



**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. 2018. hal-01918943v1

**HAL Id: hal-01918943**

**<https://hal.science/hal-01918943v1>**

Preprint submitted on 12 Nov 2018 (v1), last revised 4 Nov 2019 (v2)

**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 Bruned<sup>1,2</sup>, André Mas<sup>1</sup>, and Sylvain Wlodarczyk<sup>2</sup>

<sup>1</sup>IMAG, Univ Montpellier, CNRS, Montpellier, France

<sup>2</sup>Schlumberger Petroleum Services, Montpellier, 34000, France

## Abstract

Particle swarm optimization algorithm is a stochastic meta-heuristic solving global optimization problems appreciated for its efficacy and simplicity. It consists in a swarm of particles interacting among themselves and searching the global optimum. The trajectory of the particles has been well-studied in a deterministic case and more recently in a stochastic context. Assuming the convergence of PSO, we proposed here two CLT for the particles corresponding to two kinds of convergence behavior. These results can lead to build confidence intervals around the local minimum found by the swarm or to the evaluation of the risk. A simulation study confirms these properties.

**Keywords** Particle swarm optimization, Convergence, Central limit theorem

## 1 Introduction

Eberhart and Kennedy [1995] introduced the particle swarm optimization algorithm (PSO) based on social interactions (behaviors of birds or fishes). Since then PSO has known a great popularity in many domains and gave to birth to many variants of the original algorithm (see Zhang et al. 2015 for a survey of variants and applications of PSO). 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 among themselves. It is an iterative algorithm where the particles update at each step their position and their speed. The dynamic of the particles relies on two attractors: their personal best position (historical best position of the particle), and the neighborhood best position (corresponding to the social component of the particles). In the dynamic equation, the attractors are linked with a stochastic process to have a global exploration of 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  $g_n^s$  (neighbourhood best position).  
    Change velocity ( $v_n^s$ ) and position ( $x_n^s$ ) according to dynamic equation.  
end for
```

---

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 such convergence.
- the convergence of each particle to a point (Poli 2009, Cleghorn and Engelbrecht 2018).

Here we focus on the convergence of each particle to a point. It started with 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).

Our objective is to provide a central limit theorem for the convergence of a single particle. We distinguish two cases: a case where the particle is oscillating between  $g_n^s$  and  $p_n^s$ , with  $p_n^s \neq g_n^s$ , and another one where the particle converges to  $g_n^s = p_n^s$ . For the non-oscillating case, we propose also a study of the risk  $\mathbb{E}|x_n^s - g_n^s|^2$ .

## 2 Dynamic Equation

We consider here a cost function  $f : \mathbb{R}^d \rightarrow \mathbb{R}^+$  that should be minimized on a compact set  $\Omega$ . 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  $(r_{j,n})_{j=1,2,n \geq 1}$  be sequences of independent random vectors whose margins obey a uniform distribution over  $[0, 1]$  and denote by  $\omega$ ,  $c_1$  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 equation, Poli 2009):

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

where  $p_n^s$  and  $g_n^s$  are respectively the best personal position and the best neighborhood position of the particle  $s$ ,

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

with  $\mathcal{V}(s)$  the neighborhood of particle  $s$ . In Equation (1),  $\odot$  stands for the Hadamard product:

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

The neighbourhood is associated to the swarm's topology: if the topology is called global (all the particles communicate between each other) then  $g_n^s = g_n = \operatorname{argmin}_{t \in \{p_n^1, \dots, p_n^S\}} f(t)$ .

We take it for granted that particles are warm, reached an area of the domain were they fluctuate without exiting and where the personal and global best converge. In other words, our

final goal is to provide asymptotic confidence sets without addressing the issue of convergence. The type of convergence we have in mind will be made more precise in assumptions below.

We fix once and for all:

$$c_1 = c_2 = c$$

Without this assumption above, computations turn out to get considerably intricate.

### 3 Main results

Let us denote  $\mathcal{F}_n^s$  the filtration generated by  $\{x_0^s, \dots, x_n^s\}$ , a single particle up to step  $n$  and  $\mathcal{F}_n^S$  the filtration generated by the swarm  $\{x_0^s, \dots, x_n^s : s = 1, \dots, S\}$  up to step  $n$ . We denote  $g_n^s = \mathbb{E}g_n^s + \xi_n^s$  and  $p_n^s = \mathbb{E}(p_n^s) + \nu_n^s$  the expectation-variance decomposition of  $g_n^s$  and  $p_n^s$  where  $\xi_n^s$  and  $\nu_n^s$  are centered random vectors and support all the variability of  $g_n^s$  and  $p_n^s$  respectively.

#### Notations:

The usual euclidean norm and associated inner product for vectors in  $\mathbb{R}^d$  are denoted respectively  $\|\cdot\|$  and  $\langle \cdot, \cdot \rangle$ . We will also need norms for square matrices of size  $d$ . It is well known that all norms on the vector space of  $d \times d$  matrices are equivalent. We will use here two of them: the usual sup norm and Frobenius norm. For any square matrix  $M$  of size  $d$  with entries  $M_{i,j}$  these norms are respectively defined by:

$$\|M\|_\infty = \sup_{x \neq 0} \frac{\|Mx\|}{\|x\|} \quad \|M\|_F = \sqrt{\sum_{i,j} |M_{i,j}|^2}.$$

The bound  $\|M\|_\infty \leq \|M\|_F \leq \sqrt{d} \|M\|_\infty$  is classical.

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 by  $(u \otimes v)(x) = \langle v, x \rangle u$ . Besides  $\|u \otimes v\|_\infty = \|u \otimes v\|_F = \|u\| \|v\|$ . 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  $\eta, \varepsilon, u$  and  $v$  be four vectors in  $\mathbb{R}^d$ . Then:

$$(\eta \odot u) \otimes (\varepsilon \odot v) = (\eta \otimes \varepsilon) \odot (u \otimes v), \quad (2)$$

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.

If  $X$  is a random vectors with null expectation then  $\mathbb{E}(X \otimes X)$  is the covariance matrix of  $X$ .

Convergence in probability of  $X_n$  to  $X$  is denoted  $X_n \rightarrow_{\mathbb{P}} X$ . The arrow  $\leftrightarrow$  stands for convergence in distribution (weak convergence).

In all the sequel we consider a single particle and drop the particle index so that  $x_n^s = x_n$ ,  $p_n^s = p_n$  and  $g_n^s = g_n$ .

#### 3.1 First case: oscillatory ( $p \neq g$ )

With the simplifications defined above, the second line in (1) becomes:

$$x_{n+1} = (1 + \omega) x_n - \omega x_{n-1} + c(r_{1,n} + r_{2,n}) \odot \left( \frac{p_n + g_n}{2} - x_n \right) + c(r_{1,n} - r_{2,n}) \odot \frac{p_n - g_n}{2}.$$

Taking expectation:

$$\mathbb{E}x_{n+1} = (1 + \omega) \mathbb{E}x_n - \omega \mathbb{E}x_{n-1} + c \left( \frac{\mathbb{E}p_n + \mathbb{E}g_n}{2} - \mathbb{E}x_n \right). \quad (3)$$

By subtracting we get the following Proposition from the above.

**Proposition 1.** *Denote the centered process  $z_n = x_n - \mathbb{E}x_n$  and  $e$  the vector of  $\mathbb{R}^d$  defined by  $e = (1, 1, \dots, 1)$  then the PSO equation becomes*

$$z_{n+1} = (1 + \omega - c) z_n - \omega z_{n-1} - c \varepsilon_n \odot z_n + c \tilde{r}_n + \frac{c}{2} \eta_n \odot (p_n - g_n), \quad (4)$$

with

$$\begin{aligned} \varepsilon_n &= r_{1,n} + r_{2,n} - e & \eta_n &= r_{1,n} - r_{2,n}, \\ \tilde{r}_n &= \varepsilon_n \odot \left( \frac{\mathbb{E}p_n + \mathbb{E}g_n}{2} - \mathbb{E}x_n \right) + \frac{e + \varepsilon_n}{2} \odot (p_n - \mathbb{E}p_n + g_n - \mathbb{E}g_n). \end{aligned}$$

Equation (4) will be a starting point especially for studying single particle trajectories.

### Assumptions:

**A<sub>1</sub>** : At any step  $n$  and for all particles  $s$   $x_n^s \in \Omega$  where  $\Omega$  is a compact subset of  $\mathbb{R}^d$ .

**A<sub>2</sub>** : For both  $t_n = \xi_n^s$  and  $t_n = \nu_n^s$  :  $\frac{1}{\sqrt{N}} \sum_{n=1}^N t_n \rightarrow_{\mathbb{P}} 0$  and  $\lim_{n \rightarrow +\infty} \mathbb{E} \|t_n\| = 0$ .

**A<sub>3</sub>** : The following inequality connecting  $c$  and  $\omega$  holds

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

### 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<sub>2</sub>** is specific of what we intend by a convergent PSO.

First, notice that  $\frac{1}{\sqrt{N}} \sum_{n=1}^N t_n \rightarrow_{\mathbb{P}} 0$  does not imply  $\lim_{n \rightarrow +\infty} \mathbb{E} \|t_n\| = 0$  nor the converse. Take for instance  $t_n = (-1)^n t_0$  with  $\mathbb{E}t_0 \neq 0$  then  $\frac{1}{\sqrt{N}} \sum_{n=1}^N t_n \rightarrow_{\mathbb{P}} 0$  whereas  $\mathbb{E} \|t_n\| = \mathbb{E} \|t_0\|$ . Conversely take  $t_n = t_0 / \log(n)$   $\lim_{n \rightarrow +\infty} \mathbb{E} \|t_n\| = 0$  but  $\frac{1}{\sqrt{N}} \sum_{n=1}^N t_n = \frac{t_0}{\sqrt{N}} \sum_{n=1}^N \log^{-1}(n)$  cannot converge in probability to 0. Assumption **A<sub>2</sub>** 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 hence  $g_n = \mathbb{E}g_n$  for  $n$  in  $[\underline{N}, \overline{N}]$  supposedly. Notice however that convergence of the expectation towards  $p$  and  $g$  is not mentioned at this step.

Note that assumption **A<sub>3</sub>** holds for the classical calibration appearing in Clerc and Kennedy [2002] (constriction constraints) with  $c = 1.496172$  and  $\omega = 0.72984$ .

Let  $\delta = (\delta_1, \dots, \delta_d) \in \mathbb{R}^d$ , in the Theorem below 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  $\delta^{\odot 2} = (\delta_1^2, \dots, \delta_d^2)$ .

**Theorem 2.** *Assume that, when  $n$  goes to infinity  $\mathbb{E}p_n \rightarrow p$  and  $\mathbb{E}g_n \rightarrow g$ . Set:*

$$\mathfrak{L} = 2c \left( \frac{1 - \omega}{1 + \omega} \right) \left( 1 + \omega - \frac{c}{2} \right) - \frac{c^2}{6}, \quad \mathfrak{C} = \frac{c}{12\mathfrak{L}} \frac{1 - \omega}{1 + \omega} \left( 1 + \omega - \frac{c}{2} \right).$$

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

$$\sqrt{N} \left( \frac{1}{N} \sum_{n=1}^N x_n - \frac{1}{N} \sum_{n=1}^N \frac{\mathbb{E}p_n + \mathbb{E}g_n}{2} \right) \hookrightarrow \mathcal{N}(0, \Gamma),$$

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

**Corollary 3.** *If the convergence of  $\mathbb{E}p_n$  and  $\mathbb{E}g_n$  is fast enough, that is:*

$$\frac{1}{\sqrt{N}} \sum_{n=1}^N (|p - \mathbb{E}p_n| + |g - \mathbb{E}g_n|) \rightarrow 0,$$

then the Theorem above comes down to:

$$\sqrt{N} \left( \frac{1}{N} \sum_{n=1}^N x_n - \frac{p + g}{2} \right) \hookrightarrow \mathcal{N}(0, \Gamma).$$

**Remark 4.** *From the Corollary above we can immediately derive approximated confidence sets for  $\theta = (p + g)/2$ . We note however that the vector  $\theta$  may not be of crucial interest for the initial optimization problem conversely to  $g$ .*

### 3.2 Second case: non-oscillatory ( $p = g$ )

In this section we suppose once and for all that  $x_n \in \mathbb{R}$ . Starting from Equation (1), the PSO equation becomes this time:

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

We change our centering and consider  $x_n - g_n \equiv y_n$  instead of  $x_n - \mathbb{E}x_n$ :

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

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

Under matrix form the system is purely linear but driven by a random matrix:

$$\begin{pmatrix} y_{n+1} \\ y_n \end{pmatrix} = \begin{bmatrix} 1 + \omega - c(1 + \varepsilon_n) & -\omega \\ 1 & 0 \end{bmatrix} \begin{pmatrix} y_n \\ y_{n-1} \end{pmatrix} \quad (5)$$

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

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  $(y_n)_{n \in \mathbb{N}}$ . 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  $S$ . These assumptions are uneasy to check here that is why in section 3.2.2 (Weak convergence), we turn to a direct approach. Let us start with a result on the risk for a single particle.

In this part, we fix once for all  $\lambda = 1 + \omega - c$ .

### 3.2.1 Risk

**Proposition 5.** *Assume that  $0 \in \Omega$  and that  $x_0$  and  $x_1$  are drawn independently from the same centered distribution with variance  $\sigma^2$  then for all  $n$   $\mathbb{E}\mathbf{y}_n = 0$ . Denote  $\lambda = 1 + \omega - c$ , consider the polynomial:*

$$\mathbf{P}_M(x) = -x^3 + (\lambda^2 - \omega + c^2/6)x^2 + \omega(\omega - \lambda^2 + c^2/6)x + \omega^3,$$

and assume that:

$$\mathbf{B}_1 : \frac{c^2}{6} < \min\left(\omega - \lambda^2, 3\frac{|\lambda|}{\sqrt{\omega}}(\omega - \lambda^2), 1 + \lambda^2 - \omega^2 - 2\frac{\lambda^2}{1 + \omega}\right)$$

holds. Then  $\mathbf{P}_M$  has a single real strictly positive root denoted  $r^*$ . Besides let  $z$  and  $\bar{z}$  be the two other complex conjugate roots then the following inequality holds:

$$|z| < r^* < 1, \quad (6)$$

and almost surely:

$$\|\text{cov}(\mathbf{y}_n)\| \leq \mathbf{K} \cdot (r^*)^n,$$

where  $\mathbf{K}$  is some constant which does not depend on  $n$  and  $\|\text{cov}(\mathbf{y}_n)\|$  stands for any classical matrix norm (Schatten, entry-wise or induced by vectors) of the covariance matrix of  $\mathbf{y}_n$ .

**Corollary 6.** *As a consequence the inequality above provides an upper bound for the mean square distance between the particle and  $g_n$  in the non-oscillatory setting:*

$$\mathbb{E}|x_n - g_n|^2 \leq \mathbf{K}' \cdot (r^*)^n.$$

**Remark 7.** *Assumption  $\mathbf{B}_1$  holds for  $\omega = 0.72984$  and  $c = 1.496172$ . Then an estimate for  $r^*$  is 0.9442164. When  $|\lambda|/\sqrt{\omega} < 1/3$  condition  $\mathbf{B}_1$  comes down to:*

$$\frac{c^2}{6} < \min\left(3\frac{\lambda}{\sqrt{\omega}}(\omega - \lambda^2), 1 + \lambda^2 - \omega^2 - 2\frac{\lambda^2}{1 + \omega}\right).$$

### 3.2.2 Weak convergence

To have the weak convergence, we turn to another approach. Assuming that for all  $n$   $y_n \neq 0$ , we have successively:

$$\begin{aligned} \frac{y_{n+1}}{y_n} &= (1 + \omega - c + c\varepsilon_n) - \omega\frac{y_{n-1}}{y_n}, \\ X_{n+1} &= 1 + \omega - c - \frac{\omega}{X_n} + c\varepsilon_n. \end{aligned}$$

with  $X_n = y_n/y_{n-1}$ . Notice that  $\sum_{n=1}^N \log|X_n| = \log|y_N| - \log|y_0|$ . We focus 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  $\mu$  and  $\sigma^2$ :

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

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]. We need to check two points: existence of a small set  $\mathcal{C}$  and of a function  $g$  with a drift condition (see Meyn and Tweedie, 2012).

Let us focus on the Markov chain:

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

where  $\varepsilon_n = r_{1,n} + r_{2,n} - 1$  has a “witch hat” distribution with support  $[-1, +1]$ . Denote  $\pi$  the stationary distribution of  $X_n$  and:

$$\begin{aligned} \mu_x &= \mathbb{E}_\pi \log |X|, \\ \sigma_x^2 &= \text{Var}_\pi (\log |X_0|) + 2 \sum_{k=1}^{+\infty} \text{Cov}_\pi (\log |X_0|, \log |X_k|). \end{aligned}$$

**Theorem 8.** *Assume that  $\omega \in ]0, 1[$ ,  $c > 1$ ,  $X_n$  is Harris recurrent, for sufficiently large  $n$   $g_n = p_n$ , and that:*

$$\mathbf{B}_2 : \lambda < \omega/c < (1 + c)/4$$

holds. Then:

$$\frac{1}{\sqrt{n}} (\log |x_n - g_n| - n\mu_x) \hookrightarrow \mathcal{N} \left( 0, \sigma_x^2 \right).$$

**Remark 9.** *The mean and variance  $\mu_x$  and  $\sigma_x^2$  are usually unknown but may be approximated numerically. We refer to the simulation section for more details.*

**Corollary 10.** *If  $p_n = p$  an asymptotic confidence (non convex) region for  $p$  at level  $1 - \alpha$  denoted  $\Lambda_{1-\alpha}$  below may be derived from the preceding Theorem:*

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

### 3.3 The swarm at a fixed step

In this section we change our viewpoint but we still address only the case  $x_n \in \mathbb{R}$  even if our results may be straightforwardly generalized to  $x_n \in \mathbb{R}^d$ . 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 the whole swarm. We provide below two kinds of results. First, a Central Limit Theorem suited to the case when the number of particles in the swarm is reasonably large. Second, when typically only a few particles are generated (say less than 30) we obtain a concentration inequality available for any swarm size. In the sequel, the iteration step  $n$  is assumed to be fixed and we denote  $S$  the swarm size. In order to clarify the method, we assume below that for all particles  $x_n^i$  in the swarm  $p_n^i = g_n$ . In other words, no local minimum stands in the domain  $\Omega$ . This may be possible by a preliminary screening of the search space. 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 (5) we get with  $\mathbf{y}_n^i = x_n^i - g_n$  :

$$\begin{aligned} \mathbf{y}_n^i &= \mathbf{T}_n^i \mathbf{y}_1^i, \\ \mathbf{T}_n^i &= \Pi_{j=2}^n \mathbf{S}_j, \quad \mathbf{S}_j = \begin{bmatrix} 1 + \omega - c(1 + \varepsilon_j^i) & -\omega \\ 1 & 0 \end{bmatrix}. \end{aligned}$$

Assume that the domain  $\Omega$  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.

**Assumptions :**

$\mathbf{C}_1$  : The operational domain  $\Omega$  contains 0 (and is ideally a symmetric set).

$\mathbf{C}_2$  : The couples  $(x_0^i, x_1^i)_{i \leq S}$  are i.i.d. and centered.

$\mathbf{C}_3$  : For all  $i$  in  $\{1, \dots, S\}$   $p_n^i = g_n$ .

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

**Proposition 11.** *Under assumptions  $\mathbf{C}_{1-3}$  a CLT holds when  $S$  the number of particles in the swarm becomes large :*

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

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

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

**Remark 12.** *The convergence of  $\hat{\sigma}_n^2$  to  $\sigma_n^2$  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 Proposition above paves the way towards asymptotic confidence sets of the form :*

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

Under the previous set of assumptions we obtain as well exponential concentration inequalities (see Chapter 2 in Massart, 2007) which may be of interest when  $S$  is not large. We state below the Hoeffding version.

**Proposition 13.** *Let  $\mathbf{D}(\Omega)$  be the diameter of  $\Omega$ , under assumptions  $\mathbf{C}_{1-3}$  and for all  $S$  :*

$$\mathbb{P} \left( \left| \bar{x}^S - g \right| > \tau \right) \leq 2 \exp \left( -\frac{2S\tau^2}{\mathbf{D}^2(\Omega)} \right).$$

**Remark 14.** *In the Proposition above  $\mathbf{D}(\Omega)$  may be replaced by a sharper constant. Namely if for all  $n$  and  $i$   $x_n^i \in [\underline{c}_n, \bar{c}^n]$  almost surely then  $\mathbf{D}^2(\Omega)$  may be replaced by  $(\bar{c}^n - \underline{c}_n)^2$ . Clearly Proposition 13 does not make sense for small values of  $S/\mathbf{D}^2(\Omega)$  which may occur if the size of  $\Omega$  is large with respect to  $S$ . Notice however that in the present framework where we consider convergent PSO we can expect  $|\bar{c}^n - \underline{c}_n|$  to tend to zero quickly enough to ensure a satisfactory bound after several iteration steps, even with few particles in the swarm. As a consequence, a well-calibrated non asymptotic concentration inequality as the one just above may outperform the confidence intervals derived from the CLT.*

## 4 Simulation and numerical results

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:  $f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$ . Figure 1 illustrates the contour of this function.

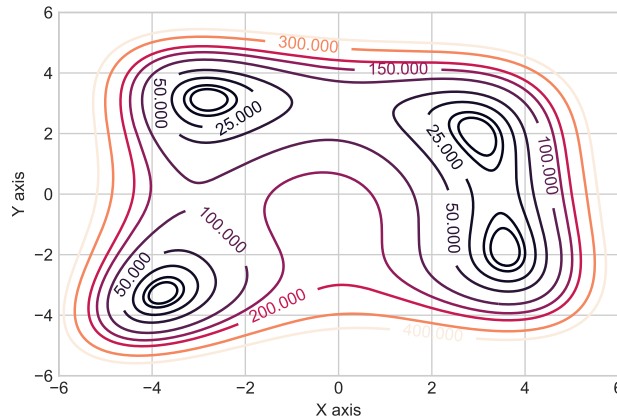


Figure 1: Contour of the Himmelblau's function in the space  $[-6, 6]^2$ .

With the Himmelblau's function, we can observe the two different behaviors of the particles: oscillatory and non-oscillatory. The Himmelblau's function has four local 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.

### 4.1 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 neighbourhood best position. In this case, the convergence of the  $p_n$  and  $g_n$  to  $(3.58, -1.84)$  or  $(3, 2)$  are satisfying the conditions of Corollary 3. We have to verify the Gaussian asymptotic behaviour of  $H_1^s(N) = \sqrt{N} \left( \frac{1}{N} \sum_{n=1}^N x_n^s - \frac{p+q}{2} \right)$  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 2 and 3 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.

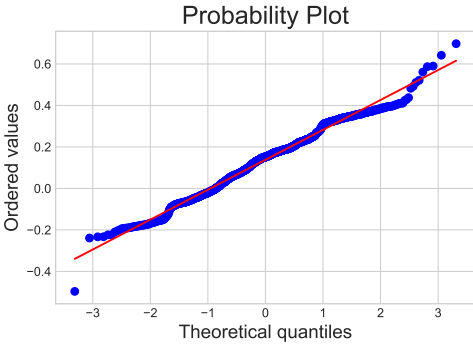


Figure 2: Normal probability plot of  $H_1^s(N)$  on the first coordinate.

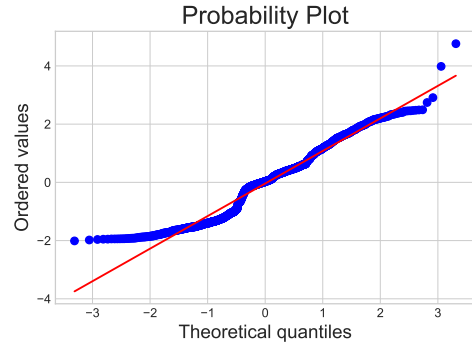


Figure 3: Normal probability plot of  $H_1^s(N)$  on the second coordinate.

To check the formula of the covariance matrix  $\Gamma$ , the confidence ellipsoid is also a good indicator to display (see Figure 4). For 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 5 shows the trajectory of  $x_n^s$  and  $H_1^s(N)$  on the y axis,  $H_1^s(N)$  remains bounded.

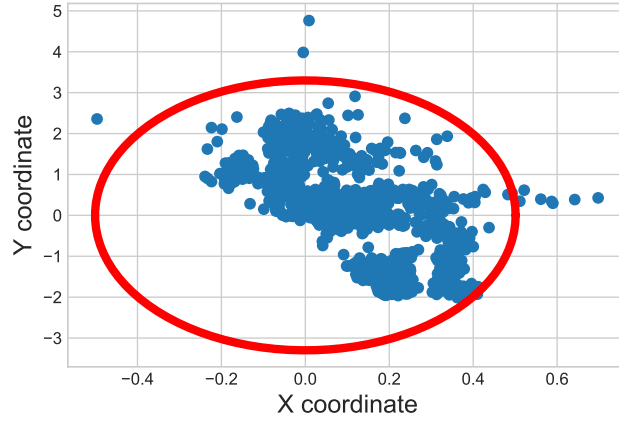


Figure 4: Trajectory of  $H_1^s(N)$  for an oscillating particle in  $[-10, 10]^2$ . The confidence ellipsoid at a level of 85% is displayed in red. Around 99% of the trajectory of  $H_1^s(N)$  is inside the ellipse.

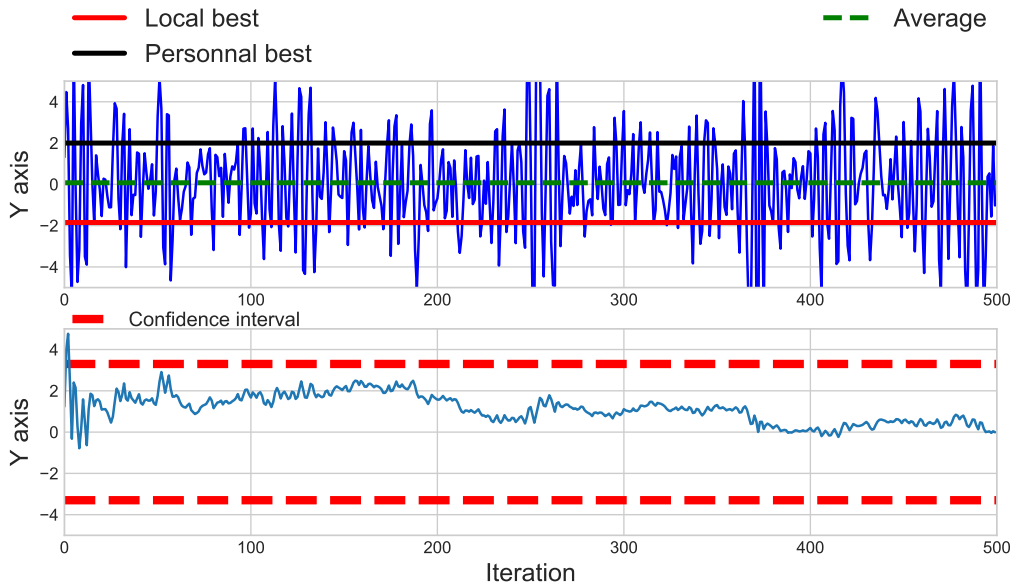


Figure 5: Top track: Trajectory of a oscillating particle on the y axis. The particle is oscillating between 2 and -1.84. Bottom track: corresponding trajectory of  $H_1^s(N)$  on the y axis, the red dot are corresponding to the 95% confidence interval. The trajectory of  $H_1^s(N)$  is well bounded.

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 6 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).

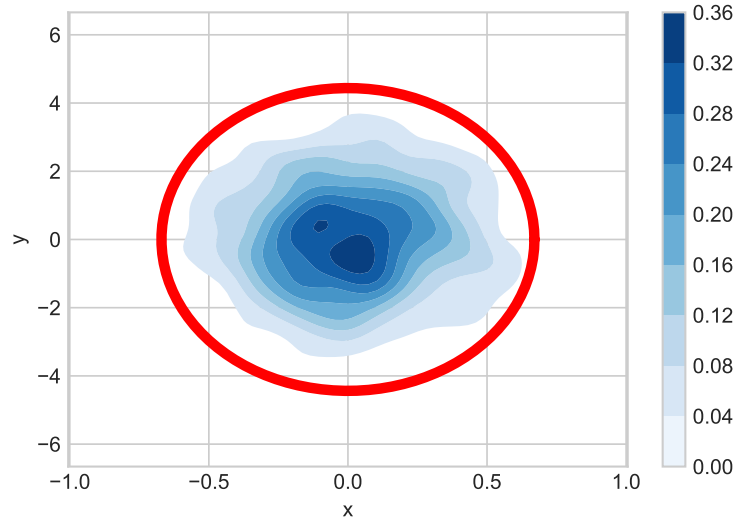


Figure 6: Density of  $H_1(2000)$  with 1150 particles issued from Monte-Carlo simulations. The red ellipse is the 95% confidence ellipsoid.

## 4.2 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  $p_n = g_n$  for a sufficiently large  $n$ .

### 4.2.1 Risk

First, Figure 7 illustrates the condition over  $\omega$  and  $c$  of Proposition 5. We notice that every values of  $c$  above 1.6 are rejected.

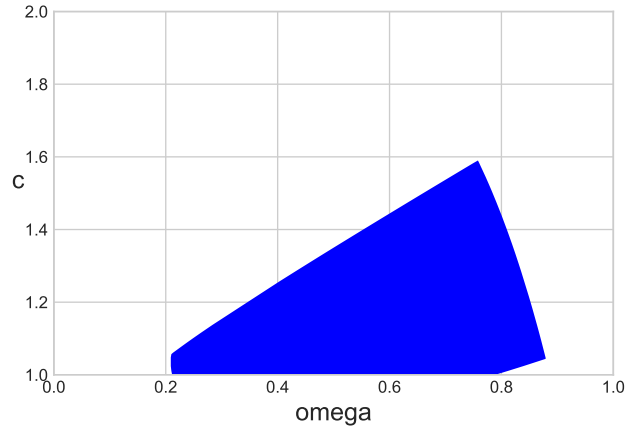


Figure 7: Condition over  $\omega$  and  $c$  of Proposition 5. The acceptance region is displayed in blue.

To verify Proposition 5, we compute the trajectory of  $\mathbb{E}|x_n - g_n|^2$  with around 200 converging particles to  $(3, 2)$ . Figure 8 displays this trajectory which always stays below  $(r^*)^n$  where  $r^* = 0.9442164$ .

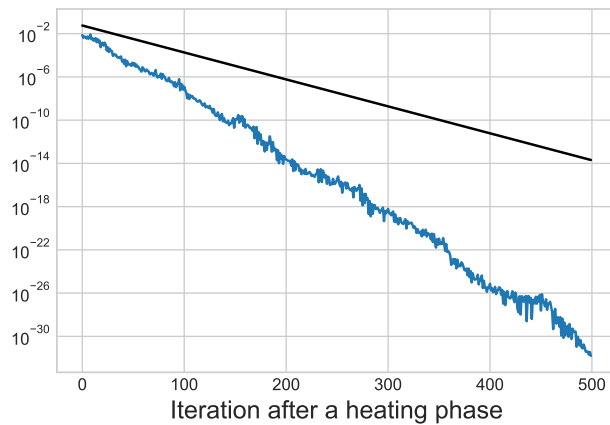


Figure 8: Trajectory of  $\mathbb{E}|x_n - g_n|^2$  (with 200 particles) in logarithmic scale after a heating phase of 50 iterations. The black line is  $(r^*)^n$ .

### 4.2.2 Weak convergence

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  $\mu_x$  can be performed thanks to a linear regression.

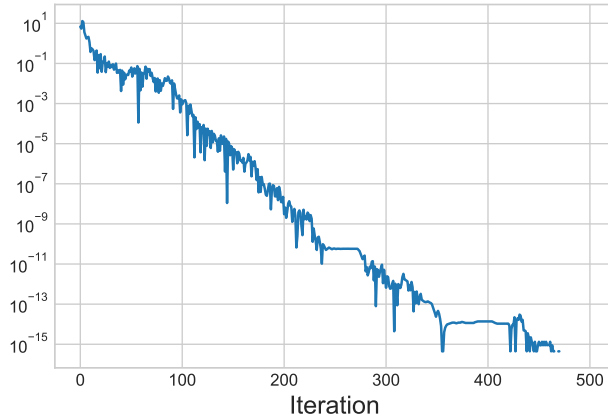


Figure 9:  $|x_n - g_n|$  over 500 iterations in a logarithmic scale. After 300 iterations, we reach the computer precision. One can notice the linear behavior of  $\log(|x_n - g_n|)$ .

Using many converging particles, a monte carlo approximation of  $\mu_x$  is done. For the approximation of  $\sigma_x$ , 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  $\mu_x$ . Figure 10 displays the normal probability plot of  $H_2(N)$  on the first coordinate, the theoretical quantiles are well fitted by  $H_2(N)$ .

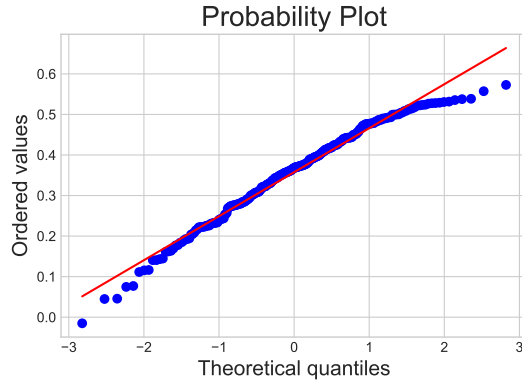


Figure 10: Normal probability plot of  $H_2(N)$  on the first coordinate.

Figure 11 illustrates different trajectories of  $H_2(N)$  on the first coordinate which are bounded by the 95% confidence interval deduced from  $\bar{\sigma}_x$ .

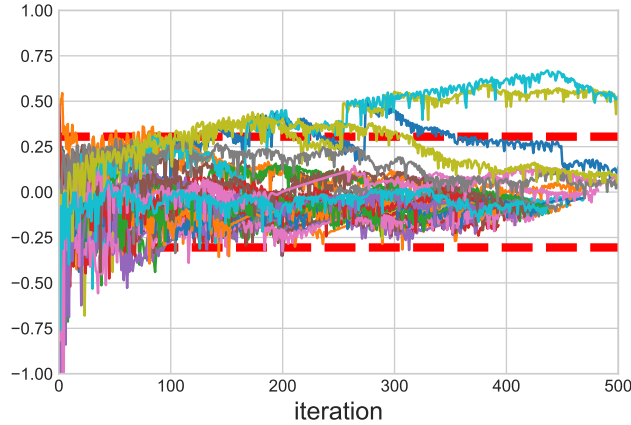


Figure 11: Trajectories of  $H_2(N)$  on the first coordinate for five particles. The red dot represents the 95% confidence interval deduced from  $\bar{\sigma}_x$ . Trajectories stop around the 400 iterations (after an heating phase) due to numerical precision.

### 4.2.3 Swarm at a fixed step

To check the Proposition 11, 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 Proposition 11 because of the convergence rate of the particles. Indeed, when  $p_n^s = g_n^s$ , 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 behaviour 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 behaviour 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, Proposition 11 seems to be verified.

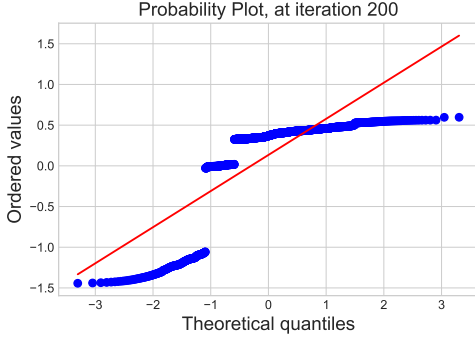


Figure 12: Normal probability plot of  $H_3^{n=200}(S)$  at the 200th iteration on the first coordinate, using all the particles. We observe a discontinuity of the probability plot due to belated particles.

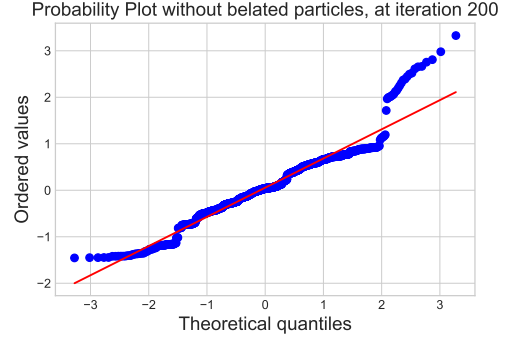


Figure 13: Normal probability plot of  $H_3^{n=200}(N)$  at the 200th iteration on the first coordinate, computed without outliers or belated particles.

## 5 Derivations of the results

### 5.1 First case: oscillatory

#### 5.1.1 Technical Lemmas

Let us start with some Lemmas who will be invoked later.

**Lemma 15.** *Let  $\varepsilon_n^{(i)}$  and  $\eta_n^{(j)}$  be any coordinate of the random vectors  $\varepsilon_n$  and  $\eta_n$ . Clearly,  $\varepsilon_n^{(i)}$  and  $\eta_n^{(j)}$  are not independent but not correlated and follow the same type of distribution. Besides:*

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

**Lemma 16.** *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 both  $\|\cdot\|_\infty$  and  $\|\cdot\|_F$  norms:*

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

The proof of the above Lemma is a straightforward application of Cauchy-Schwartz inequality.

**Lemma 17.** *When  $\mathbf{A}_{1-3}$  hold and when  $N$  goes to infinity:*

$$\frac{1}{N} \sum_{n=1}^N \left\| \frac{\mathbb{E}p_n + \mathbb{E}g_n}{2} - \mathbb{E}x_n \right\|^2 \rightarrow 0.$$



**Proof:** Denote  $q_n = \mathbb{E}x_n - (\mathbb{E}p_n + \mathbb{E}g_n)/2$ . From (3) we get in matrix form:

$$\begin{pmatrix} q_{n+1} \\ q_n \end{pmatrix} = \begin{bmatrix} 1 + \omega - c & -\omega \\ 1 & 0 \end{bmatrix} \begin{pmatrix} q_n \\ q_{n-1} \end{pmatrix} + \begin{pmatrix} K_n \\ 0 \end{pmatrix}$$

where:

$$K_{n+1} = (1 + \omega) \mathbb{E}(p_n + g_n) - \omega \mathbb{E}(p_{n-1} + g_{n-1}) - \mathbb{E}(p_{n+1} + g_{n+1}).$$

The above equation may be rewritten  $\mathbf{q}_{n+1} = \mathbf{M}\mathbf{q}_n + \mathbf{K}_{n+1}$  with obvious but shorter notations. Notice that in  $K_n$  each  $\mathbb{E}(p_k + g_k)$ ,  $k \in \{n-1, n, n+1\}$  may be replaced by  $\mathbb{E}(p_k + g_k) - (p + g)$ . Consequently, we know that  $K_n$  tends to 0. Also notice that  $\|\mathbf{M}\|_\infty < 1$  because  $\omega < 1$ .

Then we derive:

$$\mathbf{q}_n = \sum_{p=2}^n \mathbf{M}^{n-p} \mathbf{K}_p.$$

It suffices to notice now that if  $\mathbf{q}_n$  decays to 0, so does  $\|\mathbf{q}_n\|$  and Lemma 17 will be derived by a simple application of Cesaro's famous Lemma. Let us prove now that  $\mathbf{q}_n \downarrow 0$ .

Take  $\varepsilon > 0$  denote  $K_\infty = \max(\|\mathbf{K}_p\|)$  and  $C_M = [1 - \|\mathbf{M}\|_\infty]^{-1}$ . First pick  $N^*$  such that  $\sup_{p \geq N^*+1} \|\mathbf{K}_p\| < \varepsilon / (2C_M)$

$$\begin{aligned} \|\mathbf{q}_n\| &\leq \sum_{p=2}^{N^*} \|\mathbf{M}\|_\infty^{n-p} \|\mathbf{K}_p\| + \sum_{p=N^*+1}^n \|\mathbf{M}\|_\infty^{n-p} \|\mathbf{K}_p\| \\ &\leq \|\mathbf{M}\|_\infty^{n-N^*} \sum_{p=2}^{N^*} \|\mathbf{M}\|_\infty^{n-p} \|\mathbf{K}_p\| + \max_{p \geq N^*+1} \|\mathbf{K}_p\| \sum_{p=N^*+1}^n \|\mathbf{M}\|_\infty^{n-p} \\ &\leq C_M \left[ K_\infty \|\mathbf{M}\|_\infty^{n-N^*} + \sup_{p \geq N^*+1} \|\mathbf{K}_p\| \cdot \right] \end{aligned}$$

Then let  $N^\dagger$  be such that  $\|\mathbf{M}\|_\infty^{N^\dagger} < \varepsilon / (2C_M K_\infty)$ . It can be seen from the equations above that for  $n > N^\dagger + N^*$ ,  $\|\mathbf{q}_n\| < \varepsilon$ .

### 5.1.2 Proof of Theorem 2

The following Proposition is crucial:

**Proposition 18.** *Let  $S_N = \sum_{n=1}^N z_n$  then:*

$$S_N = \sum_{n=1}^N \left[ \frac{\eta_n}{2} \odot (p_n - g_n) - \varepsilon_n \odot z_n \right] + \sum_{n=1}^N \tilde{r}_n + \frac{1}{c} (z_1 - \omega z_0 + \omega z_N - z_{N+1}).$$

Under assumptions  $\mathbf{A}_2$

$$\frac{1}{\sqrt{N}} \sum_{n=1}^N \tilde{r}_n \rightarrow_{\mathbb{P}} 0.$$

Besides when  $\mathbf{A}_1$  holds (the  $x'_n$ s are almost surely bounded) then  $S_N/\sqrt{N}$  converges weakly if and only if:

$$\frac{1}{\sqrt{N}} \sum_{n=1}^N \left[ \frac{\eta_n}{2} \odot (p_n - g_n) - \varepsilon_n \odot z_n \right]$$

converges weakly too.

Proposition 18 shows that the limiting distribution of  $S_N$  is completely determined by the limiting distribution of  $\sum_{n=1}^N \left[ \frac{\eta_n}{2} \odot (p_n - g_n) - \varepsilon_n \odot z_n \right]$ . If a Central Limit Theorem holds for the previous series, then the limiting distribution will depend on the covariance matrix of  $\frac{\eta_n}{2} \odot (p_n - g_n) - \varepsilon_n \odot z_n$ . The latter will be decomposed in several terms. The next Proposition focuses on:

$$\Gamma_{N,z} = \frac{1}{N} \sum_{n=1}^N z_n \otimes z_n,$$

whose convergence is a key step towards Theorem 2.

**Proposition 19.** *Under assumption  $A_3$  denote:*

$$\mathfrak{H} = \frac{c^2}{24} \left( 2c \left( \frac{1-\omega}{1+\omega} \right) \left( 1 + \omega - \frac{c}{2} \right) - \frac{c^2}{6} \right)^{-1},$$

then when  $N$  tends to  $+\infty$

$$\Gamma_{N,z} - \frac{\mathfrak{H}}{N} \sum_{n=1}^N \text{diag} \left( \delta_n^{\odot 2} \right) \rightarrow_{\mathbb{P}} 0,$$

with  $\delta_n = p_n - g_n$ .

The proofs of proposition 18 and 19 are postponed to section 5.1.3. We are now ready to turn to the proof of Theorem 2.

**Proof of Theorem 2:** Starting from:

$$\frac{1}{N} \sum_{n=1}^N x_n - \frac{1}{N} \sum_{n=1}^N \frac{\mathbb{E}p_n + \mathbb{E}g_n}{2} = \frac{1}{N} \sum_{n=1}^N z_n + \frac{1}{N} \sum_{n=1}^N \left( \mathbb{E}x_n - \frac{\mathbb{E}p_n + \mathbb{E}g_n}{2} \right),$$

we see from the proof of Lemma 17 that  $(1/\sqrt{N}) \sum_{n=1}^N [\mathbb{E}x_n - (\mathbb{E}p_n + \mathbb{E}g_n)/2] \rightarrow 0$  as  $N$  tends to infinity. Following Proposition 18, the Theorem 2 comes down to proving:

$$\frac{1}{\sqrt{N}} \sum_{n=1}^N \left[ \frac{\eta_n}{2} \odot (p_n - g_n) - \varepsilon_n \odot z_n \right] \hookrightarrow \mathcal{N}(0, \Gamma).$$

We can first remark that  $u_n = \frac{\eta_n}{2} \odot (p_n - g_n) - \varepsilon_n \odot z_n$  is a vector-valued martingale difference sequence with respect to the filtration  $\mathcal{F}_n^s = \mathcal{F}_n$ . We confine ourselves to 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 ensuring convergence of the conditional covariance structure of  $(1/\sqrt{N}) \sum_{n=1}^N u_n$ , then checking the Lyapunov condition holds (hence the Lindeberg's uniform integrability that ensures uniform tightness of the sequence).

**First step:** The conditional covariance sequence of  $u_n$  is the sequence of matrices defined by  $\frac{1}{N} \sum_{n=1}^N \mathbb{E}(u_n \otimes u_n | \mathcal{F}_n)$ . We show in this first step that this sequence converges in probability

to  $\Gamma$ . We start with elementary calculations:

$$\begin{aligned}
\frac{1}{N} \sum_{n=1}^N \mathbb{E}(u_n \otimes u_n | \mathcal{F}_n) &= \frac{1}{N} \sum_{n=1}^N \mathbb{E} \left( \frac{\eta_n}{2} \odot (p_n - g_n) \otimes \frac{\eta_n}{2} \odot (p_n - g_n) | \mathcal{F}_n \right) \\
&+ \frac{1}{N} \sum_{n=1}^N \mathbb{E}((\varepsilon_n \odot z_n) \otimes (\varepsilon_n \odot z_n) | \mathcal{F}_n) \\
&- \frac{1}{2N} \sum_{n=1}^N \mathbb{E}([\eta_n \odot (p_n - g_n)] \otimes [\varepsilon_n \odot z_n] | \mathcal{F}_n) \\
&- \frac{1}{2N} \sum_{n=1}^N \mathbb{E}([\varepsilon_n \odot z_n] \otimes [\eta_n \odot (p_n - g_n)] | \mathcal{F}_n) \\
&= \frac{\mathbb{E}\varepsilon_1^2}{N} \mathbf{I} \odot \left\{ \sum_{n=1}^N \frac{1}{4} [(p_n - g_n) \otimes (p_n - g_n)] + z_n \otimes z_n \right\} \\
&= \frac{1}{6} \mathbf{I} \odot \left[ \frac{1}{4} \Gamma_{N,p,g} + \Gamma_{N,z} \right],
\end{aligned}$$

where we used Lemma 15 and denoted  $\Gamma_{N,p,g} = \frac{1}{N} \sum_{n=1}^N [(p_n - g_n) \otimes (p_n - g_n)]$ .

By Proposition 19,  $\Gamma_{N,z} - \frac{\mathfrak{H}}{N} \Gamma_{N,p,g}$  tends in probability to zero. Hence whenever the limit in the r.h.s below exists:

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N \mathbb{E}(u_n \otimes u_n | \mathcal{F}_n) = \left( \frac{1}{24} + \frac{\mathfrak{H}}{6} \right) \lim_{N \rightarrow +\infty} \mathbf{I} \odot \Gamma_{N,p,g}.$$

Under assumption  $\mathbf{A}_{1-2}$   $\lim_{N \rightarrow +\infty} \Gamma_{N,p,g} - \bar{\Gamma}_{N,p,g} = 0$  in matrix norm, where

$$\bar{\Gamma}_{N,p,g} = \frac{1}{N} \sum_{n=1}^N \mathbb{E}[(p_n - g_n) \otimes \mathbb{E}(p_n - g_n)].$$

To prove this it suffices to write:

$$\Gamma_{N,p,g} - \bar{\Gamma}_{N,p,g} = \frac{1}{N} \sum_{n=1}^N (\nu_n - \xi_n) \otimes (p_n - g_n) + \frac{1}{N} \sum_{n=1}^N \mathbb{E}(p_n - g_n) \otimes (\nu_n - \xi_n),$$

and apply the bounds derived from the proof of Proposition 19, for instance:

$$\begin{aligned}
\frac{1}{N} \left\| \sum_{n=1}^N (\nu_n - \xi_n) \otimes (p_n - g_n) \right\|_{\infty} &\leq \frac{1}{N} \sum_{n=1}^N \|(\nu_n - \xi_n) \otimes (p_n - g_n)\|_{\infty} \\
&\leq \frac{1}{N} \sum_{n=1}^N \|\nu_n - \xi_n\| \|p_n - g_n\| \\
&\leq \sup_n \|p_n - g_n\| \frac{1}{N} \sum_{n=1}^N (\|\nu_n\| + \|\xi_n\|) \xrightarrow{\mathbb{P}} 0.
\end{aligned}$$

Finally, the convergence of  $\bar{\Gamma}_{N,p,g}$  to  $\Gamma$  is a consequence of the adaptation of Cesaro's Lemma for tensor products.

**Second step:** A Lyapunov condition holds almost trivially here. Namely we are going to prove that:

$$\frac{1}{N^2} \sum_{n=1}^N \mathbb{E} \|u_n\|^4 \xrightarrow{N \rightarrow +\infty} 0.$$

Simple calculations provide:

$$\begin{aligned} \|u_n\|^2 &\leq 2 \left\| \frac{\eta_n}{2} \odot (p_n - g_n) \right\|^2 + 2 \|\varepsilon_n \odot z_n\|^2 \\ &\leq \frac{1}{2} \sup_{n,i} \{(p_{n,i} - g_{n,i})^2\} \|\eta_n\|^2 + 2 \sup_{n,i} \{z_{n,i}^2\} \|\varepsilon_n\|^2 \\ &\leq 2 |\Omega|_\infty^2 \left[ \|\eta_n\|^2 + \|\varepsilon_n\|^2 \right], \end{aligned}$$

with  $|\Omega|_\infty = \sup_{s \in \Omega} |s|$  hence

$$\mathbb{E} \|u_n\|^4 \leq C |\Omega|_\infty^4,$$

for some constant  $C$  then  $(1/N^2) \sum_{n \leq N} \mathbb{E} \|u_n\|^4$  tends to zero when  $N$  tends to infinity. This completes the proof of Theorem 2.

### 5.1.3 Derivations of auxiliary results

**Proof of Proposition 18:** The first equation is obtained by summing from  $n = 1$  to  $N$  both sides of Equation (4). Then we are going to prove that  $\frac{1}{\sqrt{N}} \sum_{n=1}^N \tilde{r}_n \xrightarrow{\mathbb{P}} 0$  by Markov inequality which will be enough to get the rest of the Proposition. We start from  $\tilde{r}_n = \varepsilon_n \left( \frac{\mathbb{E}p_n + \mathbb{E}g_n}{2} - \mathbb{E}x_n \right) + \frac{1+\varepsilon_n}{2} (p_n - \mathbb{E}p_n + g_n - \mathbb{E}g_n)$ . By the first part of assumption **A**<sub>2</sub>

$$\frac{1}{\sqrt{N}} \sum_{n=1}^N (p_n - \mathbb{E}p_n) \xrightarrow{\mathbb{P}} 0 \quad \frac{1}{\sqrt{N}} \sum_{n=1}^N (g_n - \mathbb{E}g_n) \xrightarrow{\mathbb{P}} 0.$$

We turn to:

$$\mathbb{E} \left\| \sum_{n=1}^N \varepsilon_n \left( \frac{\mathbb{E}p_n + \mathbb{E}g_n}{2} - \mathbb{E}x_n \right) \right\|^2 = \mathbb{E} \varepsilon_1^2 \sum_{n=1}^N \left\| \frac{\mathbb{E}p_n + \mathbb{E}g_n}{2} - \mathbb{E}x_n \right\|^2$$

because  $\varepsilon_n$  is an i.i.d. centered sequence. The proposition now follows from Lemma 17. This completes the proof of the Proposition.

**Proof of the Proposition 19:** We take advantage of Equation (4) and note that

$$\begin{aligned} (\varepsilon_n \odot z_n) \otimes (\varepsilon_n \odot z_n) &= (\varepsilon_n \otimes \varepsilon_n) \odot (z_n \otimes z_n) \\ &= \left[ (\varepsilon_n \otimes \varepsilon_n) - \sigma_\varepsilon^2 \mathbf{I} \right] \odot (z_n \otimes z_n) + \sigma_\varepsilon^2 \mathbf{I} \odot (z_n \otimes z_n), \end{aligned}$$

where  $\sigma_\varepsilon^2 = 1/6$  by Lemma 15 and  $\mathbf{I}$  stands for identity matrix. After some tedious calculations we obtain:

$$\begin{aligned} z_{n+1} \otimes z_{n+1} &= (1 + \omega - c)^2 (z_n \otimes z_n) + \omega^2 z_{n-1} \otimes z_{n-1} \\ &\quad + c^2 \sigma_\varepsilon^2 [\mathbf{I} \odot (z_n \otimes z_n)] + \frac{c^2}{4} [\eta_n \odot (p_n - g_n)] \otimes [\eta_n \odot (p_n - g_n)] \\ &\quad - \omega (1 + \omega - c) [z_n \otimes z_{n-1} + z_{n-1} \otimes z_n] + \mathcal{C}_n. \end{aligned} \tag{7}$$

For the sake of completeness, we list now all the eleven terms contained in  $\mathcal{C}_n$ . In order to simplify notations let  $[[u : v]] = u \otimes v + v \otimes u$ . Notice for further use:

$$\|[[u : v]]\|_\infty \leq 2 \|u\| \|v\|.$$

Then:

$$\begin{aligned} \mathcal{C}_n = & c(1 + \omega - c) \left\{ [[z_n : \tilde{r}_n]] - [[z_n : \varepsilon_n \odot z_n]] + \frac{1}{2} [[z_n : \eta_n \odot (p_n - g_n)]] \right\} \\ & + \omega c \left\{ [[z_{n-1} : \varepsilon_n \odot z_n]] - [[z_{n-1} : \tilde{r}_n]] - \frac{1}{2} [[z_{n-1} : \eta_n \odot (p_n - g_n)]] \right\} \\ & - c^2 \left\{ [[c\varepsilon_n \odot z_n : \tilde{r}_n]] + \frac{1}{2} [[\varepsilon_n \odot z_n : \eta_n \odot (p_n - g_n)]] - \frac{1}{2} [[\tilde{r}_n : \eta_n \odot (p_n - g_n)]] \right\} \\ & + c^2 \tilde{r}_n \otimes \tilde{r}_n + c^2 [(\varepsilon_n \otimes \varepsilon_n) - \sigma_\varepsilon^2 \mathbf{I}] \odot (z_n \otimes z_n). \end{aligned} \quad (8)$$

From 7 we sum over all indices  $n$  in  $\{1, \dots, N\}$  we get:

$$\begin{aligned} (1 - (1 + \omega - c)^2 - \omega^2) \sum_{n=1}^N z_n \otimes z_n = & c^2 \sigma_\varepsilon^2 \sum_{n=1}^N [\mathbf{I} \odot (z_n \otimes z_n)] \\ & + \frac{c^2}{4} \sum_{n=1}^N [\eta_n \odot (p_n - g_n)] \otimes [\eta_n \odot (p_n - g_n)] \\ & - \omega(1 + \omega - c) \sum_{n=1}^N [z_n \otimes z_{n-1} + z_{n-1} \otimes z_n] \\ & + \sum_{n=1}^N \mathcal{C}_n + z_1 \otimes z_1 - z_{N+1} \otimes z_{N+1} \\ & + \omega^2 (z_0 \otimes z_0 - z_N \otimes z_N). \end{aligned}$$

Now we need to go slightly further in the computations and to find a recurrent equation for  $z_n \otimes z_{n-1}$ . Let us start again from (4):

$$\begin{aligned} z_{n+1} \otimes z_n = & (1 + \omega - c) z_n \otimes z_n - \omega z_{n-1} \otimes z_n - c(\varepsilon_n \odot z_n) \otimes z_n + c\tilde{r}_n \otimes z_n \\ & + \frac{c}{2} [\eta_n \odot (p_n - g_n)] \otimes z_n, \\ z_n \otimes z_{n+1} = & (1 + \omega - c) z_n \otimes z_n - \omega z_n \otimes z_{n-1} - cz_n \otimes (\varepsilon_n \odot z_n) + cz_n \otimes \tilde{r}_n \\ & + \frac{c}{2} z_n \otimes [\eta_n \odot (p_n - g_n)]. \end{aligned}$$

Summing over all indices  $n$  in  $\{1, \dots, N\}$  above we come to

$$\begin{aligned} (1 + \omega) \sum_{n=1}^N [z_n \otimes z_{n-1} + z_{n-1} \otimes z_n] \\ = & z_1 \otimes z_0 + z_0 \otimes z_1 - z_N \otimes z_{N+1} - z_{N+1} \otimes z_N + 2(1 + \omega - c) \sum_{n=1}^N [z_n \otimes z_n] \\ & + c \sum_{n=1}^N \left\{ [[\tilde{r}_n : z_n]] - [[(\varepsilon_n \odot z_n) : z_n]] + \frac{1}{2} [[\eta_n \odot (p_n - g_n) : z_n]] \right\}. \end{aligned}$$

Plugging the last equation in line (7) we get finally:

$$\begin{aligned}
& \left(1 - (1 + \omega - c)^2 - \omega^2\right) \sum_{n=1}^N z_n \otimes z_n \\
&= c^2 \sigma_\varepsilon^2 \sum_{n=1}^N [\mathbf{I} \odot (z_n \otimes z_n)] + \frac{c^2}{4} \sum_{n=1}^N [\eta_n \odot (p_n - g_n)] \otimes [\eta_n \odot (p_n - g_n)] \\
&\quad - 2\omega \frac{(1 + \omega - c)^2}{1 + \omega} \sum_{n=1}^N [z_n \otimes z_n] + \omega c \frac{(1 + \omega - c)}{1 + \omega} \sum_{n=1}^N [[(\varepsilon_n \odot z_n) : z_n]] \\
&\quad - \omega c \frac{(1 + \omega - c)}{1 + \omega} \sum_{n=1}^N [[\tilde{r}_n : z_n]] - \omega c \frac{(1 + \omega - c)}{(1 + \omega)} \sum_{n=1}^N \frac{1}{2} [[\eta_n \odot (p_n - g_n) : z_n]] + \sum_{n=1}^N \mathcal{C}_n.
\end{aligned}$$

Denoting  $\kappa = \left(1 - (1 + \omega - c)^2 - \omega^2 + 2\omega \frac{(1 + \omega - c)^2}{1 + \omega}\right) = 2c \left(\frac{1 - \omega}{1 + \omega}\right) (1 + \omega - \frac{c}{2})$ :

$$\kappa \sum_{n=1}^N z_n \otimes z_n = \frac{c^2}{6} \sum_{n=1}^N [\mathbf{I} \odot (z_n \otimes z_n)] + \frac{c^2}{4} \sum_{n=1}^N [\eta_n \odot (p_n - g_n)] \otimes [\eta_n \odot (p_n - g_n)] + \mathfrak{R}_N,$$

with

$$\begin{aligned}
\mathfrak{R}_N &= -\omega \frac{(1 + \omega - c)}{1 + \omega} ([z_1 : z_0] - [z_N : z_{N+1}]) + z_1 \otimes z_1 - z_{N+1} \otimes z_{N+1} \\
&\quad + \omega^2 (z_0 \otimes z_0 - z_N \otimes z_N) + \sum_{n=1}^N \mathcal{C}_n \\
&\quad + \omega c \frac{(1 + \omega - c)}{1 + \omega} \sum_{n=1}^N \left\{ [[(\varepsilon_n \odot z_n) : z_n]] - [[\tilde{r}_n : z_n]] - \frac{1}{2} [[\eta_n \odot (p_n - g_n) : z_n]] \right\}.
\end{aligned} \tag{9}$$

It is proved in Lemma 20 that  $\mathfrak{R}_N/N \rightarrow_{\mathbb{P}} 0$ . First, let us unravel the matrix:

$$\mathbf{T}_N = \frac{\kappa}{N} \sum_{n=1}^N z_n \otimes z_n - \frac{c^2}{6} \frac{1}{N} \sum_{n=1}^N [\mathbf{I} \odot (z_n \otimes z_n)].$$

Denote  $[\mathbf{T}_N]_{ij}$  the  $(i, j)$  cell of matrix  $\mathbf{T}_n$ . It is simple to see that:

$$[\mathbf{T}_N]_{ii} = \left(\kappa - \frac{c^2}{6}\right) \left[\frac{1}{N} \sum_{n=1}^N (z_n \otimes z_n)\right]_{ii}, \quad [\mathbf{T}_N]_{ij} = \kappa \left[\frac{1}{N} \sum_{n=1}^N z_n \otimes z_n\right]_{i,j} \tag{10}$$

for  $i \neq j$ . Now denote:

$$\begin{aligned}
\Lambda_N &= \frac{1}{N} \sum_{n=1}^N [\eta_n \odot (p_n - g_n)] \otimes [\eta_n \odot (p_n - g_n)] \\
&= \frac{1}{N} \sum_{n=1}^N (\eta_n \otimes \eta_n) \odot [(p_n - g_n) \otimes (p_n - g_n)].
\end{aligned}$$

Taking conditional expectation w.r.t  $\mathcal{F}_n$  we get -all non diagonal term vanish:

$$\begin{aligned}
\mathbb{E}(\Lambda_N | \mathcal{F}_n) &= \frac{1}{N} \sum_{n=1}^N [\mathbb{E}(\eta_n \otimes \eta_n)] \odot [(p_n - g_n) \otimes (p_n - g_n)] \\
&= \frac{\mathbb{E}\eta_1^2}{N} \sum_{n=1}^N \mathbf{I} \odot [(p_n - g_n) \otimes (p_n - g_n)],
\end{aligned}$$

with  $\mathbf{I}$  the identity matrix. Noting that the difference:

$$\Lambda_N - \mathbb{E}(\Lambda_N | \mathcal{F}_N) = \frac{1}{N} \sum_{n=1}^N \left[ (\eta_n \otimes \eta_n) - (\mathbb{E}\eta_n^2) \mathbf{I} \right] \odot [(p_n - g_n) \otimes (p_n - g_n)]$$

vanishes when  $N$  tends to infinity by applying Lemma 16 we get with Landau notation in probability:

$$\mathbf{T}_N = \frac{c^2}{24N} \sum_{n=1}^N \mathbf{I} \odot [(p_n - g_n) \otimes (p_n - g_n)] + o_P\left(\frac{1}{N}\right).$$

From (10) we obtain simultaneously that  $[\mathbf{T}_N]_{ij} \rightarrow_{\mathbb{P}} 0$  for  $i \neq j$  and

$$[\mathbf{T}_N]_{ii} - \frac{c^2}{24\left(\kappa - \frac{c^2}{6}\right)} \frac{1}{N} \sum_{n=1}^N \mathbf{I} \odot [(p_n - g_n) \otimes (p_n - g_n)] \rightarrow_{\mathbb{P}} 0$$

which is precisely the statement of Proposition 19.

**Lemma 20.**  $\mathfrak{A}_N/N \rightarrow_{\mathbb{P}} 0$  in matrix norm.

**Proof:** Let us start from (9). First by assumption  $\mathbf{A}_1$ :

$$\frac{1}{N} ([z_1 : z_0] - [z_N : z_{N+1}]) + z_1 \otimes z_1 - z_{N+1} \otimes z_{N+1} + \omega^2 (z_0 \otimes z_0 - z_N \otimes z_N) \rightarrow_{\mathbb{P}} 0.$$

It must be also noticed that:

$$\omega c \frac{(1 + \omega - c)}{1 + \omega} \sum_{n=1}^N \left\{ [(\varepsilon_n \odot z_n) : z_n] - [\tilde{r}_n : z_n] - \frac{1}{2} [(\eta_n \odot (p_n - g_n)) : z_n] \right\}$$

involves three terms between double brackets that already appear in  $\mathcal{C}_n$ , up to constants. Proving that each term of  $(1/N) \sum_{n=1}^N \mathcal{C}_n$  tends in probability to 0 will be sufficient to complete the proof of the Lemma.

Our aim is to perform successive applications of Lemma 16. For the sake of completeness we remind now the eleven terms contained in  $\mathcal{C}_n$  and mentioned earlier.

$$\begin{aligned} \mathcal{C}_n = & c(1 + \omega - c) \left\{ [z_n : \tilde{r}_n] - [z_n : \varepsilon_n \odot z_n] + \frac{1}{2} [z_n : \eta_n \odot (p_n - g_n)] \right\} \\ & + \omega c \left\{ [z_{n-1} : \varepsilon_n \odot z_n] - [z_{n-1} : \tilde{r}_n] - \frac{1}{2} [z_{n-1} : \eta_n \odot (p_n - g_n)] \right\} \\ & - c^2 \left\{ [c\varepsilon_n \odot z_n : \tilde{r}_n] + \frac{1}{2} [(\varepsilon_n \odot z_n : \eta_n \odot (p_n - g_n))] - \frac{1}{2} [\tilde{r}_n : \eta_n \odot (p_n - g_n)] \right\} \\ & + c^2 \tilde{r}_n \otimes \tilde{r}_n + c^2 [(\varepsilon_n \otimes \varepsilon_n) - \sigma_\varepsilon^2 \mathbf{I}] \odot (z_n \otimes z_n). \end{aligned}$$

In the list above terms numbered 2, 3, 4, 6, 11 vanish by applying successively (2) and Lemma 16. Take for instance term 1, get rid of the constants and focus on the first tensor product in the double bracket namely:

$$z_n \otimes (\varepsilon_n \odot z_n) = (e \odot z_n) \otimes (\varepsilon_n \odot z_n) = (e \otimes \varepsilon_n) \odot (z_n \otimes z_n).$$

Recall that  $e$  is the vector with all components valued at 1. Then we can apply Lemma 16 which shows that  $(1/N) \sum_{n=1}^N (e \otimes \varepsilon_n) \odot (z_n \otimes z_n)$  vanishes in probability. It is not hard to see that terms 3, 4, 6 and 11 may be treated the same way.

Term number 8 namely depends on both  $\varepsilon_n$  and  $\eta_n$ . An application of (2) leads to considering, up to a constant and commutation to:

$$(\eta_n \otimes \varepsilon_n) \odot (z_n \otimes (p_n - g_n)).$$

The reader can check that the matrix  $E_n = (\eta_n \otimes \varepsilon_n)$  is stochastically independent from  $z_n \otimes (p_n - g_n)$ , that the sequence of matrices  $E_n$  are independent and centered. Lemma 16 applies here.

The terms numbered 1, 5, 7, 9, 10 depend on  $\tilde{r}_n$ . Now consider

$$\tilde{r}_n = \varepsilon_n \odot \left( \frac{\mathbb{E}p_n + \mathbb{E}g_n}{2} - \mathbb{E}x_n \right) + \frac{e + \varepsilon_n}{2} \odot (\xi_n + \nu_n) = r_{n1} + r_{n2},$$

and the five terms containing them in  $\mathcal{C}_n$ . We split  $\tilde{r}_n$  and deal separately with  $r_{n1}$  and  $r_{n2}$  to derive the needed bounds in probability.

First, let us focus on  $r_{n1} = \varepsilon_n \odot \left( \frac{\mathbb{E}p_n + \mathbb{E}g_n}{2} - \mathbb{E}x_n \right)$  only. Once again Lemma 16 may be invoked for the specific halves of terms 1, 5 and 9 by using the same methods as above. The half part of terms 7 and 10 of (8) containing  $\varepsilon_n$  may be bounded the following way (denote  $\beta_n = \frac{\mathbb{E}p_n + \mathbb{E}g_n}{2} - \mathbb{E}x_n$ ):

$$\begin{aligned} \|[\varepsilon_n \odot z_n : \varepsilon_n \odot \beta_n]\|_\infty &\leq 2 \|z_n\| \|\beta_n\|, \\ \|(\varepsilon_n \odot \beta_n) \otimes (\varepsilon_n \odot \beta_n)\|_\infty &\leq 2 \|\beta_n\|^2. \end{aligned}$$

Lemma 17 ensures that both right hand side above tend to zero (since  $\sup_n \|z_n\|$  is almost surely bounded) and Cesaro's Lemma terminates this part.

We should now inspect the terms numbered 1, 5, 7, 9, 10 in (8) with respect to  $r_{n2} = \frac{e + \varepsilon_n}{2} \odot (\xi_n + \nu_n)$ . Terms 7 and 9 may be controlled by Lemma 17 and Lemma 16 respectively. Let us deal with term 1:

$$\begin{aligned} \left\| \sum_{n=1}^N [[z_n : r_{n2}]] \right\|_\infty &\leq 2 \sum_{n=1}^N \|z_n\| \|r_{n2}\| \\ &\leq 2 \left( \sup_n \|z_n\| \right) \sum_{n=1}^N \|r_{n2}\| \\ &\leq 2 \sup_n \|z_n\| \sum_{n=1}^N [\|\nu_n\| + \|\xi_n\|]. \end{aligned}$$

Take for instance  $\sum_{n=1}^N \|\nu_n\|$ . Applying Markov inequality:

$$\mathbb{P} \left( \frac{1}{N} \sum_{n=1}^N \|\nu_n\| > \varepsilon \right) \leq \frac{1}{N} \sum_{n=1}^N \mathbb{E} \|\nu_n\|,$$

then Assumption **A**<sub>2</sub> together with Cesaro's Lemma again ensure that  $\frac{1}{N} \sum_{n=1}^N \|\nu_n\| \rightarrow_{\mathbb{P}} 0$  as well as  $\frac{1}{N} \sum_{n=1}^N \|\xi_n\|$  hence  $\frac{1}{N} \left\| \sum_{n=1}^N [[z_n : r_{n2}]] \right\|_\infty$ . Term 5 vanishes in probability with the same technique at hand. Now term 10 gives:

$$\left\| \frac{1}{N} \sum_{n=1}^N r_{2n} \otimes r_{2n} \right\|_\infty \leq \frac{1}{N} \sum_{n=1}^N \|r_{2n} \otimes r_{2n}\|_\infty = \frac{1}{N} \sum_{n=1}^N \|r_{2n}\|^2$$



which also tends to zero in probability under  $\mathbf{A}_{1-2}$ . By the way we mention that the cross-product  $[[r_{1n} : r_{2n}]]$  in term 10 vanishes due to Lemma 17. Our task is almost done. Let us deal now with the remaining terms of  $\mathfrak{R}_N$ . Both:

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N [(\varepsilon_n \odot z_n) \otimes z_n + z_n \otimes (\varepsilon_n \odot z_n)], \\ & \frac{1}{N} \sum_{n=1}^N [\eta_n \odot (p_n - g_n)] \otimes z_n + z_n \otimes [\eta_n \odot (p_n - g_n)] \end{aligned}$$

tends to 0 in probability by Lemma 16 and 2. At last, the remaining  $[[\tilde{r}_n : z_n]]$  is basically the same as term 1 in (8) and may be addressed the same way. This terminates the proof of the Lemma.

## 5.2 Second case: non-oscillatory

### 5.3 Risk

#### Proof of the Proposition 5:

The first fact  $\mathbb{E}\mathbf{y}_n = 0$  is plain since the random vector  $(x_0, x_1)$  is independent from all the  $\varepsilon_n$ 's. Basic computations give:

$$\begin{aligned} \text{cov}(\mathbf{y}_n) &= \sigma^2 \mathbb{E}[\mathbf{S}_n \mathbf{S}_{n-1} \dots \mathbf{S}_2 \mathbf{S}_2^* \dots \mathbf{S}_n^*] \\ &= \sigma^2 \mathbb{E}\{\mathbf{S}_n \dots \mathbb{E}[\mathbf{S}_3 \mathbb{E}(\mathbf{S}_2 \mathbf{S}_2^*) \mathbf{S}_3] \dots \mathbf{S}_n^*\} \end{aligned}$$

Denote

$$\text{cov}(\mathbf{y}_n) = \begin{bmatrix} A_n & B_n \\ B_n & D_n \end{bmatrix}$$

then

$$\text{cov}(\mathbf{y}_{n+1}) = \mathbb{E}[\mathbf{S}_{n+1} \text{cov}(\mathbf{y}_n) \mathbf{S}_{n+1}^*] = \begin{bmatrix} A_{n+1} & B_{n+1} \\ B_{n+1} & D_{n+1} \end{bmatrix}.$$

From which we immediately get the following linear recurrent equations (denote  $\gamma = \sqrt{(\lambda^2 + \frac{c^2}{6})}$ ):

$$\begin{cases} A_{n+1} = \gamma^2 A_n - 2\lambda\omega B_n + \omega^2 D_n \\ B_{n+1} = \lambda A_n - \omega B_n \\ D_{n+1} = A_n \end{cases}$$

who are driven by the matrix:

$$M = \begin{pmatrix} \gamma^2 & -2\lambda\omega & \omega^2 \\ \lambda & -\omega & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

We show below that the largest eigenvalue of  $M$  in modulus is bounded by  $r^*$  which will yield  $\max\{|A_n|, |B_n|, |D_n|\} \leq c \cdot (r^*)^n$  hence the desired result. It takes two steps. First, we provide a sufficient condition for  $\mathbf{P}_M$  the characteristic polynomial of  $M$  to have a single real eigenvalue. Second we ensure that this real eigenvalue is the largest of three and strictly below 1.

Simple calculations give:

$$\mathbf{P}_M(x) = \det(M - x\mathbf{I}) = -x^3 + \left(\lambda^2 - \omega + \frac{c^2}{6}\right)x^2 + \omega\left(\omega - \lambda^2 + \frac{c^2}{6}\right)x + \omega^3.$$

Denote  $\beta = \omega - \lambda^2$ . When  $\mathbf{B}_1$  within the Proposition holds,  $\beta > c^2/6$  and the discriminant of  $\mathbf{P}_M$  is:

$$\begin{aligned}\Delta_{\mathbf{P}_M} &= \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4 + \Delta_5, \\ \Delta_1 &= \omega^2 (\beta + c^2/6)^2 (-\beta + c^2/6)^2, \quad \Delta_2 = -18\omega^4 (\beta + c^2/6) (-\beta + c^2/6), \\ \Delta_3 &= -27\omega^6, \quad \Delta_4 = 4\omega^3 (\beta + c^2/6)^3, \quad \Delta_5 = -4\omega^3 (-\beta + c^2/6)^3.\end{aligned}$$

Clearly  $\Delta_1, \Delta_2, \Delta_4$  and  $\Delta_5$  are positive whereas  $\Delta_3$  is negative. A sufficient condition for a single real root for  $\mathbf{P}_M$  is that  $\Delta_{\mathbf{P}_M} < 0$  hence

$$\Delta_1 + \Delta_2 + \Delta_4 + \Delta_5 < -\Delta_3 \tag{11}$$

From

$$\begin{aligned}\Delta_1 &= \omega^2 (\beta^2 - c^4/36)^2, \quad \Delta_2 = 18\omega^4 (\beta^2 - c^4/36) \\ \Delta_4 + \Delta_5 &= 4\omega^3 \left[ (\beta + c^2/6)^3 + (\beta - c^2/6)^3 \right] = 8\beta^3\omega^3 \left[ 1 + \frac{c^4}{12\beta^2} \right],\end{aligned}$$

we compute:

$$\Delta_1 + \Delta_2 + \Delta_4 + \Delta_5 = \omega^2\beta^4 \left( 1 - \frac{c^4}{36\beta^2} \right)^2 + 18\omega^4\beta^2 \left( 1 - \frac{c^4}{36\beta^2} \right) + 8\beta^3\omega^3 \left( 1 + \frac{c^4}{12\beta^2} \right).$$

Hence

$$\begin{aligned}\frac{\Delta_1 + \Delta_2 + \Delta_4 + \Delta_5}{\omega^6} &= \left( \frac{\beta}{\omega} \right)^4 \left( 1 - \frac{c^4}{36\beta^2} \right)^2 + 18 \left( \frac{\beta}{\omega} \right)^2 \left( 1 - \frac{c^4}{36\beta^2} \right) \\ &\quad + 8 \left( \frac{\beta}{\omega} \right)^3 \left( 1 + \frac{c^4}{12\beta^2} \right),\end{aligned}$$

with  $\beta/\omega = 1 - \lambda^2/\omega$ . We just need a condition under which the right hand side in the equation above is upper-bounded by 27.

Denote  $u = \lambda^2/\omega$  and  $h = c^4/(36\beta^2)$  then under  $\mathbf{B}_1$  both  $h < 1$  and  $u < 1$  and:

$$\begin{aligned}\frac{\Delta_1 + \Delta_2 + \Delta_4 + \Delta_5}{\omega^6} &\leq (1 - u) (27 + \Lambda), \\ \Lambda &= 4h - 9u - 22uh + h^2.\end{aligned}$$

If  $\Lambda < 0$  then we are done. Suppose not, then:

$$\begin{aligned}\frac{\Delta_1 + \Delta_2 + \Delta_4 + \Delta_5}{\omega^6} &\leq 27 - 27u + \Lambda \\ -27u + \Lambda &= -36u + 4h - 22uh + h^2\end{aligned}$$

is negative whenever

$$h^2 + 4h < 36u + 22uh.$$

At this point, we could propose maybe optimal but intricate solutions by considering in the previous equations an order 2 polynomial in  $h$ . We prefer to go through a simpler though

weaker way. Assume that  $4h < 36u$  then it is easy to see that  $h^2 < 22uh$  and the condition rewrites:

$$\frac{c^4}{36\beta^2} < 9\frac{\lambda^2}{\omega} \iff \frac{c^2}{6} < 3\frac{|\lambda|}{\sqrt{\omega}}\beta$$

Now we know that  $\mathbf{P}_M$  has a single real root say  $r^*$ . We have to ensure that the three roots of  $\mathbf{P}_M$  have moduli in accordance with (6). Remind that  $z$  and  $\bar{z}$  are the two complex conjugate roots. We know that  $r^*|z|^2 = \omega^3$  hence that  $r^* > 0$ . Consequently we just have to check the two following facts in order to complete the proof of the Proposition. First that  $r^* < 1$ . It suffices to check that  $\mathbf{P}_M(1) < 0$ . Second that  $|z| < r^*$ . It is not hard to see that we may confine ourselves to proving that  $\mathbf{P}_M(\omega) = c^2/3$ . This means that  $\omega < r^*$  and from  $r^*|z|^2 = \omega^3$ ,  $|z| < \omega < r^*$ . Then we turn to:

$$\begin{aligned} \mathbf{P}_M(1) &= -1 + (\lambda^2 - \omega + c^2/6) + \omega(\omega - \lambda^2 + c^2/6) + \omega^3 \\ &= -1 + \lambda^2 - \omega + c^2/6 + \omega^2 - \omega\lambda^2 + \omega c^2/6 + \omega^3. \end{aligned}$$

The latter will be negative when:

$$\begin{aligned} \frac{c^2}{6}(1 + \omega) &< 1 - \lambda^2 + \omega - \omega^2 + \omega\lambda^2 - \omega^3 \\ &= (1 + \omega)(1 + \lambda^2) - (1 + \omega)\omega^2 - 2\lambda^2 \end{aligned}$$

which gives finally in accordance with  $\mathbf{B}_1$ :

$$\frac{c^2}{6} < 1 + \lambda^2 - \omega^2 - 2\frac{\lambda^2}{1 + \omega}.$$

This terminates the proof of the Proposition 5.

## 5.4 Weak Convergence

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  $\mathcal{E}_x = [1 + \omega - 2c - \frac{\omega}{x}, 1 + \omega - \frac{\omega}{x}]$ . The Theorem 8 is a consequence of the two Lemmas below coupled with the above-mentioned Theorem 1 p.302 in Jones [2004].

**Lemma 21.** *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 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  $\varepsilon_n$  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  $1 + \omega - 2c + \tau + c\varepsilon_n$  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 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  $x$ :

$$\int_{\mathbb{R}} g(t) P(t, x) dt \leq \rho_1 g(x) + \rho_2 1_{x \in \mathcal{C}}, \quad (12)$$

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  $x$ :

$$[\log |x|]^2 \leq g(x). \quad (13)$$

Note however that, if (12) holds for  $g$  but (13) fails, then both conditions will hold for updated function  $g^* = \eta g$  with constant  $\eta > 1$  and  $\rho'_2 = \eta \rho_2$  such that (13) holds.

**Lemma 22.** *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:*

$$\mathbf{B}_2 : \lambda < \omega/c < (1 + c)/4.$$

*Then it is always possible to choose three constants  $\tau$ ,  $C_1$  and  $C_2$  such that (12) holds for a specific choice of  $\rho_1$  and  $\rho_2$ .*

**Proof:**

The proof of the Lemma just consists in an explicit construction of the above-mentioned  $\tau$ ,  $C_1$ , 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  $\varepsilon_n$  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

$$C_2 K_1(\omega, c, \tau) \leq \rho_1 C_1 \tag{14}$$

and  $\rho_1$  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  $\tau$ :  $\{\tau \geq c - \omega\} \cup \{\tau \leq c - 1\}$ . We keep the second and assume once and for all that:

$$\tau \leq c - 1. \tag{15}$$

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

$$\begin{aligned} \frac{1}{2c} \int_{(\omega/x)-\lambda-c}^{(\omega/x)-\lambda+c} g(s) ds &= \frac{1}{2c} \int_{(\omega/x)-\lambda-c}^{M_\tau} g(s) ds + \frac{1}{2c} \int_{M_\tau}^{(\omega/x)-\lambda+c} g(s) ds \\ &\equiv \mathcal{I}_1 + \mathcal{I}_2. \end{aligned} \tag{16}$$

We want to make sure that the upper bound  $(\omega/x) - \lambda + c$  is larger than  $M_\tau$ . 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 (15) consequently (16) holds.

Focus on the first term  $\mathcal{I}_1$  in (16) 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 hand 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:

$$\frac{1}{c} \left[ \sqrt{|M_{\tau^+}(1+\omega) - \omega|} + M_{\tau^+} \right] = 1 - 2\delta < 1 \quad (17)$$

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**<sub>2</sub>) 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 (16):

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

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 (16) 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:

$$C_2 K_2(\omega, c, \tau) \leq C_1 \delta. \quad (18)$$

Finally comparing (14) and (18), we see that both conditions cannot be incompatible. Accurate choices of the couple  $(C_1^+, C_2^+)$  are given by the summary bound:

$$C_2^+ \leq C_1^+ \min \left( \frac{\delta}{K_2}, \frac{1-\delta}{K_1} \right). \quad (19)$$

It is now basic to see that the quadruple  $(C_1^+, C_2^+, \tau^+, \rho_1^+)$  yields the drift condition (12) for  $x \notin \mathcal{C}$ .

*Third step:* The remaining step is to check the inequality for some  $\rho_2$ :

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

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 (12) 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  $\tau$ . 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:

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

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  $\rho_1$  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:

$$\tau \leq \min\left(\lambda, c - \frac{\omega}{2c}\right) \quad (20)$$

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 (20) 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 re:  $K_1(\omega, c, \tau^-) = K_1^-$ ). The new condition:

$$C_2^- \leq C_1^- \min\left(\frac{\rho_1^-}{K_1^-}, \frac{1 - 1/c}{2K_2^-}\right) \quad (21)$$

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  $\rho_2^-$ .

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 (19) and (21) 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.

## References

- Benoist, Y. and J.-F. Quint (2016, 03). Central limit theorem for linear groups. *Ann. Probab.* 44(2), 1308–1340.
- Berger, M. A. (1984). Central limit theorem for products of random matrices. *Transactions of the American Mathematical Society* 285(2), 777–803.
- Bonyadi, M. R. and Z. Michalewicz (2016). Stability analysis of the particle swarm optimization without stagnation assumption. *IEEE Transactions on Evolutionary Computation* 20(5), 814–819.
- Cleghorn, C. W. and A. P. Engelbrecht (2018). Particle swarm stability: a theoretical extension using the non-stagnate distribution assumption. *Swarm Intelligence* 12(1), 1–22.
- Clerc, M. and J. Kennedy (2002). The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE transactions on Evolutionary Computation* 6(1), 58–73.
- Duflo, M. (1997). *Random Iterative Models* (1st ed.). Berlin, Heidelberg: Springer-Verlag.
- Eberhart, R. and J. Kennedy (1995). A new optimizer using particle swarm theory. In *Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on*, pp. 39–43. IEEE.
- Furstenberg, H. and H. Kesten (1960, 06). Products of random matrices. *Ann. Math. Statist.* 31(2), 457–469.
- 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(299-320), 5–1.

- Kadirkamanathan, V., K. Selvarajah, and P. J. Fleming (2006). Stability analysis of the particle dynamics in particle swarm optimizer. *IEEE Transactions on Evolutionary Computation* 10(3), 245–255.
- Lane, J., A. Engelbrecht, and J. Gain (2008). Particle swarm optimization with spatially meaningful neighbours. In *Swarm Intelligence Symposium, 2008. SIS 2008. IEEE*, pp. 1–8. IEEE.
- Le Page, É. (1982). Théoremes limites pour les produits de matrices aléatoires. In *Probability measures on groups*, pp. 258–303. Springer.
- Massart, P. (2007). *Concentration inequalities and model selection: Ecole d’Eté de Probabilités de Saint-Flour XXXIII - 2003*. Number n° 1896 in Ecole d’Eté de Probabilités de Saint-Flour. Springer-Verlag.
- Meyn, S. P. and R. L. Tweedie (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(4), 712–721.
- Schmitt, M. and R. Wanka (2015). Particle swarm optimization almost surely finds local optima. *Theoretical Computer Science* 561, 57–72.
- Van Den Bergh, F. and A. P. Engelbrecht (2006). A study of particle swarm optimization particle trajectories. *Information sciences* 176(8), 937–971.
- Van den Bergh, F. and A. P. Engelbrecht (2010). A convergence proof for the particle swarm optimiser. *Fundamenta Informaticae* 105(4), 341–374.
- Zhang, Y., S. Wang, and G. Ji (2015). A comprehensive survey on particle swarm optimization algorithm and its applications. *Mathematical Problems in Engineering* 2015.