leq d} \\ \tau_{i,n} &= \frac{c_1^2 c_2^2}{6(c_1 + c_2)^2} (p_i - g_i)^2 + \frac{(c_1^2 + c_2^2)}{12} z_{i,n}^2 - \frac{c_1 - c_2}{6} \frac{c_1 c_2}{c_1 + c_2} [(p_i - g_i) z_{i,n}]\end{aligned}$$

Averaging the previous results in n we get :

$$\begin{aligned}\Gamma_N &= \frac{c_1^2 c_2^2}{(c_1 + c_2)^2} \frac{1}{6} \mathbf{I} \odot [(p-g) \otimes (p-g)] + \frac{(c_1^2 + c_2^2)}{12} \mathbf{I} \odot \left(\frac{1}{N} \sum_{n=1}^N z_n \otimes z_n \right) \\ &\quad - \frac{c_1 - c_2}{6} \frac{c_1 c_2}{c_1 + c_2} \mathbf{I} \odot \left[(p-g) \otimes \left(\frac{1}{N} \sum_{n=1}^N z_n \right) \right]\end{aligned}$$

The last term is negligible : indeed Lemma 8 shows that $\frac{1}{N} \sum_{n=1}^N z_n \rightarrow_{\mathbb{P}} 0$. Our only task consists in studying convergence for $\mathbf{I} \odot \left(\frac{1}{N} \sum_{n=1}^N z_n \otimes z_n \right)$. More precisely we just need to prove convergence for $(1/N) \sum_{n=1}^N z_{i,n}^2$ for all $i \in \{1, \dots, d\}$.

Third (and last) step : Proof of convergence of $\frac{1}{N} \sum_{n=1}^N z_{i,n}^2$ for all i .

Let us restart from (6) :

$$z_{n+1} - (1 + \omega - c) z_n + \omega z_{n-1} = u_n + R_n$$

Clearly R_n is a negligible term due to assumption \mathbf{A}_2 . All the authors' computations were effectively carried out with R_n but in order to alleviate this proof we will proceed here as if :

$$z_{i,n+1} = (1 + \omega - c) z_{i,n} - \omega z_{i,n-1} + u_{i,n}$$

From now on we drop the dimension index i in the previous equation :

$$\begin{aligned}z_{i,n+1}^2 &= (1 + \omega - c)^2 z_{i,n}^2 + \omega^2 z_{i,n-1}^2 + u_{i,n}^2 \\ &\quad + 2[(1 + \omega - c) z_{i,n} - \omega z_{i,n-1}] u_{i,n} - 2\omega(1 + \omega - c) [z_{i,n} z_{i,n-1}]\end{aligned}$$

The following equation is derived from the preceding one after summing and considering the i^{th} component of each vector :

$$\begin{aligned}\left(1 - (1 + \omega - c)^2 - \omega^2\right) \sum_{n=1}^N z_{i,n}^2 &= \sum_{n=1}^N u_{i,n}^2 - 2\omega(1 + \omega - c) \sum_{n=1}^N z_{i,n} z_{i,n-1} \\ &\quad + 2 \sum_{n=1}^N [(1 + \omega - c) z_{i,n} - \omega z_{i,n-1}] u_{i,n} + z_{i,1}^2 - z_{i,N+1}^2 + \omega^2 (z_{i,0}^2 - z_{i,N}^2)\end{aligned} \tag{8}$$

It is easily seen that Lemma 8 may be applied to the series with term $[(1 + \omega - c) z_{i,n} - \omega z_{i,n-1}] u_{i,n}$ and its average tends to zero.

From now on we remove all indices i for components of the vectors in order to alleviate notations.

Our next task consists in studying $\sum_{n=1}^N z_n z_{n-1}$. To that purpose we turn back again to equation (6) and multiply by z_n the whole equation and sum over n :

$$\sum_{n=1}^N z_{n+1} z_n - (1 + \omega - c) \sum_{n=1}^N z_n^2 + \omega \sum_{n=1}^N z_n z_{n-1} = \sum_{n=1}^N z_n u_n$$

hence

$$(1 + \omega) \sum_{n=1}^N z_n z_{n-1} = (1 + \omega - c) \sum_{n=1}^N z_n^2 + \sum_{n=1}^N z_n u_n + z_1 z_0 - z_{N+1} z_N$$

The same arguments as above show that $(1/N) \sum_{n=1}^N z_n u_n$ tends to 0 in probability by the weak law of large numbers. This proves that when $N \rightarrow +\infty$:

$$\frac{1}{N} \left(\sum_{n=1}^N z_n z_{n-1} - \frac{1 + \omega - c}{1 + \omega} \sum_{n=1}^N z_n^2 \right) \rightarrow_{\mathbb{P}} 0$$

Plugging this result in (8) and taking into account the remarks above :

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \left\{ \left[\left(1 - (1 + \omega - c)^2 - \omega^2 \right) + 2\omega \frac{(1 + \omega - c)^2}{1 + \omega} \right] \sum_{n=1}^N z_n^2 - \sum_{n=1}^N u_n^2 \right\} = 0 \quad \text{jdg (9)}$$

Our last task consists in finding $\lim_{N \rightarrow +\infty} (1/N) \sum_{n=1}^N u_n^2$. Calculations are closed to those carried out at the beginning of the second step but this time without conditioning with respect to \mathcal{F}_n . Remind that the index i was removed for the sake of clarity :

$$u_n^2 = \left[c_1 \varepsilon_{1,n} \left(\frac{c_2}{c_1 + c_2} (p - g) - z_n \right) - c_2 \varepsilon_{2,n} \left(\frac{c_1}{c_1 + c_2} (p - g) + z_n \right) \right]^2 \quad \text{jdg (10)}$$

Developping the square into bracket and taking into account that the cross product will tend to zero after averaging namely :

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N \varepsilon_{1,n} \varepsilon_{2,n} \left(\frac{c_2}{c_1 + c_2} (p - g) - z_n \right) \left(\frac{c_1}{c_1 + c_2} (p - g) + z_n \right) = 0$$

in probability, we can focus only on :

$$v_{1,N} = \frac{c_1^2}{N} \sum_{n=1}^N \varepsilon_{1,n}^2 \left(\frac{c_2}{c_1 + c_2} (p - g) - z_n \right)^2 \quad \text{and} \quad v_{2,N} = \frac{c_2^2}{N} \sum_{n=1}^N \varepsilon_{2,n}^2 \left(\frac{c_1}{c_1 + c_2} (p - g) + z_n \right)^2$$

Both $v_{1,N}$ and $v_{2,N}$ may be managed the same way. We develop only for $v_{1,N}$:

$$v_{1,N} = \frac{c_1^2}{N} \sum_{n=1}^N (\varepsilon_{1,n}^2 - \mathbb{E} \varepsilon_{1,n}^2) \left(\frac{c_2}{c_1 + c_2} (p - g) - z_n \right)^2 + \frac{c_1^2}{N} \sum_{n=1}^N \mathbb{E} \varepsilon_{1,n}^2 \left(\frac{c_2}{c_1 + c_2} (p - g) - z_n \right)^2$$

The first series above tends in probability to 0 again by the weak law of large numbers for martingale difference arrays already mentioned several times earlier in the proof. The second series depends only on the z_i 's. We finally get :

$$\lim_{N \rightarrow +\infty} v_{1,N} = \lim_{N \rightarrow +\infty} \frac{c_1^2}{12N} \sum_{n=1}^N \left(\frac{c_2}{c_1 + c_2} (p - g) - z_n \right)^2$$

Similarly :

$$\lim_{N \rightarrow +\infty} v_{2,N} = \lim_{N \rightarrow +\infty} \frac{c_2^2}{12N} \sum_{n=1}^N \left(\frac{c_1}{c_1 + c_2} (p - g) + z_n \right)^2$$

Putting the two last equations with (9) and (10) we conclude that :

$$\lim_{N \rightarrow +\infty} \left\{ \left(1 - (1 + \omega - c)^2 - \omega^2 \right) + 2\omega \frac{(1 + \omega - c)^2}{1 + \omega} - \frac{c_1^2 + c_2^2}{12} \right\} \frac{1}{N} \sum_{n=1}^N z_{i,n}^2 = \frac{1}{6} \left(\frac{c_1 c_2}{c_1 + c_2} \right)^2 (p - g)^2$$

However the previous equation is valid only if the big term between brackets in the right hand side is strictly positive. This term depends only on c_1 , c_2 and ω . This compatibility condition ensures just that a non-null limit for $\sum_{n=1}^N z_{i,n}^2$ exists. Besides :

$$1 - (1 + \omega - c)^2 - \omega^2 + 2\omega \frac{(1 + \omega - c)^2}{1 + \omega} - \frac{c_1^2 + c_2^2}{12} = 2c \frac{1 - \omega}{1 + \omega} \left(1 + \omega - \frac{c}{2} \right) - \frac{c_1^2 + c_2^2}{12}$$

and the positivity condition matches assumption \mathbf{A}_3 within the Theorem.

This yields :

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N z_{i,n}^2 = \frac{1}{6} \left(\frac{c_1 c_2}{c_1 + c_2} \right)^2 \left\{ 2c \frac{1 - \omega}{1 + \omega} \left(1 + \omega - \frac{c}{2} \right) - \frac{c_1^2 + c_2^2}{12} \right\}^{-1} (p - g)^2$$

hence from (??) and with simple algebra :

$$\Gamma_N \rightarrow_{\mathbb{P}} \mathfrak{K} \text{diag} (p - g)^{\odot 2}$$

with

$$\mathfrak{K} = \frac{1}{24} \left(\frac{c_1 c_2}{c} \right)^2 \left[2c \frac{1 - \omega}{1 + \omega} \left(1 + \omega - \frac{c}{2} \right) \left\{ 2c \frac{1 - \omega}{1 + \omega} \left(1 + \omega - \frac{c}{2} \right) - \frac{c_1^2 + c_2^2}{12} \right\} \right]^{-1}$$

All that was done until this point ensures that $\frac{1}{N} \sum_{n=1}^N u_n \xrightarrow{\mathbb{P}} \mathbf{Z}$ where \mathbf{Z} is a centered Gaussian random vector in \mathbb{R}^d with covariance operator $\mathfrak{K} \text{diag} (p - g)^{\odot 2}$.

At last (7) leads to the desired result : $\frac{1}{N} \sum_{n=1}^N z_n \xrightarrow{\mathbb{P}} \mathbf{Z}/c$.

5.2 Second case: non-oscillatory and stagnant

We start from (3) and set $X_n = y_n/y_{n-1}$. Then we get :

$$X_{n+1} = 1 + \omega - c - \frac{\omega}{X_n} + c\varepsilon_n,$$

where $\varepsilon_n = r_{1,n} + r_{2,n} - 1$ has a ‘‘witch hat’’ distribution (convolution of two uniform distributions) with support $[-1, +1]$. We focus now on the above homogeneous Markov chain X_n and we aim at proving that a CLT holds for $h(X_n) = \log |X_n|$ namely that for some μ and σ^2 :

$$\sqrt{N} \left[\frac{1}{N} \sum_{n=1}^N \log |X_n| - \mu \right] \hookrightarrow \mathcal{N}(0, \sigma^2),$$

which will yield:

$$\sqrt{N} \left[\frac{1}{N} \log |y_N| - \mu \right] \hookrightarrow \mathcal{N}(0, \sigma^2).$$

We aim at applying Theorem 1 p. 302 in Jones (2004) (see also Meyn and Tweedie (2012), section 17.5 for similar results). We need to check three points : (i) X_n is Harris ergodic, (ii) the existence of a small set \mathcal{C} and (iii) of a function g with a drift condition (see Meyn and Tweedie (2012)) such that (12) below holds. Following Hairer and Mattingly (2011), section 3.2 it turns out that proving (ii) and (iii) above with a drift function g but without the minoration bound (12) is sufficient to ensure (i).

Denote $P(t, x)$ the transition kernel of X_n . It is plain that $P(t, x)$ coincides with the density of the uniform distribution on the set :

$$\mathcal{E}_x = \left[1 + \omega - 2c - \frac{\omega}{x}, 1 + \omega - \frac{\omega}{x} \right].$$

Theorem 3 is a consequence of the two Lemmas below coupled with the above-mentioned Theorem 1 p.302 in Jones (2004).

Lemma 9. *Take $M_\tau = \omega/(c - \tau)$ with any $0 < \tau < c$ then the set $\mathcal{C} = (-\infty, -M_\tau] \cup [M_\tau, +\infty)$ is a smallset small set for the transition kernel of X_n .*

Proof :

We have to show that for all $x \in \mathcal{C}$ and Borel set A in \mathbb{R} :

$$P(A, x) \geq \varepsilon Q(A),$$

where $\varepsilon > 0$ and Q is a probability distribution. The main problem here comes from the compact support of $P(t, x)$. Take x such that $|x| \geq M$ then:

$$1 + \omega - c - \frac{\omega}{M} + c\varepsilon_n \leq 1 + \omega - c - \frac{\omega}{x} + c\varepsilon_n \leq 1 + \omega - c + \frac{\omega}{M} + c\varepsilon_n,$$

where ε_n has compact support $[-1, +1]$. It is simple to see that with $M = M_\tau = \omega/(c - \tau)$ the above bound becomes:

$$1 + \omega - 2c + \tau + c\varepsilon_n \leq 1 + \omega - c - \frac{\omega}{x} + c\varepsilon_n \leq 1 + \omega + \tau + c\varepsilon_n.$$

The intersection of the supports of $1 + \omega - 2c + \tau + c\varepsilon_n$ and $1 + \omega + \tau + c\varepsilon_n$ is the set $[1 + \omega - c - \tau, 1 + \omega - c + \tau]$ whatever the value of x in \mathcal{C} . The probability measure Q mentioned above may be chosen as the uniform distribution with support $[1 + \omega - c - \tau, 1 + \omega - c + \tau]$.

Now we turn to the drift condition. Our task consists in constructing a function $g : \mathbb{R} \rightarrow [1, +\infty[$ such that for all x :

$$\int_{\mathbb{R}} g(t) P(t, x) dt \leq \rho_1 g(x) + \rho_2 1_{x \in \mathcal{C}}, \quad \text{drift_condition} \quad (11)$$

where $0 < \rho_1 < 1$ and $\rho_2 \geq 0$. Besides, in order to get a CLT on $\log |X_n|$ we must further ensure that for all x :

$$\log^2 |x| \leq g(x). \quad \text{maj_log} \quad (12)$$

Note however that, if (11) holds for g but (12) fails, then both conditions will hold for updated function $g^* = \eta g$ with constant $\eta > 1$ and $\rho'_2 = \eta \rho_2$ such that (12) holds.

The next Lemma constructs the function g mentioned above.

Lemma 10. *Take for g the even function defined by $g(x) = C_1/\sqrt{|x|}$ for $|x| \leq M_\tau$ and $g(x) = C_2(\log |x|)^2$ for $|x| > M_\tau$. Assume that \mathbf{B}_1 holds.*

Then it is always possible to choose three constants τ, C_1 and C_2 such that (11) holds for a specific choice of ρ_1 and ρ_2 .

The proof of Lemma 10 is postponed to the Appendix.

References

- Benoist, Y., Quint, J.F., 2016. Central limit theorem for linear groups. *Ann. Probab.* 44, 1308–1340. doi:10.1214/15-AOP1002.
- Berger, M.A., 1984. Central limit theorem for products of random matrices. *Transactions of the American Mathematical Society* 285, 777–803.
- Van den Bergh, F., Engelbrecht, A.P., 2010. A convergence proof for the particle swarm optimiser. *Fundamenta Informaticae* 105, 341–374.
- Billingsley, P., 1999. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. second ed., John Wiley & Sons Inc. A Wiley-Interscience Publication.
- Bonyadi, M.R., Michalewicz, Z., 2016. Stability analysis of the particle swarm optimization without stagnation assumption. *IEEE Transactions on Evolutionary Computation* 20, 814–819.
- Cleghorn, C.W., Engelbrecht, A.P., 2018. Particle swarm stability: a theoretical extension using the non-stagnate distribution assumption. *Swarm Intelligence* 12, 1–22.
- 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.
- Dufo, M., 1997. *Random Iterative Models*. 1st ed., Springer-Verlag, Berlin, Heidelberg.
- Eberhart, R., Kennedy, J., 1995. A new optimizer using particle swarm theory, in: *Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on*, IEEE. pp. 39–43.
- Furstenberg, H., Kesten, H., 1960. Products of random matrices. *Ann. Math. Statist.* 31, 457–469. doi:10.1214/aoms/1177705909.
- Hairer, M., Mattingly, J.C., 2011. Yet another look at harris' ergodic theorem for markov chains, in: *Seminar on Stochastic Analysis, Random Fields and Applications VI*. Springer Basel, pp. 109–117. doi:10.1007/978-3-0348-0021-1_7.
- Hennion, H., 1997. Limit theorems for products of positive random matrices. *The Annals of Probability* , 1545–1587.
- Jones, G.L., 2004. On the markov chain central limit theorem. *Probability surveys* 1, 5–1.
- Kadirkamanathan, V., Selvarajah, K., Fleming, P.J., 2006. Stability analysis of the particle dynamics in particle swarm optimizer. *IEEE Transactions on Evolutionary Computation* 10, 245–255.
- Lane, J., Engelbrecht, A., Gain, J., 2008. Particle swarm optimization with spatially meaningful neighbours, in: *Swarm Intelligence Symposium, 2008. SIS 2008. IEEE, IEEE*. pp. 1–8.
- Le Page, É., 1982. Théoremes limites pour les produits de matrices aléatoires, in: *Probability measures on groups*. Springer, pp. 258–303.
- Meyn, S.P., Tweedie, R.L., 2012. *Markov chains and stochastic stability*. Springer Science and Business Media.
- Poli, R., 2009. Mean and variance of the sampling distribution of particle swarm optimizers during stagnation. *IEEE Transactions on Evolutionary Computation* 13, 712–721.
- Schmitt, M., Wanka, R., 2015. Particle swarm optimization almost surely finds local optima. *Theoretical Computer Science* 561, 57–72.
- Van Den Bergh, F., Engelbrecht, A.P., 2006. A study of particle swarm optimization particle trajectories. *Information sciences* 176, 937–971.
- Xu, G., Yu, G., 2018. On convergence analysis of particle swarm optimization algorithm. *Journal of Computational and Applied Mathematics* 333, 65–73. doi:10.1016/j.cam.2017.10.026.
- Zhang, Y., Wang, S., Ji, G., 2015. A comprehensive survey on particle swarm optimization algorithm and its applications. *Mathematical Problems in Engineering* 2015.

Appendices

This appendix provides mathematical derivations of Lemma 10.

A Proof of Lemma 10

1

The proof of the Lemma just consists in an explicit construction of the above-mentioned τ , C_1 , and C_2 . This construction is detailed for the sake of completeness.

At this point and in order to simplify the computations below we will assume that the distribution of ε_n is uniform on $[-1, +1]$ instead of the convolution of two $\mathcal{U}_{[-1/2, 1/2]}$ distributions.

Set $\lambda = 1 + \omega - c$, assume that $\lambda > 0$ (the case $\lambda < 0$ follows the same lines) and notice that:

$$\int_{\mathbb{R}} g(t) P(t, x) dt = \frac{1}{2c} \int_{\lambda - (\omega/x) - c}^{\lambda - (\omega/x) + c} g(s) ds = \frac{1}{2c} \int_{(\omega/x) - \lambda - c}^{(\omega/x) - \lambda + c} g(s) ds,$$

the last inequality stemming from parity of g . We should consider two cases $x > 0$ and $x \leq 0$.

The proof takes 2 parts ($x > 0$ and $x < 0$ respectively). Both are given again for completeness and because the problem is not symmetric. Each part is split in three steps: the two first steps deal with $x \notin \mathcal{C}$, the third with $x \in \mathcal{C} = (-\infty, -M_\tau] \cup [M_\tau, +\infty)$.

Part 1: $x > 0$

First step: We split $[0, M_\tau]$ in two subsets, $[0, M_\tau] = [0, A_\tau] \cup [A_\tau, M_\tau]$ with:

$$A_\tau = \omega / (M_\tau + 1 + \omega)$$

is chosen such that $0 \leq x \leq A_\tau$ implies the following inequality on the lower bound of the integral: $(\omega/x) - \lambda - c > M_\tau$. Clearly $A_\tau \leq M_\tau$ because $\lambda > 0 > -\tau - M_\tau$. Then:

$$\frac{1}{2c} \int_{(\omega/x) - \lambda - c}^{(\omega/x) - \lambda + c} g(s) ds = \frac{C_2}{2c} \int_{(\omega/x) - \lambda - c}^{(\omega/x) - \lambda + c} \log^2 |s| ds \leq C_2 \log^2 |(\omega/x) - \lambda + c|.$$

Let:

$$\sup_{0 \leq x \leq \omega / (c + M_\tau)} \sqrt{|x|} (\log |(\omega/x) - \lambda + c|)^2 = K_1(\omega, c, \tau) < +\infty.$$

The strictly positive $K_1(\omega, c, \tau)$ exists because $\sqrt{|x|} \log^2 |(\omega/x) - \lambda + c|$ is bounded on $[0, A_\tau]$. The first condition reads:

$$\frac{1}{2c} \int_{(\omega/x) - \lambda - c}^{(\omega/x) - \lambda + c} g(s) ds \leq \rho_1 C_1 / \sqrt{|x|}, \quad 0 \leq x \leq A_\tau$$

whenever

$$C_2 K_1(\omega, c, \tau) \leq \rho_1 C_1 \tag{13} \text{ cond1}$$

and ρ_1 will be fixed after the second step.

Second step: Now we turn to $A_\tau \leq x \leq M_\tau$. We still have $g(x) = C_1 / \sqrt{|x|}$ but we need to focus on the bounds of the integral.

This time the lower bound of the integral $(\omega/x) - \lambda - c \in [-\lambda - \tau, M_\tau]$ and the upper bound $(\omega/x) - \lambda + c \in [2c - \lambda - \tau, 2c + M_\tau]$. We are going to require that $(\omega/x) - \lambda - c \geq -M_\tau$ it suffices to take $\lambda + \tau \leq M_\tau$ and this comes down to the following set of constraint on τ : $\{\tau \geq c - \omega\} \cup \{\tau \leq c - 1\}$. We keep the second and assume once and for all that:

$$\tau \leq c - 1. \tag{14} \text{ constr-tau}$$

Then for $x \in [A_\tau, M_\tau]$,

$$\begin{aligned} \frac{1}{2c} \int_{(\omega/x) - \lambda - c}^{(\omega/x) - \lambda + c} g(s) ds &= \frac{1}{2c} \int_{(\omega/x) - \lambda - c}^{M_\tau} g(s) ds + \frac{1}{2c} \int_{M_\tau}^{(\omega/x) - \lambda + c} g(s) ds \\ &\equiv \mathcal{I}_1 + \mathcal{I}_2. \end{aligned} \tag{15} \text{ calc}$$

We want to make sure that the upper bound $(\omega/x) - \lambda + c$ is larger than M_τ . This will hold if $(\omega/M_\tau) - \lambda + c \geq M_\tau$ hence if $2c - \tau - \lambda \geq M_\tau$. We imposed previously that $\lambda + \tau \leq M_\tau$. So:

$$\tau < c - \frac{\omega}{c} \Rightarrow M_\tau < c \Rightarrow 2c - \tau - \lambda \geq M_\tau$$

but the constraint $\tau < c - \omega/c$ is weaker than (14) consequently (15) holds.

Focus on the first term \mathcal{I}_1 in (15) and consider:

$$\mathcal{I}_1 = \frac{1}{2c} \int_{(\omega/x)-\lambda-c}^{M_\tau} g(s) ds = \frac{1}{2c} \int_{(\omega/x)-\lambda-c}^{M_\tau} \frac{C_1}{\sqrt{|s|}} ds.$$

Consider the (only) two situations on the sign of $(\omega/x) - \lambda - c = (\omega/x) - (1 + \omega)$.

If $x < \omega/(1 + \omega)$ then $(\omega/x) - (1 + \omega) > 0$ and $\mathcal{I}_1 \leq \frac{C_1}{c} \sqrt{M_\tau}$ Notice by the way and for further purpose that:

$$\sup_{x \in [A_\tau, \omega/(1+\omega)]} \sqrt{|x|} \mathcal{I}_1 \leq \frac{C_1}{c} \sqrt{\frac{\omega}{1+\omega}} M_\tau.$$

If $x \geq \omega/(1 + \omega)$ then $(\omega/x) - (1 + \omega) \leq 0$ and:

$$\begin{aligned} \mathcal{I}_1 &= \frac{1}{2c} \int_{(\omega/x)-\lambda-c}^{M_\tau} g(s) ds = \frac{1}{2c} \int_{(\omega/x)-(1+\omega)}^0 \frac{C_1}{\sqrt{|s|}} ds + \frac{1}{2c} \int_0^{M_\tau} \frac{C_1}{\sqrt{|s|}} ds \\ &= \frac{C_1}{c} \left[\sqrt{|(\omega/x) - (1 + \omega)|} + \sqrt{M_\tau} \right]. \end{aligned}$$

Again:

$$\sqrt{|x|} \mathcal{I}_1 \leq \frac{C_1}{c} \left[\sqrt{|x(1 + \omega) - \omega|} + \sqrt{x M_\tau} \right].$$

From the bounds above we see that:

$$\sup_{x \in [A_\tau, M_\tau]} \sqrt{|x|} \mathcal{I}_1 \leq \frac{C_1}{c} \left[\sqrt{|M_\tau(1 + \omega) - \omega|} + M_\tau \right].$$

The reader will soon understand why we need to make sure that the right had side in equation above is strictly under C_1 . It is not hard to see that the function $\tau \mapsto \sqrt{|M_\tau(1 + \omega) - \omega|} + M_\tau$ is increasing and continuous on $[0, c - 1]$. If we prove that for some $\delta \in]0, 1[$:

$$\frac{1}{c} \left[\sqrt{|M_0(1 + \omega) - \omega|} + M_0 \right] = 1 - 3\delta < 1,$$

then the existence of some $\tau^+ > 0$ such that:

$$\frac{1}{c} \left[\sqrt{|M_{\tau^+}(1 + \omega) - \omega|} + M_{\tau^+} \right] = 1 - 2\delta < 1$$

delta
(16)

will be granted. But $\frac{1}{c} \left[\sqrt{|M_0(1 + \omega) - \omega|} + M_0 \right] = \frac{1}{c} \left[\sqrt{\frac{\omega}{c} \lambda + \frac{\omega}{c}} \right]$.

If we assume that $\lambda < \omega/c < (1 + c)/4$ (assumption **B**₁) then since $c > 1$:

$$\frac{1}{c} \left[\sqrt{\frac{\omega}{c} \lambda + \frac{\omega}{c}} \right] < \frac{1}{2} \left(1 + \frac{1}{c} \right) < 1.$$

We turn to \mathcal{I}_2 in (15):

$$\mathcal{I}_2 = \frac{1}{2c} \int_{M_\tau}^{(\omega/x)-\lambda+c} g(s) ds = \frac{C_2}{2c} \int_{M_\tau}^{(\omega/x)-\lambda+c} (\log s)^2 ds \leq \frac{C_2}{\sqrt{|x|}} K_2(\omega, c, \tau^+),$$

where:

$$K_2(\omega, c, \tau^+) = \sup_{x \in [A_\tau, M_\tau]} \frac{\sqrt{|x|}}{2c} \int_{M_\tau}^{(\omega/x)-\lambda+c} (\log s)^2 ds.$$

Set finally $\rho_1^+ = 1 - \delta < 1$.

From (15) we get:

$$\begin{aligned} \frac{1}{2c} \int_{(\omega/x)-\lambda-c}^{(\omega/x)-\lambda+c} g(s) ds &\leq \frac{C_1}{\sqrt{|x|}} (1 - 2\delta) + \frac{C_2}{\sqrt{|x|}} K_2(\omega, c, \tau^+) \\ &\leq \rho_1^+ \frac{C_1}{\sqrt{|x|}}, \end{aligned}$$

whenever holds the new condition:

$$C_2 K_2(\omega, c, \tau) \leq C_1 \delta.$$

cond2
(17)

Finally comparing (13) and (17), we see that both conditions cannot be incompatible. Accurate choices of the couple (C_1^+, C_2^+) are given by the summary bound:

$$C_2^+ \leq C_1^+ \min\left(\frac{\delta}{K_2}, \frac{1-\delta}{K_1}\right). \quad \text{condPart1 (18)}$$

It is now basic to see that the quadruple $(C_1^+, C_2^+, \tau^+, \rho_1^+)$ yields the drift condition (11) for $x \notin \mathcal{C}$.

Third step: The remaining step is to check the inequality for some ρ_2 :

$$\int_{\mathbb{R}} g(t) P(t, x) dt \leq \rho_1^+ g(x) + \rho_2,$$

for any x in \mathcal{C} -that is any $|x| > M_\tau$ (rather $x > M_\tau$ here as explained above since $x > 0$). We see that:

$$0 \leq \frac{\omega}{x} \leq \frac{\omega}{M_\tau},$$

and:

$$\frac{1}{2c} \int_{(\omega/x)-\lambda-c}^{(\omega/x)-\lambda+c} g(s) ds \leq \frac{1}{2c} \int_{-\lambda-c}^{2c-\tau-\lambda} g(s) ds \leq \frac{1}{2c} \int_{-(1+\omega)}^{3c-(1+\omega)} g(s) ds.$$

The values of the constants C_1 and C_2 were fixed above. Then denote:

$$\rho_2^+ = \frac{1}{2c} \int_{-(1+\omega)}^{3c-(1+\omega)} g(s) ds > 0,$$

then clearly for any x in \mathcal{C} :

$$\int_{\mathbb{R}} g(t) P(t, x) dt \leq \rho_2^+,$$

so that (11) holds.

Part 2 ($x \leq 0$)

We go on with $x < 0$ and $\lambda > 0$, set $y = -x \geq 0$,

$$\int_{\mathbb{R}} g(t) P(t, x) dt = \frac{1}{2c} \int_{\lambda-(\omega/x)-c}^{\lambda-(\omega/x)+c} g(s) ds = \frac{1}{2c} \int_{(\omega/y)+\lambda-c}^{(\omega/y)+\lambda+c} g(s) ds.$$

Since g is even and in view of the proposed \mathcal{C} we just have to prove exactly the following drift condition with $x > 0$:

$$\frac{1}{2c} \int_{(\omega/x)+\lambda-c}^{(\omega/x)+\lambda+c} g(s) ds \leq \rho_1 g(x) + \rho_2 1_{x \in \mathcal{C}}.$$

First step: Take $x \notin \mathcal{C}$. We split $[0, M_\tau]$ in two subsets, $[0, M_\tau] = [0, B_\tau] \cup [B_\tau, M_\tau]$ with $B_\tau = \omega / (M_\tau - \lambda + c)$ is chosen

such that $0 \leq x \leq B_\tau$ implies the following inequality on the lower bound of the integral: $(\omega/x) + \lambda - c > M_\tau$. Clearly $B_\tau \leq M_\tau$ for all τ . Then:

$$\frac{1}{2c} \int_{(\omega/x)+\lambda-c}^{(\omega/x)+\lambda+c} g(s) ds = \frac{C_2}{2c} \int_{(\omega/x)+\lambda-c}^{(\omega/x)+\lambda+c} \log^2 |s| ds \leq C_2 \log^2 |(\omega/x) + \lambda + c|.$$

Let:

$$\sup_{0 \leq x \leq B_\tau} \sqrt{|x|} \log^2 |(\omega/x) + \lambda + c| = K_1(\omega, c, \tau) < +\infty.$$

The strictly positive $K_1(\omega, c, \tau)$ exists because $\sqrt{|x|} \log^2 |(\omega/x) + \lambda + c|$ is bounded on $[0, B_\tau]$. The initial condition reads:

$$\frac{1}{2c} \int_{(\omega/x)+\lambda-c}^{(\omega/x)+\lambda+c} g(s) ds \leq \rho_1 C_1 / \sqrt{|x|}, \quad 0 \leq x \leq B_\tau$$

whenever $C_2 K_1(\omega, c, \tau) \leq \rho_1 C_1$ and ρ_1 will be fixed later.

Second step: Now we turn to $B_\tau \leq x \leq M_\tau$. We still have $g(x) = C_1 / \sqrt{|x|}$ but we need to focus on the bounds of the integral.

This time the lower bound of the integral $(\omega/x) + \lambda - c \in [\lambda - \tau, M_\tau]$ and the upper bound $(\omega/x) + \lambda + c \in [\lambda + 2c - \tau, M_\tau + 2c]$. If we assume that $\tau \leq \lambda$ then $(\omega/x) + \lambda - c \geq 0 \geq -M_\tau$. Besides in order that $\lambda + 2c - \tau \geq M_\tau$ we just have need that $2c \geq M_\tau$ or $\tau \leq c - \omega/(2c)$. As a consequence the assumption:

$$\tau \leq \min\left(\lambda, c - \frac{\omega}{2c}\right) \quad \text{condthau2} \quad (19)$$

allows to write for $x \in [B_\tau, M_\tau]$,

$$\begin{aligned} \frac{1}{2c} \int_{(\omega/x)+\lambda-c}^{(\omega/x)+\lambda+c} g(s) ds &= \frac{1}{2c} \int_{(\omega/x)+\lambda-c}^{M_\tau} g(s) ds + \frac{1}{2c} \int_{M_\tau}^{(\omega/x)+\lambda+c} g(s) ds \\ &\equiv \mathcal{I}_1 + \mathcal{I}_2, \end{aligned}$$

with non-null \mathcal{I}_1 and \mathcal{I}_2 . Focus on:

$$\begin{aligned} \mathcal{I}_1 &= \frac{C_1}{2c} \int_{(\omega/x)+\lambda-c}^{M_\tau} \frac{1}{\sqrt{s}} ds = \frac{C_1}{c} \left[\sqrt{M_\tau} - \sqrt{(\omega/x) + \lambda - c} \right] \\ \sup_{x \in [B_\tau, M_\tau]} \sqrt{x} \mathcal{I}_1 &\leq \frac{C_1}{c} M_\tau. \end{aligned}$$

At last we see that for $M_\tau \leq 1$ i.e. $\tau \leq c - \omega$, $\sup_{x \in [B_\tau, M_\tau]} \sqrt{x} \mathcal{I}_1 \leq C_1/c$. This condition combined with (19) let us set in the sequel:

$$\tau \leq \tau^- = \min(\lambda, c - \omega).$$

We turn to \mathcal{I}_2 :

$$\frac{1}{2c} \int_{M_\tau}^{(\omega/x)+\lambda+c} g(s) ds = \frac{C_2}{2c} \int_{M_\tau}^{(\omega/x)+\lambda+c} \log^2 |s| ds \leq \frac{C_2}{\sqrt{x}} K_2(\omega, c, \tau^-),$$

with:

$$K_2^- = K_2(\omega, c, \tau^-) = \sup_{x \in [B_\tau, M_\tau]} \frac{\sqrt{|x|}}{2c} \int_{M_\tau}^{(\omega/x)+\lambda+c} (\log s)^2 ds.$$

Set finally $\rho_1^- = (1 + 1/c)/2 < 1$. From all that was done above we get:

$$\int_{\mathbb{R}} g(t) P(t, x) dt \leq \frac{C_1}{\sqrt{|x|}} \frac{1}{c} + \frac{C_2}{\sqrt{|x|}} K_2^- \leq \rho_1^- \frac{C_1}{\sqrt{|x|}},$$

whenever:

$$C_2 K_2(\omega, c, \tau^-) \leq \frac{1-1/c}{2} C_1.$$

This will be combined with the constraint of the first step $C_2 K_1^- \leq \rho_1 C_1$ (we denoted $K_1(\omega, c, \tau^-) = K_1^-$). The new condition:

$$C_2^- \leq C_1^- \min\left(\frac{\rho_1^-}{K_1^-}, \frac{1-1/c}{2K_2^-}\right) \quad \text{condPart2} \quad (20)$$

ensures that

$$\frac{1}{2c} \int_{(\omega/x)+\lambda-c}^{(\omega/x)+\lambda+c} g(s) ds \leq \rho_1^- g(x) \text{ for } x \in [0, M_\tau].$$

Third step: The remaining step is to check the inequality:

$$\int_{\mathbb{R}} g(t) P(t, x) dt \leq \rho_1^- g(x) + \rho_2,$$

for any x in \mathcal{C} -that is here any $x > M_\tau$. Adapting the method given above is straightforward and leads to the desired result with a given ρ_2^- .

We are ready to conclude. Take

$$C_2^* = C_1^* \min\left(\frac{\delta}{K_2^+}, \frac{1-\delta}{K_1^+}, \frac{\rho_1^-}{K_1^-}, \frac{1-1/c}{2K_2^-}\right).$$

Conditions (18) and (20) hold for the couple (C_1^*, C_2^*) . For such a couple we have:

$$\begin{aligned} \int_{\mathbb{R}} g(t) P(t, x) dt &\leq \rho_1^+ g(x) + \rho_2^+, \quad x > 0, \\ \int_{\mathbb{R}} g(t) P(t, x) dt &\leq \rho_1^- g(x) + \rho_2^-, \quad x \leq 0, \end{aligned}$$

and for all x :

$$\int_{\mathbb{R}} g(t) P(t, x) dt \leq \max(\rho_1^+, \rho_1^-) g(x) + \max(\rho_2^+, \rho_2^-).$$

This finishes the proof of the Lemma.