Particle Swarm Stability: A Theoretical Extension using the Non-Stagnate Distribution Assumption

Christopher W. Cleghorn ·
Andries P. Engelbrecht

Abstract This paper presents an extension of the state of the art theoretical model utilized for understanding the stability criteria of the particles in particle swarm optimization algorithms. Conditions for order-1 and order-2 stability are derived by modeling, in the simplest case, the expected value and variance of a particle's personal and neighborhood best positions as convergent sequences of random variables. Furthermore, the condition that the expected value and variance of a particle's personal and neighborhood best positions are convergent sequences is shown to be a necessary condition for order-1 and order-2 stability. The theoretical analysis presented is applicable to a large class of particle swarm optimization variants.

Keywords Particle Swarm Optimization · Stability Analysis · Stability Criteria

1 Introduction

Particle swarm optimization (PSO), originally developed by Kennedy and Eberhart (1995), has become a widely used optimization technique (Poli, 2008a). Given PSO's success, a substantial amount of theoretical work has been performed on the stochastic search algorithm to try and predict and understand its underlying behavior (Ozcan and Mohan, 1998; Clerc and Kennedy, 2002; Jiang et al., 2007; García-Gonzalo and Fernández-Martinez, 2014a; Cleghorn and Engelbrecht, 2014; Liu, 2015).

Almost all theoretical research performed has to some extent relied upon the stagnation assumption, whereby the personal and neighborhood best positions of a particle are assumed to be fixed, which is not a true reflection of the behavior of PSO algorithms. Recently, it was shown by Bonyadi and Michalewicz (2016b) that

Christopher W. Cleghorn
Department of Computer Science, University of Pretoria
E-mail: ccleghorn@cs.up.ac.za

Andries P. Engelbrecht
Department of Computer Science, University of Pretoria Tel.: +27 12 420 3578
E-mail: engel@cs.up.ac.za

if it is assumed that the personal and neighborhood best positions are random variables with well defined expectations and variances, criteria for order-1 and order-2 stability could be derived.

In this paper, criteria for order-1 and order-2 stability are derived under a weaker assumption: It is assumed, in the simplest case where there are only two particle informers, that the personal and neighborhood best positions' distributions are allowed to vary over time, provided that their expectation and variance are in fact convergent. Future personal and neighborhood best positions are affected by information obtained during the search process, implying that the distribution from which they are sampled should be time dependent. Therefore, allowing the distributions of the personal and neighborhood best positions to change over time is more representative of a PSO's real run time behavior. It is also shown that the assumption used in this paper for the derivation of the order-1 and order-2 stable regions is in fact the necessary condition. The theoretical model presented in this paper is currently a more representative model of the PSO than the model utilized in previous theoretical studies.

In order to provide the most general result possible, rather than focusing on just the original or the canonical PSO (CPSO), a large class of PSO variants are considered. In the general case, all positional memory, such as the personal and neighborhood best positions, are modeled as sequences of random variables. This generality implies that variants such as the fully informed PSO (Kennedy and Mendes, 2003) and the unified PSO (Parsopoulos and Vrahatis, 2004), to name a few, are automatically included in the theoretical model presented in this paper. Therefore, the theoretical model presented in this paper has further utility above its application to the canonical PSO.

It has been shown that PSO particle stability (order-1 and order-2) has a substantial impact on performance (Cleghorn and Engelbrecht, 2016). Specifically, it was shown by Cleghorn and Engelbrecht (2016) that parameter configurations that resulted in particle instability almost always caused PSO to perform worse than random search. Given the relationship between particle stability and performance, it is important to understand the criteria that will ensure particle stability in PSO variants.

A brief description of PSO is given in section 2, followed by a summary of the relevant theoretical work in PSO in section 3. The theoretical derivations of criteria for order-1 and order-2 stability are presented in section 4. Section 5 presents an application of the theoretical stability results. A summary of the paper's findings along with future work is presented in section 6.

2 Particle swarm optimization

Particle swarm optimization (PSO) was originally originally inspired by the complex movement of birds in a flock. The variant of PSO this section focuses on uses the inertia coefficient proposed by Shi and Eberhart (1998), which is referred to as the canonical PSO (CPSO) in this paper.

The PSO algorithm is defined as follows: Let $f : \mathbb{R}^d \to \mathbb{R}$ be the objective function that the PSO algorithm aims to find an optimum for, where d is the dimensionality of the objective function. For the sake of simplicity, a minimization problem is assumed from this point onwards. Specifically, an optimum $o \in \mathbb{R}^d$ is

defined such that, for all $\boldsymbol{x} \in \mathbb{R}^d$, $f(\boldsymbol{o}) \leq f(\boldsymbol{x})$. In this paper the analysis focus is on objective functions where the optima exist. Let $\Omega(t)$ be a set of N particles in $\mathbb{R}^d$ at a discrete time step t. Then $\Omega(t)$ is said to be the particle swarm at time t. The position $\boldsymbol{x}_i$ of particle i is updated using

$$\boldsymbol{x}_i(t+1) = \boldsymbol{x}_i(t) + \boldsymbol{v}_i(t+1), \tag{1}$$

where the velocity update, $\boldsymbol{v}_i(t+1)$, is defined as

$$\boldsymbol{v}_i(t+1) = w\boldsymbol{v}_i(t) + c_1\boldsymbol{r}_1(t) \otimes (\boldsymbol{y}_i(t) - \boldsymbol{x}_i(t)) + c_2\boldsymbol{r}_2(t) \otimes (\hat{\boldsymbol{y}}_i(t) - \boldsymbol{x}_i(t)), \tag{2}$$

where $r_{1,k}(t), r_{2,k}(t) \sim U(0,1)$ for all t and $1 \leq k \leq d$. The operator $\otimes$ is used to indicate component-wise multiplication of two vectors. The position $\boldsymbol{y}_i(t)$ represents the "best" position that particle i has visited, where "best" means the location where the particle had obtained the lowest objective function evaluation. The position $\hat{\boldsymbol{y}}_i(t)$ represents the "best" position that the particles in the neighborhood of the i-th particle have visited. The coefficients c_1, c_2, and w are the cognitive, social, and inertia weights, respectively. A full algorithm description is presented in algorithm 1.

Algorithm 1 PSO algorithm

Create and initialize a swarm, $\Omega(0)$, of N particles uniformly within a predefined hypercube of dimension d.
Let f be the objective function.
Let $\boldsymbol{y}_i$ represent the personal best position of particle i, initialized to $\boldsymbol{x}_i(0)$.
Let $\hat{\boldsymbol{y}}_i$ represent the neighborhood best position of particle i, initialized to $\boldsymbol{x}_i(0)$.
Initialize $\boldsymbol{v}_i(0)$ to $\boldsymbol{0}$.
Let $t = 0$
repeat
    for all particles $i = 1, \cdots, N$ **do**
        if $f(\boldsymbol{x}_i) < f(\boldsymbol{y}_i)$ **then**
            $\boldsymbol{y}_i = \boldsymbol{x}_i$
        end if
        for all particles $\hat{i}$ with particle i in their neighborhood **do**
            if $f(\boldsymbol{y}_i) < f(\hat{\boldsymbol{y}}_{\hat{i}})$ **then**
                $\hat{\boldsymbol{y}}_{\hat{i}} = \boldsymbol{y}_i$
            end if
        end for
    end for
    $t = t + 1$
    for all particles $i = 1, \cdots, N$ **do**
        update the velocity of particle i using equation (2)
        update the position of particle i using equation (1)
    end for
until stopping condition is met

A primary feature of the PSO algorithm is social interaction, specifically the way in which knowledge about the search space is shared amongst the particles in the swarm. In general, the social topology of a swarm can be viewed as a graph, where nodes represent particles, and the edges are the allowable direct communication routes. The social topology chosen has a direct impact on the behaviour of the swarm as a whole (Kennedy, 1999; Kennedy and Mendes, 2002; Engelbrecht, 2013). The fixed topologies, star, ring, and Von Neumann, are frequently used in

PSO. A number of dynamic topologies have also been proposed. The interested reader is referred to the work of Bonyadi and Michalewicz (2016a) for an in-depth discussion on dynamic topologies.

3 Theoretical background

The focus of this section is on the existing theoretical stability results for PSO. Section 3.1 presents the commonly used assumptions in existing theoretical studies on particle stability. Section 3.2 presents the definitions for order-1 and order-2 stability. The relevant theoretical findings for PSO are presented in section 3.3.

3.1 Common assumptions

This section briefly discusses the commonly utilized theoretical assumptions in PSO stability analysis. Where and when each assumption was made in the theoretical literature will be stated in detail in section 3.3.

The primary assumptions that occur in the theoretical PSO research are as follows:

Definition 3.1 Deterministic assumption:
It is assumed that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1(t) = c_1 \boldsymbol{r}_1(t)$, and $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2(t) = c_2 \boldsymbol{r}_2(t)$, for all t (Ozcan and Mohan, 1998).

Definition 3.2 Stagnation assumption:
It is assumed that $\boldsymbol{y}_i(t) = \boldsymbol{y}_i$, and $\hat{\boldsymbol{y}}_i(t) = \hat{\boldsymbol{y}}_i$, for all t sufficiently large (Ozcan and Mohan, 1998).

Definition 3.3 Weak chaotic assumption:
It is assumed that both $\boldsymbol{y}_i(t)$ and $\hat{\boldsymbol{y}}_i(t)$ will occupy an arbitrarily large finite number of unique positions (distinct positions), ψ_i and $\hat{\psi}_i$, respectively (Cleghorn and Engelbrecht, 2014).

Definition 3.4 Weak stagnation assumption:
It is assumed that $\boldsymbol{y}_{\hat{i}}(t) = \boldsymbol{y}_{\hat{i}}$, for all t sufficiently large, where $\hat{i}$ is the index of the particle that has obtained the best objective function evaluations (Liu, 2015).

Definition 3.5 Stagnant distribution assumption:
It is assumed that both $\boldsymbol{y}_i(t)$ and $\hat{\boldsymbol{y}}_i(t)$ are random variables sampled from a fixed distribution, such that both $\boldsymbol{y}_i(t)$ and $\hat{\boldsymbol{y}}_i(t)$ have well defined expectations and variances (Bonyadi and Michalewicz, 2016b).

Each of the assumptions mentioned in this section simplifies the PSO algorithm in order to allow for mathematical analysis to be performed. However, the accuracy of the mathematical model is directly related to the number of PSO's behaviors that are removed due to simplifications. In recent literature, as will be discussed in section 3.3, the deterministic assumption has been successfully removed, which means that the stochastic aspect to PSO is now catered for. However, some form of assumption is still placed on all particles' personal and neighborhood best positions in all existing theoretical stability studies. For this reason, it is important

to understand the relative ordering, in terms of strength of assumption, of the existing assumptions on particles' personal and neighborhood best positions.

The stagnation assumption is the strongest assumption on particles' personal and neighborhood best positions, as the assumption keeps the positions completely fixed. The stagnation assumption implies that no particle ever finds a better position, and as a direct result the swarm will never optimize. This implies that the stability criteria derived under the stagnation assumption are only guaranteed after the swarm has stopped optimizing.

All three of the remaining assumptions on particles' personal and neighborhood best position are weaker than the stagnation assumption. However, there is no clear ordering between the assumptions, since each weakens the assumption on particles' personal and neighborhood best positions in a slightly different way. The weak chaotic assumption allows the particles' personal and neighborhood best positions to occupy an arbitrarily large finite number of distinct positions. The weak stagnation assumption, in essence, assumes stagnation on the global best position, but allows a potentially infinite number of personal best positions, provided they are not more optimal than the stagnant global best position. The most recently used assumption is the stagnant distribution assumption (Bonyadi and Michalewicz, 2016b), which allows a potentially infinite number of different personal and neighborhood best positions. However, there is a superficial restriction on the personal and neighborhood best positions, namely that they are sampled from stationary distributions. During a run, PSO gains information about the search space from which it would derive potentially new personal and neighborhood best positions. This implies that the personal and neighborhood best positions can not be accurately derived from stationary distributions, as clearly the distributions should be functions of the PSO's state at the current iteration, implying the distributions should at least be iteration dependent. In this paper the following weaker assumption is utilized:

Definition 3.6 Non-stagnant distribution assumption:
It is assumed that both $\boldsymbol{y}_i(t)$ and $\hat{\boldsymbol{y}}_i(t)$ are random variables sampled from a time dependent distribution, such that both $\boldsymbol{y}_i(t)$ and $\hat{\boldsymbol{y}}_i(t)$ have well defined expectations and variances for each t and that $\lim_{t\to\infty} E[\boldsymbol{y}_i(t)], \lim_{t\to\infty} E[\hat{\boldsymbol{y}}_i(t)], \lim_{t\to\infty} V[\boldsymbol{y}_i(t)],$ and $\lim_{t\to\infty} V[\hat{\boldsymbol{y}}_i(t)]$ exist.

It should be noted that the non-stagnant distribution assumption is a weaker assumption than all previously made assumptions placed on the particles' personal and neighborhood best positions, as each of the mentioned assumptions can be constructed as a specialization of the non-stagnant distribution assumption. Furthermore, it is shown in section 5 that the non-stagnant distribution assumption is in fact a necessary assumption for order-1 and order-2 stability as defined in the next section.

3.2 Order-1 and order-2 stability

This section discusses the types of convergence used in the stability analysis of PSO.

In the context on a deterministic PSO model, that is a theoretical PSO model that is utilizing the deterministic assumption, the aim is to prove convergence of particle positions. Specifically, convergence is defined in the traditional sense as

Definition 3.7 Convergent sequence
The sequence (s_t) in $\mathbb{R}^n$ is convergent if there exists an $s \in \mathbb{R}^n$ such that

$$\lim_{t \to \infty} s_t = s \qquad (3)$$

It should be made clear that the convergence as defined in equation (3) does not imply convergence to an optimum. The convergence, as described in definition 3.7, can be seen as complete stability, in that the particle's position completely ceases to move as t approach infinity. In a stochastic context it is seldom that complete stability is obtained, and other, more appropriate forms of stability are often considered when working with stochastic sequences. The first form of stability is order-1 stability, defined as

Definition 3.8 Order-1 stability
The sequence (s_t) in $\mathbb{R}^n$ is order-1 stable if there exists an $s_E \in \mathbb{R}^n$ such that

$$\lim_{t \to \infty} E[s_t] = s_E \qquad (4)$$

where $E[s_t]$ is the expectation of s_t.

While order-1 stability is useful, it does not necessarily provide a very strong level of stability. For example, it is possible for a stochastic sequence to have a stationary mean while having a variable or even increasing variance. This leads to the next type of stability, namely

Definition 3.9 Order-2 stability[1]
The sequence (s_t) in $\mathbb{R}^n$ is order-2 stable if there exists a $s_V \in \mathbb{R}^n$ such that

$$\lim_{t \to \infty} V[s_t] = s_V \qquad (5)$$

where $V[s_t]$ is the variance of s_t.

While it is possible to study higher-order stability, such as those related to skewness or kurtosis of a random sequence as was considered by Poli (2008b), it is generally the aim of PSO stability analysis to obtain the criteria needed to ensure order-1 and order-2 stability (Poli, 2009; Blackwell, 2012; García-Gonzalo and Fernández-Martinez, 2014b).

It should be noted that order-1 and order-2 stability does not not imply the deterministic convergence of definition 3.7, except if $s_V = 0$. Some PSO researchers have attempted to provide criteria to ensure that $s_V = 0$ (Poli, 2009). However, it was shown in the work of Poli (2009) that $s_V = 0$ occurs only if order-2 stability as defined in equation (5) occurs in addition to the condition that the personal and neighborhood best positions are equal for all particles, which is not a condition that can be guaranteed to occur. In the work of Bonyadi and Michalewicz

---

[1] For the sake of completeness, it should be noted, that a subset of PSO researchers have utilized the term, order-2 stability, to indicate that the variance of the sequence (s_t) converges to $\mathbf{0}$. However, definition 3.9 is in agreement with the original theoretical PSO research.

(2016b), where the personal best and neighborhood best positions are modeled as random variables, one of the following restrictive cases is required in addition to order-2 stability as defined in equation (5) for $s_V = 0$ to occur in CPSO:

1. $V[\mathbf{y}] = 0$, $V[\hat{\mathbf{y}}] > 0$, $E^2[c_2 r_2] = V[c_2 r_2] = 0$, and $E^2[c_1 r_1] \neq 0$ or
2. $V[\mathbf{y}] > 0$, $V[\hat{\mathbf{y}}] = 0$, $E^2[c_1 r_1] = V[c_1 r_1] = 0$, and $V[c_2 r_2] \neq 0$ or
3. $V[\mathbf{y}] = V[\hat{\mathbf{y}}] = 0$ and $E[\mathbf{y}] = E[\hat{\mathbf{y}}]$ or
4. $V[\mathbf{y}] = V[\hat{\mathbf{y}}] = 0$ and $E[\mathbf{y}] \neq E[\hat{\mathbf{y}}]$ and
   (a) $E^2[c_2 r_2] = V[c_2 r_2] = 0$ and $E^2[c_1 r_1] \neq 0$ or
   (b) $E^2[c_1 r_2] = V[c_1 r_1] = 0$ and $E^2[c_2 r_2] \neq 0$.

Each of the aforementioned cases requires some prior knowledge of the exact expectation and variance of the personal and neighborhood best positions. Furthermore, the cases where $E[\mathbf{y}] = E[\hat{\mathbf{y}}]$ and $V[\mathbf{y}] = V[\hat{\mathbf{y}}] = 0$ are not required, have either the cognitive component or the social component of the PSO's update equation turned off. Specifically,

– for case 1 to hold, $E^2[c_2 r_2] = 0$, which implies $c_2 = 0$;
– for case 2 to hold, $E^2[c_1 r_1] = 0$, which implies $c_1 = 0$;
– for case 4 to hold, either $E^2[c_1 r_1] = 0$ or $E^2[c_2 r_2] = 0$, which again implies that either $c_1 = 0$ or $c_2 = 0$.

It therefore follows that if both the cognitive and social components of the PSO's update equation are used, then the personal and neighborhood best positions must be equal, which cannot be guaranteed to occur in practice.

3.3 Theoretical results for PSO

This section presents each theoretically derived region that is sufficient and/or necessary for particle convergence in the PSO algorithm, along with the corresponding assumptions utilized in the region's derivation.

There are numerous important early contributions to the theoretical understanding of PSO particle trajectories (Ozcan and Mohan, 1998, 1999; Clerc and Kennedy, 2002). However, the first studies to explicitly derive the criteria needed to ensure particle convergence of PSO with the inclusion of the inertia weight was the work of Van den Bergh and Engelbrecht (2006), Van den Bergh (2002), and that of Trelea (2003). Both Van den Bergh and Trelea derived the necessary and sufficient criteria for particle convergence under the deterministic assumption and the stagnation assumption.

The region derived by Van den Bergh and Engelbrecht (2006) is

$$0 < c_1 + c_2 < 2\left(1 + w\right), \quad |w| < 1, \tag{6}$$

whereas the region derived by Trelea (2003) is

$$0 < c_1 + c_2 < 4\left(1 + w\right), \quad |w| < 1, \tag{7}$$

The discrepancy between equations (6) and (7) is due to how the stochastic components were handled. In the work of Van den Bergh and Engelbrecht the stochastic components of the PSO update equation were set such that $r_{1,k} = r_{2,k} = 1$.

This can be seen as a more conservative approach than that used by Trelea, were $r_{1,k} = r_{2,k} = 1/2$, which is the expected value of the random components. Equation (6) and (7) are illustrated in figure 1 as the triangles AFB and ACB, respectively. It should be noted that, from a theoretical perspective, the constriction PSO is an equivalent model to the inertia PSO under certain conditions, so many findings from the earlier work of Clerc and Kennedy (2002) are relevant to the inertia PSO.

More recently, under the deterministic and weak chaotic assumptions, Cleghorn and Engelbrecht (2014) derived the same region as equations (6) or (7), depending on how the stochastic variables $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are set.

Under the stagnation assumption only, Kadirkamanathan et al. (2006) derived the following sufficient region for order-2 stability:

$$\begin{cases} c_1 + c_2 < 2\,(1 + w) & \text{for} \quad w \in (-1, 0] \\ c_1 + c_2 < \frac{2(1-w)^2}{1+w} & \text{for} \quad w \in (0, 1)\,. \end{cases} \tag{8}$$

Still under the stagnation assumption, Gazi (2012) expanded the derived region of equation (8), resulting in the region

$$\begin{cases} c_1 + c_2 < \frac{24(1+w)}{7} & \text{for} \quad w \in (-1, 0] \\ c_1 + c_2 < \frac{24(1-w)^2}{7(1+w)} & \text{for} \quad w \in (0, 1)\,. \end{cases} \tag{9}$$

The regions corresponding to equations (8) and (9) are illustrated in figure 1 as triangle like regions ADB and AEB, respectively. Unfortunately, both equations (8) and (9) are very conservative regions, as they were derived utilizing the Lyapunov condition (Kisacanin and Agarwal, 2001).

Without the use of the Lyapunov condition or the stochastic assumption, Poli and Broomhead (2007) and Poli (2009) derived the necessary and sufficient criteria for order-1 and order-2 stability. The order-2 region's sufficient condition was partially obtained via experimental means. The region derived by Poli for order-1 stability is the same as the region derived by Trelea in equation (7). The region derived for order-2 stability is as follows:

$$c_1 + c_2 < \frac{24\left(1 - w^2\right)}{7 - 5w} \quad \text{for} \quad w \in [-1, 1]\,. \tag{10}$$

The region defined by equation (10) is illustrated in figure 1 as the curved line segment AGB. The region defined by equation (10) was also independently derived by Jiang et al. (2007) under the stagnation assumption.

Blackwell (2012) showed that the criteria of equation (10) are a necessary condition for order-2 stability, utilizing an approach that was both computationally simpler than Poli's approach while also being applicable to a range of PSO variants. García-Gonzalo and Fernández-Martinez (2014a) also derived necessary conditions for order-1 and order-2 stability, allowing w, $c_1\boldsymbol{r}_1$, and $c_2\boldsymbol{r}_2$ to be random variables with well defined expectations and variances. García-Gonzalo and Fernández-Martinez utilized similar techniques to their earlier contribution which relied on modeling the PSO as a stochastic damped mass-spring system (García-Gonzalo and Fernández-Martinez, 2011).

Liu (2015) rederived, under the weak stagnation assumption, the same necessary and sufficient conditions for order-2 stability as Poli. The work of Liu (2015)
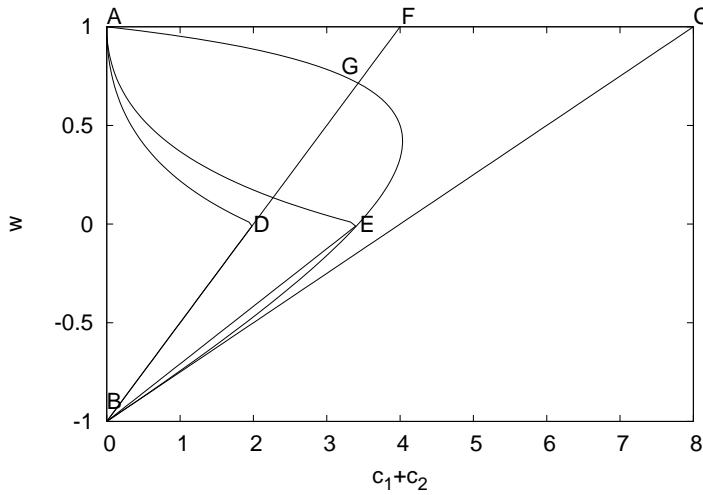
Fig. 1: Theoretically derived regions sufficient for particle convergence

also implies that the convergence region of equation (10) is the same irrespective of the social network topology utilized by PSO.

Recently, under the stagnation distribution assumption, Bonyadi and Michalewicz (2016b) were able to derive criteria for order-1 and order-2 stability, while also allowing w, $c_1 r_1$, and $c_2 r_2$ to be random variables with well defined expectations and variances.

The work of Poli (2009) and Bonyadi and Michalewicz (2016b) both rely on a first order, non-homogeneous recurrence relation. Specifically, a first order, non-homogeneous recurrence relation is defined as a sequence (z_t) in $\mathbb{R}^q$, constructed from

$$z_t = M z_{t-1} + b, \tag{11}$$

where b is a constant offset in $\mathbb{R}^q$ and M is a $q \times q$ matrix. The recurrence relation's initial term is defined as z_1. The work of Poli (2009) and Bonyadi and Michalewicz (2016b) utilize a well known theorem from analysis, namely,

Theorem 3.1 *The sequence (z_t) converges for any initial condition $z_0 \in \mathbb{R}^q$ and offset $b \in \mathbb{R}^q$ if and only if $\rho(M) < 1$, where ρ denotes the spectral radius* (Atkinson and Han, 2009).

In order to utilize the non-stagnate distribution assumption, as defined in section 3.1, a generalization of the sufficient condition of theorem 3.1 is needed. Such a generalization is presented in the next section.

4 Stability proof

This section presents a derivation of necessary and sufficient conditions for order-1 and order-2 stability of PSO under the non-stagnant distribution assumption. The

section starts with a fundamental definition from analysis followed by two small lemma's which are used to improve the flow of the more substantial lemma 4.3. Lemma 4.3 is an extension of a classic theorem from mathematical analysis. The main result is then presented in theorem 4.1.

Definition 4.1 Uniform Operator Convergence
A sequence $(\boldsymbol{T}_n)$ of operators $\boldsymbol{T}_n : \mathbb{R}^p \to \mathbb{R}^q$ is said to be uniformly operator convergent to the operator $\boldsymbol{T}$ if $\|\boldsymbol{T}_n - \boldsymbol{T}\| \to 0$, where $\|\cdot\|$ is an operator norm. (Kreyszig, 1978)

Lemma 4.1 *Let $(\boldsymbol{T}_n)$ be a sequence of bounded linear operators* (Kreyszig, 1978) *from $\mathbb{R}^p$ to $\mathbb{R}^p$, all with equal spectral radius, and let $(\boldsymbol{T}_n)$ be uniformly operator convergent. If $\rho(\boldsymbol{T}_1) < 1$, then $\lim\limits_{n\to\infty} \boldsymbol{T}_n\boldsymbol{T}_{n-1}\cdots\boldsymbol{T}_1 = \Theta$, where Θ is the null operator; $\rho(\boldsymbol{T}_1)$ is used to indicate the spectral radius of $\boldsymbol{T}_1$.*

Proof It is known that for any $\delta > 0$, there exists a norm $\|\cdot\|_\delta$ such that $\rho(\boldsymbol{T}_1) \leq \|\boldsymbol{T}_1\|_\delta \leq \rho(\boldsymbol{T}_1) + \delta$ (Kreyszig, 1978), and since $\rho(\boldsymbol{T}_1) < 1$, a δ can be selected such that there exists a σ where $\|\boldsymbol{T}_1\|_\delta \leq \sigma < 1$. The same is true for any element of the sequence $(\boldsymbol{T}_n)$. Since $(\boldsymbol{T}_n)$ is uniformly operator convergent, and each $\rho(\boldsymbol{T}_n) < 1$, there exists a N, σ, and norm $\|\cdot\|_\delta$ such that $\|\boldsymbol{T}_n\| \leq \sigma < 1$ for all $n \geq N$. Furthermore, since $(\boldsymbol{T}_n)$ is uniformly operator convergent there exists a bound such that $\|\boldsymbol{T}_n\| \leq \xi$ for all n. It then follows that, for all $n > N$,

$$\|\boldsymbol{T}_n\boldsymbol{T}_{n-1}\cdots\boldsymbol{T}_N\boldsymbol{T}_{N-1}\cdots\boldsymbol{T}_1\|_\delta \leq \|\boldsymbol{T}_n\boldsymbol{T}_{n-1}\cdots\boldsymbol{T}_{N+1}\|_\delta\|\boldsymbol{T}_N\boldsymbol{T}_{N-1}\cdots\boldsymbol{T}_1\|_\delta \quad (12)$$
$$\leq \sigma^{n-N}\xi^N$$

Given that N and ξ are finite, $\sigma^{n-N}\xi^N \to 0$ as $n \to \infty$. Since $\|\boldsymbol{T}_n\boldsymbol{T}_{n-1}\cdots\boldsymbol{T}_1\|_\delta \to 0$, $\boldsymbol{T}_n\boldsymbol{T}_{n-1}\cdots\boldsymbol{T}_1 \to \Theta$, as was to be proved. $\qquad\square$

Lemma 4.2 *Let $(\boldsymbol{T}_n)$ be a sequence of bounded linear operators from $\mathbb{R}^p$ to $\mathbb{R}^p$, and let $(\boldsymbol{T}_n)$ be uniformly operator convergent. Then for any finite j the sequence $(\boldsymbol{T}_n\cdots\boldsymbol{T}_{n-j})$ is also uniformly operator convergent.*

Proof An inductive argument is used. For $j = 0$ operator convergence is directly obtained from the given assumption. The inductive step is as follows: assume that $(\boldsymbol{T}_n\cdots\boldsymbol{T}_{n-j})$ is convergent. Since both $(\boldsymbol{T}_{n-j-1})$ and $(\boldsymbol{T}_n\cdots\boldsymbol{T}_{n-j})$ are convergent, they are bounded, so there exists an η_1 and η_2 such that $\|\boldsymbol{T}_n\cdots\boldsymbol{T}_{n-j}\| \leq \eta_1$ and $\|\boldsymbol{T}_{n-j-1}\| \leq \eta_2$, for all n. It also follows that for any $\epsilon > 0$ there exists a $N_{\epsilon,j-1}$ such that, if $m > n \geq N_{\epsilon,j-1}$ for $m, n \in \mathbb{N}$, then $\|\boldsymbol{T}_{m-j-1} - \boldsymbol{T}_{n-j-1}\| < \frac{\epsilon}{2\eta_1}$. Similarly, there exists an N_ϵ such that, if $m > n \geq N_\epsilon$, then $\|\boldsymbol{T}_m\cdots\boldsymbol{T}_{m-j} - \boldsymbol{T}_n\cdots\boldsymbol{T}_{n-j}\| < \frac{\epsilon}{2\eta_2}$. Let $N = max\{N_\epsilon, N_{\epsilon,j-1}\}$. If $m > n > N$, then

$$\begin{aligned}
&\|\boldsymbol{T}_m\cdots\boldsymbol{T}_{m-j-1} - \boldsymbol{T}_n\cdots\boldsymbol{T}_{n-j-1}\| \\
&\leq \|\boldsymbol{T}_m\cdots\boldsymbol{T}_{m-j-1} - \boldsymbol{T}_m\cdots\boldsymbol{T}_{m-j}\boldsymbol{T}_{n-j-1}\| \\
&\quad + \|\boldsymbol{T}_m\cdots\boldsymbol{T}_{m-j}\boldsymbol{T}_{n-j-1} - \boldsymbol{T}_n\cdots\boldsymbol{T}_{n-j-1}\| \\
&\leq \|\boldsymbol{T}_m\cdots\boldsymbol{T}_{m-j}\|\|\boldsymbol{T}_{m-j-1} - \boldsymbol{T}_{n-j-1}\| \\
&\quad + \|\boldsymbol{T}_m\cdots\boldsymbol{T}_{m-j} - \boldsymbol{T}_n\cdots\boldsymbol{T}_{n-j}\|\|\boldsymbol{T}_{n-j-1}\| \\
&\leq \eta_1\|\boldsymbol{T}_{m-j-1} - \boldsymbol{T}_{n-j-1}\| + \eta_2\|\boldsymbol{T}_m\cdots\boldsymbol{T}_{m-j} - \boldsymbol{T}_n\cdots\boldsymbol{T}_{n-j}\| \\
&< \epsilon
\end{aligned} \qquad (13)$$

Therefore, $(\boldsymbol{T}_m \cdots \boldsymbol{T}_{m-j-1})$ is Cauchy, and therefore convergent, proving the inductive step. This concludes the proof of the lemma. $\square$

Lemma 4.3 *Let $(\boldsymbol{x}_n)$ be a sequence in $\mathbb{R}^p$, defined as*

$$\boldsymbol{x}_n = \boldsymbol{T}_n \boldsymbol{x}_{n-1} + \boldsymbol{b}_{n-1} \tag{14}$$

where $(\boldsymbol{T}_n)$ is a sequence of bounded linear operators from $\mathbb{R}^p$ to $\mathbb{R}^p$ which is uniformly operator convergent, with each element having equal spectral radius, and $(\boldsymbol{b}_n)$ is a sequence in $\mathbb{R}^p$. The term $\boldsymbol{x}_1$ is used to indicate the initial condition. Now, if $\rho(\boldsymbol{T}_n) < 1$ for all n and $(\boldsymbol{b}_n)$ converges, then $(\boldsymbol{x}_n)$ converges.

Proof: Firstly, it is assumed that $\rho(\boldsymbol{T}_n) < 1$ for all n and that $(\boldsymbol{b}_n)$ converges. As a notational convenience, let $\boldsymbol{C}_{(s,e)} = \boldsymbol{T}_e \boldsymbol{T}_{e-1} \cdots \boldsymbol{T}_s$ when $s \leq e$, and $\boldsymbol{C}_{(s,e)} = I$ when $s > e$, where I is the identity operator. As in Lemma 4.1 it is known that there exists a norm $\|\cdot\|$ such that $\|\boldsymbol{T}_n\| \leq \sigma < 1$. This norm will be used for the remainder of the proof.

Since the lemma is set in a finite dimensional space, it is sufficient to prove that $(\boldsymbol{x}_n)$ is Cauchy in order to prove convergence (Kreyszig, 1978), which is now done. Let $m, n \in \mathbb{N}$ and $m > n$. Unwinding $\boldsymbol{x}_n$ leads to

$$\begin{aligned}
\boldsymbol{x}_n &= \boldsymbol{T}_n \boldsymbol{x}_{n-1} + \boldsymbol{b}_{n-1} \\
&= \boldsymbol{T}_n(\boldsymbol{T}_{n-1}\boldsymbol{x}_{n-2} + \boldsymbol{b}_{n-2}) + \boldsymbol{b}_{n-1} \\
&= \cdots \\
&= \boldsymbol{C}_{(1,n-1)}\boldsymbol{x}_1 + \sum_{i=0}^{n-2} \boldsymbol{C}_{(n+1-i,n)}\boldsymbol{b}_{n-1-i}
\end{aligned} \tag{15}$$

Now, using equation (15),

$$\begin{aligned}
&\|\boldsymbol{x}_m - \boldsymbol{x}_n\| \\
&= \|\boldsymbol{C}_{(1,m-1)}\boldsymbol{x}_1 + \sum_{i=0}^{m-2} \boldsymbol{C}_{(m+1-i,m)}\boldsymbol{b}_{m-1-i} \\
&\quad - \boldsymbol{C}_{(1,n-1)}\boldsymbol{x}_1 - \sum_{i=0}^{n-2} \boldsymbol{C}_{(n+1-i,n)}\boldsymbol{b}_{n-1-i}\| \\
&\leq \|\left(\boldsymbol{C}_{(1,m-1)} - \boldsymbol{C}_{(1,n-1)}\right)\boldsymbol{x}_1\| \\
&\quad + \|\sum_{i=0}^{m-2} \boldsymbol{C}_{(m+1-i,m)}\boldsymbol{b}_{m-1-i} - \sum_{i=0}^{n-2} \boldsymbol{C}_{(n+1-i,n)}\boldsymbol{b}_{n-1-i}\|
\end{aligned} \tag{16}$$

Considering the first term of the summation in equation (16), it is seen that

$$\|\left(\boldsymbol{C}_{(1,m-1)} - \boldsymbol{C}_{(1,n-1)}\right)\boldsymbol{x}_1\| \leq \|\left(\boldsymbol{C}_{(1,m-1)} - \boldsymbol{C}_{(1,n-1)}\right)\| \, \|\boldsymbol{x}_1\| \tag{17}$$

since $\rho(\boldsymbol{T}_n) < 1$ for all n, and using lemma 4.1, the sequence of operators $(\boldsymbol{C}_{(1,n)})$ is convergent, and therefore Cauchy. It follows that $\|\left(\boldsymbol{C}_{(1,m-1)} - \boldsymbol{C}_{(1,n-1)}\right)\|$ convergences to zero, and trivially so does equation (17).

Focusing on the second term of the summation in equation (16),

$$
\begin{aligned}
\| \sum_{i=0}^{m-2} & \boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{m-1-i} - \sum_{i=0}^{n-2} \boldsymbol{C}_{(n+1-i,n)} \boldsymbol{b}_{n-1-i} \| \\
\leq & \| \sum_{i=n-1}^{m-2} \boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{m-1-i} \| + \\
& \| \sum_{i=0}^{n-2} \left(\boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{m-1-i} - \boldsymbol{C}_{(n+1-i,n)} \boldsymbol{b}_{n-1-i} \right) \|
\end{aligned} \tag{18}
$$

Note that, for the first term in equation (18), since $(\boldsymbol{b}_n)$ is convergent, there exists a ζ such that $\|\boldsymbol{b}_n\| < \zeta$ for every n. It therefore follows that

$$
\begin{aligned}
\| \sum_{i=n-1}^{m-2} \boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{m-1-i} \| & \leq \sum_{i=n-1}^{m-2} \| \boldsymbol{C}_{(m+1-i,m)} \| \, \| \boldsymbol{b}_{m-1-i} \| \\
& \leq \sum_{i=n-1}^{m-2} \| \boldsymbol{C}_{(m+1-i,m)} \| \, \zeta \leq \sum_{i=n-1}^{m-2} \prod_{j=m}^{m-1-i} \| \boldsymbol{T}_j \| \, \zeta
\end{aligned} \tag{19}
$$

It is also known that $\|\boldsymbol{T}_j\| \leq \sigma < 1$. It therefore follows from equation (19) that

$$
\sum_{i=n-1}^{m-2} \prod_{j=m}^{m-1-i} \| \boldsymbol{T}_j \| \, \zeta \leq \sum_{i=n-1}^{m-2} \prod_{j=m}^{m-1-i} \sigma \, \zeta = \zeta \sum_{i=n-1}^{m-2} \sigma^{i+1} \tag{20}
$$

Since $\sigma < 1$, the elementary geometric series formula can be used to transform equation (20) to

$$
\zeta \sum_{i=n-1}^{m-2} \sigma^{i+1} = \zeta \frac{\sigma^{n-1} - \sigma^{m-1}}{1 - \sigma} \ \to 0 \text{ as } n, m \to \infty
$$

Focusing on the remaining term in equation (18),

$$
\begin{aligned}
\| \sum_{i=0}^{n-2} & \left(\boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{m-1-i} - \boldsymbol{C}_{(n+1-i,n)} \boldsymbol{b}_{n-1-i} \right) \| \\
\leq & \| \sum_{i=0}^{n-2} \left(\boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{m-1-i} - \boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{n-1-i} \right) \| \\
& + \| \sum_{i=0}^{n-2} \left(\boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{n-1-i} - \boldsymbol{C}_{(n+1-i,n)} \boldsymbol{b}_{n-1-i} \right) \|
\end{aligned} \tag{21}
$$

Both terms of equation (21) require a more subtle mathematical treatment, given the complexity of the internal terms. The standard epsilon-n approach from anal-

ysis will be used. Consider the first term of equation (21),

$$
\begin{aligned}
\| \sum_{i=0}^{n-2} & \left(\boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{m-1-i} - \boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{n-1-i} \right) \| \\
& \leq \sum_{i=0}^{n-2} \| \boldsymbol{C}_{(m+1-i,m)} \| \, \| \boldsymbol{b}_{m-1-i} - \boldsymbol{b}_{n-1-i} \| \\
& \leq \sum_{i=0}^{n-2} \sigma^i \, \| \boldsymbol{b}_{m-1-i} - \boldsymbol{b}_{n-1-i} \| \| \\
& \leq \sum_{i=0}^{n-2} \sigma^i \, \| \boldsymbol{b}_{m-1-i} - \boldsymbol{b} \| + \sum_{i=0}^{n-2} \sigma^i \, \| \boldsymbol{b}_{n-1-i} - \boldsymbol{b} \|
\end{aligned}
\tag{22}
$$

Let $\boldsymbol{z}_i = \boldsymbol{b}_{i+1} - \boldsymbol{b}$. Since $\boldsymbol{b}_i \to \boldsymbol{b}$, $\|\boldsymbol{z}_i\| \to 0$, there exists a $\tau \in \mathbb{R}$ such that $\|\boldsymbol{z}_i\| < \tau$ for all i, because $(\boldsymbol{z}_i)$ is convergent. Now there exists a n_ϵ, such that $\sigma^n < \epsilon$ and $\|\boldsymbol{z}_{n+1}\| < \epsilon$ for all $n > n_\epsilon$. So, for the second term of equation (22),

$$
\begin{aligned}
\sum_{i=0}^{n-2} \sigma^i \, \| \boldsymbol{b}_{n-1-i} - \boldsymbol{b} \| & = \sum_{i=0}^{n-2} \sigma^{n-2-i} \, \| \boldsymbol{b}_{i+1} - \boldsymbol{b} \| \\
& \leq \tau \sum_{i=0}^{n_\epsilon - 2} \sigma^{n-2-i} + \epsilon \sum_{i=n_\epsilon - 1}^{n-2} \sigma^{n-2-i} \\
& \leq \frac{\sigma^{n-n_\epsilon} - \sigma^{n-1}}{1-\sigma} \tau + \frac{1 - \sigma^{n-n_\epsilon}}{1-\sigma} \epsilon \\
& \leq \frac{\sigma^{n-n_\epsilon}}{1-\sigma} \tau + \frac{\epsilon}{1-\sigma} \\
& \leq \frac{\tau + 1}{1-\sigma} \epsilon
\end{aligned}
\tag{23}
$$

Now, since $m > n$, the same argument can be made for the first term of equation (22) as was used for the second term. This implies that, for a large enough n and m, equation (22) can be made less than an arbitrarily small $\epsilon > 0$.

The last remaining term requiring analysis is the second term of equation (21). Note that, since $\sigma^n \to 0$ for any $\epsilon > 0$, there exists a n_ϵ such that $\sigma^n <$

$\epsilon\,(1-\sigma)\,/(2\zeta)$ if $n \geq n_\epsilon$. Equation (21) can then be handled as follows:

$$\| \sum_{i=0}^{n-2} \left(\boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{n-1-i} - \boldsymbol{C}_{(n+1-i,n)} \boldsymbol{b}_{n-1-i} \right) \|$$

$$\leq \sum_{i=0}^{n-2} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| \| \boldsymbol{b}_{n-1-i} \|$$

$$\leq \zeta \sum_{i=1}^{n_\epsilon-1} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| + \zeta \sum_{i=n_\epsilon}^{n-2} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \|$$

$$\leq \zeta \sum_{i=1}^{n_\epsilon-1} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| + 2\zeta \sum_{i=n_\epsilon}^{n-2} \sigma^i$$

$$\leq \zeta \sum_{i=1}^{n_\epsilon-1} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| + 2\zeta \frac{\sigma^{n_\epsilon}}{1-\sigma}$$

$$\leq \zeta \sum_{i=1}^{n_\epsilon-1} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| + \epsilon \qquad (24)$$

Now that m and n have been decoupled from the summation limit of the first term in equation (24), the limit can be dealt with directly. It is known from lemma 4.2 that, since (T_n) was convergent, then $(\boldsymbol{T}_n \cdots \boldsymbol{T}_{n-i})$ is also convergent for a finite i. It then follows that for every $\delta > 0$ and for each sequence $(\boldsymbol{C}_{(n+1-i,n)})$ where $0 \leq i \leq n_\epsilon$ there exists a $n_{\delta,i}$ such that, if $m > n \geq n_{\delta,i}$, then $\| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| < \delta/(\zeta n_\epsilon)$. Let $n_\delta = max\{n_{\delta,i} | 0 \leq i \leq n_\epsilon\}$. Then equation (24) becomes

$$\zeta \sum_{i=1}^{n_\epsilon-1} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| + \epsilon \leq \delta + \epsilon \qquad (25)$$

Since ϵ and δ can be made arbitrarily small, this completes the proof. $\qquad \square$

Now that lemma (4.3) has been proved, the focus is moved to the main result on PSO stability. In this paper all PSO variants with update equations of the form

$$x_k(t+1) = x_k(t)\alpha + x_k(t-1)\beta + \gamma_t \qquad (26)$$

are considered, where k indicates the vector component, α and β are well defined random variables, and (γ_t) is a sequence of well defined random variables. In the context of this paper a random variable is said to be well defined if it has an expectation and variance.

This class of PSOs includes CPSO, fully informed PSO (Kennedy and Mendes, 2003), and unified PSO (Parsopoulos and Vrahatis, 2004), though many others exist. Furthermore, this class also allows arbitrary distributions to be utilized for all vector components, provided they are dimension independent and have well defined expectations and variances. It should be noted that the considered class of PSOs does not cater for PSO variants with time dependent random variables α and β, or for PSO variants where α or β do not have well defined expectations and variances. Both of these mentioned variants are beyond the scope of this paper.

Theorem 4.1 *The following properties hold for all PSO variants of the form described in equation (26):*

1. *Assuming i_t converges, particle positions are order-1 stable for every initial condition if and only if $\rho(A) < 1$, where*

$$A = \begin{bmatrix} E[\alpha] & E[\beta] \\ 1 & 0 \end{bmatrix} \ and \ \boldsymbol{i}_t = \begin{bmatrix} E[\gamma_t] \\ 0 \end{bmatrix} \tag{27}$$

2. *The particle positions are order-2 stable if $\rho(\boldsymbol{B}) < 1$ and $(\boldsymbol{j}_t)$ converges, where*

$$\boldsymbol{B} = \begin{bmatrix} E[\alpha] & E[\beta] & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & E[\alpha^2] & E[\beta^2] & 2E[\alpha\beta] \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & E[\alpha] & 0 & E[\beta] \end{bmatrix}$$

and

$$\boldsymbol{j}_t = \begin{bmatrix} E[\gamma_t] \\ 0 \\ E[\gamma_t^2] \\ 0 \\ 0 \end{bmatrix} \tag{28}$$

under the assumption that the limits of $(E[\gamma_t\alpha])$ and $(E[\gamma_t\beta])$ exist.

3. *Assuming that $x(t)$ is order-1 stable, then the following is a necessary condition for order-2 stability:*

$$1 - E[\alpha] - E[\beta] \neq 0 \tag{29}$$

$$1 - E\left[\alpha^2\right] - E\left[\beta^2\right] - \left(\frac{2E[\alpha\beta]\,E[\alpha]}{1 - E[\beta]}\right) > 0 \tag{30}$$

4. *The convergence of $E[\gamma_t]$ is a necessary condition for order-1 stability, and the convergence of both $E[\gamma_t]$ and $E[\gamma_t^2]$ is a necessary condition for order-2 stability.*

Proof It should first be noted that there is no coupling between dimensions in the PSO variants considered in this theorem. Therefore, analysis can be performed in one dimension only without loss of generality. This is possible because each dimension can be modeled as an independent problem. Furthermore, since the coefficients and distributions used are the same in each dimension, the stability criteria for one dimension is the same for all dimensions. As a result the dimension subscript k is dropped.

Property 1 is proved first. The application of the expectation operator to equation (26) yields

$$E[x(t+1)] = E[x(t)]E[\alpha] + E[x(t-1)]E[\beta] + E[\gamma_t] \tag{31}$$

which is reformulated to

$$\boldsymbol{u}_t = \boldsymbol{A}\boldsymbol{u}_{t-1} + \boldsymbol{i}_t \tag{32}$$

where $\boldsymbol{u}_t = \begin{bmatrix} E[x(t+1)] \\ E[x(t)] \end{bmatrix}$, $\boldsymbol{A} = \begin{bmatrix} E[\alpha] & E[\beta] \\ 1 & 0 \end{bmatrix}$, and $\boldsymbol{i}_t = \begin{bmatrix} E[\gamma_t] \\ 0 \end{bmatrix}$. Direct application of lemma 4.3 shows that $(\boldsymbol{u}_t)$ converges if $\rho(\boldsymbol{A}) < 1$ and $(\boldsymbol{i}_t)$ is convergent, implying order-1 stability of particle positions. From theorem 3.1 it is known that if $\boldsymbol{i}_t$ is constant, then $\rho(\boldsymbol{A}) < 1$ is a necessary condition for convergence. Since a constant $\boldsymbol{i}_t$ is a special case of equation (32), $\rho(\boldsymbol{A}) < 1$ is also a necessary condition for convergence of $\boldsymbol{u}_t$ (specifically, $\rho(\boldsymbol{A}) < 1$ ensures convergence for all possible initial conditions).

Now, consider property 2. In order to study the variance of equation (26), defined as,

$$V[x(t+1)] = E[x^2(t+1)] + E[x(t+1)]^2 \tag{33}$$

the dynamics of $E[x^2(t+1)]$ and $E[x(t)x(t-1)]$ need to be considered.

The term $x^2(t+1)$ is calculated as

$$\begin{aligned} x^2(t+1) &= x^2(t)\alpha^2 + x^2(t-1)\beta^2 + \gamma_t^2 + 2x(t)\gamma_t\alpha \\ &\quad + 2x(t)x(t-1)\alpha\beta + 2x(t-1)\beta\gamma_t \end{aligned} \tag{34}$$

Application of the expectation operator produces

$$\begin{aligned} E[x^2(t+1)] &= E[x^2(t)]E[\alpha^2] + E[x^2(t-1)]E[\beta^2] + E[\gamma_t^2] \\ &\quad + 2E[x(t)]E[\gamma_t\alpha] + 2E[x(t)x(t-1)]E[\alpha\beta] \\ &\quad + 2E[x(t-1)]E[\beta\gamma_t] \end{aligned} \tag{35}$$

The expectation of $x(t)x(t-1)$ is obtained by multiplying equation (26) by $x(t)$ and applying the expectation operator to yield

$$E[x(t+1)x(t)] = E[\alpha]E[x^2(t)] + E[\beta]E[x(t)x(t+1)] + E[x(t)]E[\gamma_t] \tag{36}$$

Given equations (31), (35), and (36), the dynamics of $E[x^2(t+1)]$ and $E[x(t)x(t-1)]$ are derived by relying on a 5-dimensional recurrence relation, as there are five unknowns in the system, namely $E[x(t)]$, $E[x(t-1)]$, $E[x^2(t)]$, $E[x^2(t-1)]$, and $E[x(t)x(t-1)]$. If the recurrence relation has a limit, then so does $V[x(t)]$, implying order-2 stability. The specific recurrence relation under consideration is

$$\boldsymbol{g}_t = \boldsymbol{B}_t\boldsymbol{g}_{t-1} + \boldsymbol{j}_t \tag{37}$$

where

$$\boldsymbol{B}_t = \begin{bmatrix} E[\alpha] & E[\beta] & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 2E[\gamma_t\alpha] & 2E[\gamma_t\beta] & E[\alpha^2] & E[\beta^2] & 2E[\alpha\beta] \\ 0 & 0 & 1 & 0 & 0 \\ E[\gamma_t] & 0 & E[\alpha] & 0 & E[\beta] \end{bmatrix} \tag{38}$$

$$\boldsymbol{g}_t = \begin{bmatrix} E[x(t)] \\ E[x(t-1)] \\ E[x^2(t)] \\ E[x^2(t-1)] \\ E[x(t)x(t-1)] \end{bmatrix}, \qquad \boldsymbol{j}_t = \begin{bmatrix} E[\gamma_t] \\ 0 \\ E[\gamma_t^2] \\ 0 \\ 0 \end{bmatrix} \tag{39}$$

Since the limits $(E[\gamma_t\alpha])$, $(E[\gamma_t\beta])$ and $(E[\gamma_t])$ exist by assumption 3.6, then so does the limit of $(\boldsymbol{B}_t)$. One of the conditions of lemma 4.3 is that the spectral radius of $\boldsymbol{B}_t$ must be the same for all t. The eigenvalues of $\boldsymbol{B}_t$ were calculated using Matlab's symbolic toolbox, and are given in appendix A. The eigenvalues actually do not contain the terms $E[\gamma_t\alpha]$, $E[\gamma_t\beta]$, or $E[\gamma_t]$ at all. The absence of γ_t in any term implies that the spectral radius of $\boldsymbol{B}_t$ is constant, and therefore, direct application of lemma 4.3 shows that $\boldsymbol{g}_t$ converges if $\rho(\boldsymbol{B}_t) < 1$ and $\boldsymbol{j}_t$ is convergent, implying order-2 stability of PSO particles as was to be proved. Furthermore, since the spectral radius of $\boldsymbol{B}_t$ does not depend on $E[\gamma_t\alpha]$, $E[\gamma_t\beta]$, or $E[\gamma_t]$, the spectral radius of $\boldsymbol{B}$, with $E[\gamma_t\alpha]$, $E[\gamma_t\beta]$ and $E[\gamma_t]$ all set to zero, is equivalent to the spectral radius of $\boldsymbol{B}_t$. Therefore, the conditions under which $\rho(\boldsymbol{B}_t) < 1$ are the same as the conditions under which $\rho(\boldsymbol{B}) < 1$.

The proof of property 3 follows directly from the work of Blackwell (2012). It was shown by Blackwell that if γ_t is a constant random variable and that if $x(t)$ is order-1 stable, then equations (29) and (30) are necessary conditions for order-2 stability. Note that a constant γ_t is simply a special case of equation (26), which implies that the necessary conditions hold equally for the class of PSOs under consideration in equation (26).

Property 4 is now proved. First note that, trivially, $(E[\gamma_t])$ converges if and only if $(\boldsymbol{i}_t)$ converges and that $(E[\gamma_t])$ and $(E[\gamma_t^2])$ converge if and only if $(\boldsymbol{j}_t)$ converges. The approach taken is to prove property 3 by contradiction. Assume that $(\boldsymbol{i}_t)$ diverges, but that $(\boldsymbol{u}_t)$ converges. Equation (32) can now be reformulated to

$$\boldsymbol{u}_t - \boldsymbol{A}\boldsymbol{u}_{t-1} = \boldsymbol{i}_t \tag{40}$$

Because $(\boldsymbol{u}_t)$ converges and $\boldsymbol{A}$ is continuous, the summation $(\boldsymbol{u}_t - \boldsymbol{A}\boldsymbol{u}_{t-1})$ is also convergent. But, this is impossible as $(\boldsymbol{i}_t)$ is divergent by assumption, hence a contradiction. The same approach can be used to show that $(\boldsymbol{j}_t)$ must be convergent if $(\boldsymbol{y}_t)$ is, and therefore the convergence of both $(E[\gamma_t])$ and $(E[\gamma_t^2])$ are necessary conditions.

This completes the proof of properties 1, 2, 3 and 4. $\qquad\square$

5 Direct application of stability theory

This section provides an illustrative example using theorem 4.1 to provide the reader with a simple procedure for using theorem 4.1 to derive stability criteria for new PSO variants that comply with the formulation given in equation (26).

A general version of CPSO is considered. Specifically, the components $c_1 r_1 = \theta_1$, $c_2 r_2 = \theta_2$, and w are allowed to be arbitrary independent random variables with well defined expectations and variances, as considered in (Bonyadi and Michalewicz, 2016b). It is shown that the same criteria as in (Bonyadi and Michalewicz, 2016b) can be derived for both order-1 and order-2 stability utilizing theorem 4.1 under the less restrictive non-stagnant distribution assumption. It should be noted that while $E[y_t]$, $E[\hat{y}_t]$, $V[y_t]$, and $E[\hat{y}_t]$ are assumed to exist under the non-stagnate distribution assumption, the moment's explicit values are never needed in the derivation of order-1 and order-2 stability criteria in this section.

Rewriting the general version of CPSO in the form of equation (26) is achieved by setting the following terms:

$$\alpha = (1 + w) - \theta_1 - \theta_2$$
$$\beta = -w$$
$$\gamma_t = \theta_1 y_t + \theta_2 \hat{y}_t \tag{41}$$

where y_t and $\hat{y}_t$, the personal and neighborhood best positions respectively, are modeled as sequences of random variables which are convergent in expectation and variance. In order to use theorem 4.1 it should first be verified if $(\boldsymbol{i}_t)$ and $(\boldsymbol{j}_t)$ are convergent. Note that, because $E[\theta_1]$ and $E[\theta_2]$ are constant and the limit of $(E[y_t])$ exists, the limit of $(E[\gamma_t])$ also exists, where $E[\gamma_t] = E[\theta_1]E[y_t] + E[\theta_2]E[\hat{y}_t]$. The existence of the limit of $(E[\gamma_t])$ implies that $(\boldsymbol{i}_t)$ is convergent. In order for $(\boldsymbol{j}_t)$ to be convergent, the limit of $(E[\gamma_t^2])$ must also exist. Observe that

$$E[\gamma_t^2] = E[\theta_2^2]E[y_t^2] + 2E[\theta_1]E[\theta_2]E[y_t]E[\hat{y}_t] + E[\theta_2^2]E[\hat{y}_t^2]$$

Because $V[\chi] = E[\chi^2] - (E[\chi])^2$, with χ an arbitrary random variable, it directly follows that $E[\theta_1^2]$ and $E[\theta_1^2]$ exist. Similarly, the limit of $(E[y_t^2])$ exists, and as a result so does the limit of $(E[\gamma_t^2])$.

Now applying properties 1 of theorem 4.1, along with the existence of the limit of $(E[\gamma_t])$, the criteria for order-1 stability are directly calculated as

$$-1 < E[w] < 1 \qquad \text{and} \qquad 0 < \frac{E[\theta_1] + E[\theta_2]}{E[w] + 1} < 2 \tag{42}$$

Using the criteria for order-1 stability of equation (42), and property 4 of theorem 4.1, along with the assistance of Matlab's symbolic toolbox, and using similar steps to that of Bonyadi and Michalewicz (2016b), the following necessary criteria for order-2 stability are obtained:

$$-1 < \frac{E[w]}{\sqrt{1 - V[w]}} < 1 \tag{43}$$

$$0 < E[\theta_1] + E[\theta_2] < \frac{-2(E[w]^2 + V[w] - 1)}{1 - E[w] + \frac{(V[\theta_1] + V[\theta_2])(1 + E[w])}{(E[\theta_1] + E[\theta_2])^2}} \tag{44}$$

It should be noted that the expected values of $E[\alpha]$, $E[\alpha^2]$, $E[\beta]$, $E[\beta^2]$, and $E[\alpha\beta]$ are needed for this calculation. The detailed calculation of these expected values can be found in Bonyadi and Michalewicz (2016b).

The next step is to verify that equations (43) and (44) are in fact sufficient conditions for order-2 stability, using properties 3 of theorem 4.1. The existence of the limits of $(E[\gamma_t \alpha])$ and $(E[\gamma_t \beta])$ must first be shown. Observe that

$$E[\gamma_t \alpha] = E[\theta_1]E[y_t](1 + E[w]) + E[\theta_2]E[\hat{y}_t](1 + E[w]) - E[\theta_1^2]E[y_t]$$
$$- 2E[\theta_1]E[\theta_2]E[y_t]E[\hat{y}_t] - E[\theta_2^2]E[\hat{y}_t] \tag{45}$$

and

$$E[\gamma_t \beta] = -E[w]E[\theta_1]E[y_t] - E[w]E[\theta_2]E[\hat{y}_t] \tag{46}$$

Both the limits of $(E[\gamma_t\alpha])$ and $(E[\gamma_t\beta])$ clearly exist since the limits of $(E[y_t])$ and $(E[\hat{y}_t])$ exist. In order to obtain the sufficient conditions for convergence, the condition $\rho(\boldsymbol{B}) < 1$ must be simplified. Unfortunately, due to the generality of the considered CPSO variant, the simplification of the condition $\rho(\boldsymbol{B}) < 1$ becomes intractable. However, the conditions in equations (43) and (44) can be empirically verified to be sufficient for convergence using an empirical approach similar to Bonyadi and Michalewicz (2016b). The experimental procedure is as follows: 10^{12} random combinations of the form $\{E[w], E[\theta_1], E[\theta_2], V[w], V[\theta_1], V[\theta_2]\}$ were constructed such that equations (43) and (44) were satisfied. It was then tested whether or not $\rho(\boldsymbol{B}) < 1$. It was found that in 100% of the cases, if equations (43) and (44) were satisfied, then the condition $\rho(\boldsymbol{B}_t) < 1$ held. This provides strong evidence that equations (43) and (44) are in fact also the sufficient conditions for order-2 stability.

It is also seen via the application of property 4 of theorem 4.1 that convergence of $(E[y_t])$, $(E[\hat{y}_t])$, $(V[y_t])$, and $(V[\hat{y}_t])$ are in fact necessary conditions for order-1 and order-2 stability.

It is informative to note that the criteria for order-1 and order-2 stability of the regular CPSO algorithm can be directly obtained from equations (42), (43), and (44). Let w be a constant, and let $\theta_1 = c_1 r_1$, $\theta_2 = c_2 r_2$ as in the regular CPSO algorithm. Then it follows that $E[w] = w$, $E[\theta_1] = \frac{c_1}{2}$, $E[\theta_2] = \frac{c_2}{2}$, $V[w] = 0$, $V[\theta_1] = \frac{c_1^2}{12}$, and $V[\theta_2] = \frac{c_2^2}{12}$. Substituting the calculated expectations and variances into equation (42) leads to the following criteria for order-1 stability:

$$-1 < w < 1 \qquad \text{and} \qquad 0 < c_1 + c_2 < 4(w + 1) \qquad (47)$$

The criteria for order-1 stability in equation (47) agrees with the criteria for order-1 stability as derived by (Poli, 2009) under the more restrictive stagnation assumption. Furthermore, substituting the calculated expectations and variances into equations (43) and (44) leads to the following criteria for order-2 stability:

$$-1 < w < 1 \qquad \text{and} \qquad 0 < c_1 + c_2 < \frac{24(1 - w^2)}{7 - 5w} \qquad (48)$$

The criteria in equation (48) for order-2 stability are in exact agreement with the criteria derived by (Poli, 2009) under the more restrictive stagnation assumption.

While this section focused on a general version of CPSO, the same procedure can be followed with any new or existing PSO variant that is contained in the class of positional updates described by equation (26).

It should be noted that the use of theorem 4.1 does still rely on a simplifying assumption, specifically, the non-stagnate distribution assumption. In order to verify whether or not newly derived stability criteria are truly representative of the unsimplified PSO variant under consideration, it is still recommended to perform some form of empirical verification of the criteria in an assumption free context. Such an empirical approach is detailed in Cleghorn and Engelbrecht (2015).

6 Conclusions and future work

This paper provided a meaningful extension to the theoretical stability analysis currently performed on PSO. The criteria for order-1 and order-2 particle stability

were provided for a large class of PSO variants. The stability criteria were derived by modeling, in the simplest case, the personal and neighborhood best positions as convergent sequences of random variables. It was also shown that the non-stagnant distribution assumption is a necessary condition for order-1 and order-2 stability.

In terms of potential future work, there are still a few relatively unexplored areas of PSO stability analysis. The first is the theoretical derivation of stability criteria for PSO variants where the control coefficients are time dependent, specifically the following class of PSO update equations could be considered

$$x_k(t+1) = x_k(t)\alpha_t + x_k(t-1)\beta_t + \gamma_t \tag{49}$$

where k indicates the vector component, (α_t), (β_t), and (γ_t) are sequences of random variables. The class of PSOs described by equation (49) includes numerous PSO variants where the inertia, cognitive and/or social coefficients are altered over time, as in many self-adaptive PSOs (Naka et al., 2001; Ratnaweera et al., 2003; Suganthan, 1999; Yoshida et al., 1999; Perman et al., 2003; Harrison et al., 2016). The second relatively unexplored area is to perform theoretical stability analysis on PSO variants where the particle position update equation does not operate on dimensions independently. A good example of a PSO variant with this coupling between dimensions is standard PSO 2011 (SPSO2011), as proposed by Clerc (2011), as well as the PSO variant proposed by Bonyadi et al. (2014). The required mathematical techniques needed to perform this type of analysis in a tractable fashion are, unfortunately, not immediately apparent. The last interesting areas of research is the development of an approach to predict long term behavior of PSO variants which rely on random variables without well defined order-1 and order-2 moments. An example of such a PSO variant would be that of Miranda and Fonseca (2002) if a Cauchy or Lévy distribution was sampled from for the inertia perturbation, or even for the perturbation of c_1 and c_2.

References

Atkinson, K. and Han, W. (2009). *Theoretical Numerical Analysis: A Functional Analysis Framework*. Springer, Heidelberg, Berlin.

Blackwell, T. (2012). A study of collapse in bare bones particle swarm optimzation. *IEEE Transactions on Evolutionary Computation*, 16(3):354–372.

Bonyadi, M. and Michalewicz, Z. (2016a). Particle swarm optimization for single objective continuous space problems: a review. *Evolutionary Computation*, 25(1):1–54.

Bonyadi, M. and Michalewicz, Z. (2016b). Stability analysis of the particle swarm optimization without stagnation assumption. *IEEE Transactions on Evolutionary Computation*, 20(5):814–819.

Bonyadi, M., Michalewicz, Z., and Li, X. (2014). An analysis of the velocity updating rule of the particle swarm optimization algorithm. *Journal of Heuristics*, 20(4):417–452.

Cleghorn, C. and Engelbrecht, A. (2014). A generalized theoretical deterministic particle swarm model. *Swarm Intelligence*, 8(1):35–59.

Cleghorn, C. and Engelbrecht, A. (2015). Particle swarm variants: Standardized convergence analysis. *Swarm Intelligence*, 9(2–3):177–203.

Cleghorn, C. and Engelbrecht, A. (2016). Particle swarm optimizer: The impact of unstable particles on performance. In *Proceedings of the IEEE Symposium Series on Swarm Intelligence*, pages 1–7, Piscataway, NJ. IEEE Press.

Clerc, M. (2011). Standard particle swarm optimisation. Technical report, http://clerc.maurice.free.fr/pso/SPSO_descriptions.pdf.

Clerc, M. and Kennedy, J. (2002). The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, 6(1):58–73.

Engelbrecht, A. (2013). Particle swarm optimization: Global best or local best? In *Proceedings of the 1st BRICS Countries Congress on Computational Intelligence*, pages 124–135, Piscataway, NJ. IEEE Press.

García-Gonzalo, E. and Fernández-Martinez, J. (2011). Stochastic stability analysis of the linear continuous and discrete pso models. *IEEE Transactions on Evolutionary Computation*, 15(3):405–422.

García-Gonzalo, E. and Fernández-Martinez, J. (2014a). Convergence and stochastic stability analysis of particle swarm optimization variants with generic parameter distributions. *Applied Mathematics and Computation*, 249:286–302.

García-Gonzalo, E. and Fernández-Martinez, J. (2014b). Convergence and stochastic stability analysis of particle swarm optimization variants with generic parameter distributions. *Applied Mathematics and Computation*, 249:286–302.

Gazi, V. (2012). Stochastic stability analysis of the particle dynamics in the PSO algorithm. In *Proceedings of the IEEE International Symposium on Intelligent Control*, pages 708–713, Piscataway. IEEE Press.

Harrison, K., Engelbrecht, A., and Ombuki-Berman, B. (2016). Inertia weight control strategies for particle swarm optimization. *Swarm Intelligence*, 10(4):267–305.

Jiang, M., Luo, Y., and Yang, S. (2007). Stochastic convergence analysis and parameter selection of the standard particle swarm optimization algorithm. *Information Processing Letters*, 102(1):8–16.

Kadirkamanathan, V., Selvarajah, K., and Fleming, P. (2006). Stability analysis of the particle dynamics in particle swarm optimizer. *IEEE Transactions on Evolutionary Computation*, 10(3):245–255.

Kennedy, J. (1999). Small worlds and mega-minds: effects of neighborhood topology on particle swarm performance. In *Proceedings of the IEEE Congress on Evolutionary Computation*, volume 3, pages 1931–1938, Piscataway, NJ. IEEE Press.

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, pages 1942–1948, Piscataway, NJ. IEEE Press.

Kennedy, J. and Mendes, R. (2002). Population structure and particle performance. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 1671–1676, Piscataway, NJ. IEEE Press.

Kennedy, J. and Mendes, R. (2003). Neighborhood topologies in fully-informed and best-of-neighborhood particle swarms. In *Proceedings of the IEEE International Workshop on Soft Computing in Industrial Applications*, pages 45–50, Piscataway, NJ. IEEE Press.

Kisacanin, B. and Agarwal, G. (2001). *Linear Control Systems: With Solved Problems and Matlab Examples*. Springer, New York.

Kreyszig, E. (1978). *Introductory Functional Analysis With Applications*. Wiley, Wiltshire, UK.

Liu, Q. (2015). Order-2 stability analysis of particle swarm optimization. *Evolutionary Computation*, 23(2):187–216.

Miranda, V. and Fonseca, N. (2002). New evolutionary particle swarm algorithm (epso) applied to voltage/ var control. In *Proceedings of the 14th power systems computation conference (PSCC)*, pages 1–6, Red Hook, NY. Curran associates inc.

Naka, S., Genji, T., Yara, T., and Fukuyama, Y. (2001). Practical distribution state estimation using hybrid particle swarm optimization. In *IEEE Power Engineering Society Winter Meeting*, volume 2, pages 815–820, Piscataway, NJ. IEEE Press.

Ozcan, E. and Mohan, C. (1998). Analysis of a simple particle swarm optimization system. *Intelligent Engineering Systems through Artificial Neural Networks*, volume 8:253–258.

Ozcan, E. and Mohan, C. (1999). Particle swarm optimization: Surfing the waves. In *Proceedings of the IEEE Congress on Evolutionary Computation*, volume 3, pages 1939–1944, Piscataway, NJ. IEEE Press.

Parsopoulos, K. and Vrahatis, M. (2004). UPSO: A unified particle swarm optimization scheme. In *Proceedings of the International Conference on Computational Methods in Sciences and Engineering*, pages 868–873, Netherlands. VSP International Science Publishers.

Perman, T., Veeramachaneni, K., and Mohan, C. (2003). Fitness-distance-ratio based particle swarm optimization. In *Proceedings of the IEEE Swarm Intelligence Symposium*, pages 174–181, Piscataway, NJ. IEEE Press.

Poli, R. (2008a). Analysis of the publications on the applications of particle swarm optimisation. *Journal of Artificial Evolution and Applications*, 2008:1–10.

Poli, R. (2008b). Dynamics and stability of the sampling distribution of particle swarm optimisers via moment analysis. *Journal of Artificial Evolution and Applications*, 2008:1–10.

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.

Poli, R. and Broomhead, D. (2007). Exact analysis of the sampling distribution for the canonical particle swarm optimiser and its convergence during stagnation. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 134–141, New York, NY. ACM Press.

Ratnaweera, A., Halgamuge, S., and Watson, H. (2003). Particle swarm optimization with self-adaptive acceleration coefficients. In *Proceedings of the First International Conference on Fuzzy Systems and Knowledge Discovery*, pages 264–268, Heidelberg, Berlin. Springer.

Shi, Y. and Eberhart, R. (1998). A modified particle swarm optimizer. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 69–73, Piscataway, NJ. IEEE Press.

Suganthan, P. (1999). Particle swarm optimiser with neighborhood operator. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 1958–1962, Piscataway, NJ. IEEE Press.

Trelea, I. (2003). The particle swarm optimization algorithm: Convergence analysis and parameter selection. *Information Processing Letters*, 85(6):317–325.

Van den Bergh, F. (2002). *An Analysis of Particle Swarm Optimizers*. PhD thesis, Department of Computer Science, University of Pretoria, Pretoria, South Africa.

Van den Bergh, F. and Engelbrecht, A. (2006). A study of particle swarm optimization particle trajectories. *Information Sciences*, 176(8):937–971.

Yoshida, H., Fukuyama, Y., Takayama, S., and Nakanishi, Y. (1999). A particle swarm optimization for reactive power and voltage control in electric power systems considering voltage security assessment. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, volume 6, pages 497–502, Piscataway, NJ. IEEE Press.

A Eigenvalues from Theoretical Analysis

For the sake of completeness, this appendix contains the five eigenvalues of matrix $\boldsymbol{B}_t$, from section 4. The presented eigenvalues were generated using Matlab's symbolic toolbox. Given the length of some of the eigenvalues, they have been typeset in a compact manner.

Eigenvalue 1:

$E[\alpha]/2 - (E[\alpha]\hat{} 2 + 4E[\beta])\hat{} (1/2)/2$

Eigenvalue 2:

$E[\alpha]/2 + (E[\alpha]\hat{} 2 + 4E[\beta])\hat{} (1/2)/2$

Eigenvalue 3:

$E[\beta]/3 + E[\alpha^2]/3 + (E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{} 2/9$ - $(E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)/((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2 + (((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2)\hat{} 2$ - $(E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{} 2/9$ - $(E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)\hat{} 3)\hat{} (1/2))\hat{} (1/3) + ((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2 + (((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2)\hat{} 2$ - $(E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{} 2/9$ - $(E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)\hat{} 3)\hat{} (1/2))\hat{} (1/3)$

Eigenvalue 4:

$E[\beta]/3 + E[\alpha^2]/3$ - $(3\hat{} (1/2)((E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{} 2/9$ - $(E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)/((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2 + (((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2)\hat{} 2$ - $(E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{} 2/9$ - $(E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)\hat{} 3)\hat{} (1/2))\hat{} (1/3)$ - $((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2 + (((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2)\hat{} 2$ - $(E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{} 2/9$ - $(E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)\hat{} 3)\hat{} (1/2))\hat{} (1/3))/(2((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2 + (((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2)\hat{} 2$ - $(E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{} 2/9$ - $(E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)\hat{} 3)\hat{} (1/2))\hat{} (1/3))$ - $((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2 + (((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2)\hat{} 2$ - $(E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{} 2/9$ - $(E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)\hat{} 3)\hat{} (1/2))\hat{} (1/3)/2$

Eigenvalue 5:

$E[\beta]/3 + E[\alpha^2]/3 + (3\hat{} (1/2)((E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{} 2/9$ - $(E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)/((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2 + (((E[\beta] + E[\alpha^2])\hat{} 3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2]$ - $E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2$ - $(E[\beta]E[\beta^2])/2)\hat{} 2$ - $(E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{} 2/9$ - $(E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)\hat{} 3)\hat{} (1/2))\hat{} (1/3)$ - $((E[\beta] + E[\alpha^2])\hat{} 3/27 +$

$((E[\beta] + E[\alpha^2])(E[\beta^2] - E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2 - (E[\beta]E[\beta^2])/2 + (((E[\beta] + E[\alpha^2])\hat{}3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2] - E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2 - (E[\beta]E[\beta^2])/2)\hat{}2 - (E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{}2/9 - (E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)\hat{}3)\hat{}(1/2))\hat{}(1/3))1)/2 - (E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{}2/9 - (E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)/(2((E[\beta] + E[\alpha^2])\hat{}3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2] - E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2 - (E[\beta]E[\beta^2])/2 + (((E[\beta] + E[\alpha^2])\hat{}3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2] - E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2 - (E[\beta]E[\beta^2])/2)\hat{}2 - (E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{}2/9 - (E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)\hat{}3)\hat{}(1/2))\hat{}(1/3)) - ((E[\beta] + E[\alpha^2])\hat{}3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2] - E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2 - (E[\beta]E[\beta^2])/2 + (((E[\beta] + E[\alpha^2])\hat{}3/27 + ((E[\beta] + E[\alpha^2])(E[\beta^2] - E[\beta]E[\alpha^2] + 2E[\alpha]E[\alpha\beta]))/6 + (E[\alpha]E[\beta^2])/2 - (E[\beta]E[\beta^2])/2)\hat{}2 - (E[\beta^2]/3 + (E[\beta] + E[\alpha^2])\hat{}2/9 - (E[\beta]E[\alpha^2])/3 + (2E[\alpha]E[\alpha\beta])/3)\hat{}3)\hat{}(1/2))\hat{}(1/3)/2$