UNIVERSITY OF PRETORIA

Particle Swarm Optimization: Empirical and Theoretical Stability Analysis

by

Christopher Wesley Cleghorn

Submitted in partial fulfillment of the requirements for the degree
Philosophiae Doctor (Computer Science)

in the Faculty of
Engineering, Built Environment and Information Technology

June 2017

"Machines take me by surprise with great frequency."

Alan Turning

UNIVERSITY OF PRETORIA

Particle Swarm Optimization: Empirical and Theoretical Stability Analysis

Abstract

By

Christopher Wesley Cleghorn

Particle swarm optimization (PSO) is a well-known stochastic population-based search algorithm, originally developed by Kennedy and Eberhart in 1995. Given PSO's success at solving numerous real world problems, a large number of PSO variants have been proposed. However, unlike the original PSO, most variants currently have little to no existing theoretical results. This lack of a theoretical underpinning makes it difficult, if not impossible, for practitioners to make informed decisions about the algorithmic setup. This thesis focuses on the criteria needed for particle stability, or as it is often refereed to as, particle convergence.

While new PSO variants are proposed at a rapid rate, the theoretical analysis often takes substantially longer to emerge, if at all. In some situation the theoretical analysis is not performed as the mathematical models needed to actually represent the PSO variants become too complex or contain intractable subproblems. It is for this reason that a rapid means of determining approximate stability criteria that does not require complex mathematical modeling is needed. This thesis presents an empirical approach for determining the stability criteria for PSO variants. This approach is designed to provide a real world depiction of particle stability by imposing absolutely no simplifying assumption on the underlying PSO variant being investigated. This approach is utilized to identify a number of previously unknown stability criteria.

This thesis also contains novel theoretical derivations of the stability criteria for both the fully informed PSO and the unified PSO. The theoretical models are then empirically validated utilizing the aforementioned empirical approach in an assumption free context.

The thesis closes with a substantial theoretical extension of current PSO stability research. It is common practice within the existing theoretical PSO research to assume that, in the simplest case, the personal and neighborhood best positions are stagnant. However, in this thesis, stability criteria are derived under a mathematical model where by the personal best and neighborhood best positions are treated as convergent sequences of random variables. It is also proved that, in order to derive stability criteria, no weaker assumption on the behavior of the personal and neighborhood best positions can be made. The theoretical extension presented caters for a large range of PSO variants.

Supervisor : Professor Andries P. Engelbrecht

Department : Department of Computer Science

Degree : Doctor of Philosophy

Acknowledgements

This work would not have seen completion without the following people:

- My supervisor, Prof. Andries Engelbrecht, for his crucial academic and financial support.
- My parents, Robert and Joy Cleghorn, for there unwavering support and encouragement.
- My partner, Belinda Stapelberg, for all her love and support.

This work has been partly supported by bursaries from:

- The National Research Foundation (http://www.nrf.ac.za) through the Computational Intelligence Research Group (http://cirg.cs.up.ac.za).
- The University of Pretoria.

Contents

Chapter 1

Introduction

1.1 Motivation

In the field of computational intelligence there exists an incredibly vast number of optimization algorithms. Many optimization algorithms are designed based on analogies, often inspired by nature, but the resulting algorithmic behavior is often ignored in favor of a simple empirical performance analysis. As a result, most computational intelligence based optimization algorithms are not truly understood from a theoretical perspective. This leaves two important questions open: Firstly, does the algorithm actually exhibit behavior in line with its initial intent? Secondly, is the algorithm's behavior predictable, and if so what kind of guarantees can be made about the algorithm? This thesis focuses on answering the second question in the context of particle swarm optimization (PSO).

There are numerous aspects of an optimization technique that can be investigated, ranging from an algorithm's computational complexity, to the more specific property of rotational invariance. This thesis focuses on the stability of PSO and its variants' particles.

While there exists a number of theoretical results about PSO, most of PSO's variants have undergone little or no analysis. The reason for this is two fold: Firstly, PSO variants are introduced at such a rapid rate that it is very difficult for the theory to catch up, given the complexity involved in modeling the algorithm. The second reason is that some PSO variants contain internal mechanisms that are fundamentally hard to model

in practice, for example, dependencies between multiple particles or even sub-swarms, possible dependencies between previous time steps, the influence of the underlying objective function on the algorithm behavior, to name but a few. The task of constructing a model of an algorithm, that is both reflective enough of the actual algorithm and simple enough to allow for meaningful conclusions to be derived from that model, is far from straightforward. It is for this reason that an rapid empirical approach for investigating stability criteria for PSO variants is needed. This thesis provides such an approach.

The empirical approach should have an additional use, namely the verification of theoretically derived stability criteria. All stability analysis performed on PSOs rely on a variant of the stagnation assumption, where by the updating of the personal and neighborhood best positions is heavily restricted. In the most extreme case, the personal and neighborhood best positions are held constant. While restriction of the personal and neighborhood best positions' ability to update can allow for effective modeling of the PSO's behavior, all derived results should still be empirically verified against a complete, and unsimplified PSO. It is for this reason that a simple method for verifying the criteria for stability is needed.

The use of a stagnation assumption is common place in theoretical analysis of PSOs. An immediate question is how weak of a stagntion assumption can be made, while still allowing for stability criteria to be derived. This thesis investigates this important question for a wide class of PSO variants.

1.2 Objectives

The primary objectives of this thesis are summarized as follows: To

- provide an empirical approach for exploratory stability analysis of new PSO variants;
- provide an empirical approach that can be utilized to verify theoretically derived stability criteria in an assumption free context;

- make use of the proposed empirical verification approach to verify newly derived stability criteria for PSO variants;
- provide a theoretical extension to the existing PSO stability research that relies on the weakest possible assumption on the behavior of the personal best and neighborhood best positions. The provided extension should be sufficiently general to cater for a large class of PSO variants.

1.3 Publication Derived from the Thesis

This section provides a list of all published and submitted conference and journal article derived from the content of this thesis.

The following is a list of published work:

- C.W. Cleghorn and A.P. Engelbrecht. Particle swarm convergence: An empirical investigation. In Proceedings of the IEEE Congress on Evolutionary Computation, pages 2524–2530, Piscataway, NJ, 2014. IEEE Press.
- C.W. Cleghorn and A.P. Engelbrecht. Particle swarm convergence: Standardized analysis and topological influence. In Proceedings of International Swarm Intelligence Conference (ANTS), Swarm Intelligence, pages 134–145, Switzerland, 2014. Springer International Publishing.
- C.W. Cleghorn and A.P. Engelbrecht. Fully informed particle swarm optimizer: Convergence analysis. In Proceedings of the IEEE Congress on Evolutionary Computation, pages 164–170, Piscataway, NJ, 2015. IEEE Press.
- C.W. Cleghorn and A.P. Engelbrecht. Particle swarm variants: Standardized convergence analysis. Swarm Intelligence, 9(2–3):177–203, 2015.
- C.W. Cleghorn and A.P. Engelbrecht. Unified particle swarm optimizer: Convergence analysis. In Proceedings of the IEEE Congress on Evolutionary Computation, pages 448–454, Piscataway, NJ, 2016. IEEE Press.

- C.W. Cleghorn and A.P. Engelbrecht. 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, 2016. IEEE Press.

The following work has been submitted for review:

- Particle Swarm Stability: Weakest Allowable Assumption on Particle Informers. Swarm Intelligence Journal, 2017

1.4 Thesis Outline

Chapter 2 presents an algorithmic description of PSO and the PSO variants directly relevant to this thesis. Specifically, the original PSO as developed by Kennedy and Eberhart [31], with the inclusion of the inertia coefficient as proposed by Shi and Eberhart [56] is discussed in detail. A detailed explanation of fully informed PSO (FIPS), unified PSO (UPSO), bare bones PSO (BPSO), and standard PSO 2011 (SPSO2011) is also given.

Chapter 3 presents existing theoretical stability results for PSO and PSO variants. Commonly utilized assumptions present in PSO stability analysis literature are presented along with an overview of the types of convergence relevant to PSO stability analysis. Existing theoretical stability results for PSO, FIPS, BPSO, and SPSO2011 are presented.

Chapter 4 investigates, using a novel empirical approach, which of the existing theoretically derived stability regions for PSO accurately predicts the true unsimplified PSO stability region.

Chapter 5 provides a description of a standardized approach for performing stability analysis on PSO variants. The effectiveness of a specifically designed objective function for stability analysis is shown. The empirical approach for stability analysis is utilized to test the influence of the social network structure on PSO's stability criteria. The stability criteria for FIPS, BPSO, and SPSO2011 are analyzed.

Chapter 6 focuses on the stability criteria of the fully informed PSO. A novel theoretical derivation of the stability criteria for FIPS is provided followed by an experimental verification of the derived stability criteria.

The stability criteria of the unified PSO is the focus of chapter 7. The theoretical derivation of the stability criteria for UPSO is presented as well as an experimental verification of the derived stability criteria.

Chapter 8 provides an empirical study of the relationship between particle stability and PSO's performance.

Chapter 9 presents an extension to the state of the art theoretical model utilized for understanding the stability of PSO's particles. A novel theorem for deriving stability criteria for a large class of PSO variants, under what is shown to be the weakest allowable assumption on the personal and neighborhood best positions, is proved. The application of the proved theorem is shown by deriving stability criteria for a generalized PSO where all coefficients can be modeled as random variables with arbitrary distributions.

Chapter 10 presents a summary of the findings of this thesis and topics for future research are also discussed.

Chapter 2

Particle Swarm Optimization

This chapter presents a detailed discussion and description of PSO and popular PSO variants. Section 2.1 presents the PSO algorithm, and variants of PSO are discussed in section 2.2.

2.1 The Particle Swarm Optimization Algorithm

This section introduces the particle swarm optimization (PSO) algorithm. Section 2.1.1 provides a detailed algorithmic description of PSO, followed by a discussion of PSO's social network structure in 2.1.2.

2.1.1 Algorithmic Description and Origin of PSO

Particle swarm optimization (PSO) was originally developed by Kennedy and Eberhart [31] to simulate the complex movement of birds in a flock. The variant of PSO this section focuses on includes the inertia coefficient proposed by Shi and Eberhart [56], which, for simplicity is referred to as the PSO algorithm in this thesis.

The PSO algorithm is defined as follows: Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ 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. An optimum $\boldsymbol{o} \in \mathbb{R}^d$ is defined such that, for all $\boldsymbol{x} \in \mathbb{R}^d$, $f(\boldsymbol{o}) \leq f(\boldsymbol{x})$. In this thesis, 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

$$\boldsymbol{x}_i(t+1) = \boldsymbol{x}_i(t) + \boldsymbol{v}_i(t+1), \tag{2.1}$$

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

$$\boldsymbol{v}_i(t+1) = w\boldsymbol{v}_i(t) + c_1\boldsymbol{r}_1(t) \otimes (\boldsymbol{y}_i(t) - \boldsymbol{x}_i(t)) + c_2\boldsymbol{r}_2(t) \otimes (\hat{\boldsymbol{y}}_i(t) - \boldsymbol{x}_i(t)), \tag{2.2}$$

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.

The PSO algorithm is summarized in algorithm 1.

2.1.2 Social Network Structure

One of the driving features of the PSO is social interaction, specifically the way in which knowledge is shared amongst the particles in the swarm. In general, the social network structure of a swarm can be viewed as a graph, where nodes represent particles, and the edges are the allowable direct communication routes. Direct knowledge sharing is only performed if the particles are directly connected. The group of particles that particle i is directly connected to is called the neighbourhood $\mathcal{N}_i$ of particle i. For example, consider the graph in figure 2.1 with four particles, then $\mathcal{N}_1 = \{1,2\}$, $\mathcal{N}_2 = \{1,2,3\}$, $\mathcal{N}_3 = \{2,3,4\}$, and $\mathcal{N}_4 = \{3,4\}$. It is also possible to have the social network set up in such a manner that the communication link is not bi-directional. The graph then becomes directional, as in the work of Mohais *et al* [40].

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 $\mathbf{0}$.
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
  for all particles $i = 1, \cdots, N$ **do**
    update the velocity of particle i using equation (2.2)
    update the position of particle i using equation (2.1)
  end for
until stopping condition is met

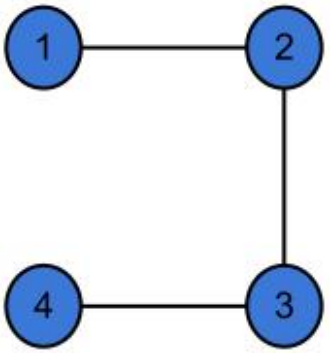


FIGURE 2.1: Simple Social Network.

The social network structure chosen has a direct impact on the behaviour of the swarm as a whole [29, 32, 46]. When a social network is highly connected, knowledge of the best particle (its best location) is quickly propagated through the social network. From an optimization perspective, this implies a faster rate of convergence to an optimum, as opposed to that of a less connected social network structure. This faster rate of convergence to an optimum comes at the cost of a greater susceptibility to getting stuck in local optima. There are many factors that play a role in the flow of information through a social network. These factors are well explained by Watts and Strogatz [62].

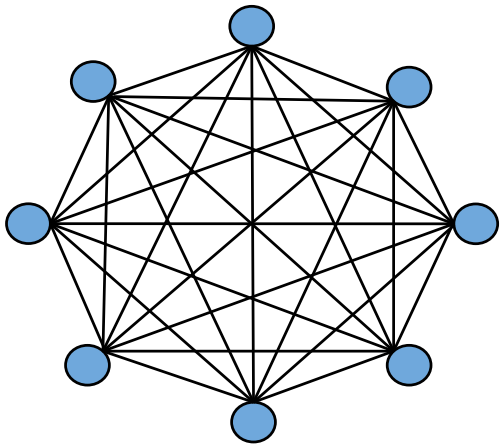
Some of the most frequently used social network structures are discussed below:

- **Star**: The star network structure is one where all the particles in the swarm are interconnected as illustrated in figure 2.2a. The original implementation of the PSO algorithm utilized the star network structure. A PSO utilizing the star network structure is commonly referred to as the Gbest PSO.
- **Ring**: The ring network structure is one where each particle is in a neighbourhood with only two other particles, with the resulting structure forming a ring as illustrated in figure 2.2b. The ring network structure can be generalized to a network structure where larger neighbourhoods are used.
- **Von Neumann**: The Von Neumann network structure is one where the particles are arranged in a grid-like structure. The 2-D variant is illustrated in figure 2.2c, and the 3-D variant is illustrated in figure 2.2d.

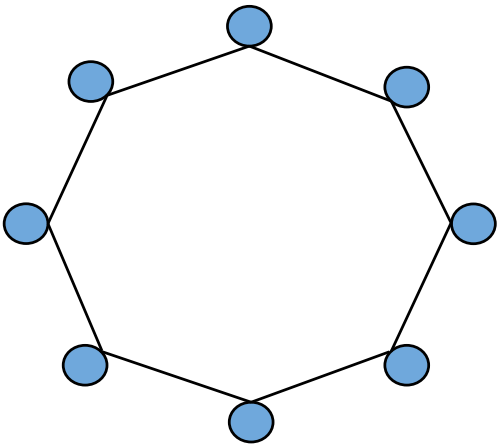
There are many other available social network structures, such as the wheel, pyramid, four cluster, random, amongst others [32]. Each social network structure has certain advantages and disadvantages, with no social neighbourhood structure that can be labelled as the "best". The optimal social network structure choice is problem dependent [21, 52]. That being said, a number of empirical studies have shown that the Von Neumann network structure outperforms other network structure on a large array of problems [32, 46].

2.2 Popular Particle Swarm Optimization Variants

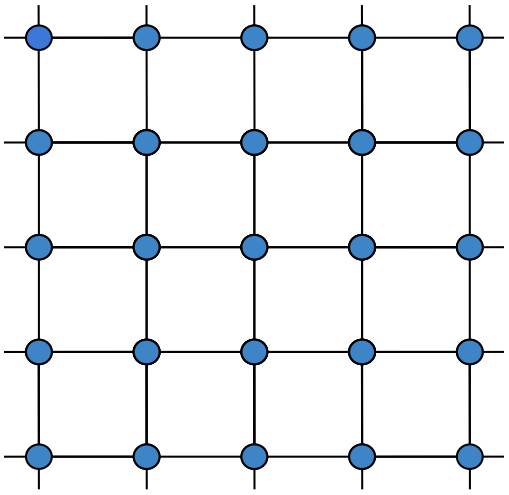
There exists a large number of PSO variants [6, 19, 20]. The simplest variants alter one or more of the PSO velocity update equation's coefficients to be a function of time [42, 47, 55, 57, 64], in an attempt to control the exploration-exploitation behavior of the swarm over the course of the search. There are also more sophisticated PSO variants that substantially alter the PSO's behavior. This section focuses on four of these variants that are commonly used: the fully informed PSO (FIPS), unified PSO (USPO), bare



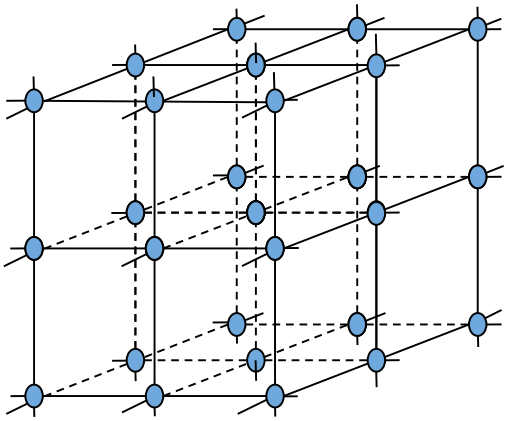
(a) Star topology.



(b) Ring topology.



(c) 2-D von Neumann topology.



(d) 3-D von Neumann topology.

FIGURE 2.2: Common social topologies

bones PSO (BPSO), and the standard PSO 2011 (SPSO2011). All variants discussed are either empirically or theoretically analyzed within this thesis.

2.2.1 Fully Informed PSO

The FIPS algorithm was inspired by the observation made by Kennedy and Mendes [33] that human individuals are not influenced by only a single individual, but rather by a statistical summary of the state of their neighborhood. Based on this observation, the velocity equation is altered such that each particle is influenced by the successes of all its neighbors, and not by the performance of only one other individual in the neighborhood.

The velocity update equation of FIPS is defined as follows:

$$\boldsymbol{v}_i(t+1) = w\boldsymbol{v}_i(t) + \sum_{m=1}^{|\mathcal{N}_i|} \boldsymbol{\gamma}_m(t) \otimes \frac{(\boldsymbol{y}_m(t) - \boldsymbol{x}_i(t))}{|\mathcal{N}_i|}, \tag{2.3}$$

where $\mathcal{N}_i$ is set of particles in particle i's neighborhood, $|\mathcal{N}_i|$ is the cardinality of $\mathcal{N}_i$, and $\gamma_{m,k}(t) \sim U(0, c_1 + c_2)$ for $1 \leq k \leq d$.

The FIPS algorithm was originally proposed using constriction. For a detailed explanation of constriction the reader is referred to [17]. However, the velocity update equation (2.3) can be rewritten to utilize constriction instead of an inertia weight by setting the constriction factor $\mathcal{X}$ equal to w, and $(c_1 + c_2)/\mathcal{X} = \hat{c}_1 + \hat{c}_2$, where $\hat{c}_1$ and $\hat{c}_2$ are coefficients chosen for a PSO using constriction. From a theoretical perspective, the models are equivalent.

2.2.2 Unified PSO

The UPSO algorithm was developed by Parsopoulos and Vrahatis [45]. The authors argued that it would be beneficial to be able to utilize the exploitative nature of global best PSO (Gbest PSO) and the exploratory nature of local best PSO (Lbest PSO) in one unified scheme. Gbest PSO is defined as a PSO that draws its social knowledge from the best performing particle within its swarm. On the other hand, Lbest PSO rather draws its social knowledge from the best performing particle within its neighborhood, often a subset of the swarm connected in a ring structure [31].

Based on this idea, the velocity update equations of PSO was altered to be a combination of the Gbest and Lbest velocity update equations. The weighting of Lbest and Gbest is controlled by an additional control parameter, $u \in [0, 1]$, called the unification factor.

The velocity and position update equations of UPSO are defined as follows:

$$\boldsymbol{g}_i(t+1) = w\boldsymbol{v}_i(t) + c_1\boldsymbol{r_1} \otimes (\boldsymbol{y}_i(t) - \boldsymbol{x}_i(t)) + c_2\boldsymbol{r_2} \otimes (\mathbf{g}(t) - \boldsymbol{x}_i(t))$$

$$\boldsymbol{l}_i(t+1) = w\boldsymbol{v}_i(t) + c_1\boldsymbol{r'_1} \otimes (\boldsymbol{y}_i(t) - \boldsymbol{x}_i(t)) + c_2\boldsymbol{r'_2} \otimes (\hat{\boldsymbol{y}}_i(t) - \boldsymbol{x}_i(t))$$

$$\boldsymbol{v}_i(t+1) = u\boldsymbol{g}_i(t+1) + (1-u)\,\boldsymbol{l}_i(t+1) \tag{2.4}$$

$$\boldsymbol{x}_i(t+1) = \boldsymbol{x}_i(t) + \boldsymbol{v}_i(t+1), \tag{2.5}$$

where $\boldsymbol{r_1}, \boldsymbol{r_2}, \boldsymbol{r'_1}, \boldsymbol{r'_2} \sim U(0,1)^d$, with d the dimension of the problem PSO is attempting to solve. The positions $\boldsymbol{y}_i$ and $\hat{\boldsymbol{y}}_i$ are as defined for PSO. The position $\mathbf{g}$ is the "best" position found by the swarm so far. If the star network structure is utilized $\mathbf{g}$ and $\hat{\boldsymbol{y}}_i(t)$ represent the same position for all particles, however, if the network structure is not fully connected, $\mathbf{g}$ and $\hat{\boldsymbol{y}}_i(t)$ may be different. The coefficients c_1, c_2, and w are the cognitive, social, and inertia weights respectively.

Parsopoulos and Vrahatis [45] also proposed the two variants to the velocity update equation, namely

$$\mathbf{v}_i(t+1) = u\mathbf{r_3} \otimes \mathbf{g}_i(t+1) + (1-u)\,\mathbf{l}_i(t+1) \tag{2.6}$$

and

$$\mathbf{v}_i(t+1) = u\mathbf{g}_i(t+1) + \mathbf{r_3} \otimes (1-u)\,\mathbf{l}_i(t+1) \tag{2.7}$$

where $\boldsymbol{r}_3 \sim N(\mu, \Sigma)$, and $\Sigma = \sigma^2\boldsymbol{I}$, with $\boldsymbol{I}$ the identity matrix. This thesis focuses on the standard version of UPSO using update equations (2.4) and (2.5).

The UPSO algorithm was also originally proposed using constriction. However, the velocity update in equation (2.4) can be rewritten to utilize constriction utilizing the same approach as discussed in section 2.2.1. From a theoretical perspective the models are equivalent.

2.2.3 Bare Bones PSO

Kennedy [30] proposed the BPSO algorithm based on the empirical observation, of PSO under the star topology, that the distribution of particle positions are centered around the weighted average between the personal best and global best positions, specifically,

$$\frac{\boldsymbol{y}_i(t) + \hat{\boldsymbol{y}}_i(t)}{2} \tag{2.8}$$

assuming that c_1 and c_2 are equal. In a generalized context, the particle positions are centered around

$$\frac{c_1\boldsymbol{y}_i(t) + c_2\hat{\boldsymbol{y}}_i(t)}{c_1 + c_2}. \tag{2.9}$$

This observation was later supported by the theoretical work of Van den Bergh and Engelbrecht [61] and Trelea [59], where it was shown under the deterministic and stagnation assumption, as defined in section 3.1, that each particle converges to the point defined in equation (2.9) (assuming a star neighborhood topology).

For BPSO, the velocity update equation changes to

$$v_{i,k}\left(t+1\right) \quad \sim \quad N\left(\zeta_{i,k}, \phi_{i,k}\left(t\right)\right), \tag{2.10}$$

where $\phi_{i,k}\left(t\right) = |y_{i,k}(t) - \hat{y}_{i,k}(t)|$, and $\boldsymbol{\zeta}_i$ is equal to equation (2.9) . The position update equation is changed to

$$\boldsymbol{x}_i\left(t+1\right) \quad = \quad \boldsymbol{v}_i\left(t+1\right). \tag{2.11}$$

In the standard implementation of BPSO [30], equation (2.8) is used as the point of convergence, $\boldsymbol{\zeta}_i$. For the purposes of this thesis, the case where c_1 and c_2 are equal is treated as a special case of the BPSO in all subsequent stability analysis.

2.2.4 Standard PSO 2011

Clerc [16] developed SPSO2011 in an attempt to define a new baseline for future PSO improvements. The two primary benefits of the SPSO2011 are stated to be rotational invariance and an adaptive topology. The first published work on SPSO2011 was by Zambrano-Bigiarini and Clerc [65]. The particle velocity update equation is defined as follows:

$$\boldsymbol{v}_i(t+1) = w\boldsymbol{v}_i(t) + \mathcal{H}_i(\boldsymbol{c}_i(t), ||\boldsymbol{c}_i - \boldsymbol{x}_i||_2) - \boldsymbol{x}_i(t), \tag{2.12}$$

where $\boldsymbol{c}_i$ is defined as

$$\boldsymbol{c}_i(t) = \frac{\boldsymbol{x}_i(t) + \boldsymbol{a}_i(t) + \boldsymbol{b}_i(t)}{3}, \tag{2.13}$$

with $\boldsymbol{a}_i(t)$ and $\boldsymbol{b}_i(t)$ are defined as

$$\boldsymbol{a}_i(t) = \boldsymbol{x}_i(t) + c_1\boldsymbol{r}_1 \otimes (\boldsymbol{y}_i(t) - \boldsymbol{x}_i(t)), \tag{2.14}$$

$$\boldsymbol{b}_i(t) = \boldsymbol{x}_i(t) + c_2\boldsymbol{r}_2 \otimes (\hat{\boldsymbol{y}}_i(t) - \boldsymbol{x}_i(t)). \tag{2.15}$$

The function $\mathcal{H}_i(\boldsymbol{c}_i(t), ||\boldsymbol{c}_i - \boldsymbol{x}_i||_2)$ returns a uniformly sampled random position from a hyper-sphere centered at $\boldsymbol{c}_i(t)$ with a radius of $||\boldsymbol{c}_i - \boldsymbol{x}_i||_2$.

The samples from $\mathcal{H}_i$ are obtained using the following approach: Construct a random k dimensional vector, $\boldsymbol{rv}$, whose scalar components are sampled from the normal distribution, $N(0,1)$. The random vector must then be normalized, and multiplied by a random scalar sampled uniformly from 0 to the hypersphere's radius. The random vector $\boldsymbol{rv}$ must then be translated to the specified center point.

In the work of Zambrano-Bigiarini and Clerc [65], no special consideration was explicitly made for the case where $\boldsymbol{y}_i(t) = \hat{\boldsymbol{y}}_i(t)$. However, in the original work, Clerc [16] replaced equation (2.13) with

$$\boldsymbol{c}_i(t) = \frac{\boldsymbol{x}_i(t) + (\boldsymbol{x}_i(t) + c(\boldsymbol{y}_i(t) - \boldsymbol{x}_i(t)))}{2}. \tag{2.16}$$

This thesis uses the following equation for the center of gravity, which applies the same principle used by Zambrano-Bigiarini and Clerc [65] for the case where $\boldsymbol{y}_i(t) \neq \hat{\boldsymbol{y}}_i(t)$:

$$\boldsymbol{c}_i\left(t\right) = \frac{\boldsymbol{x}_i\left(t\right) + \boldsymbol{a}_i\left(t\right)}{2}. \tag{2.17}$$

The topology used by SPSO2011 is a particular case of the stochastic star topology proposed by [38]. On initialization, each particle's neighborhood is constructed by selecting three particles randomly from the swarm and the particle itself (the same particle is allowed to be chosen several times). If an unsuccessful iteration occurs, the neighborhoods are reconstructed. An unsuccessful iteration is defined as an iteration where no new position was found that improved the previous best objective evaluation obtained by the whole swarm.

In the work of Zambrano-Bigiarini and Clerc [65], the SPSO2011 algorithm prevents particles from leaving the search space by setting the component of the particle that breached the search space boundary to the boundary value and the particle's whole velocity to zero.

2.3 Summary

This chapter discussed the PSO algorithm and PSO variants that are used in this thesis. The next chapter provides a detailed discussion on the existing theoretical work on stability analysis for both PSO and PSO variants.

Chapter 3

Review of Existing Theoretical Stability Analysis of Particle Swarm optimization

The focus of this chapter is on the exiting theoretical stability results for PSO and PSO variants. Section 3.1 presents the commonly used assumptions in existing theoretical studies on particle stability. Section 3.2 presents an overview of the types of convergence and their relationship to stability analysis of PSO. The relevant theoretical findings for PSO, FIPS, BPSO, and SPSO2011 are presented in sections 3.3, 3.4, 3.5, 3.6, and 3.7 respectively. A summary of this chapter is presented in section 3.8.

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 section 3.3.

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

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.

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.

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.

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 value.

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.

Each of the assumptions mentioned in this section simplifies the PSO algorithm in order to allow for a 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 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 not a very 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 better than the stagnant global best position. The most recently used assumption is the stagnant distribution assumption, which allows a potentially infinite number of different personal and neighborhood best position. 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.

3.2 Convergence and Stability

This section discusses the types of convergence used in the stability analysis of PSO and PSO's variants.

In the context of 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 3.1. Convergent sequence

The sequence $(\boldsymbol{s}_t)$ in $\mathbb{R}^n$ is convergent if there exists an $\boldsymbol{s} \in \mathbb{R}^n$ such that

$$\lim_{t \to \infty} \boldsymbol{s}_t = \boldsymbol{s} \tag{3.1}$$

It should be made clear that the convergence as defined in equation (3.1) does not imply convergence to an optimum. The convergence, as described in definition 3.1, 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. Consider the following illustrative example of a stochastic sequence,

$$\upsilon_t \sim U\left(\frac{1}{t}, \frac{1}{t} + 1\right). \tag{3.2}$$

Clearly, (υ_t) does not convergence according to definition 3.1. However, the sequence does converge in some weaker sense. Observe that, given enough time, the expected value, $E[\upsilon]$, of the sequence converges. Specifically,

$$\lim_{t\to\infty} E[\upsilon_t] = \lim_{t\to\infty}\left(\frac{1}{t} + \frac{1}{2}\right) = \frac{1}{2} \tag{3.3}$$

This type of convergence is now defined as

Definition 3.2. Order-1 stability

The sequence $(\boldsymbol{s}_t)$ in $\mathbb{R}^n$ is order-1 stable if there exists an $\boldsymbol{s}_E \in \mathbb{R}^n$ such that

$$\lim_{t\to\infty} E[\boldsymbol{s}_t] = \boldsymbol{s}_E \tag{3.4}$$

where $E[\boldsymbol{s}_t]$ is the expectation of $\boldsymbol{s}_t$.

While order-1 stability is useful, it does not necessarily provide a strong enough level of stability, which is now illustrated: Consider the random sequence, defined as

$$\hat{r}_t \sim U(-t, t). \tag{3.5}$$

Now, the expectation of $\hat{r}_t$ is zero for every t, which implies the sequence $(\hat{r}_t)$ is order-1 stable. However, the variance of the sequence $(\hat{r}_t)$ is increasing over time, which does not appear particularly stable. Hence the introduction of the following type of convergence:

Definition 3.3. Order-2 stability

The sequence $(\boldsymbol{s}_t)$ in $\mathbb{R}^n$ is order-2 stable if there exists a $\boldsymbol{s}_V \in \mathbb{R}^n$ such that

$$\lim_{t\to\infty} V[\boldsymbol{s}_t] = \boldsymbol{s}_V \tag{3.6}$$

where $V[\boldsymbol{s}_t]$ is the variance of $\boldsymbol{s}_t$.

The earlier sequence in equation (3.2) is order-2 stable, because

$$\lim_{t\to\infty} V[\upsilon_t] = \lim_{t\to\infty} (E[\upsilon_t^2] - (E[\upsilon_t])^2) = \frac{1}{12}. \tag{3.7}$$

However, the sequence in equation (3.5) is not order-2 stable, because

$$\lim_{t\to\infty} V[\hat{r}_t] = \lim_{t\to\infty} (E[\hat{r}_t{}^2] - (E[\hat{r}_t])^2) = \lim_{t\to\infty} \frac{t^2}{3} \to \infty \tag{3.8}$$

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 [49], it is generally the aim of PSO stability analysis to obtain the criteria needed to ensure order-1 and order-2 stability. If a particle position is order-1 and order-2 stable, it is often referred to as convergent. However, this does not imply the deterministic convergence of definition 3.1.

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 trajectory [15, 17, 43, 44]. 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 [60, 61], and that of Trelea [59]. Both 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 is

$$0 < c_1 + c_2 < 2\,(1 + w)\,, \quad |w| < 1, \tag{3.9}$$

whereas the region derived by Trelea is

$$0 < c_1 + c_2 < 4\,(1 + w)\,, \quad |w| < 1, \tag{3.10}$$

The discrepancy between equation (3.9) and (3.10) is due to how the stochastic components where handled. In the work of Van den Bergh and Engelbrecht the stochastic components of the PSO update equation where set such that $r_{1,k} = r_{2,k} = 1$. This can be seen as a more conservative approach than that used by Trelea, where $r_{1,k} = r_{2,k} = 1/2$, which is the expected value of the random components. Equation (3.9) and (3.10) are illustrated in figure 3.1 as the triangles AFB and ACB, respectively.

More recently, under the deterministic and weak chaotic assumption Cleghorn and Engelbrecht [8] derived the same region as equation (3.9) or (3.10), depending on how the stochastic variables $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are set.

Under the stagnation assumption only, Kadirkamanathan *et al* [27] derived the following sufficient region for order-2 stability:

$$\begin{cases} c_1 + c_2 < 2\,(1 + w) & \text{for} \quad w \in (-1, 0] \\ c_1 + c_2 < \frac{2(1-w)^2}{1+w} & \text{for} \quad w \in (0, 1)\,. \end{cases} \tag{3.11}$$

Still under the stagnation assumption, Gazi [25] expanded the derived region of equation (3.11), resulting in the region

$$\begin{cases} c_1 + c_2 < \frac{24(1+w)}{7} & \text{for} \quad w \in (-1, 0] \\ c_1 + c_2 < \frac{24(1-w)^2}{7(1+w)} & \text{for} \quad w \in (0, 1)\,. \end{cases} \tag{3.12}$$

The regions corresponding to equations (3.11) and (3.12) are illustrated in figure 3.1 as triangle like regions ADB and AEB, respectively. Unfortunately, both equations (3.11) and (3.12) are very conservative regions, as they were derived utilizing the Lyapunov condition [34].

Without the use of the Lyapunov condition or the stochastic assumption, Poli and Broomhead [51] and Poli [50] 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 (3.10). The region derived for order-2 stability is as follows:

$$c_1 + c_2 < \frac{24\left(1 - w^2\right)}{7 - 5w} \quad \text{for} \quad w \in [-1, 1]\,. \tag{3.13}$$

The region defined by equation (3.13) is illustrated in figure 3.1 as the curved line segment AB. The region defined by equation (3.13) was also independently derived by Jiang [26] under the stagnation assumption.

Blackwell [30] showed that the criteria of equation (3.13) 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 [24] 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 the modeling the PSO as a stochastic damped mass-spring system [23].

Liu [37] rederived, under the weak stagnation assumption, the same necessary and sufficient conditions for order-2 stability as Poli. The work of Liu [37] also implies that the convergence region of equation (3.13) is the same irrespective of the social network topology utilized by PSO.

Recently, under the stagnation distribution assumption, Bonyadi and Michalewicz [7] were able to derive criteria for order-1 and order-2 stability, while also allowing w, $c_1\boldsymbol{r}_1$, and $c_2\boldsymbol{r}_2$ to be random variables with well defined expectations and variances. Specifically,

- order-1 stability:

$$-1 < E[w] < 1 \qquad \text{and} \qquad 0 < \frac{E[c_1 r_1] + E[c_2 r_2]}{E[w] + 1} < 2 \tag{3.14}$$

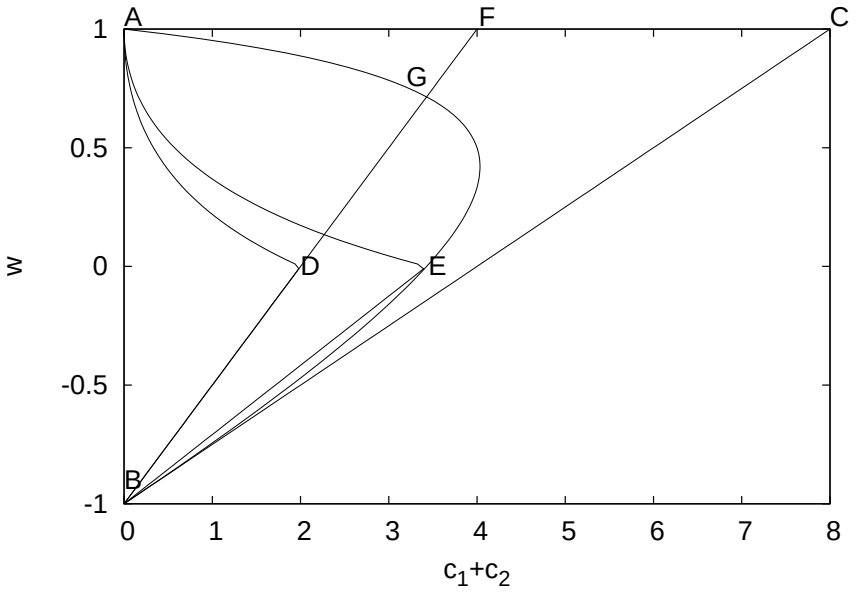


FIGURE 3.1: Theoretically derived regions sufficient for particle convergence

- order-2 stability:

$$-1 < \frac{E[w]}{\sqrt{1 - V[w]}} < 1 \tag{3.15}$$

$$0 < E[\theta_1] + E[\theta_1] < \frac{-2(E[w]^2 + V[w] - 1)}{1 - E[w] + \frac{(V[c_1 r_1] + V[c_1 r_2])(1 + E[w])}{(E[c_1 r_1] + E[c_2 r_2])^2}} \tag{3.16}$$

3.4 Theoretical Analysis of the Fully Informed Particle Swarm Optimizer

The FIPS algorithm has undergone far less theoretical investigation than the canonical PSO has. The primary contributions with regard to convergence analysis are discussed in this section.

Poli [48] was the first to analyze the FIPS algorithm from a theoretical perspective. The analysis was performed under the stagnation assumption, and focused on the case where the neighborhood size was three (i.e. FIPS3). The analysis compared the order 1, 2, 3 and 4 (mean, deviation, skewness, kurtosis) stability of the social-only and canonical PSO with that of FIPS3. It was found that FIPS3 was, surprisingly, the most stable

of the three, despite the FIPS3 algorithm containing more sources of randomness. No general region of convergence was provided for an arbitrary choice of neighborhood size.

In the study performed by de Oca and Stützle [41] it was shown that particles are attracted to the centroid of the particle's neighborhood best found positions, given that coefficients were selected that satisfy the constriction conditions as defined in [17]. The centroid is defined as the average over all the neighborhood best positions. The study was performed under the stagnation and deterministic assumptions, where γ_m of equation (2.3) was replaced with the expected value $\frac{c_1+c_2}{2}$. A general region for particle convergence was not presented. The study did, however, empirically determine that a more connected swarm topology resulted in a smaller region of the search space being explored. It was also found that the FIPS algorithm has a very strong bias to the centroid of each particle's previously found neighborhood best positions.

García-Gonzalo and Fernández-Martinez [24] derived the following necessary conditions for order-2 stability under the stagnation assumption

$$c_1 + c_2 < \frac{12|\mathcal{N}_i|\left(1-w^2\right)}{3|\mathcal{N}_i|+1+w\left(1-3|\mathcal{N}_i|\right)} \tag{3.17}$$

where $|\mathcal{N}_i|$ is the neighborhood size of particle i. It should be noted at this point that the contribution found in chapters 5 and 6 were accepted for publication before the work of García-Gonzalo and Fernández-Martine was published and available.

3.5 Theoretical Analysis of the Bare Bones Particle Swarm Optimizer

Despite the BPSO algorithm being well supported by theoretical convergence results of PSO, the algorithm itself has not undergone much theoretical study. The primary contribution is that of Blackwell [3]. The study focused on a generalized class of PSO update equations. The class of PSOs considered were those with update equations that could be applied component wise, and that can be rearranged in the form

$$x_{ij}(t+1) + a x_{ij}(t) + b(t) x_{ij}(t-1) = c(\mathcal{N}_i), \tag{3.18}$$

were a and b are random variables, and $c(\mathcal{N}_i)$ is a random variable that also depends on constant neighborhood positions. Blackwell [3] was able to show that, under the stagnation assumption, the sequence of particle positions is weakly stationary for BPSO. If a series is weakly stationary, it is by implication order-2 stable as shown by [28]. Alterations of the BPSO algorithm with theoretically derived non-collapse conditions were also presented.

Blackwell [3] used the same approach to derive necessary criteria for order-2 stability of PSO particles, as stated in section 3.3.

3.6 Theoretical Analysis of the Unified Particle Swarm Optimizer 2011

The only theoretical analysis performed on UPSO was in its introduction by Parsopoulos and Vrahatis [45]. The analysis was on a weak type of local convergence to an optimal value on a highly oversimplified version of UPSO. The analysis performed was also only valid for convex objective functions.

3.7 Theoretical Analysis of the Standard Particle Swarm Optimizer 2011

To date, there have been only two theoretical studies on SPSO2011. Both of which where performed by Bonyadi and Michalewicz [4, 5]. Both analyses where completed under the stagnation assumption, and without the special treatment of $\boldsymbol{g}_i(t)$ when $\boldsymbol{y}_i(t) = \hat{\boldsymbol{y}}_i(t)$ as defined in equation (2.17). It was shown that SPSO2011 was not locally convergent to an optimum. However, SPSO2011 was shown to be rotationally invariant. The convergent (stable) region of SPSO2011 was plotted using an empirical technique under forced stagnation, with each particle's personal best and neighborhood best positions set to be equal. Forced stagnation is a situation when neither the personal and neighborhood best positions are allowed to update, or the objective function is a constant. The former was used in the studies by Bonyadi and Michalewicz [4, 5]. It was shown that the size of the

convergence region appeared to decrease as the dimensionality of the problem increased. However, no explicit conditions for convergence were presented.

3.8 Summary

This chapter presented an overview of existing PSO stability analyses. The assumptions that have been utilized in existing PSO stability analyses were discussed. An overview of order-1 and order-2 stability was given. Existing stability studies of PSO and PSO's variants were also discussed.

The next chapter presents a novel empirical approach that is used to verify the theoretically derived criteria for order-2 stability in an assumption free context.

Chapter 4

Finding the True Convergence Region for PSO

As discussed in chapter 3, a number of theoretical studies on PSO stability exist. However, these studies rely on some form of simplifying assumption. This chapter aims to empirically verify which criteria for particle stability are an accurate representation of PSO particle behavior in an unsimplified context. Content from this chapter was published in the proceeding of the 2014 IEEE Congress on Evolutionary Computation [9].

Section 4.1 presents the experimental procedure and setup for testing PSO particle stability. The experimental results and discussion are presented in section 4.2. A summary of this chapter's findings is given in section 4.3.

4.1 Experimental Setup

The experiment conducted in this chapter is designed to illustrate under what parameter settings the PSO algorithm will actually exhibit convergent behavior. There is an inherent difficulty in empirically analyzing the convergence behavior of PSO particles, specifically with regards to understanding the influence of the underlying objective function's landscape on the PSO algorithm. In an attempt to try and mitigate this issue,

the following objective function that will make it "hard" for PSO to become stagnant, is used:

$$CF(\boldsymbol{x}) \sim U(-1000, 1000). \tag{4.1}$$

The value of CF, for each $\boldsymbol{x}$ in the domain of CF, is calculated and stored the first time it is required in the execution of the PSO algorithm. The calculated value for each $\boldsymbol{x}$ in the domain of CF remains static after its initial computation. Objective function values are generated anew for each independent run of the PSO algorithm. In other words, for each evaluated $\boldsymbol{x}$ in the domain of CF a unique random value from the uniform distribution $U(-1000, 1000)$ is assigned. What the objective function in equation (4.1) provides, is an environment that is rife with discontinuities and completely unstructured, resulting in a search space in which the PSO algorithm will be highly unlikely to enter a state of full stagnation. If particles are seen to be convergent (not necessarily to the same point) with such a degenerate objective function as equation (4.1), it is reasonable to assume that convergent behavior will hold with less degenerate objective function where stagnations is more likely.

The experiment utilizes the following static parameters: Population size of 64, 2000 iterations, a 50-dimensional search space, the star neighborhood structure (gbest). Particle positions are initialized within $(-1000, 1000)^d$ and velocities are initialized to $\mathbf{0}$. Equation (4.1) is utilized as the objective function. A population size of 64 is utilized to allow for easier future comparison of differing PSO neighborhood structures. A population size of 64 allows for complete 2-D and 3-D von Neumann neighborhood structures.

The measure of convergence is as follows:

$$\Delta(t+1) = \frac{1}{N} \sum_{i=1}^{N} \|\boldsymbol{x}_i(t+1) - \boldsymbol{x}_i(t)\|_2, \tag{4.2}$$

which is the average Euclidean distance moved by each particle in the swarm from iteration t to iteration $t+1$.

The test is conducted over the following parameter region:

$$w \in [-1.1, 1.1] \text{ and } c_1 + c_2 \in (0, 4.4], \tag{4.3}$$

where $c_1 = c_2$, with a sample point every 0.1 along w and $c_1 + c_2$. A total of 1012 sample points are used. The region defined by equation (4.3) is slightly larger than that needed to include all the regions presented in figure 3.1 in section 3.3, except the region defined by equation (3.10) which is partially omitted. The sub-region

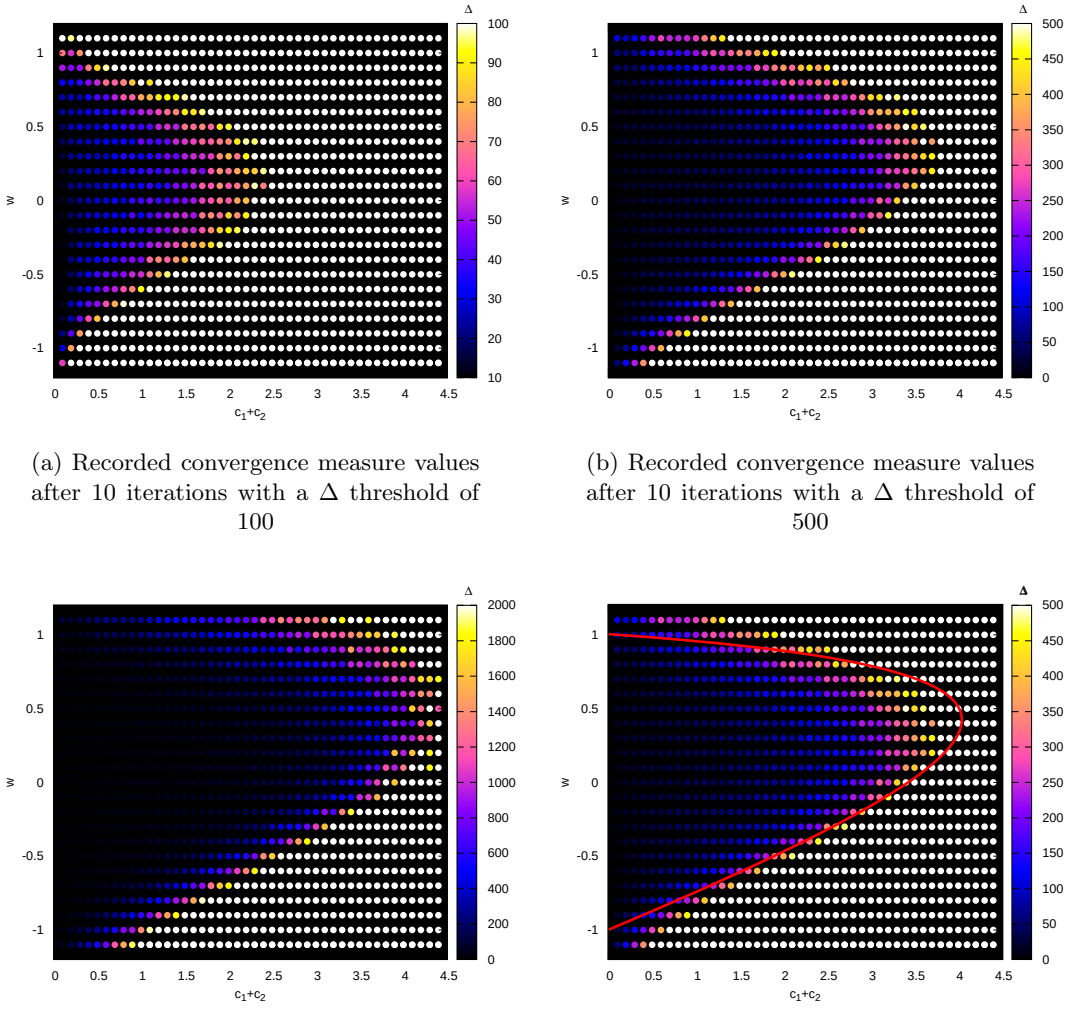
$$c_1 + c_2 < 4\,(1 + w)\,, \quad c_1 + c_2 > 4.4, \tag{4.4}$$

of the region defined by equation (3.10) is omitted from the analysis due to the lack of any convergent behavior within the omitted region (this fact will become obvious in section 4.2). The results reported in Section 4.2 are averages over 35 independent runs for each sample point.

4.2 Experimental Results and Discussion

This section presents the results of the experiment described in section 4.1. A snapshot of all parameter configurations' resulting convergence measure values are presented at the PSO's 10th, 500th and 2000th iteration. For each iteration, a snapshot figure with the convergence measure value capped at 100, 500, and 2000 is presented, so as to prevent very large convergence measure values from obscuring the important observations.

At iteration 10, a relatively clear picture of particle behavior is already developing. Even with the low convergence measure limit of 100, figure 4.1a illustrates parameter settings that are more conducive to convergence behavior, even at this low iteration count of 100. More specifically, the parameter values within the curved region with the apex at $w = 0.2$ and $c_1 + c_2 = 2.3$ are clearly exhibiting convergence behavior. In figure 4.1b, the region with the most convergent behavior is already starting to show similarity to the region of equation (3.13). In figure 4.1c, there is a large number of parameter settings exhibiting convergent like behavior. This is not surprising given the early iteration count and the substantial large convergence measure threshold. What is quite surprising, however, is the large number of parameter settings that have resulted in convergence measure values in excess of 2000 after only 10 iterations. As an illustrative example, the exact convergence measure value at $w = -0.8$, $c_1 + c_2 = 3.5$ in figure 4.1c



(a) Recorded convergence measure values after 10 iterations with a Δ threshold of 100

(b) Recorded convergence measure values after 10 iterations with a Δ threshold of 500

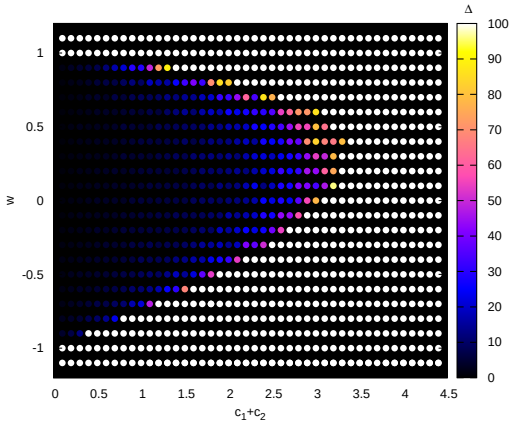
(c) Recorded convergence measure values after 10 iterations with a Δ threshold of 2000

(d) Recorded convergence measure values after 10 iterations with a Δ threshold of 500 with equation (3.13) overlaid

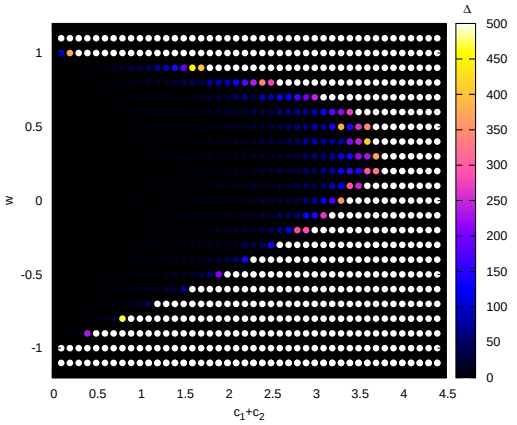
FIGURE 4.1: Convergence measure values at the 10th iteration

is 153101. Such large convergence measure values could be a serious hindrance on PSO's search capability, and in the extreme case, the PSO's search is surely useless [22].

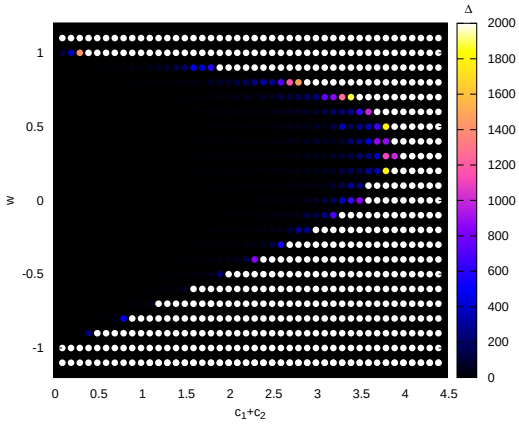
At iteration 500 the particle behavior is already substantially more stable than at iteration 10. This is illustrated by the close similarity between figures 4.2b and 4.2c. In figure 4.2a, the region of convergent behavior is slightly narrower than that of figure 4.1a, but the apex is further out. However, the number of particles with a convergence measure value below 100 has increased over the course of 490 iterations. In figures 4.2b and 4.2c, the correspondence between equation (3.13) and the convergent behavior is becoming very clear, as illustrated in figure 4.2d. The only discrepancy is that the apex of the



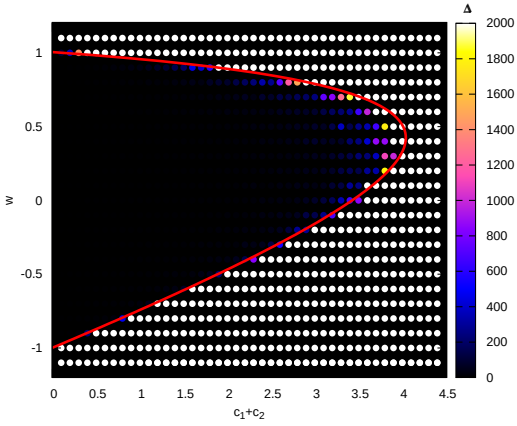
(a) Recorded convergence measure values after 500 iterations with a Δ threshold of 100



(b) Recorded convergence measure values after 500 iterations with a Δ threshold of 500



(c) Recorded convergence measure values after 500 iterations with a Δ threshold of 2000

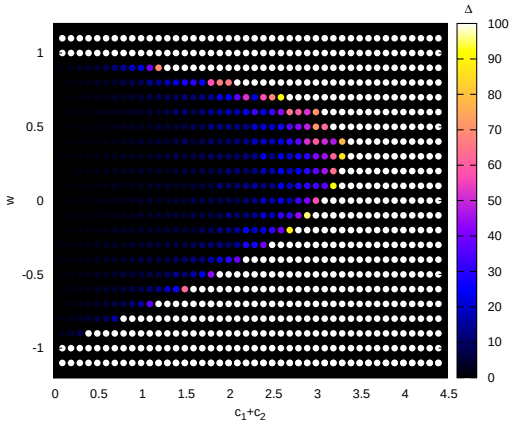


(d) Recorded convergence measure values after 500 iterations with a Δ threshold of 2000 with equation (3.13) overlaid

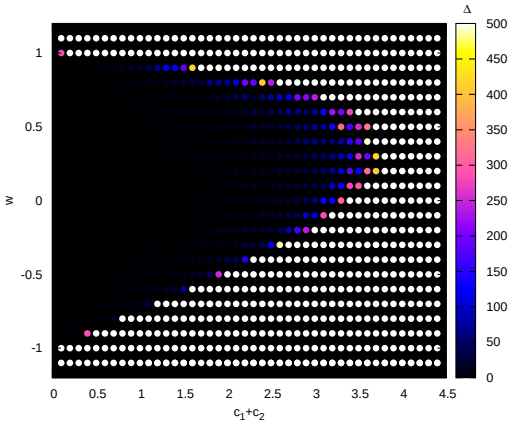
FIGURE 4.2: Convergence measure values at the 500th iteration

convergence region of figures 4.2b and 4.2c is slightly less than that of equation (3.13). However, with this correlation in mind, there are still convergence measure values of over 2000 corresponding to parameters that are technically in the region of equation (3.13).

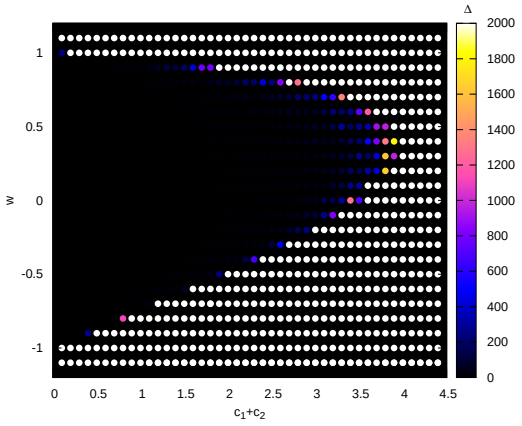
At iteration 1000 the region of convergent behavior is nearly identical to the corresponding figures for iteration 500, as can bee seen in figures 4.3a, 4.3b and 4.3c. There is, however, a small decrease in convergence measure values across the region of convergent behavior. For example, when $w = 0.5$ and $c_1 + c_2 = 3$, the convergence measure changed from 79.689 at 500 iterations to 70.8303 at 1000 iterations. This decrease in convergence measure values is not negligible. However, the rate of decrease is particularly slow.



(a) Recorded convergence measure values after 1000 iterations with a Δ threshold of 100



(b) Recorded convergence measure values after 1000 iterations with a Δ threshold of 500

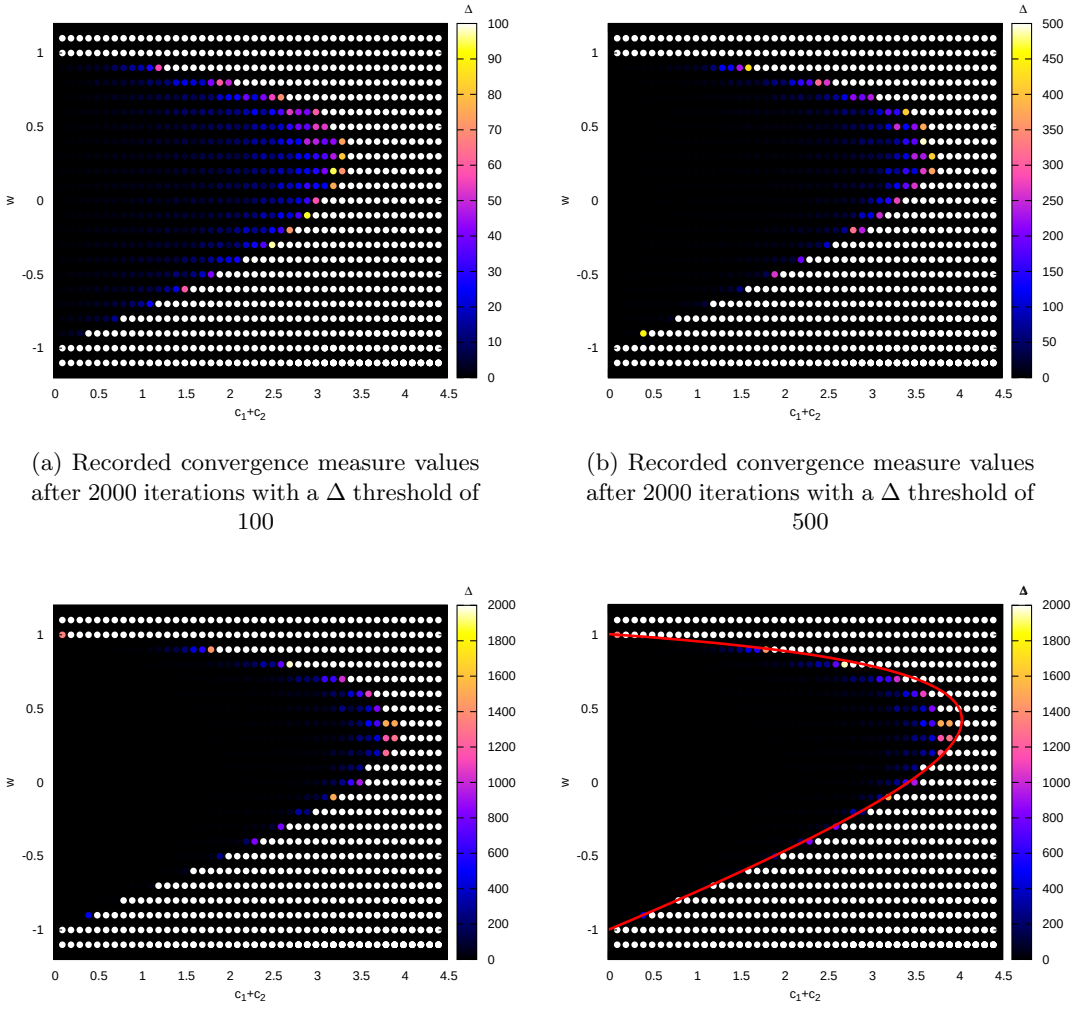


(c) Recorded convergence measure values after 1000 iterations with a Δ threshold of 2000

FIGURE 4.3: Convergence measure values at the 1000th iteration

At iteration 2000 the same phenomenon as from 500 to 1000 iterations occurred, namely, figures 4.4a, 4.4b, and 4.4c are nearly identical to the corresponding figures for 1000 iterations. Again, a small decrease in convergence measure values across the region of convergent behavior occurred. For example, when $w = 0.5$ and $c_1 + c_2 = 3$, the convergence measure changed from 70.8303 at 1000 iterations to 53.8461 at 2000 iterations. As illustrated in figure 4.4d, equation (3.13) matches almost perfectly with the convergence region of figure 4.4c.

In general, the convergence results correspond well with the derived region of equation (3.13). The regions defined by equations (3.11) and (3.12) both result in convergent



(a) Recorded convergence measure values after 2000 iterations with a Δ threshold of 100

(b) Recorded convergence measure values after 2000 iterations with a Δ threshold of 500

(c) Recorded convergence measure values after 2000 iterations with a Δ threshold of 2000

(d) Recorded convergence measure values after 2000 iterations with a Δ threshold of 2000 with equation (3.13) overlaid

FIGURE 4.4: Convergence measure values at the 200th iteration

behaviour, as they are subsets of the region defined in equation (3.13). However, the regions defined by equations (3.11) and (3.12) excludes a large number convergent parameter settings. The extended region defined by equation (3.10) (with $|w| < 1$) is far larger than the actual observed convergence region of parameter values. The extended region defined by equation (3.9) (with $|w| < 1$) contains a large number of convergent parameter settings. However, there are a number of excluded parameters, namely those within the region BEG of figure 3.1. The extended region defined by equation (3.9) also includes a small number of parameters that do not fall within the actual convergence region, namely, those within the region AGF of figure 3.1.

There are, however, two additional observations: Firstly, parameters that reside very near to the apex of the region defined in equation (3.13) do not exhibit a fast convergent trend. For example, the parameter setting $w = 0.5$ and $c_1 + c_2 = 3.9$, which is within the derived region of equation (3.13), has a convergence measure value of 2576.85 after 2000 iterations. While this parameter configuration may actually result in convergence, the rate of convergence is prohibitively slow in practise. The second observation is that the closer $c_1 + c_2$ is to zero and w to roughly 0.4, the quicker particle convergence occurs.

Given these observations, when utilizing PSO in practice, selecting parameters from a region of the same form as that of equation (3.13), but excluding configurations within 0.1 of equation (3.13)'s boundary, will yield a reasonable convergence rate.

4.3 Summary

The aim of this chapter was to perform an experiment that clearly shows which theoretically derived convergence region is most applicable to practical PSO use. It was found that there is a very strong correlation between the convergence behavior of the PSO and the parameter region defined by Poli [50, 51]. Despite this very strong correlation, the empirical results also shows that, when PSO parameter values are near the edge of the region defined by equation (3.13), convergence is incredibly slow, with some parameter settings still having an average particle movement of over 2500 after 2000 iterations. From these observations it is concluded that, in practice, PSO parameter setting should be selected from a slightly smaller region than that of equation (3.13), if an unreasonably slow rate of particle convergence is to be avoided.

The next chapter expands the empirical approach to stability analysis and the effectiveness of the objective function as defined in equation (4.1) is demonstrated.

Chapter 5

Empirical Stability Analysis Approach

This chapter has two main aims. Firstly, to demonstrate that the use of a specially designed objective function for convergence analysis, as defined in equation (4.1), is both a simple and valid method for performing assumption free stability analysis. Secondly, to utilize the standardized stability analysis approach to investigate the stability criteria for PSO variants. Content from this chapter was published in the Proceedings of the 2014 International Swarm Intelligence Conference (ANTS) [10] and in the Swarm Intelligence journal [12].

A thorough empirical analysis on the effectiveness of using the objective function, as defined in equation (4.1), for PSO stability analysis is presented in Section 5.1. The influence of the social network structure on the stability criteria of PSO is investigated in section 5.2. Sections 5.3, 5.4, and 5.5 presents an empirical investigation of the stability criteria of FIPS, BPSO and SPSO 2011, respectively. A summary of the chapter's fundings are given in section 5.7.

5.1 Custom Objective Function for Convergence Analysis

As discussed in chapter 4, there is an inherent difficulty in empirically analyzing the convergence behavior of PSO particles. This difficulty arises from the potential dependency between the particle exhibiting convergent behavior, for given parameter configurations, and the underlying objective function's landscape. Chapter 4 proposed that the following objective function, as originally defined in equation (4.1),

$$CF(\boldsymbol{x}) \sim U(-1000, 1000)$$

is an effective choice of objective function when performing stability analysis of PSO and PSO variants.

The aim of most recent theoretical convergence research performed on PSO and PSO variants [3, 7, 26, 50, 51] was to prove, for a simplified version of the optimization algorithm, the following theorem, with the hope that the derived results are applicable to the unsimplified version of the optimization algorithm:

Theorem 5.1. *There exists a set C of PSO parameter values such that if parameters are selected from C then for all objective functions $f : \mathbb{R}^n \to \mathbb{R}$ there exists an iteration T such that for all iterations $t > T$ each particle's position is order-1 and order-2 stable.*

The objective function in equation (4.1) is designed to be an ideal counter example to theorem 5.1. The premise is that if a PSO variant can converge for a given parameter configuration using equation (4.1) as an objective function, then the parameter configuration is very likely to be a truly convergent parameter configuration for all objective functions.

The experiment conducted in this section aims to justify the use of a specifically designed objective function for the convergent parameter region analysis. The experimental setup and results are presented in sections 5.1.1 and 5.1.2, respectively.

5.1.1 Experimental Setup

The measure of convergence used in this section is that of equation (4.2), as defined in chapter 4. Equation (4.2) is chosen as the measure of convergence because if any particle is divergent, the convergence measure value will reflect this divergence within the swarm. A swarm can only be classified as convergent if every particle in the swarm exhibits convergent behavior.

The experiment utilizes the following static parameters: Swarm size of 64 particles, 5000 iterations, and a 50-dimensional search space. A swarm size of 64 particles is utilized to allow for all the social topologies tested to be complete structures. Particle positions are initialized within $(-100, 100)^d$ and velocities are initialized to $\mathbf{0}$.

The experiment is conducted over the following parameter region:

$$w \in [-1.1, 1.1] \text{ and } c_1 + c_2 \in (0, 4.3]\,, \tag{5.1}$$

where $c_1 = c_2$, with a sample point every 0.1 along w and $c_1 + c_2$. The experiment is performed for each of the following neighborhood topologies: star, ring, 2-D and 3-D von Neumann. The experiment is conducted using CF and 11 base objective functions from the CEC 2014 problem set [36]. The functions are as follows: Ackley, High Conditioned Elliptic, Bent Cigar, Discus, Rosenbrock, Griewank, Rastrigin, HappyCat, HGBat, Katsuura, and Expanded Griewank and Rosenbrock composition. The explicit equations for each objective function is given in appendix 8.1. The region of equation (4.3) contains exactly 504 points that satisfy equation (3.13). A total of 989 sample points from the region defined in equation (5.1) are used per objective function and topology pair. The results reported in Section 5.1.2 are averages over 35 independent runs for each sample point. It should be noted that, for all PSO variants used in this chapter, no form of search space bounding is performed. Any attempt to force particles to remain within a given bounded area would seriously hinder the ability to perform an empirical analysis, and implicitly impose a form of order-2 stability on the swarm.

In order to allow for a sensible comparison of convergence properties, the convergence measure values are bounded as follows: If $[l, u]^d$ is the initial domain of the objective

function, then

$$\Delta_{max} = \sqrt{d\,(l-u)^2}, \tag{5.2}$$

where l and u are the lower and upper bounds of the domain per dimension, respectively. Δ_{max} is the maximum distance of two points in the initialized search space. For this section, $\Delta_{max} = 1414.214$, as $d = 50$. Utilizing Δ_{max} to bound the presented results is reasonable as any swarm that has the average particle movement exceeding the maximum initial distance possible between two particles in the search space after 5000 iterations cannot be thought of as convergent in a practical context. The convergence measure values are bounded instead of log scaled as multiple parameter configurations resulted in particle movement so extreme that a 64-bit floating point number was experiencing overflow. The value of Δ_{max} also has a secondary purpose as the classification boundary between convergent and divergent particle movement. While Δ_{max} appears to be a large allowance for convergence, it allows for particles that are converging at a very slow rate to still be classified as converging. However, utilizing Δ_{max} is not a hindernace to correctly classify particles that are slowly diverging, as it is easy for particles to exceed Δ_{max}, and be classified as divergent after swarm initialization, due to the well known phenomenon of particle velocity explosion [22].

5.1.2 Experimental Results and Discussion

This subsection presents a table per PSO social topology containing the following measurements per objective function:

- **Measurement A**: The number of PSO parameter configurations that resulted in a final convergence measure value less than or equal to the final convergence measure obtained if the CF objective function was used instead.
- **Measurement B**: The number of PSO parameter configurations that resulted in a final convergence measure value greater than the final convergence measure obtained if the CF objective function was used instead.
- **Measurement C**: The number of PSO parameter configurations that resulted in a final convergence measure greater than or equal to Δ_{max}.

- **Measurement D**: The number of PSO parameter configurations that resulted in a final convergence measure less than Δ_{max}.
- **Measurement E**: The number of PSO parameter configurations that satisfied equation (3.13) and resulted in a final convergence measure less than Δ_{max}.
- **Measurement F**: The number of PSO parameter configurations that satisfied equation (3.13) and did not result in a final convergence measure less than Δ_{max}.
- **Measurement G**: The average convergence measure value across all parameter configurations, with all elements bounded at Δ_{max}. It should be noted that reported averages for this measurement are calculated after bounding of the convergence measure has occurred, so as to prevent divergent configurations from radically affecting the results.

Measurements A and B provide a concise way of seeing, per objective function, how much better or worse the CF objective function performs as a reference convergence analysis function. An ideal convergence analysis function is one that, in general, will yield the highest resulting convergence measure for all possible parameter configurations. The higher the resulting convergence measure value is, the harder it was for the PSO to have converged under a given objective function. Measurements C and D give a clear picture of how effectively the underlying objective function highlights possible divergent particle behavior. Given the tested region of equation (5.1), there are a total of 504 parameter configurations that satisfy equation (3.13), leaving 485 parameter configurations that should produce divergent behavior. Ideally, an objective function utilized for convergence analysis should result in a value for measurement C as close as possible to 485, and a value for measurement D as close as possible to 504. Measurements E and F are an extension of measurements C and D, in that an objective function should have at most 504 parameter configurations that both satisfy equation (3.13) and have a convergence measure value not exceeding Δ_{max}. An objective function with a measurement E value smaller than 504 is more conservative in assigning the label of a convergent particle. A slightly conservative assignment is a positive feature of an objective function being used for convergence analysis, as falsely classifying a parameter configuration as convergent could lead to a PSO user obtaining radically unexpected results when utilizing the

parameter configuration in practice. Measurement G provides an overall view of how difficult the used objective function has made it for the PSO algorithm to converge.

A snapshot of the convergence measure values are presented for four cases:

- **Case A**: For each parameter configuration the maximum convergence measure value across all 11 objective functions and topologies is reported.
- **Case B**: For each parameter configuration the maximum convergence measure value across all topologies using only the CF objective function is reported.

In order to deduce the convergence region from the empirical data of all 11 base functions and all topologies, the largest recorded convergence measure value of each parameter configuration is reported in case A. Case B is presented to illustrate the similarity between the mapped out convergence region of the PSO algorithm using the CF objective function to the mapped out convergence region of the PSO algorithm in case A, which is constructed using the complete pool of gathered data of the 11 objective functions.

- **Cases C and D**: For each parameter configuration the maximum convergence measure value across all topologies using the two objective functions which have the most similar resulting measurements to case B is reported.

Cases C and D are presented to illustrate that the mapped out convergence region of cases A and B are not identical to the convergence regions of any arbitrary objective function. In particular, cases A and B should result in a subset of the region produced by an arbitrary objective function.

Measurements A and B in table 5.1 show that the Gbest PSO applied to the CF objective function resulted in higher convergence measures than 9 of the 11 other objective functions for nearly all parameter configurations. For the two remaining objective functions, Katsuura is the only objection function close to the CF objective function in terms of measurement A. However, Katsuura has an average convergence measure of 49.672 less than CF has, making CF the better objective function for convergence analysis. The CF objective function also obtained the largest number of parameter configurations that

resulted in a convergence measure that exceeded the bound of Δ_{max}, and the highest average convergence measure evaluations. These measurements indicate the effectiveness of CF as an objective function for convergence analysis. The CF objective function, under the star topology, provides an environment that is the hardest for PSO particles to converge in.

Measurements A and B in table 5.2 show that the Lbest PSO applied to the CF objective function resulted in higher convergence measures than 9 of the 11 other objective functions for nearly all parameter configurations. Once again, Katsuura provided the second lowest value for measurement A, while CF provided the best results for all other measurements, applying the same analysis logic used for the star topology. Though inferior, Ackley resulted in values for C, D and G very close to that obtained by the CF objective function. However, CF provided far better results in terms of measurement A, making CF the best choice as an objective function for convergence analysis.

Measurements A through G in tables 5.3 and 5.4 show for both the 2-D and 3-D von Neumann topologies that the results remain almost identical to those of the ring and star topologies. This provides evidence that the topology has a negligible impact on the effectiveness of CF as an objective function for convergence analysis.

For case A, the convergence region as illustrated in figure 5.1a matches the derived region of equation (3.13) almost perfectly, as does the region seen in figure 5.1b for case B. While there exists a slight difference between figures 5.1a and 5.1b in terms of convergence measure values, the overall convergence regions are nearly identical. The similarity

TABLE 5.1: Convergence properties per objective function under the Star topology

Function \ Measurement	A	B	C	D	E	F	G
CF	–	–	467	522	504	0	683.437
Ackley	879	110	464	525	502	2	676.293
High Conditioned Elliptic	989	0	400	589	504	0	573.601
Bent Cigar	989	0	412	577	504	0	598.593
Discus	989	0	409	580	504	0	592.545
Rosenbrock	988	1	424	565	504	0	622.009
Griewank	989	0	412	577	504	0	596.772
Rastrigin	989	0	411	578	504	0	596.909
HappyCat	989	0	411	578	504	0	595.375
HGBat	989	0	412	577	504	0	595.366
Katsuura	507	482	416	573	504	0	623.765
Expanded Griewank and Rosenbrock	989	0	416	573	504	0	603.981

TABLE 5.2: Convergence properties per objective function under the Ring topology

Function \ Measurement	A	B	C	D	E	F	G
CF	–	–	473	516	503	1	690.797
Ackley	912	77	469	520	504	0	682.570
High Conditioned Elliptic	989	0	400	589	504	0	574.659
Bent Cigar	989	0	415	574	504	0	602.194
Discus	989	0	414	575	504	0	597.668
Rosenbrock	989	0	417	572	504	0	613.094
Griewank	989	0	412	577	504	0	603.304
Rastrigin	989	0	412	577	504	0	601.111
HappyCat	989	0	415	574	504	0	603.536
HGBat	989	0	414	575	504	0	601.403
Katsuura	509	480	413	576	504	0	623.710
Expanded Griewank and Rosenbrock	989	0	416	573	504	0	609.277

TABLE 5.3: Convergence properties per objective function under the 2-D von Neumann topology

Function \ Measurement	A	B	C	D	E	F	G
CF	–	–	480	509	500	4	704.946
Ackley	915	74	475	514	501	3	692.301
High Conditioned Elliptic	989	0	402	587	504	0	577.036
Bent Cigar	989	0	413	576	504	0	600.998
Discus	989	0	414	575	504	0	598.234
Rosenbrock	988	1	415	574	504	0	616.365
Griewank	989	0	414	575	504	0	600.839
Rastrigin	989	0	412	577	504	0	597.999
HappyCat	989	0	414	575	504	0	600.869
HGBat	989	0	413	576	504	0	599.576
Katsuura	525	464	415	574	504	0	622.108
Expanded Griewank and Rosenbrock	988	1	416	573	504	0	608.545

TABLE 5.4: Convergence properties per objective function under the 3-D von Neumann topology

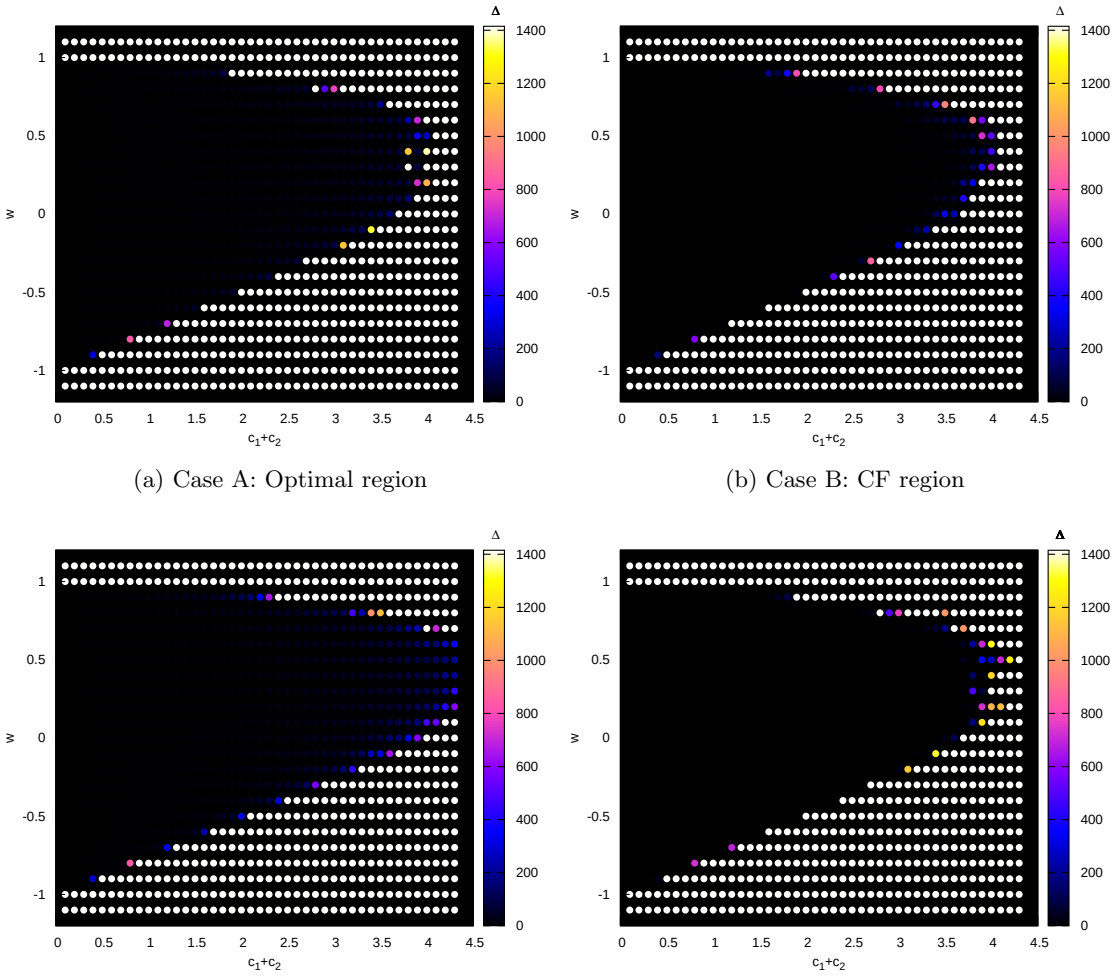
Function \ Measurement	A	B	C	D	E	F	G
CF	–	–	479	510	500	4	704.173
Ackley	925	64	473	516	503	1	691.705
High Conditioned Elliptic	989	0	401	588	504	0	576.575
Bent Cigar	989	0	415	574	504	0	601.344
Discus	989	0	416	573	504	0	600.027
Rosenbrock	989	0	416	573	504	0	615.662
Griewank	989	0	417	572	504	0	602.236
Rastrigin	989	0	413	576	504	0	601.200
HappyCat	988	1	415	574	504	0	603.712
HGBat	988	1	415	574	504	0	600.504
Katsuura	532	457	417	572	504	0	624.483
Expanded Griewank and Rosenbrock	988	1	418	571	504	0	610.503

observed between figures 5.1a and 5.1b indicates that the utilization of the CF function is sufficient for the purpose of empirical convergence analysis. The similarity between figures 5.1a and 5.1b is not observed for the other objective functions. For example in case C, the Katsuura function, when used with PSO, resulted in properties similar to the PSO using CF in tables 5.1 through 5.4. However, Katsuura has a substantially different convergence region to both figures 5.1a and 5.1b, with an apex extending past $c_1 + c_2 = 4.5$, as illustrated in figure 5.1c. For Case D, the convergence region obtained when using the Ackley objective function is illustrated in figure 5.1d. The obtained convergence region is substantially closer to the convergence region obtained in case A and B. However, the apex of the convergence region obtained when using the Ackley objective function is quite jagged in comparison to cases A and B.

A very promising feature of the convergence analysis approach presented in this paper is the high level of accuracy that can be obtained when using CF as an objective function and Δ_{max} as a classification boundary between convergent and divergent parameter configurations. Specifically, with the PSO algorithm, if every parameter configuration with a convergence measure value below Δ_{max} is classified as convergent and every parameter configuration with a convergence measure value above or equal to Δ_{max} is classified as divergent, a total accuracy of 98.79% is obtained when compared to the region derived by [50], with only 12 of the 989 parameter settings misclassified (10 falsely classified as divergent, and 2 falsely classified as convergent).

5.2 PSO, Convergence Analysis of Topological Influence

This section aims to verify that the theoretically derived region of Poli for order-2 stability, as described in equation (3.13), remains valid under multiple social topologies. Considering the results of subsection 5.1, it is clear that the topology does not have a very meaningful impact on the convergence results. This is seen in the similarity between tables 5.1 to 5.4 where, under all measurements, there is minimal to no change, implying that the topology has no real influence on the parameter region corresponding to particle convergence.



(a) Case A: Optimal region

(b) Case B: CF region

(c) Case C: Katsuura region

(d) Case D: Ackley region

FIGURE 5.1: Convergence measure values at the 5000th iteration

A snapshot of all parameter configurations' resulting convergence measure values is presented in figure 5.2 for the following situation:

Topology influence: The optimal convergence region is constructed for each investigated topology, using the same method as explained in case A of section 5.1. The resulting optimal region that has the greatest distance in terms of convergence measure values from case A is reported.

The snapshot in figure 5.2 illustrates the maximum deviation between the convergent parameter region under multiple topologies. If the convergent parameter regions between the presented snapshot and that of case A from section 5.1 are identical, then the topological choice has no influence on the convergent parameter regions.

The ring topology had the greatest Euclidean distance of 1738.23 from the optimal region of case A from subsection 5.1, and therefore experienced the greatest topological influence. While an Euclidean distance of 1738.23 appears large, in the context of a 989 dimensioned vector, it is actually relatively negligible. The convergence region is illustrated in figure 5.2. Despite the ring topology having the greatest Euclidean distance from the optimal region of case A, figure 5.2 appears identical to the region of figure 5.1a, as the difference in convergence measure values are very small. The close similarity between figures 5.1a and 5.2 is a clear indication that the topology used within the PSO algorithm has no meaningful impact on the convergence region of a PSO. The conclusion that PSO's convergence region is independent of the social topology used is supported by the analysis done by Liu [37].

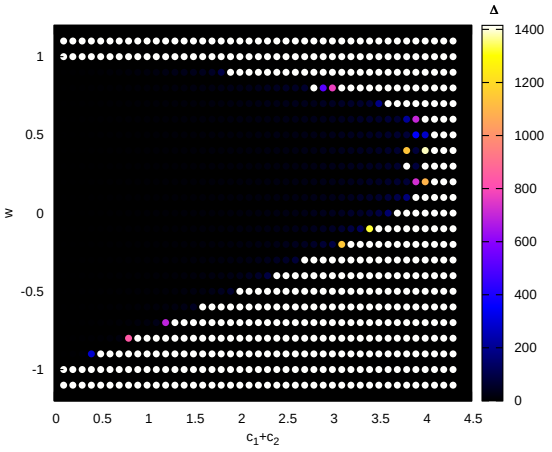


FIGURE 5.2: Topology influence: Convergence measure values for the ring topology at the 5000th iteration

5.3 Fully Informed PSO Convergence Analysis

This section aims to empirically obtain the convergence criteria for FIPS utilizing the proposed objective function CF as defined in equation (4.1).

The FIPS algorithm is implemented using the PSO description in algorithm 1 with the velocity update in equation (2.2) replaced with equation (2.3). The experimental setup and results are presented in sections 5.3.1 and 5.3.2 respectively.

5.3.1 Experimental Setup

Upon inspection of the FIPS update equations defined in equation (2.3), it is clear that the neighborhood size might be a contributing factor in the convergence criteria of FIPS. As a result, the convergence criteria are investigated for neighborhood sizes 2, 4, 8, 16, 32, and 64.

The experiment utilizes the same static parameters as subsection 5.1, except that the LBest topology is used. The analysis is done in one dimension, because the components of the update equation are independent of one another, resulting in $\Delta_{max} = 200$ for this subsection. The experiment of this subsection utilizes the convergence measure of equation (4.2) and the objective function defined in equation (4.1). The use of only the objective of equation (4.1) is validated by the experimental results of subsection 5.1, which proved the effectiveness of the objective function for convergence analysis.

The experiment is conducted over the following parameter region:

$$w \in [-1.1, 1.1] \text{ and } c_1 + c_2 \in (0, 7]\,, \tag{5.3}$$

where $c_1 = c_2$, with a sample point every 0.1 along w and $c_1 + c_2$. The parameter region was empirically determined by increasing the values of $c_1 + c_2$ and w until the complete convergent subregion was contained. A total of 1610 sample points from the region defined in equation (5.3) are used. The results reported in subsection 5.3.2 are the averages over 35 independent runs for each sample point.

5.3.2 Experimental Results and Discussion

A snapshot of the resulting convergence measure values across the region defined in equation (5.3) under varying neighborhood sizes is presented in figure 5.3, for 2, 4, 8, 16, 32, and 64 dimensions.

The convergence region obtained in figure 5.3a is very similar to the convergence region found for the PSO algorithm in subsection 5.1. The similarity of the convergence regions is to be expected given that, if the neighborhood size of 2 is used and $c_1 = c_2$, FIPS

can be shown to be the PSO algorithm assuming that each particle is within its own neighborhood.

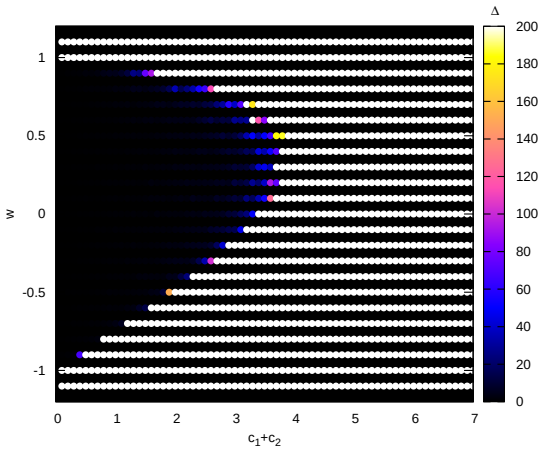
The convergence region found for FIPS with a neighborhood size of 4, as seen in figure 5.3b, is larger than that of FIPS with a neighborhood size of 2. The general form of the convergence region is close to that of FIPS with a neighborhood size of 2.

Upon inspection of FIPS with the neighborhood sizes of 8 and 16 in figure 5.3c and 5.3d, it is clear that the neighborhood size has a meaningful effect on the convergence region. The convergence region continues to allow for greater values of $c_1 + c_2$ as the neighborhood size increases. There is also a clear favoring of positive inertia weight values for larger values of $c_1 + c_2$.

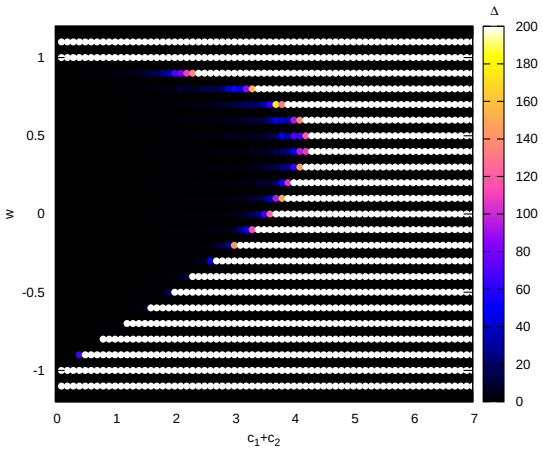
The increase in the size of the convergence region continues with neighborhood sizes 32 and 64, as is seen in figures 5.3e and 5.3f. The convergence region's shape is substantially different from the convergence region obtained when a small neighborhood size is utilized. The convergence region in figure 5.3f is almost triangular in shape.

It is clear that the larger the neighborhood size, the larger the convergence region becomes. The region also appears to extend indefinitely while simultaneously becoming more triangular in shape. This finding is inline with the observation made by [48] that the FIPS algorithm appears to be more stable with the larger neighborhood size of 3 than if a neighborhood size of 2 was used. The idea of increased stability of FIPS is empirically supported in this subsection for larger neighborhood sizes.

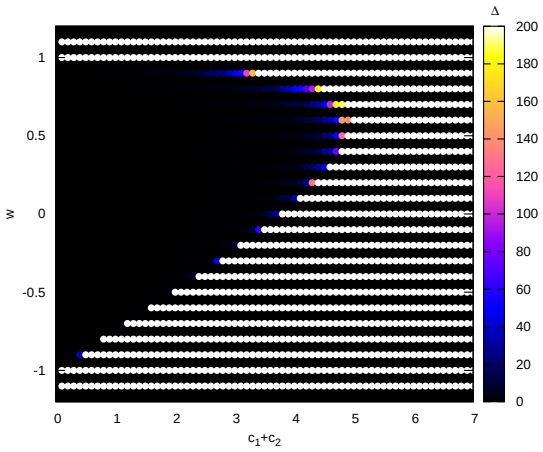
From a practical perspective, it is informative to note that the convergence region for FIPS with neighborhood size n is a subset of the convergence region for FIPS with neighborhood size $n + 1$. Given this knowledge, if the neighborhood size changes over time, convergence parameters should be selected from the region matching the lowest possible neighborhood size, if the PSO user wishes to guarantee convergent particle behavior.



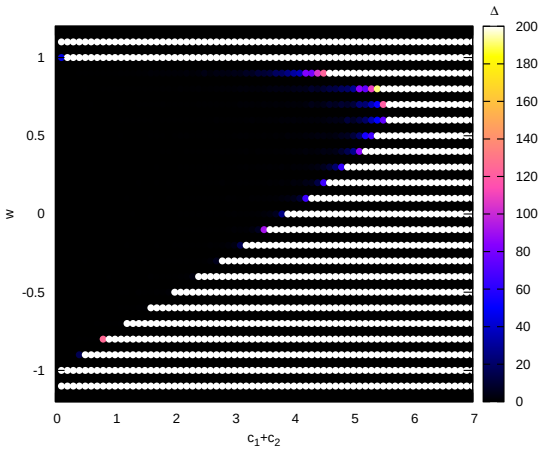
(a) FIPS with neighborhood of size 2.



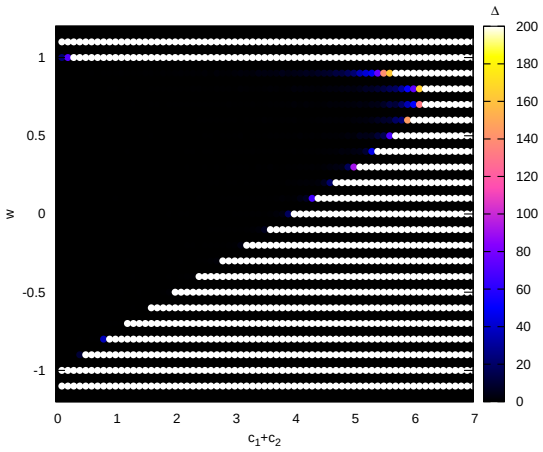
(b) FIPS with neighborhood of size 4.



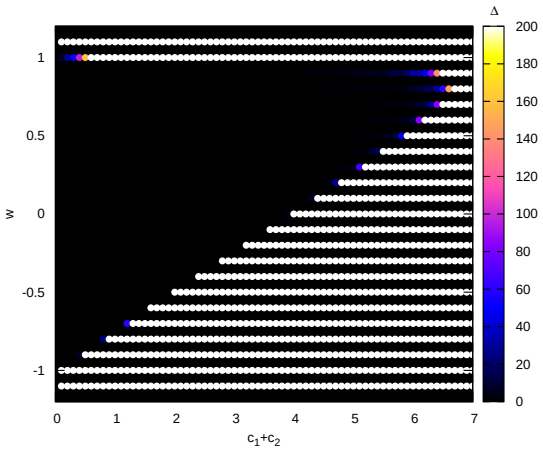
(c) FIPS with neighborhood of size 8.



(d) FIPS with neighborhood of size 16.



(e) FIPS with neighborhood of size 32.



(f) FIPS with neighborhood of size 64.

FIGURE 5.3: Fully Informed PSO convergence results

5.4 Bare Bones PSO Convergence Analysis

This section focuses on the convergence criteria for the BPSO algorithm. The experimental setup and results are presented in sections 5.4.1 and 5.4.2 respectively.

5.4.1 Experimental Setup

The experiment utilizes the same static parameters as subsection 5.3, except that the star topology is utilized.

The experiment is conducted over the following parameter region:

$$c_1 \in (0, 7] \text{ and } c_2 \in (0, 7]\,, \tag{5.4}$$

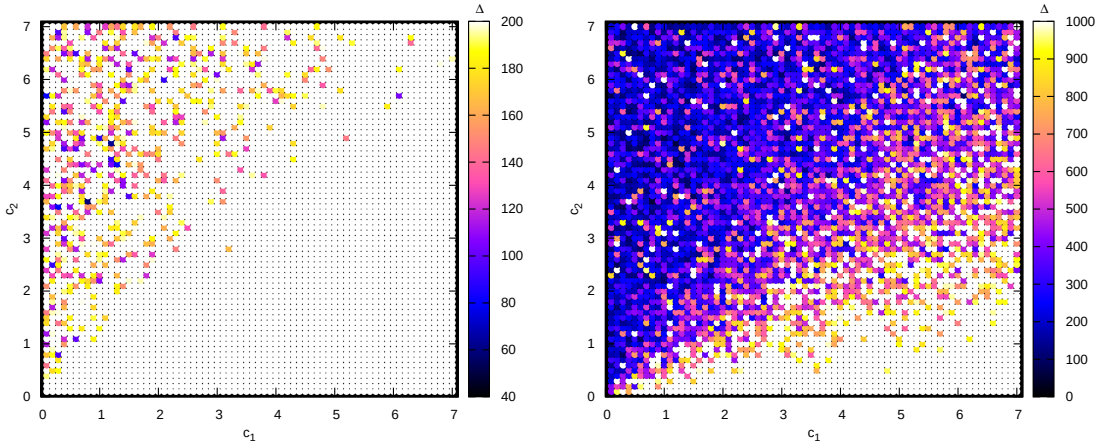
with a sample point every 0.1 along c_1 and c_2. The parameter region was selected as it allows for values of c_1 and c_2 twice as large as the region needed to contain all of FIPS's convergence regions, as determined in section 5.3. The theoretical work of [3] predicts that at least the line where $c_1 = c_2$ should exhibit convergent particle behavior. A meaningful segment of this line is therefore included in the investigated parameter region. A total of 4900 sample points from the region defined in equation (5.4) are used.

The results reported in subsection 5.4.2 are obtained from averaging over either 35, 70, or 1000 independent runs for each sample point. A differing number of independent runs are used to illustrate the amount of noise present in the BPSO experimentation.

5.4.2 Experimental Results and Discussion

Snapshots of the resulting convergence measure values across the region defined in equation (5.4), using a convergence measure bound of Δ_{max} and $5 * \Delta_{max}$, are presented in figures 5.4, 5.5, and 5.6, respectively for 35, 70 and 1000 sample runs. Two different convergence measure bounds are utilized to distinguish the degree of found divergence behavior in BSPO.

Figure 5.4a reports the convergence measure values bounded at Δ_{max} based on 35 independent runs. The convergence results of BPSO are somewhat surprising, as there are very few parameter choices that potentially indicate some level of convergent behavior. Further more, there are no parameter configurations that are clearly convergent, with the smallest reported convergence measure being 52.32. BPSO is actually divergent regardless of parameter choice. If one considers parameter settings where c_1 is larger than c_2, then more divergent behavior is indicated. When the bound on the reported convergence measure is increased to $5 * \Delta_{max}$ (see figure 5.4b), it is more clearly seen that greater divergence occurs when c_1 is larger than c_2. The results of both figures 5.4a and 5.4b have a large degree of noise present despite being the result of 35 independent runs on each sample point.



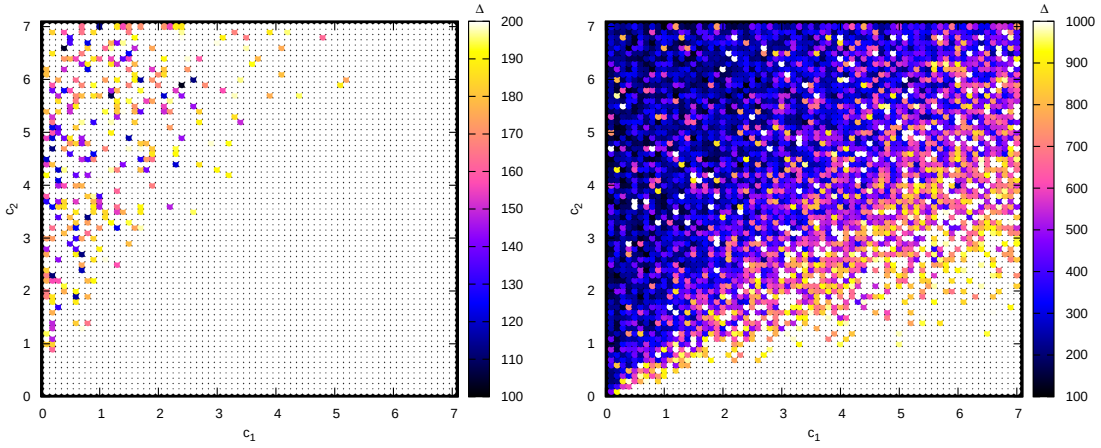
(a) BPSO: 35 samples, bounded at 200.

(b) BPSO: 35 samples, bounded at 1000.

FIGURE 5.4: BPSO convergence results for 35 samples.

In an attempt to reduce the level of noise, the same experiment was run 70 independent times as illustrated in figures 5.5a and 5.5b. In figure 5.5a, the number of parameter choices that do not result in the bound of Δ_{max} to be exceeded has been reduced slightly. Despite the increased number of independent runs, the results clearly still contain a large amount of noise, indicating a large level of unpredictability in BPSO's behavior between runs. This unpredictability is attributed to the heavy reliance of BPSO on the normal distribution as defined in section 2.2.3.

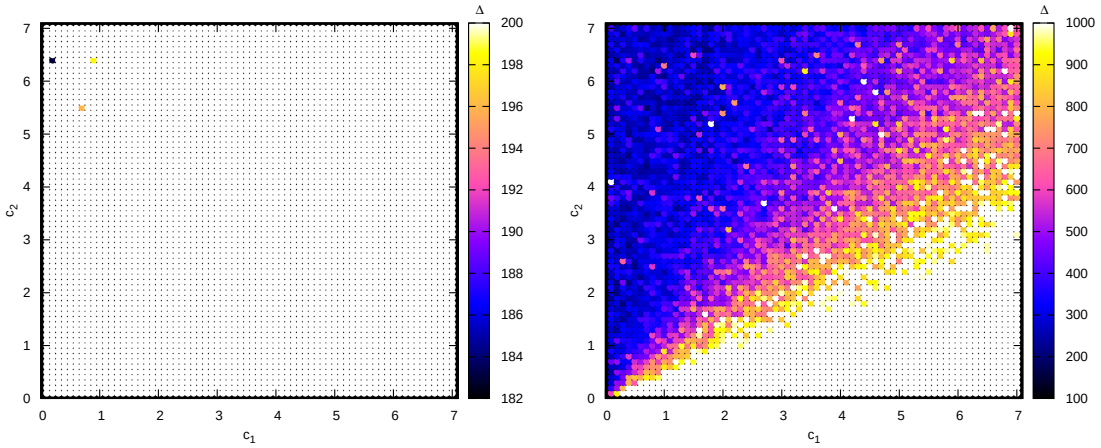
The results presented in figures 5.6a and 5.6b are the result of 1000 independent runs. In figure 5.6a, there are only three parameter settings that did not exceed the convergence



(a) BPSO: 70 samples, bounded at 200. (b) BPSO: 70 samples, bounded at 1000.

FIGURE 5.5: BPSO convergence results for 70 samples.

measure bound, Δ_{max}. There is substantially less noise present in figure 5.6b, making the early mentioned trend that the divergent behavior is more severe if c_1 is greater than c_2 clearer. It is also seen in figure 5.6b that, while the divergent behavior is more severe if c_1 is greater then c_2, the amount by which this affects the divergent behavior decreases as c_1 and c_2 increase.



(a) BPSO: 1000 samples, bounded at 200. (b) BPSO: 1000 samples, bounded at 1000.

FIGURE 5.6: BPSO convergence results for 1000 samples.

The convergence measure values are also not remaining constant over the latter part of the search. In fact, the convergence measure values increase with respect to the increase in t, as illustrated in figure 5.7 where the convergence measure is reported over the course of 5000 iterations of a BPSO algorithm with $c_1 = c_2$.

From the presented results it is clear that BPSO is in fact not guaranteed to converge. Even for the standard BPSO model, where it is assumed that c_1 and c_2 are equal, convergence does not occur. If convergence were to occur when c_1 is equal to c_2, a straight line ($c_1 = c_2$) of low convergence measure values would have been present in figure 5.6b. It is also worth noting that BPSO's particle movement is fairly unpredictable given how much noise is present in the results even with 1000 independent runs, which implies a large amount of unpredictability in the behavior of BPSO.

The theoretical finding of Blackwell [3] that BPSO is order-2 stable is clearly not an accurate representation of the algorithm as the increase in the convergence measure over time in figure 5.7 implies that convergence in standard deviation to a fixed value is not occurring in practice. The nature of the normal distribution may well imply that very high, though statistically unlikely, particle movement will be seen periodically. However, if this was the sole reason for high convergence measure values, the continued increase of the convergence measure as seen in figure 5.7 would not have occurred. Even if particles' personal and neighborhood positions stagnated at a great distance form each other, the convergence measure would only be high, and not increasing.

Given the simple structure of the BPSO's update equation (2.10), it is easy to see that at least one of the following two particle interactions occurs if the convergence measure is increasing:

- The midpoint between the personal best and the neighborhood best positions is moving through the search space at an increasing velocity. This implies that at least one of the personal best position or the neighborhood best position is moving at an increasing velocity.
- The component wise distance between the personal best and the neighborhood best positions is increasing. This implies that at least one of the personal best position or the neighborhood best position is moving.

From this analysis it is clear that the swarm is not entering a state of stagnation.

The only possible justification for the discrepancy between the theoretical findings and the empirical results of this subsection is due to the theoretical work being performed

under the stagnation assumption. In order to verify that the discrepancy is in fact caused by the stagnation assumption, the BPSO was rerun, but with stagnation forced from iteration 20 onwards, by not updating any personal or neighborhood best positions. The results are illustrated in figure 5.8. When stagnation is forced the results match the theoretical derivations of Blackwell [3] perfectly. The difference between figure 5.7 and 5.8 shows that the stagnation assumption resulted in an inaccurate theoretical model for the BPSO algorithm. The reason why stagnation is not occurring is not immediately apparent, and warrants further investigation.

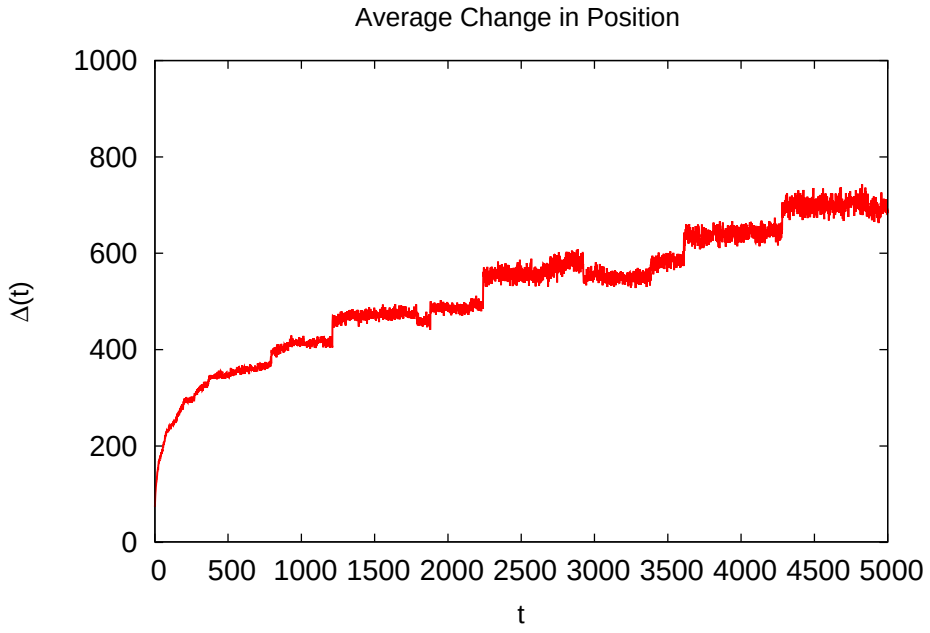


FIGURE 5.7: BPSO average change in particle position (1000 samples).

5.5 Standard PSO 2011 Convergence Analysis

This subsection focuses on the convergence criteria for the SPSO2011 algorithm. The experimental setup and results are presented in sections 5.5.1 and 5.5.2, respectively.

5.5.1 Experimental Setup

The SPSO2011 algorithm's velocity update equation (2.12) cannot be analyzed in one dimension and then generalized to an arbitrary dimension as in the PSO, FIPS, and

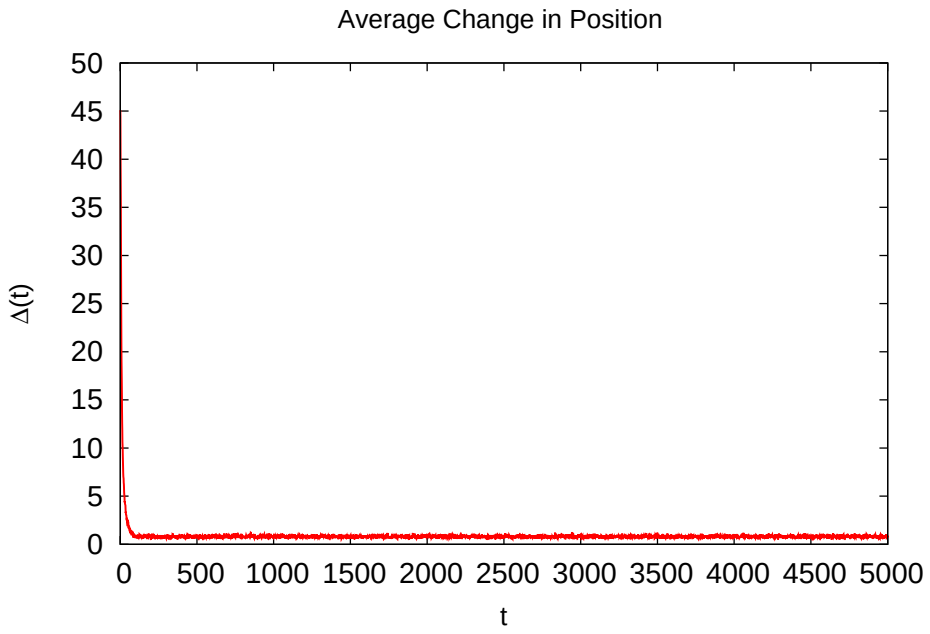


FIGURE 5.8: BPSO average change in particle position (1000 samples) with forced stagnation from iteration 20.

BPSO algorithms, since the function that generates a random point in a hypersphere cannot be investigated component wise [4]. As a result, the convergence criteria are investigated for dimension sizes 1, 10, 20, and 50.

The experiment utilizes the same static parameters as in subsection 5.3 except that the SPSO2011 topology, as defined in subsection 2.2.4, is utilized. The SPSO2011 algorithm is analyzed with and without the special treatment of the center of gravity calculation. The two cases for the center of gravity calculation are

- **Case** 1 uses only the center of gravity equation (2.13), as suggested by Zambrano-Bigiarini and Clerc [65].
- **Case** 2 uses the center of gravity equation (2.13) if the particle's personal best and neighborhood best positions are different. If the particle's personal best and neighborhood best positions are the same, the center of gravity equation (2.17) is used.

The experiment is conducted over the following parameter region:

$$w \in [-1.1, 1.1] \text{ and } c_1 + c_2 \in (0, 13]\,, \tag{5.5}$$

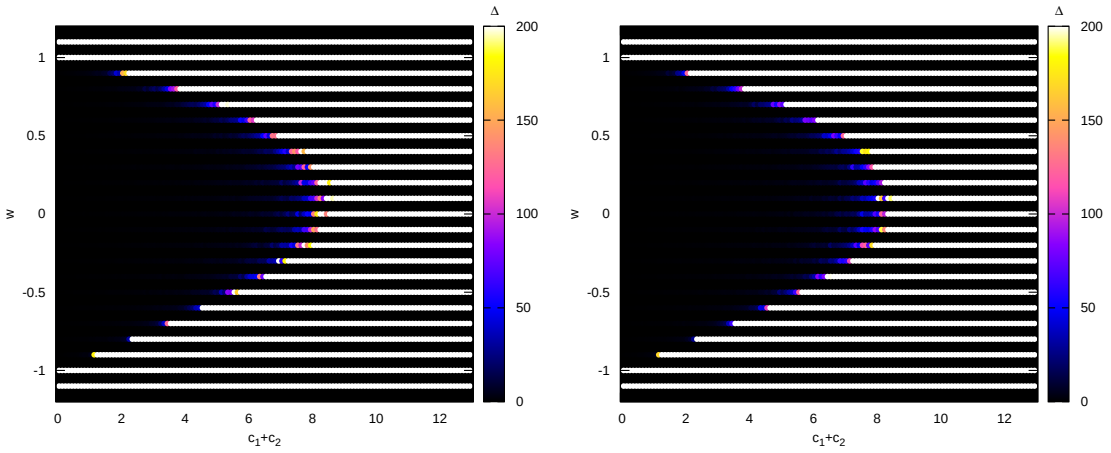
where $c_1 = c_2$, with a sample point every 0.1 along w and $c_1 + c_2$. The parameter region of equation (5.5) was selected as it contains the convergent parameter regions reported by [4]. A total of 2990 sample points from the region defined in equation (5.5) are used. The results reported in subsection 5.5.2 are the averages over 35 independent runs for each sample point. Analysis is done with the convergence measure bounded at the corresponding Δ_{max} values, namely 200, 632.456, 894.427, and 1414.214, respectively.

5.5.2 Experimental Results and Discussion

Snapshots of the resulting convergence measure values across the region defined in equation (5.5) for SPSO2011 are presented under varying dimensionality for cases 1 and 2 in figures 5.9, 5.10, 5.11, and 5.12, respectively for 1, 10, 20 and 50 dimensions.

The convergence region for SPSO2011 in one dimension is presented in figures 5.9a and 5.9b. SPSO2011 has a clear boundary between convergent parameter settings and divergent parameter settings, unlike BPSO. In figure 5.9a, the convergence region is symmetrical around $w = 0$, which is substantially different from the convergence regions of FIPS and PSO where there is a preference towards a positive inertia weight selection for particle convergence. The convergence regions for both cases 1 and 2, as seen in figures 5.9a and 5.9b, are very similar. However, the convergence region for case 2 is slightly larger. Note that, with or without the special treatment of the center of gravity calculation, there is no substantial change in the convergence region. The found convergence regions in both figures 5.9a and 5.9b are substantially different in size from the found region of Bonyadi and Michalewicz [4]. The convergence region found by Bonyadi and Michalewicz has its apex at $c_1 + c_2 = 12$ as opposed to the apex in figure 5.9a at around 8.5. At present it is not completely clear what the exact source of the discrepancy is. However, possible sources present in the work of Bonyadi and Michalewicz [4] are as follows:

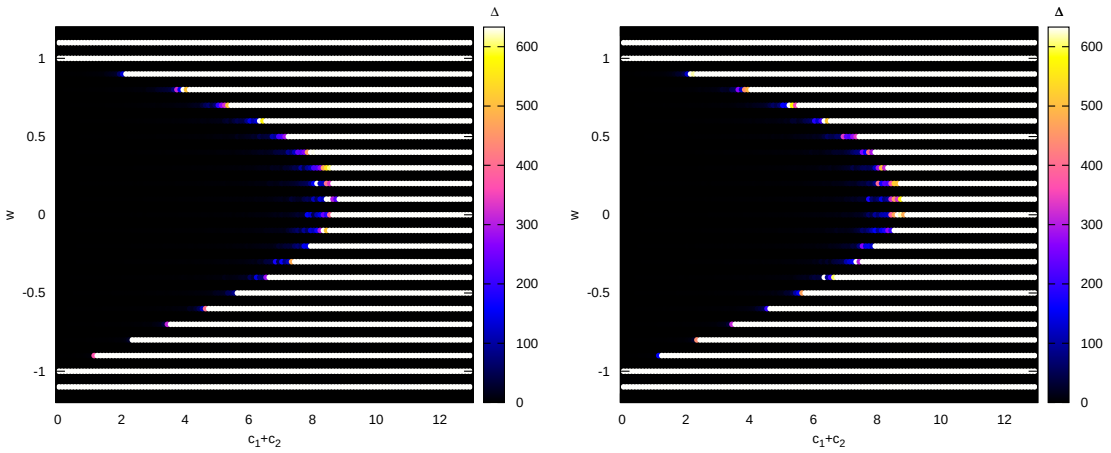
- the study is performed under the presence of forced stagnation;
- the particles' personal and neighborhood best positions are set to be equal;



(a) SPSO2011: 1 dimension, case 1.

(b) SPSO2011: 1 dimension, case 2.

FIGURE 5.9: SPSO2011 convergence results for 1 dimension.



(a) SPSO2011: 10 dimensions, case 1.

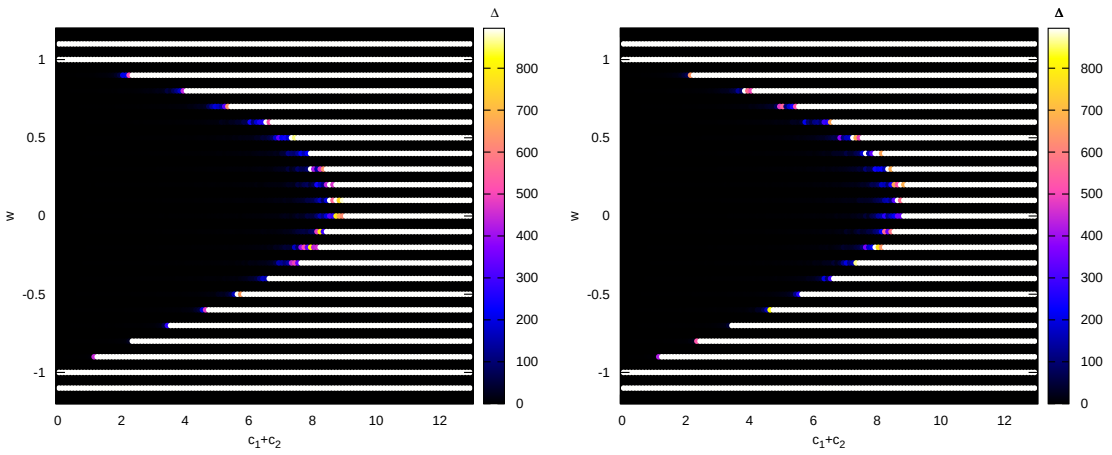
(b) SPSO2011: 10 dimensions, case 2.

FIGURE 5.10: SPSO2011 convergence results for 10 dimensions.

- there is a linear increase in the number of iterations used based on the dimensionality of the search space; however, the maximum distance between two points in a search space only increases sublinearly; and
- it is not stated how the spherical distribution, $\mathcal{H}$, is calculated.

The analysis done in this section made use of neither of the two mentioned simplifications. As a result, the regions presented in figures 5.9a and 5.9b should be a more accurate representation of SPSO2011's convergent behavior.

The convergence regions for SPSO2011 in 10 and 20 dimensions are presented in figures 5.10a, 5.10b, 5.11a and 5.11b. In both 10 and 20 dimensions, the difference in the



(a) SPSO2011: 20 dimensions, case 1.

(b) SPSO2011: 20 dimensions, case 2.

FIGURE 5.11: SPSO2011 convergence results for 20 dimensions.

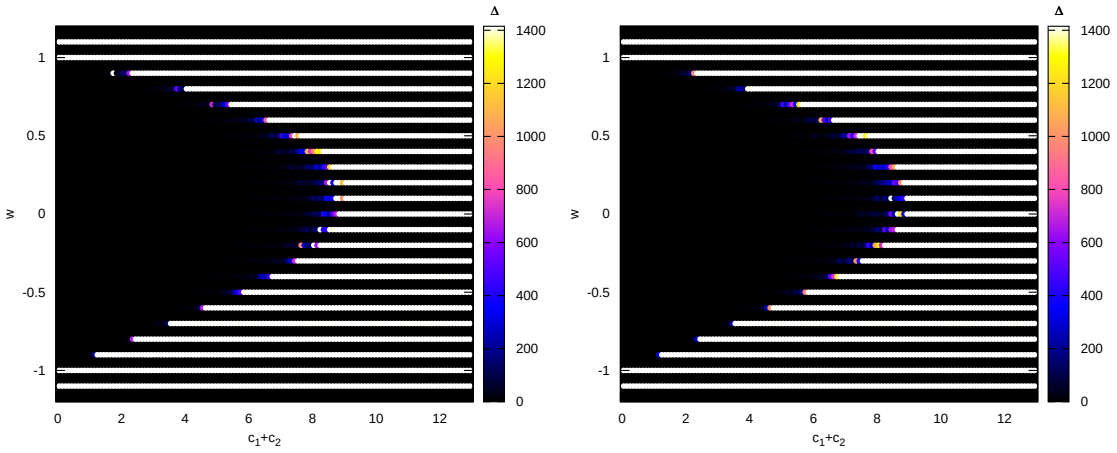
convergence regions for case 1 and case 2 is relatively minor. However, the convergence regions for case 2 are slightly larger than those of case 1. The convergence regions appear to be stable under an increase in dimension, as the regions plotted for 1, 10, and 20 dimensions appear unchanged for both cases 1 and 2.

The finding that the convergence region of SPSO2011 does not depend upon the dimensionality of the search space is in opposition to the results of Bonyadi and Michalewicz [4] . Even in 50 dimensions, as seen in figure 5.12a and 5.12b, the convergence region does not appear to change. In the work of Bonyadi and Michalewicz, the apex of the found convergence region decreases by 33.3% with the increase from an one dimensional search space to a 50 dimensional search space. This trend is clearly not present in the results of this subsection.

5.6 Performing Empirical Verification

This section presents a simple method for performing empirical verification of criteria for particle convergence.

Empirical verification of the criteria for particle convergence for PSO and PSO variants is performed by utilizing CF as the objective function for the PSO variant, along with



(a) SPSO2011: 50 dimensions, case 1.

(b) SPSO2011: 50 dimensions, case 2.

FIGURE 5.12: SPSO2011 convergence results for 50 dimensions.

Δ_{max} as a classification boundary between convergent and divergent particle behavior. The data collection approach is summarized in algorithm 7.2.

Once the necessary data is collected using algorithm 7.2, the empirical classification of each parameter tuple can be performed and compared against the theoretically obtained stability criteria (in PSO the parameter tuples are of the form (d, w, c_1, c_2), where d is the dimension of search space. A smaller tuple may be justifiable in practice if there is a good theoretical justification for why the parameter cannot effect the result). Each parameter tuple has a set of convergence measures associated with it, one for each sample run. In order for the classification between divergent and convergent to be made, the convergence measures must be summarized in some manner. The most direct approach would be to average the convergence measures across all sample runs. However, utilizing the maximum recorded convergence measures is also a sensible approach. If the maximum is used, the occurrence of divergent particle behavior in any sample run, for a parameter tuple, would result in the parameter tuple being classified as divergent. On the other hand, the mean could classify the parameter tuple as convergent if divergent behavior happens very infrequently. The use of the maximum means that if a parameter tuple is classified as causing convergent behavior, the parameter tuple always results in convergent behavior and not just convergent behavior most of the time.

Once the convergence measures over all sample runs are summarized for each parameter tuple, classification can be performed. If a parameter tuple results in a convergent

measure greater or equal to Δ_{max}, then that parameter tuple is classified as divergent. If a parameter tuple results in a convergent measure less than Δ_{max}, then that parameter tuple is classified as convergent. Once the classification of all parameter tuples is complete, the resulting classification can be compared with the theoretical prediction.

Algorithm 2 Data Collection for Empirical Verification of Stability Criteria

```
Let CF be the objective function as defined in equation (4.1)
Let s be the number of sample runs used
Let C be the set of parameter tuples under consideration
for all sample runs 1 ··· s do
    for all p ∈ C do
        Run PSO variant using the parameter tuple p for I iterations
        Record the convergence measure of the swarm utilizing equation (4.2)
    end for
end for
```

5.7 Summary

This chapter had two primary aims: The first was to show that the objective function, $CF(\boldsymbol{x}) \sim U(-1000, 1000)$, is an effective objective function to utilize for convergence region analysis. The second was to analyze the parameter region needed to ensure convergent particle behavior of particle swarm variants utilizing the proposed objective function, CF.

It was found that the CF objective function was able to capture the convergent behavior of the PSO, as the found convergence regions matched both the theoretically derived region of Poli [50], as defined in equation (3.13), and the "optimal" region, where the "optimal" region was constructed using the maximum convergence measure value across all topologies and objective functions used (excluding CF). It was also found that the social topology used by PSO had no meaningful impact on the convergence region.

Using the CF objective function, the convergence region was empirically obtained for FIPS, BPSO, and SPSO2011. It was observed that FIPS's convergence region grows with an increase in neighborhood size. It was shown that BPSO does not converge for any choice of c_1 and c_2. More specifically, in practice it was shown that BPSO is not order-2 stable, despite theoretical findings [3]. The discrepancy is linked to the

theoretical work being performed under the stagnation assumption. For SPSO2011 it was found that the convergence region does not depend on the dimensionality of the problem, as previously observed by Bonyadi and Michalewicz [4]. The region needed to ensure convergent particle behavior in SPSO2011 is also different from those obtained by Bonyadi and Michalewicz. The discrepancies are attributed to the simplifications used by Bonyadi and Michalewicz.

The next two chapters utilize the empirical approach discussed in this chapter as a means of verifying novel theoretically derived stability criteria for both the FIPS and the UPSO algorithms.

Chapter 6

Fully Informed Particle Swarm Optimization: Stability Analysis

This chapter presents novel theoretical derivations of the order-1 and order-2 stable regions for the fully informed PSO algorithm in section 6.1. The accuracy of the derived region is then empirically tested in an assumption free context in section 6.2 utilizing the approach presented in section 5.6. The summary of this chapter's findings are presented in section 6.3. Content from this chapter was published in the proceedings of the 2015 IEEE Congress on Evolutionary Computation [11].

6.1 Theoretical Derivation

This section presents theoretical derivations of the order-1 and order-2 stable regions for the FIPS algorithm.

Firstly, the stagnation assumption is used: specifically, it is assumed that when analyzing particle i that all $\boldsymbol{y}_m \in \mathcal{N}_i$ are stagnate. That is, all particles in particle i's neighborhood have a stagnant personal best position. Given that there is no dependence between the vector components of the update equation used by FIPS, it is possible to, without loss of generality, focus on an one dimensional particle trajectory. For the analysis in this

section it is assumed that the neighborhood size is the same for each particle and is denoted by $|\mathcal{N}|$.

The position update equation of FIPS, merged with the velocity update equation (2.3), is rearranged to the form

$$x_i(t+1) + \alpha_1 x_i(t) + \alpha_2 x_i(t-1) = \alpha_3 \tag{6.1}$$

where

$$\alpha_1 = -(1+w) + \frac{1}{|\mathcal{N}|}\sum_{i=1}^{|\mathcal{N}|}\theta_i \tag{6.2}$$

$$\alpha_2 = w \tag{6.3}$$

$$\alpha_3 = \frac{1}{|\mathcal{N}|}\sum_{i=1}^{|\mathcal{N}|}\theta_i y_i(t) \tag{6.4}$$

where $|\mathcal{N}|$ is the size of particle's neighborhood and $\theta_i \sim U(0, c_1 + c_2)$.

The criteria for order-1 stability are derived first. Let $c_1 + c_2 = \check{c}$. Then, application of the expectation operator to equation (6.1) leads to

$$E[x_i(t+1)] + E[\alpha_1]E[x_i(t)] + E[\alpha_2]E[x_i(t-1)] = E[\alpha_3] \tag{6.5}$$

where

$$E[\alpha_1] = -(1+w) + \frac{1}{|\mathcal{N}|}\sum_{i=1}^{|\mathcal{N}|}E[\theta_i] = -(1+w) + \frac{\check{c}}{2} \tag{6.6}$$

$$E[\alpha_2] = w \tag{6.7}$$

$$E[\alpha_3] = \frac{1}{|\mathcal{N}|}\sum_{i=1}^{|\mathcal{N}|}E[\theta_i]y_i = \frac{\check{c}}{2|\mathcal{N}|}\sum_{i=1}^{|\mathcal{N}|}y_i \tag{6.8}$$

In order to obtain the order-1 stable region, equation (6.5) is rewritten into the following equivalent matrix form:

$$\begin{bmatrix} x_i(t+1) \\ x_i(t) \end{bmatrix} = M \begin{bmatrix} x_i(t) \\ x_i(t-1) \end{bmatrix} + \begin{bmatrix} E[\alpha_3] \\ 0 \end{bmatrix} \tag{6.9}$$

where

$$M = \begin{bmatrix} -E[\alpha_1] & -E[\alpha_2] \\ 1 & 0 \end{bmatrix} \tag{6.10}$$

Now, if $\rho(M) < 1$, then order-1 stability is obtained [35], where ρ is the spectral radius of the matrix. The two eigenvalues of M are

$$\frac{-E[\alpha_1] \pm \sqrt{E[\alpha_1]^2 - 4E[\alpha_2]}}{2} \tag{6.11}$$

which means that the simplified conditions needed for $\rho(M) < 1$ are

$$|w| < 1 \text{ and } 0 < c_1 + c_2 < 4(w+1) \tag{6.12}$$

which are the same criteria for order-1 stability of PSO, as discussed in section 3.3. Given order-1 stability, it is possible to derive the fixed point of equation (6.5). Specifically, γ_{fips} can be calculated as

$$\gamma_{fips} = \frac{\sum_{i=1}^{|\mathcal{N}|} y_i}{|\mathcal{N}|} \tag{6.13}$$

by setting

$$E[x_i(t-1)] = E[x_i(t)] = E[x_i(t+1)] = \gamma_{fips}$$

in equation (6.5) and solving for γ_{upso}.

Focus is now shifted to the derivation of criteria for order-2 stability. The proof technique utilizes a method proposed by Blackwell [3]. Blackwell considered a class of PSO algorithms of following form:

$$x_{ij}(t+1) + a(t)x_{ij}(t) + b(t)x_{ij}(t-1) = c(t, \mathcal{N}_i), \tag{6.14}$$

were $a(t)$ and $b(t)$ are random variables, and $c(t, \mathcal{N}_i)$ is a random variable that also depends on stagnant neighborhood position information. The three random variables do not depend on t in any way other than the sampling of stochastic components.

Blackwell showed that, if the particles are order-1 stable, then the following conditions are necessary for order-2 stability:

$$1 + E[a] + E[b] \neq 0 \tag{6.15}$$

$$1 - E\left[a^2\right] - E\left[b^2\right] + \left(\frac{2E[ab]\,E[a]}{1 + E[b]}\right) > 0 \tag{6.16}$$

Blackwell's conditions for order-2 stability can be directly used for FIPS by setting $a = \alpha_1$, $b = \alpha_2$ and $c = \alpha_3$. In order to utilize Blackwell's conditions, a few extra expected values are required. Each of the required expectations are calculated below:

$$E\left[\alpha_2^2\right] = w^2 \tag{6.17}$$

$$E[\alpha_1\alpha_2] = -w(1+w) + \frac{w\check{c}}{2} \tag{6.18}$$

$$\begin{aligned} E\left[\alpha_1^2\right] &= E\left[\left(-(1+w) + \frac{1}{|\mathcal{N}|}\sum_{i=1}^{|\mathcal{N}|}\theta_i\right)^2\right] \\ &= E\left[(1+w)^2 - \frac{2}{|\mathcal{N}|}(1+w)\sum_{i=1}^{|\mathcal{N}|}\theta_i + \frac{1}{|\mathcal{N}|^2}\left(\sum_{i=1}^{|\mathcal{N}|}\theta_i\right)\left(\sum_{i=1}^{|\mathcal{N}|}\theta_i\right)\right] \\ &= (1+w)^2 - (1+w)\check{c} + \frac{1}{|\mathcal{N}|^2}E\left[\left(\sum_{i=1}^{|\mathcal{N}|}\theta_i\right)\left(\sum_{i=1}^{|\mathcal{N}|}\theta_i\right)\right] \end{aligned} \tag{6.19}$$

where

$$\begin{aligned} E\left[\left(\sum_{i=1}^{|\mathcal{N}|}\theta_i\right)\left(\sum_{i=1}^{|\mathcal{N}|}\theta_i\right)\right] &= E\left[\sum_{i=1}^{|\mathcal{N}|}{\theta_i}^2 + \sum_{i=1}^{|\mathcal{N}|}\sum_{j=1,j\neq i}^{|\mathcal{N}|}\theta_i\theta_j\right] \\ &= E\left[\sum_{i=1}^{|\mathcal{N}|}{\theta_i}^2\right] + E\left[\sum_{i=1}^{|\mathcal{N}|}\sum_{j=1,j\neq i}^{|\mathcal{N}|}\theta_i\theta_j\right] \\ &= \frac{|\mathcal{N}|c^2}{3} + \sum_{i=1}^{|\mathcal{N}|}\sum_{j=1,j\neq i}^{|\mathcal{N}|}E[\theta_i\theta_j] \\ &= \frac{|\mathcal{N}|c^2}{3} + P(|\mathcal{N}|,2)\frac{\check{c}^2}{4} \\ &= |\mathcal{N}|\check{c}^2\left(\frac{1+3|\mathcal{N}|}{12}\right) \end{aligned} \tag{6.20}$$

Merging equation (6.20) with equation (6.19) leads to

$$E\left[\alpha_1^2\right] = (1+w)^2 - (1+w)\check{c} + \frac{1}{12|\mathcal{N}|}\check{c}^2 + \frac{3}{12}\check{c}^2 \tag{6.21}$$

The first condition necessary for order-2 stability of FIPS, derived using equation (6.15), is:

$$1 - (1+w) + \frac{c_1 + c_2}{2} + w \neq 0 \tag{6.22}$$

which implies that

$$c_1 + c_2 \neq 0 \tag{6.23}$$

The second condition necessary for order-2 stability of FIPS, derived using equation (6.16), is:

$$\begin{aligned} 1 - \left((1+w)^2 - (1+w)\check{c} + \frac{1}{12|\mathcal{N}|}\check{c}^2 + \frac{3}{12}\check{c}^2\right) - w^2 \\ + \left(\frac{2w((1+w)^2 - (1+w)\check{c} + \frac{\check{c}^2}{4})}{1+w}\right) > 0 \end{aligned} \tag{6.24}$$

The condition can be simplified using the fact that $|w| < 1$ and $\check{c} = c_1 + c_2 > 0$ (from the order-1 stability condition from equation (6.12)) to the following:

$$c < \frac{12|\mathcal{N}|\left(1 - w^2\right)}{3|\mathcal{N}| + 1 + w\left(1 - 3|\mathcal{N}|\right)} \tag{6.25}$$

Equation (6.25) implies that the maximum convergence region is

$$c < \lim_{|\mathcal{N}|\to\infty} \frac{12|\mathcal{N}|\left(1 - w^2\right)}{3|\mathcal{N}| + 1 + w\left(1 - 3|\mathcal{N}|\right)} = 4\left(w + 1\right). \tag{6.26}$$

The convergence regions for FIPS with neighborhood size 1, 2, 4, 8, 16, 32, and 64 are illustrated in figure 6.1. The maximum convergence region of equation (6.26) is illustrated in figure 6.1 as the straight line ending at L. What is somewhat surprising is that the maximum convergence region as presented in equation (6.26) is the same as the order-1 stable region of equation (6.12).

If the neighborhood size is set to 2, then equation (6.25) leads to the following condition:

$$c < \frac{24\left(1 - w^2\right)}{7 - 5w} \tag{6.27}$$

which is in exact agreement with the order-2 stability region derived by Poli [51] for the PSO algorithm.

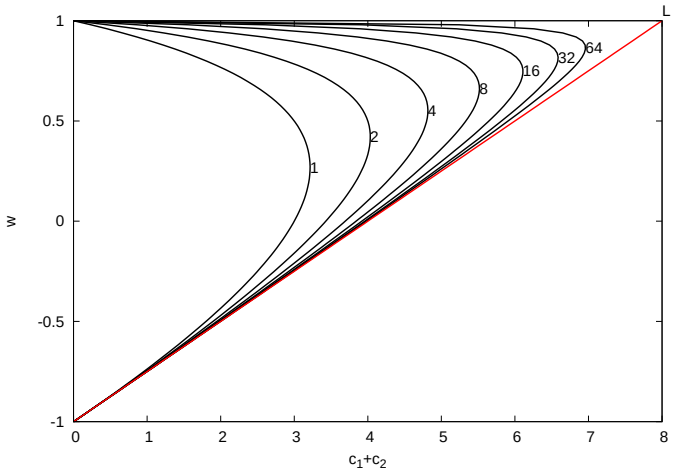


FIGURE 6.1: Derived convergence regions for $|\mathcal{N}| = 1, 2, 4, 8, 16, 32, 64$, and the maximum convergence region

It should be noted that in order for a particle's movement to be seen as convergent it should exhibit both order-1 and order-2 stability. Specifically, the conditions from both equation (6.12) and equation (6.25) should be satisfied to obtain convergent particle behavior. However, since equation (6.25) is a subset of the region $0 < c_1 + c_2 < 4(w+1)$ the criteria for order-1 and order-2 stability are given as

$$|w| < 1 \tag{6.28}$$

$$0 < c_1 + c_2 < \frac{12|\mathcal{N}|\left(1 - w^2\right)}{3|\mathcal{N}| + 1 + w\left(1 - 3|\mathcal{N}|\right)} \tag{6.29}$$

6.2 Empirical Validation

This section utilizes the method for empirically investigating the convergence region of PSO variants as proposed in section 5.6. Section 6.2.1 presents the experimental setup followed by the experimental results and discussion in section 6.2.2.

6.2.1 Empirical Setup

The experiment in this section utilizes a population size of 64, and 5000 iterations. Particle positions were initialized within $(-100, 100)$ and velocities were initialized to $\mathbf{0}$. The analysis is in 50 dimensions, resulting in the maximum possible distance, Δ_{max}, of 1414.214 between particles in the initial search space. Reported results were bounded at Δ_{max} to prevent highly divergent parameter configurations from obscuring the data. The LBest topology was used. The neighborhood sizes 2, 4, 8, 16, 32, and 64 were considered. For all neighborhood sizes used, each particle is set to be within its own neighborhood.

The empirical measure of convergence used in this chapter is that of equation (4.2), and the objective function used is CF, as defined in equation (4.1). The experiment was conducted over the following parameter region:

$$w \in [-1.1, 1.1] \text{ and } c_1 + c_2 \in (0, 7]\,, \tag{6.30}$$

where $c_1 = c_2$, with a sample point every 0.1 along w and $c_1 + c_2$. A total of 1610 sample points from the region defined in equation (6.30) were used. The region of equation (6.30) was chosen as it contains the derived regions for neighborhood sizes $|\mathcal{N}| <= 64$. The results reported in section 5.1.2 are averages over 35 independent runs for each sample point.

6.2.2 Experimental Results and Discussion

This section presents the results of the experiments described in section 6.2.1.

A snapshot of all parameter configurations' resulting convergence measure values are presented in figures 6.2a to 6.2f for the 5000th iteration for FIPS with neighborhood sizes 2, 4, 8, 16, 32, and 64. The number of parameter configuration that empirically agree or disagree with the convergence/divergence behavior predicted by the theoretically derived convergence region of equation (6.16) is presented in table 6.1. Eight measurements are given in table 6.1: the number of parameter configurations that are

theoretically convergent (TC) and divergent (TD), the number of parameter configurations that were empirically convergent (EC) and divergent (ED), the number of parameter configurations that were found to be empirically convergent despite the theory predicting divergence (EC & TD), the number of parameter configurations that were found to be empirically divergent despite the theory predicting convergence (ED & TC), and lastly the percentage error and agreement between the theoretical derivation and the empirical finding.

A parameter configuration is classified to be convergent if the value of the recorded convergence measure of equation (4.2) is less than Δ_{max}, and divergent if greater than or equal to Δ_{max}. A summary of the empirical convergence results are presented in figure 6.3, where each parameter configuration is labeled with a symbol corresponding to the lowest neighborhood size for which that parameter configuration was classified as convergent.

The empirically obtained convergence region of FIPS with a neighborhood size of 2 is illustrated in figure 6.2a. The general shape and size of the convergence region is in agreement with that of equation (6.16), as can be seen by comparing figure 6.1 with figure 6.2a. The theoretically derived convergence region agrees with the empirical findings for 99.255% of the tested parameter configurations, with only 0.621% of parameter configurations incorrectly classified as divergent by the theory, and 0.124% incorrectly classified as convergent by the theory, as seen in table 6.1.

TABLE 6.1: Theoretical prediction versus empirical findings

NBH size	TC	TD	EC	ED	EC & TD	ED & TC	Error	Agreement
2	504	1106	512	1098	10	2	0.745%	99.255%
4	585	1025	561	1049	1	25	1.615%	98.385%
8	719	891	626	984	3	96	6.150%	93.850%
16	760	850	675	935	5	90	5.901%	94.099%
32	786	824	711	899	8	83	5.652%	94.348%
64	800	810	737	873	13	76	5.528%	94.472%

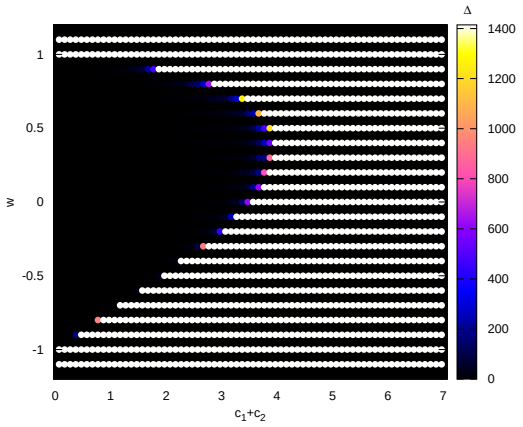
The mesaurements presented are, the number of parameter configurations that are theoretically convergent (TC) and divergent (TD), the number of parameter configurations that where empirically convergent (EC) and divergent (ED), the number of parameter configurations that were found to be empirically convergent despite the theory predicting divergence (EC & TD), the number of parameter configurations that were found to be empirically divergent despite the theory predicting convergence (ED & TC), and lastly the percentage error and agreement between the theoretically derivation and the empirical finding.

The empirically obtained convergence region of FIPS with a neighborhood size of 4 is illustrated in figure 6.2a. Visually, the convergence region is in agreement with that of equation (6.16). The theoretically derived convergence region agrees with the empirical findings for 98.385% of the tested parameter configurations. Unlike FIPS with $|\mathcal{N}| = 2$, when $|\mathcal{N}| = 4$ there are substantially more configurations incorrectly classified as convergent by the theory than divergent, with 1.553% and 0.06% respectively. However, the theory is still nearly perfectly inline with the empirical findings.

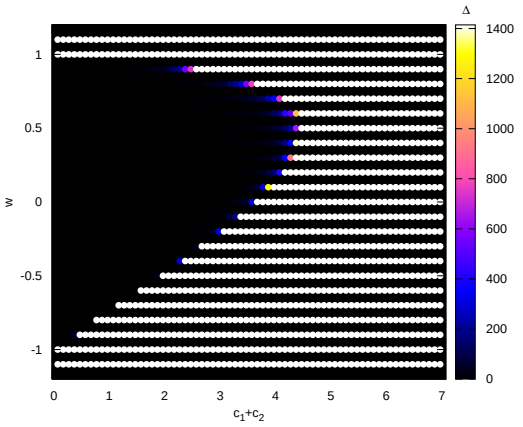
The empirically obtained convergence region of FIPS with a neighborhood size of 8 and 16 is illustrated in figures 6.2c and 6.2d respectively. The theoretically derived convergence region for $k = 8$ and $k = 16$ agrees with the empirical findings for 93.850% and 94.099% of the tested parameter configurations respectively. While the level of accuracy is less than that of $|\mathcal{N}| = 2, 4$ the theory still provides a very close fit with the empirical results. The decrease in accuracy from $|\mathcal{N}| = 2, 4$ to $|\mathcal{N}| = 8, 16$ can be attributed to two things: firstly the increase in the number of sources of randomness, and secondly that the stagnation assumption becomes stronger the larger k becomes. Despite this justification it is unclear why the accuracy for $|\mathcal{N}| = 8$ is lower than that of $|\mathcal{N}| = 16$. It appears that a specific neighborhood size exists where the effect of the multiple sources of randomness is maximal, because all considered neighborhood sizes greater than 8 have a better level of agreement with the theory as seen in table 6.1. Visually, the found convergence regions for neighborhood size 8 and 16 are in agreement with that of figure 6.1, with the convergence regions clearly becoming more triangular in shape as $|\mathcal{N}|$ increases.

The empirically obtained convergence region of FIPS with a neighborhood size of 32 and 64 is illustrated in figures 6.2e and 6.2f respectively. The theoretically derived convergence region for $k = 32$ and $|\mathcal{N}| = 64$ agrees with the empirical findings for 94.348% and 94.472% of the tested parameter configurations respectively. The convergence region obtain in figure 6.2f for a neighborhood size of 64 is nearly completely triangular in shape as predicted by equation (6.16).

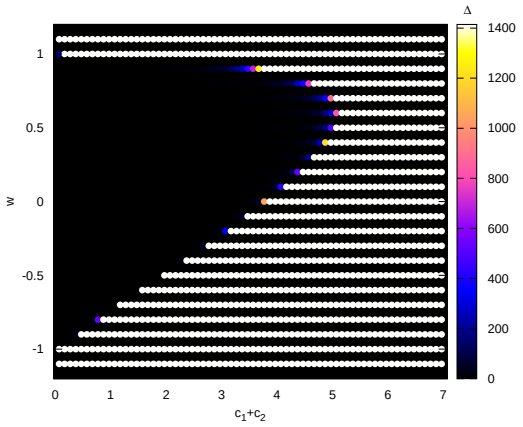
The change in convergence region with the increase in neighborhood size is clearly illustrated in figure 6.3, and is in agreement with the theoretical derivation as illustrated in



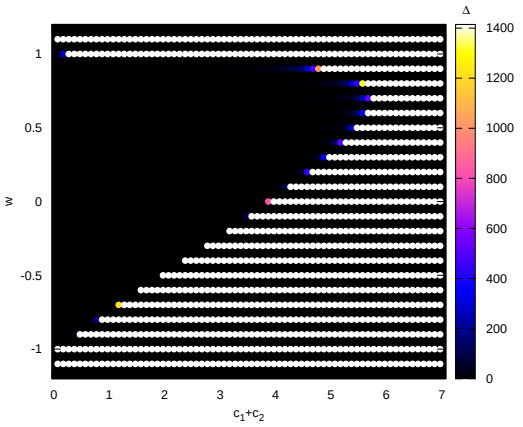
(a) FIPS convergence results for $|\mathcal{N}| = 2$ in 50 dimensions.



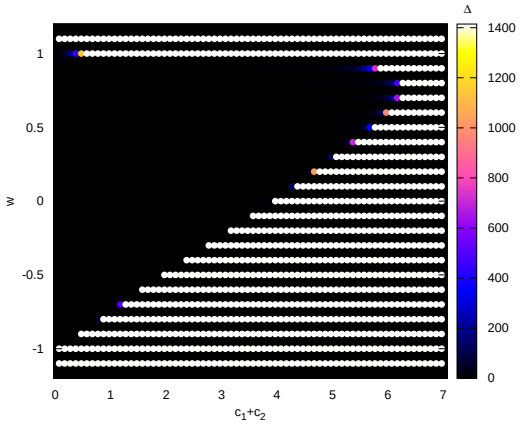
(b) FIPS convergence results for $|\mathcal{N}| = 4$ in 50 dimensions.



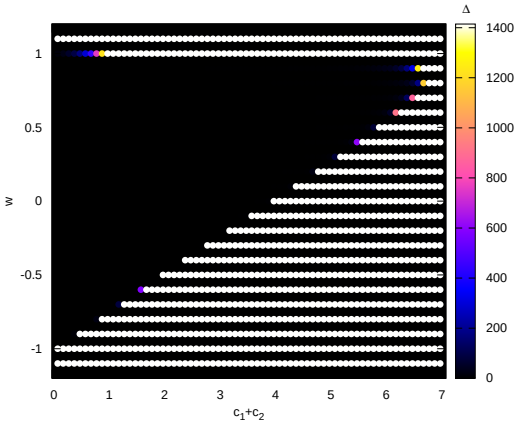
(c) FIPS convergence results for $|\mathcal{N}| = 8$ in 50 dimensions.



(d) FIPS convergence results for $|\mathcal{N}| = 16$ in 50 dimensions.



(e) FIPS convergence results for $|\mathcal{N}| = 32$ in 50 dimensions.



(f) FIPS convergence results for $|\mathcal{N}| = 64$ in 50 dimensions.

FIGURE 6.2: FIPS convergence results

figure 6.1.

It should be noted that for all neighborhood sizes other than two, the empirically obtained convergence regions were slightly smaller than the theoretically derived region, as can be seen in table 6.1 by the larger number of misclassification of configurations as being convergent by the theory. This behavior is almost certainly due to an increase in difficultly for the swarm to become stagnate when parameter configurations near the border of equation (6.16) are used.

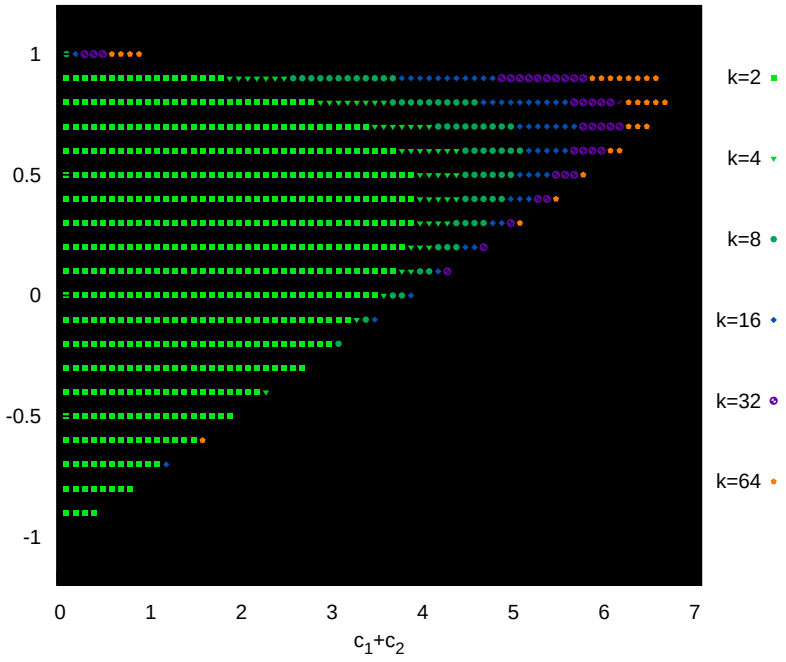


FIGURE 6.3: FIPS convergence results summary

Given the results of this section, it is clear that the convergence criteria derived in section 6.1 are an accurate representation of an assumption free FIPS algorithm. However, in practice, it is advisable to avoid parameter configurations that are particularly close to the boundary of equation (6.16), given the results of this section.

6.3 Summary

The convergence criteria for FIPS were theoretically derived under the stagnation assumption. The convergence criteria where validated empirically utilizing the method

presented in Chapter 5, were no simplifying assumptions were made on the FIPS algorithm. Given the empirical validation, the theoretical derivation is an accurate representation of FIPS's convergence region despite the region being derived under the stagnation assumption. It should be noted that, when the content of this chapter was accepted for publication, it was the first theoretical derivation of explicit stability criteria for FIPS using an arbitrary neighborhood size.

The next chapter performs stability analysis of the UPSO algorithm utilizing the same approach as used in this chapter.

Chapter 7

Unified Particle Swarm Optimization: Stability Analysis

This chapter presents novel theoretical derivations of the order-1 and order-2 stable regions for the unified PSO (UPSO) algorithm in section 7.1. The accuracy of the derived regions is then empirically tested in an assumption free context in section 7.2 utilizing the approach presented in section 5.6. A summary of this chapter's findings is presented in section 7.3. Content from this chapter was published in the proceeding of the 2016 IEEE Congress on Evolutionary Computation [14].

7.1 Theoretical Derivation

This section presents the theoretical derivation of the order-1 and order-2 stable regions for the UPSO algorithm. The point of particle convergence is also derived.

Firstly, the stagnation assumption is used: In UPSO's update equation, as defined in equation (2.5), it is assumed that $\boldsymbol{g}_i(t) = \boldsymbol{g}$, $\hat{\boldsymbol{y}}_i(t) = \hat{\boldsymbol{y}}$ and $\boldsymbol{y}_i(t) = \boldsymbol{y}$ for all t. Given that there is no dependence between the vector components of the update equation, it is possible to, without loss of generality, focus on a one dimensional particle trajectory.

he position update equation of UPSO, merged with the velocity update equation (2.5), is rearranged to the form

$$x_i(t+1) + \alpha_1 x_i(t) + \alpha_2 x_i(t-1) = \alpha_3 \tag{7.1}$$

where

$$\begin{aligned}
\alpha_1 &= -(1+w) + u\left(r_1 - r_1'\right)c_1 + u(r_2 - r_2')c_2 + c_1 r_1' + c_2 r_2' \\
\alpha_2 &= w \\
\alpha_3 &= u\left(r_1 - r_1'\right)c_1 y + uc_2(r_2 g - r_2'\hat{y}) + c_1 r_1' y + c_2 r_2'\hat{y}.
\end{aligned} \tag{7.2}$$

Application of the expectation operator to equation (7.1) results in the following equation:

$$E[x_i(t+1)] + E[\alpha_1]E[x_i(t)] + E[\alpha_2]E[x_i(t-1)] = E[\alpha_3] \tag{7.3}$$

where

$$\begin{aligned}
E[\alpha_1] &= -(1+w) + \frac{c_1}{2} + \frac{c_2}{2} \\
E[\alpha_2] &= w \\
E[\alpha_3] &= \frac{uc_2(g-\hat{y})}{2} + \frac{c_1 y}{2} + \frac{c_2\hat{y}}{2}.
\end{aligned} \tag{7.4}$$

In order to obtain the order-1 region, equation (7.3) is rewritten into the following matrix form:

$$\begin{vmatrix} x_i(t+1) \\ x_i(t) \end{vmatrix} = M \begin{vmatrix} x_i(t) \\ x_i(t-1) \end{vmatrix} + \begin{vmatrix} E[\alpha_3] \\ 0 \end{vmatrix} \tag{7.5}$$

where

$$M = \begin{vmatrix} -E[\alpha_1] & -E[\alpha_2] \\ 1 & 0 \end{vmatrix} \tag{7.6}$$

Now, if $\rho(M) < 1$, then order-1 stability is obtained [8], where ρ is the spectral radius of the matrix. The two eigenvalues of M are

$$\frac{-E[\alpha_1] \pm \sqrt{E[\alpha_1]^2 - 4E[\alpha_2]}}{2} \tag{7.7}$$

which means that the simplified conditions needed for $\rho(M) < 1$ are

$$|w| < 1 \text{ and } 0 < c_1 + c_2 < 4(w + 1) \tag{7.8}$$

So the order-1 stable region for UPSO is the same as the order-1 stable region found for both PSO and FIPS as discussed in chapters 3 and 6.

In the presence of order-1 stability, a fixed point γ_{upso} exists for equation (7.3). Specifically, γ_{upso} is calculated as

$$\gamma_{upso} = \frac{uc_2(g - \hat{y}) + c_1 y + c_2 \hat{y}}{c_1 + c_2} \tag{7.9}$$

by setting

$$E[x_i(t-1)] = E[x_i(t)] = E[x_i(t+1)] = \gamma_{upso}$$

in equation (7.3) and solving for γ_{upso}.

If $u = 0$ or $u = 1$ in equation (7.9) the same fixed point is obtained for the canonical PSO as found in [51][61] for the LBest and GBest PSO respectively.

In order to obtain the conditions necessary for order-2 stability, once again the proof technique derived by Blackwell [3] is used. The following conditions are necessary for order-2 stability:

$$1 + E[a] + E[b] \neq 0 \tag{7.10}$$

$$1 - E[a^2] - E[b^2] + \left(\frac{2E[ab]\,E[a]}{1 + E[b]}\right) > 0 \tag{7.11}$$

In the case of this chapter $a = \alpha_1$, $b = \alpha_2$.

It is not clear in the original paper [45] if r_i' and r_i are in fact distinct for $i = 1, 2$. As a result, this chapter distinguishes the cases

- $r_1' = r_1$ and $r_2' = r_2$
- $r_1' \neq r_1$ and $r_2' \neq r_2$

Both of the cases result in the same order-1 stable region, so it was not necessary to distinguish the cases earlier. If the first case is considered, UPSO's update equation can be simplified such that

$$\begin{aligned}\alpha_1 &= -(1+w) + c_1 r_1 + c_2 r_2 \\ \alpha_2 &= w \\ \alpha_3 &= u c_2 r_2 (g - n) + c_1 r_1 y + c_2 r_2 n\end{aligned} \tag{7.12}$$

Now it should be noted that the conditions necessary for order-2 stability in equations (7.10) and (7.11) do not depend on α_3. The terms α_1 and α_2 are exactly the same as in the case of the PSO algorithm, which means that the derivation of PSO's order-2 stable region in [3] is the same as that of UPSO when $r_1' = r_1$ and $r_2' = r_2$. Therefore, it is not necessary to proceed any further with this case.

The second case where $r_1' \neq r_1$ and $r_2' \neq r_2$ is more interesting. In order to use equations (6.15) and (6.16), a number of expected value are needed:

$$\begin{aligned}E[\alpha_1] &= -(1+w) + uE[(r_1 - r_1')]c_1 + uE[(r_2 - r_2')]c_2 + c_1 E[r_1'] + c_2 E[r_2'] \\ &= -(1+w) + u(0)c_1 + u(0)c_2 + c_1\frac{1}{2} + c_2\frac{1}{2} \\ &= -(1+w) + c_1\frac{1}{2} + c_2\frac{1}{2}\end{aligned} \tag{7.13}$$

$$E[\alpha_1]^2 = (1+w)^2 - (1+w)(c_1+c_2) + \frac{(c_1+c_2)^2}{4}, \tag{7.14}$$

$$E[\alpha_2] = w \tag{7.15}$$

$$E[\alpha_2]^2 = w^2 \tag{7.16}$$

$$E[\alpha_1\alpha_2] = E[\alpha_1]E[\alpha_2] = -w(1+w) + \frac{c_1 w}{2} + \frac{c_2 w}{2}. \tag{7.17}$$

Moving on to the more lengthy calculation, $E[\alpha_1^2]$ needs to be calculated. Now,

$$\begin{aligned}
\alpha_1^2 &= (-(1+w) + c_1u(r_1 - r_1') + c_2u(r_2 - r_2') + c_1r_1' + c_2r_2')^2 \\
&= (1+w)^2 - (1+w)c_1u(r_1 - r_1') - (1+w)c_2u(r_2 - r_2') \\
&- (1+w)c_1r_1' - (1+w)c_2r_2' - (1+w)c_1u(r_1 - r_1') \\
&+ c_1^2u^2(r_1 - r_1')^2 + c_1c_2u^2(r_1 - r_1')(r_2 - r_2') + c_1^2ur_1'(r_1 - r_1') \\
&+ c_1c_2ur_2'(r_1 - r_1') - (1+w)c_2u(r_2 - r_2') \\
&+ c_1c_2u^2(r_1 - r_1')(r_2 - r_2') + c_2^2u^2(r_2 - r_2')^2 + c_1c_2ur_1'(r_2 - r_2') \\
&+ c_2^2r_2'u(r_2 - r_2') - (1+w)c_1r_1' + c_1^2ur_1'(r_1 - r_1') \\
&+ c_1c_2r_1'u(r_2 - r_2') + c_1^2(r_1')^2 + c_1c_2r_1'r_2' - (1+w)c_2r_2' \\
&+ c_1c_2r_2'u(r_1 - r_1') + c_2^2r_2'u(r_2 - r_2') + c_1c_2r_1'r_2' + c_2^2(r_2')^2
\end{aligned} \tag{7.18}$$

Application of the expectation operator leads to

$$\begin{aligned}
E[\alpha_1^2] &= (1+w)^2 - (1+w)c_1uE[r_1 - r_1'] - (1+w)c_2uE[r_2 - r_2'] \\
&- (1+w)c_1E[r_1'] - (1+w)c_2E[r_2'] - (1+w)c_1uE[r_1 - r_1'] \\
&+ c_1^2u^2E[(r_1 - r_1')^2] + c_1c_2u^2E[(r_1 - r_1')(r_2 - r_2')] \\
&+ c_1^2uE[r_1'(r_1 - r_1')] + c_1c_2uE[r_2'(r_1 - r_1')] \\
&- (1+w)c_2uE[r_2 - r_2'] + c_1c_2u^2E[(r_1 - r_1')(r_2 - r_2')] \\
&+ c_2^2u^2E[(r_2 - r_2')^2] + c_1c_2uE[r_1'(r_2 - r_2')] + c_2^2uE[r_2'(r_2 - r_2')] \\
&- (1+w)c_1E[r_1'] + c_1^2uE[r_1'(r_1 - r_1')] + c_1c_2uE[r_1'(r_2 - r_2')] \\
&+ c_1^2E[(r_1')^2] + c_1c_2E[r_1'r_2'] - (1+w)c_2E[r_2'] \\
&+ c_1c_2uE[r_2'(r_1 - r_1')] + c_2^2uE[r_2'(r_2 - r_2')] \\
&+ c_1c_2E[r_1'r_2'] + c_2^2E[(r_2')^2]
\end{aligned} \tag{7.19}$$

where

$$E[r_1 - r_1'] = E[r_2 - r_2'] = 0 \tag{7.20}$$

$$E[(r_1 - r_1')^2] = E[(r_1)^2 - 2r_1r_1' + (r_1')^2] = \frac{1}{6} = E[(r_2 - r_2')^2] \tag{7.21}$$

$$E[(r_1 - r_1')(r_2 - r_2')] = E[r_1r_2 - r_1r_2' - r_1'r_2 + r_1'r_2'] = 0 \tag{7.22}$$

$$E[r_2'(r_1 - r_1')] = 0 = E[r_1'(r_2 - r_2')] \tag{7.23}$$

$$E[r_1'(r_1 - r_1')] = E[r_1'r_1 - r_1'r_1'] = -\frac{1}{12} = E[r_2'(r_2 - r_2')] \tag{7.24}$$

After substituting back these expected values into equation (7.19) and simplifying, the following is obtained:

$$E[\alpha_1^2] = (1+w)^2 - (1+w)(c_1+c_2) + \frac{c_1^2+c_2^2}{6}(u^2-u) + \frac{c_1^2}{3} + \frac{c_1c_2}{2} + \frac{c_2^2}{3} \tag{7.25}$$

Using the calculated expected values it is now possible to obtain the conditions necessary for order-2 stability. The first condition obtained from equation (7.10) is

$$1 - (1+w) + \frac{c_1}{2} + \frac{c_2}{2} + w \neq 0 \implies c_1 + c_2 \neq 0 \tag{7.26}$$

The second condition obtained from equation (7.11) is calculated as follows:

$$\begin{aligned} 1 - \left((1+w)^2 - 2(1+w)\tilde{c} + \frac{2\tilde{c}^2}{6}(u^2-u) + \frac{7\tilde{c}^2}{6}\right) - w^2 \\ + \left(\frac{2w\left((1+w)^2 - 2(1+w)\tilde{c} + \tilde{c}^2\right)}{1+w}\right) > 0 \end{aligned} \tag{7.27}$$

where $\tilde{c} = c_1 = c_2$. Using $|w| < 1$ and $c_1 + c_2 > 0$ from the order-1 stable region of equation (7.8), equation (7.27) can be simplified to

$$\tilde{c} \;<\; \frac{12(1-w^2)}{7 - 5w + 2(u^2-u)(1+w)} \tag{7.28}$$

In order for a particle's movement to be seen as convergent, it should exhibit order-1 and order-2 stability. As such, the criteria for particle convergence is merged as

$$|w| < 1 \tag{7.29}$$

$$0 < c_1 + c_2 < \frac{24(1 - w^2)}{7 - 5w + 2(u^2 - u)(1 + w)} \tag{7.30}$$

This merger is possible because the region defined by (7.28) is a subset of the region defined by $0 < c_1 + c_2 < 4(w + 1)$.

The manner in which the choice of u affects the convergence region is illustrated in figure 7.1. Clearly, the closer u gets to 0.5, the more the convergence region's apex extends. As u approaches 0 or 1 from 0.5 the convergence region shrinks in a symmetric fashion, with both $u = 0$ and $u = 1$ resulting in the identical convergence region of the PSO algorithm, as defined in equation (3.13).

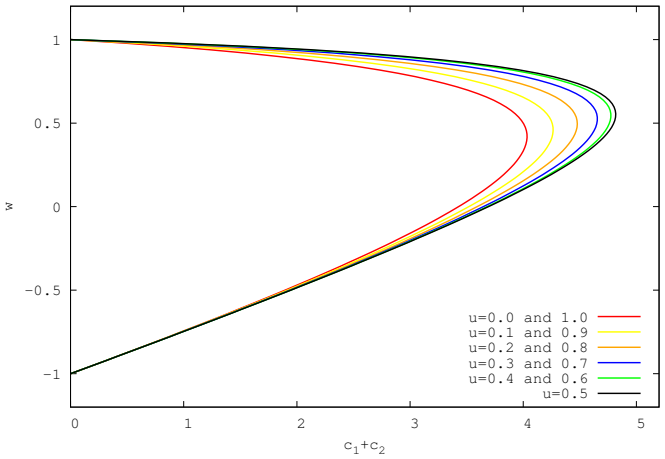


FIGURE 7.1: UPSO convergence regions for $u = 0$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1

7.2 Empirical Validation

This section utilizes the method for empirically investigating the convergence region of PSO variants as proposed in section 5.6. Section 7.2.1 presents the experimental setup followed by the experimental results and discussion in section 7.2.2.

7.2.1 Empirical Setup

The experiment of this chapter follows the same structure as was used in chapter 6. A population size of 64, and 5000 iterations were uses. Particle positions were initialized within $(-100, 100)$ and velocities were initialized to $\mathbf{0}$. The analysis is in 50 dimensions, resulting in $\Delta_{max} = 1414.214$. Reported results were bounded at Δ_{max} to prevent highly divergent parameter configurations from obscuring the data. The ring topology was used. The unification factors $u = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ and 0.9 are considered

The empirical measure of convergence used in this chapter is that of equation (4.2), and the objective function used is CF, as defined equation (4.1). The experiment was conducted over the following parameter region:

$$w \in [-1.1, 1.1] \text{ and } c_1 + c_2 \in (0, 5.2]\,, \tag{7.31}$$

where $c_1 = c_2$, with a sample point every 0.1 along w and $c_1 + c_2$. A total of 1196 sample points from the region defined in equation (7.31) were used. The region of equation (7.31) was chosen as it contains the derived regions for all values of the unification factor $u \in [0, 1]$. The results reported in section 7.2.2 are derived from 35 independent runs for each sample point.

7.2.2 Experimental Results and Discussion

This section presents the results of the experiments described in section 7.2.1.

A snapshot of all parameter configurations' resulting convergence measure values are presented in figures 7.2a to 7.4c for the 5000th iteration for UPSO with unification factors $u = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ and 0.9. The reported convergence measures are the maximum recorded over the 35 independent runs. The maximum was utilized in this chapter as apposed to the mean because the differing convergence regions (as defined by equation (7.30)) are more closely spaced than the convergence regions of FIPS (as defined in equation (6.25)) were in chapter 6.

The number of parameter configurations that empirically agree or disagree with the convergence/divergence behavior predicted by the theoretically derived convergence region of equations (7.29) and (7.30) is presented in table 7.1. The same eight measurements as used in chapter 6 are given in table 7.1.

A parameter configuration is classified to be convergent if the value of the recorded convergence measure of equation (4.2) is less than Δ_{max}, and divergent if greater than or equal to Δ_{max}.

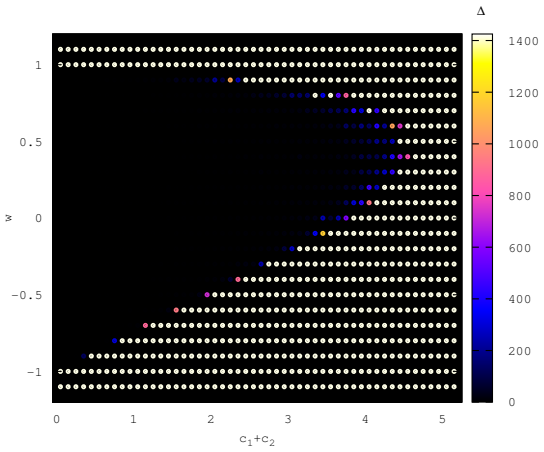
TABLE 7.1: Theoretical prediction versus empirical findings

u	TC	TD	EC	ED	EC & TD	ED & TC	Error	Agreement
0.1	532	664	578	618	46	0	3.8462%	96.1538%
0.2	552	644	597	599	45	0	3.7625%	96.2375%
0.3	571	625	618	578	47	0	3.9298%	96.0702%
0.4	581	615	626	570	45	0	3.7625%	96.2375%
0.5	587	609	633	563	46	0	3.8462%	96.1538%
0.6	581	615	621	575	40	0	3.3445%	96.6555%
0.7	571	625	613	586	42	0	3.5117%	96.4883%
0.8	552	644	603	593	41	0	4.2642%	95.7358%
0.9	532	664	580	616	48	0	4.0134%	95.9866%

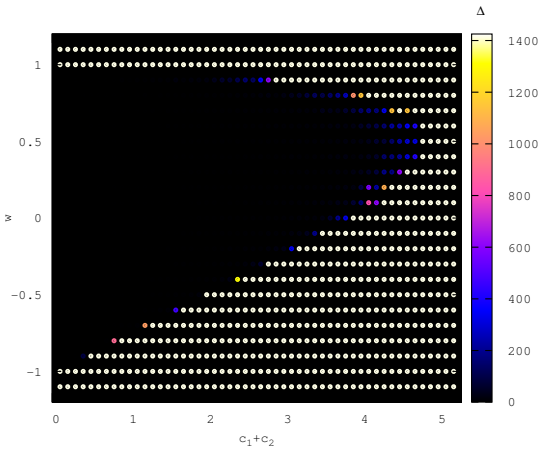
The mesaurements presented are the number of parameter configurations that are theoretically convergent (TC) and divergent (TD), the number of parameter configurations that were empirically convergent (EC) and divergent (ED), the number of parameter configurations that were found to be empirically convergent despite the theory predicting divergence (EC & TD), the number of parameter configurations that were found to be empirically divergent despite the theory predicting convergence (ED & TC), and lastly the percentage error and agreement between the theoretically derivation and the empirical finding.

The empirically obtained convergence region of UPSO with a unification factor of 0.1 is illustrated in figure 7.2a. The general shape and size of the convergence region is in agreement with that of equations (7.29) and (7.30), as can be seen by comparing figure 7.1 with figure 7.2a. The theoretically derived convergence region agrees with the empirical findings for 96.1538% of the tested parameter configurations as seen in table 7.1.

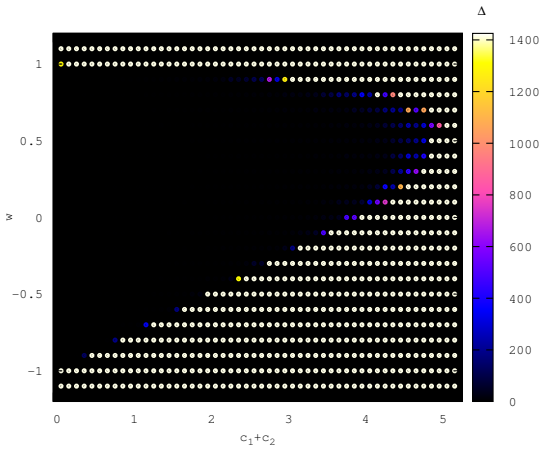
The empirically obtained convergence region of UPSO with a unification factor of 0.9 was in agreement with the theoretically derived region for 95.9866% of the parameter configurations, and is nearly identical to that of UPSO with a unification factor of 0.1 as can be seen in figures 7.4c and 7.2a, as was predicted by the theoretically derived region. As expected, the empirically obtained convergence region of UPSO with a unification factor of 0.2 is nearly identical to that of 0.8 as illustrated in figures 7.2b and 7.4b.



(a) UPSO convergence results for $u = 0.1$ in 50 dimensions.



(b) UPSO convergence results for $u = 0.2$ in 50 dimensions.

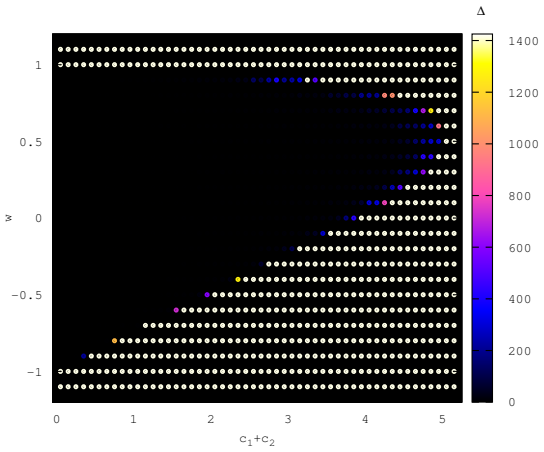


(c) UPSO convergence results for $u = 0.3$ in 50 dimensions.

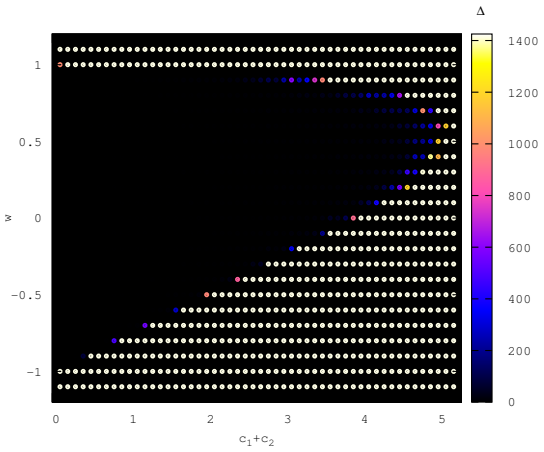
FIGURE 7.2: UPSO convergence results for $u = 0.1$, 0.2, and 0.3.

The level of agreement with the theoretical derivation was 96.2375% and 95.7358% for a unification factor of 0.2 and 0.8 respectively.

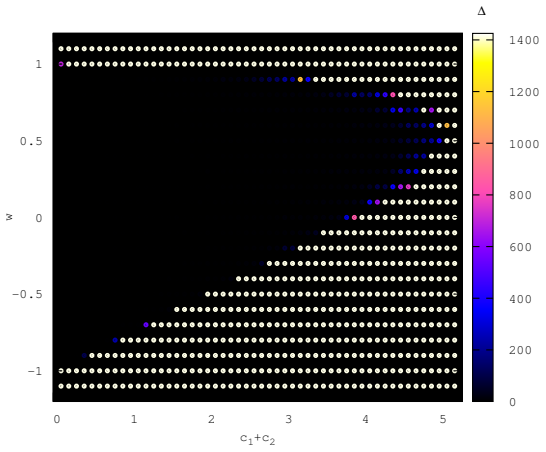
Figures 7.2c and 7.3a show that, for the unification factors 0.3 and 0.4, the apex of the convergence region extends more the closer the unification factor gets to 0.5, as was predicted by the theory. The level of agreement with the theoretically derived region is 96.0702% and 96.2375% for unification factors 0.3 and 0.4 respectively. The convergence region obtained for the unification factors 0.3 and 0.4 are nearly identical to that of the convergence region obtained for the unification factors 0.7 and 0.6 respectively, as shown in figures 7.2c, 7.3a, 7.3c, and 7.4a. The identical convergence regions are, once



(a) UPSO convergence results for $u = 0.4$ in 50 dimensions.



(b) UPSO convergence results for $u = 0.5$ in 50 dimensions.

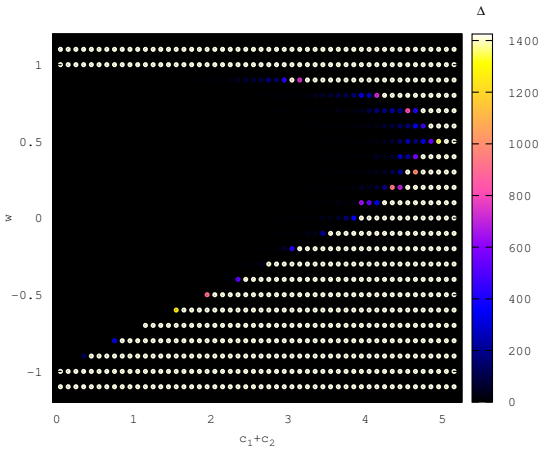


(c) UPSO convergence results for $u = 0.6$ in 50 dimensions.

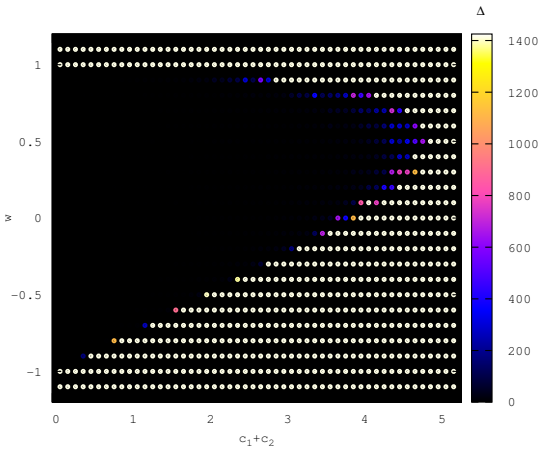
FIGURE 7.3: UPSO convergence results for $u = 0.4$, 0.5, and 0.6.

again, in complete agreement with the theoretically derivation. The empirically obtained convergence region of UPSO with a unification factor of 0.6 and 0.7 also had a very high level of agreement with the theoretical derived region with a 96.6555% and 96.4883% agreement respectively.

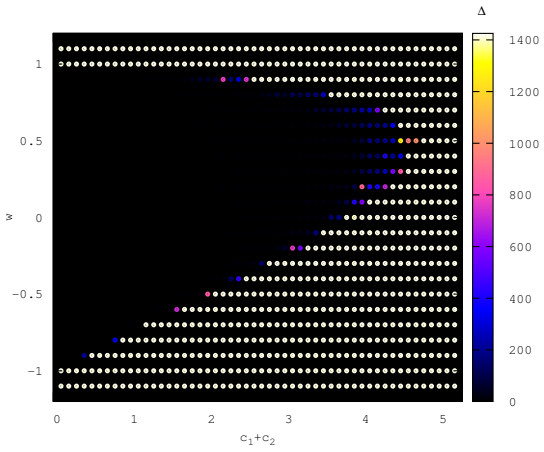
The empirically obtained convergence region of UPSO with a unification factor of 0.5 is the largest of all the convergence regions as illustrated in figure 7.3b, which was predicted by the theory. The level of agreement with the theory for a unification factor of 0.5 was 96.1538%. There is, however, very little difference between the convergence region for a unification factor of 0.5 and both 0.4 and 0.6, as is to be expected given the small



(a) UPSO convergence results for $u = 0.7$ in 50 dimensions.



(b) UPSO convergence results for $u = 0.8$ in 50 dimensions.



(c) UPSO convergence results for $u = 0.9$ in 50 dimensions.

FIGURE 7.4: UPSO convergence results for $u = 0.7$, 0.8, and 0.9.

predicted difference by the theory as illustrated in figure 7.1.

What is interesting is that there are absolutely no cases where a parameter configuration is predicted to be convergent by the theory but that was empirically detected to be divergent. This lack of false positives (false convergent configurations) is very indicative of the fact that the utilization of the theoretically derived region will completely avoid unwanted divergent behavior. It should also be noted that, in most cases, as is visible in figures 7.2a to 7.4c, there are a few parameter configurations with a relatively high convergent measure value around the boundary of the convergence regions that are less than Δ_{max}. These parameter configurations are the primary contributing factor to the

"EC & TD" measurement in table 7.1. Given this information and the comparatively large convergent measure values, it would be reasonable to assume that given enough iterations, the particle movement would in fact breach the classification threshold of Δ_{max}, and agree with the theoretical derivation.

It is clear that the theoretically derived region for particle convergence described by equations (7.29) and (7.30) is an accurate representation of the real world parameter configurations necessary to ensure particle convergence.

7.3 Summary

This chapter derived the order-1 and order-2 stable regions for UPSO, along with the point of particle convergence. The derived criteria for convergence where validated empirically utilizing the method described in chapter 5, were no simplifying assumptions were made on the UPSO algorithm. Given the empirical validation, the theoretical derivation is an accurate representation of UPSO's convergence region despite the region being derived under the stagnation assumption.

The next chapter investigates the impact that particle stability has on PSO performance.

Chapter 8

Influence of Particle Stability on Performance

This chapter investigates the effect of particle stability on the ability of PSO to find optimal solutions. Specifically, the influence of order-2 stability on performance is investigated. The performance of PSO is tested across a wide range of objective functions in varying dimensionalities. Performance is also considered across differing iteration counts. Content from this chapter was accepted for publication in the proceedings of the 2016 IEEE Symposium Series on Swarm Intelligence [13].

Section 8.1 provides a simple theoretical explanation for why unstable particles are likely to result in poor performance. Section 8.2 presents the empirical setup. The experimental results and discussion of the results are presented in section 8.3. The chapter's findings are summarized section 8.4.

8.1 Why Particle Stability Should Effect Performance

This section defines what is meant by particle stability, and discusses the implications of particle stability.

A particle is considered stable if the sequence generated from the particle's positions is order-1 and order-2 stable, as defined in definitions 3.2 and 3.3 respectively. A particle

is termed unstable if the sequence generated from the particle's positions is not both order-1 and order-2 stable.

A large body of theoretical analysis of PSO exists as discussed in 3. Currently, the most general theoretical PSO model is from Bonyadi and Michalewicz [7], which is an extension of the work of Poli [50]. This general model relies 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

$$\boldsymbol{z}_t = \boldsymbol{M}\boldsymbol{z}_{t-1} + \boldsymbol{b}, \tag{8.1}$$

where $\boldsymbol{b}$ is a constant offset in $\mathbb{R}^q$ and $\boldsymbol{M}$ a matrix from $\mathbb{R}^q$ to $\mathbb{R}^q$. The recurrence relation's initial term is defined as $\boldsymbol{z}_1$. The sequence (z_t) converges if and only if $\rho(\boldsymbol{M}) < 1$ [1], where ρ denotes the spectral radius as defined in definition C.11.

The exact values used for $\boldsymbol{M}$ and $\boldsymbol{b}$ by Bonyadi and Michalewicz [7] and Poli [50] can be found in the respective works. The fully generalized values for $\boldsymbol{M}$ and $\boldsymbol{b}$ can also be found in chapter 9. However, the exact values of $\boldsymbol{M}$ and $\boldsymbol{b}$ are not the focus of this section, rather the focus is on what the violation of $\rho(\boldsymbol{M}) < 1$ implies for PSO particle movement.

If $\rho(\boldsymbol{M}) < 1$, then PSO particles are order-2 stable, and the condition $\rho(\boldsymbol{M}) < 1$ corresponds directly to the following criteria for order-2 stability:

$$c_1 + c_2 > 0, \qquad |w| < 1, \qquad c_1 + c_2 < \frac{24(1-w^2)}{7-5w}, \tag{8.2}$$

as explained in chapter 3.

The rate at which the variance, $V[x_t]$, of particle positions increases if $\rho(M) > 1$, is now illustrated. For the sake of simplicity, consider M as symmetric with $\rho(M) > 1$ (what follows still holds if M is not symmetric, as is proven in [1]; however, the technical detail would detract from the discussion, as the proof is quite intricate). Now, unwinding of

equation (8.1) leads to

$$\boldsymbol{z}_t = \boldsymbol{M}^{t-1}\boldsymbol{z}_1 + \sum_{j=0}^{t-2} \boldsymbol{M}^j \boldsymbol{b} \tag{8.3}$$

where $\boldsymbol{M}^j$ indicates the jth matrix product, with $\boldsymbol{M}^0 = \boldsymbol{I}$ the identity matrix. Since $\boldsymbol{M}$ is symmetric there exist scalar values η_1 to η_q such that an initial term $\boldsymbol{z}_1$ can be represented as a weighted sum of $\boldsymbol{M}$'s eigenvectors, specifically, $\boldsymbol{z}_1 = \sum_{i=0}^{q} \eta_i \boldsymbol{e}_i$, where $\boldsymbol{e}_i$ are the eigenvectors of M, which have the corresponding eigenvalues λ_i. Consider the first term of equation (8.3):

$$\boldsymbol{M}^{t-1}\boldsymbol{z}_1 = \boldsymbol{M}^{t-1} \sum_{i=0}^{q} \eta_i \boldsymbol{e}_i = \sum_{i=0}^{q} \eta_i \lambda_i^{t-1} \boldsymbol{e}_i \tag{8.4}$$

because $\boldsymbol{M}\boldsymbol{e}_i = \lambda_i$. Since $\rho(\boldsymbol{M}) > 1$, at least one eigenvalue, $\lambda_{\hat{i}}$, of $\boldsymbol{M}$ is greater than one, so the summation in equation (8.4) will diverge (assuming the corresponding $\eta_i \neq 0$). Not only will the term diverge, but it will do so exponentially since $\lambda_{\hat{i}}^t$ diverges exponentially. This implies that the variance of the particle positions will increase exponentially during the PSO run. The exponential divergence is very important to note as small increases in $\rho(\boldsymbol{M})$ could drastically increase the long term particle trajectory.

The immediate questions that arise from the preceding discussion are

- What is the relationship between PSO particle stability and PSO's real world performance?
- How tolerant is PSO to having control parameters selected that result in a spectral radius slightly larger than one?

These questions are answered in the subsequent sections.

8.2 Empirical Setup

This section summarizes the experimental procedure used for this chapter. The experiment aims to illustrate that most unstable parameter configurations actually result in

such poor performance of PSO that a trivial random search can outperform them. Parameter configurations are seen as unstable if they do not satisfy the criteria for order-2 stability, as defined in equation (8.2).

The performance of the PSO was measured for each parameter configuration across the following region:

$$w \in [-1.1, 1.1] \text{ and } c_1 + c_2 \in (0, 5.5] \tag{8.5}$$

where step sizes of 0.1 were used for w and $c_1 + c_2$. This results in 1264 parameter configurations, of which 761 are unstable and 504 are stable according to equation (8.2). A star neighborhood structure was used. Velocities were initialized to $\mathbf{0}$. A population size of 20 was used. The results for each configuration were derived from 35 independent runs.

The performance of the original PSO, with the inclusion of inertia, is compared to that of a random search. If a specific configuration of PSO does worse than a random search, the PSO is not effectively optimizing. The random search algorithm used is defined in algorithm 3. For the sake of comparison, each iteration of the PSO algorithm is seen as comparable to 20 random samples of the search space, one for each of the particles in the PSO swarm.

Algorithm 3 Random Search

Let f be the objective function to be minimized, and $\boldsymbol{S}$ the search space
Let s be the maximum number of samples, and s_{count} is initialized to 0
Let $\boldsymbol{p}_{best}$ represent the position with the current lowest objective function value
while $s_{count} < s$ **do**
  Sample $\boldsymbol{c}$ from a uniform distribution over S.
  if $f(\boldsymbol{c}) < f(\boldsymbol{p}_{best})$ **then**
    $\boldsymbol{p}_{best} = \boldsymbol{c}$
  end if
end while

The objective functions used in this chapter are listed in table 8.1. Full definitions of the objective functions are given in appendix B. Each objective function is tested in 5, 10, and 30 dimensions. The performance is measured at 500, 1000, 2000, and 5000 iterations.

TABLE 8.1: Objective Functions

	Function name		Function Name
F1	Absolute Value	F15	Norwegian
F2	Ackley	F16	Quadric
F3	Alpine	F17	Quartic
F4	Bent Cigar	F18	Rastrigin
F5	Discus	F19	Rosenbrock
F6	Egg Holder	F20	Salomon
F7	Elliptic	F21	Schaffer 6
F8	Expanded F9 and F19	F22	Schwefel 2.21
F9	Griewank	F23	Schwefel 2.22
F10	HappyCat	F24	Shubert
F11	HGBat	F25	Spherical
F12	Hyper Ellipsoid	F26	Step
F13	Katsuura	F27	Vincent
F14	Michalewicz	F28	Weierstrass

8.3 Empirical Results and Discussion

This section presents the results of the experiments described in section 8.2.

For each parameter configuration, Mann-Whitney U tests, using a confidence level of 95%, were performed to determine whether a given parameter configuration resulted in a PSO that was in fact better or worse than a random search, or if the PSO showed no statistical difference with the performance of the random search. This information is summarized into four performance indicators for each objective function, and each considered maximum number of iterations:

- S_BR: The percentage of theoretically stable parameter configurations that resulted in the PSO performing better than random search.
- S_NDR: The percentage of theoretically stable parameter configurations that resulted in the PSO performing with no statistically significant difference to random search.
- US_BR: The percentage of theoretically unstable parameter configurations that resulted in the PSO performing better than random search.

- US_NDR: The percentage of theoretically unstable parameter configurations that resulted in the PSO performing with no statistically significant difference to random search.

For each dimensionality tested, a table summarizing the performance information is given. The performance results for 500, 1000, 2000, and 5000 iterations are reported in each table.

The first thing to note is that, over all the results shown in tables 8.2, 8.3, and 8.4 for 5, 10, and 30 dimensions respectively, the percentage of unstable parameter choices that were able to outperform random search was significantly low. The highest percentage of unstable parameter configurations able to outperform random search was 34.67% for Griewank in 5 dimensions, at 500 iterations, though the performance of unstable parameter configurations on Griewank decreases with an increase in iteration count, down to 27.2% at 5000 iterations. The worsening of performance of PSO, when using unstable parameter configurations, is expected given the known exponential growth in the variance of particle positions, as discussed in section 8.1. The exact parameter configurations that failed to beat random search are illustrated in figure 8.1a. What is interesting is that all the unstable parameter configurations that did in fact outperform random search are in a region which is an extension of the stability boundary of equation (3.13), with an extended apex. The extension corresponds with a spectral radius of M that is slightly larger than 1. Similar results can be seen for both Solomon and Rosenbrock in 5-dimensions at 5000 iterations as illustrated figures 8.1b and 8.1c respectively.

While it appears in figures 8.1a, 8.1b, and 8.1c as if there is a fair degree of tolerance on the stability boundary of equation (3.13), a quick scan of tables 8.2, 8.3, and 8.4 shows numerous cases where the performance of unstable parameter configurations are terrible. For example, only 6.18% of the unstable parameter configurations were able to outperform random search on Michalewicz in 30 dimensions at 1000 iterations, whereas 100% of the stable parameter configurations outperformed random search. Figure 8.1d shows that a near perfect relationship exists between stable parameter configurations and the ability of the PSO to outperform random search when optimizing Michalewicz.

There are actually two objective functions that illustrate how finely tuned PSO sometimes needs to be, namely Egg Holder and Elliptic (while Egg Holder is known to have better solutions outside of the search space, this does not impact the results as the PSO was not allowed to update particles' personal or neighborhood positions best if they were outside of the allowable search space). Over all tested dimensions even stable parameter configurations were more often than not outperformed by random search. This requirement of fine tuning is evident in figure 8.1e for the Elliptic objective function in 5 dimensions at 5000. What is worth noting in figure 8.1e is that there are unstable parameter configurations that outperform stable parameters. However, all unstable parameter configurations that were successful are still near the apex of the stable region of equation (3.13). The performance of PSO was very poor for Egg Holder, with only less than 12% of the parameter configurations (stable and unstable) outperforming random search in 5 dimensions at 500 iterations, and decreasing to less than 11% of the parameter configurations. Figure 8.1f shows that only a very small number of parameter configurations are effective at optimizing the Egg Holder objective function. What is interesting is that once again the optimal parameter configurations are clustered around the boundary of the stable region. However, in the case of the Egg Holder objective function, most parameter configurations are just slightly outside the stable region.

There is a very clear trend throughout all the data, namely that the performance of stable parameter configurations improve as the dimensionality of the objective function increases. This behavior is observed for all tested objective functions except for the Happy Cat objective function. What is interesting is that the performance across nearly all the objective functions worsens, with an increase in dimensionality for unstable parameter configurations. This implies that the higher the dimensionality, the more important it becomes to select stable parameter configurations.

TABLE 8.2: Performance of PSO versus Random Search: 5-Dimensions

	500 Iterations				1000 Iterations				2000 Iterations				5000 Iterations			
F	S_BR	S_NDR	US_BR	US_NDR	SP_BR	SP_NDR	US_BR	US_NDR	S_BR	S_NDR	US_BR	US_NDR	S_BR	S_NDR	US_BR	US_NDR
F1	100.00	0.00	32.72	9.46	100.00	0.00	30.62	8.67	100.00	0.00	29.70	4.99	99.21	0.79	27.07	4.86
F2	100.00	0.00	32.59	9.99	100.00	0.00	30.09	8.67	99.80	0.20	28.12	6.70	97.62	2.38	26.68	5.52
F3	98.21	1.79	21.81	4.73	86.31	13.49	20.24	3.42	79.17	20.24	19.97	2.89	72.42	24.40	19.97	2.50
F4	100.00	0.00	32.19	9.72	99.80	0.20	29.96	6.83	98.21	1.79	26.94	5.52	95.44	4.56	25.10	5.52
F5	41.87	50.20	20.50	4.34	35.32	31.15	20.24	2.89	30.75	15.28	19.58	2.23	28.97	9.52	19.84	1.58
F6	3.77	8.93	7.88	6.57	1.39	6.94	7.62	6.96	0.60	5.16	9.20	5.26	0.20	2.98	10.38	4.34
F7	30.75	5.95	17.87	2.76	30.56	5.36	18.40	1.45	29.37	6.15	18.27	1.18	28.97	4.56	18.40	0.92
F8	98.81	1.19	31.14	8.67	95.63	4.17	27.46	6.83	92.66	6.94	26.68	4.73	87.10	10.52	24.57	3.94
F9	100.00	0.00	34.69	9.72	100.00	0.00	30.62	6.96	99.40	0.60	29.30	5.78	97.22	2.78	27.20	4.86
F10	86.90	13.10	19.19	3.81	85.32	14.29	19.45	3.15	77.18	22.02	19.05	2.50	68.85	29.56	19.05	1.58
F11	14.48	41.87	13.67	6.57	9.52	20.24	14.19	4.20	8.53	9.52	14.32	3.29	6.94	5.95	14.98	2.50
F12	100.00	0.00	32.46	7.75	99.21	0.79	30.62	6.70	98.61	1.39	28.78	5.65	94.05	5.75	26.54	5.39
F13	100.00	0.00	28.78	6.70	100.00	0.00	26.02	5.78	99.40	0.60	24.84	4.99	92.06	7.94	24.05	3.29
F14	82.14	17.86	22.08	4.99	69.84	26.39	21.29	3.15	65.28	27.78	21.29	2.37	55.95	24.01	20.50	2.37
F15	29.56	29.96	15.24	5.39	29.96	27.58	16.82	2.89	27.58	24.60	17.74	1.31	25.60	21.63	17.87	1.58
F16	99.21	0.79	28.91	6.04	96.43	3.57	26.81	4.99	90.48	9.52	25.76	4.99	76.39	23.02	24.84	3.68
F17	98.41	1.59	29.57	8.54	95.63	3.97	27.99	5.12	94.64	4.76	27.07	3.81	88.49	9.52	24.97	3.81
F18	98.21	1.79	22.08	5.52	95.44	4.56	21.02	4.47	92.46	5.95	20.11	3.42	82.54	13.89	19.58	1.84
F19	99.40	0.60	28.25	8.94	97.42	2.58	26.81	5.65	94.44	5.56	25.49	4.73	84.92	14.09	23.39	3.94
F20	100.00	0.00	31.67	10.64	99.80	0.20	29.04	6.96	97.62	2.38	26.54	5.91	94.05	5.95	24.70	4.07
F21	100.00	0.00	19.32	3.42	99.80	0.20	18.92	2.50	97.02	2.98	18.92	1.97	85.91	14.09	18.79	1.84
F22	96.63	3.37	29.96	8.80	92.46	6.75	26.68	6.18	90.08	8.33	25.23	5.26	86.90	9.92	24.44	3.94
F23	100.00	0.00	31.80	8.67	100.00	0.00	29.30	6.96	99.21	0.79	27.60	4.60	96.03	3.97	25.49	4.20
F24	61.31	24.60	16.69	2.37	53.57	16.27	17.21	1.31	51.19	14.48	17.48	1.31	43.06	12.30	17.48	0.92
F25	100.00	0.00	32.72	9.59	99.21	0.79	30.35	7.49	97.62	2.38	28.91	5.91	95.24	4.37	26.94	4.60
F26	100.00	0.00	33.77	8.54	98.81	1.19	31.01	7.36	97.22	2.78	28.91	6.18	94.05	5.95	26.68	3.94
F27	68.65	16.07	19.45	2.89	66.87	11.71	19.19	2.50	65.48	9.13	19.19	1.84	63.29	5.75	18.66	1.58
F28	100.00	0.00	29.30	9.59	100.00	0.00	28.52	5.91	100.00	0.00	27.60	4.60	100.00	0.00	25.49	4.07

TABLE 8.3: Performance of PSO versus Random Search: 10-Dimensions

	500 Iterations				1000 Iterations				2000 Iterations				5000 Iterations			
F	S_BR	S_NDR	US_BR	US_NDR	SP_BR	SP_NDR	US_BR	US_NDR	S_BR	S_NDR	US_BR	US_NDR	S_BR	S_NDR	US_BR	US_NDR
F1	100.00	0.00	33.25	9.33	100.00	0.00	30.22	8.28	100.00	0.00	27.33	6.04	100.00	0.00	25.10	4.86
F2	100.00	0.00	29.43	7.36	100.00	0.00	27.73	5.65	100.00	0.00	26.15	4.99	100.00	0.00	24.70	4.73
F3	100.00	0.00	19.45	6.70	100.00	0.00	18.40	4.47	100.00	0.00	17.35	3.94	100.00	0.00	16.95	2.50
F4	100.00	0.00	27.99	7.62	100.00	0.00	25.62	6.04	100.00	0.00	24.57	5.39	98.81	1.19	23.52	3.55
F5	85.91	14.09	16.95	4.20	59.52	37.70	16.03	2.76	44.05	34.13	15.90	1.97	36.11	20.04	15.77	1.71
F6	23.61	21.23	9.99	1.58	13.89	17.06	10.38	1.71	11.51	13.69	11.70	1.31	7.74	7.34	11.83	1.84
F7	29.96	8.13	12.88	1.71	28.37	6.75	13.53	1.05	27.18	5.75	13.80	1.31	25.20	6.15	14.72	0.66
F8	100.00	0.00	27.60	7.49	99.80	0.20	25.62	6.44	99.60	0.40	25.10	6.04	96.83	2.98	22.60	4.20
F9	100.00	0.00	32.19	8.02	100.00	0.00	29.43	7.49	100.00	0.00	28.91	5.26	100.00	0.00	26.54	5.39
F10	72.62	20.04	14.32	3.55	53.97	34.92	13.80	2.76	44.05	38.49	13.80	3.15	36.71	36.31	14.72	2.23
F11	100.00	0.00	25.89	8.94	99.60	0.40	20.89	7.62	98.61	1.39	18.92	6.04	48.21	51.59	15.77	3.81
F12	100.00	0.00	29.57	8.02	100.00	0.00	27.46	5.91	100.00	0.00	25.76	5.26	100.00	0.00	24.84	4.20
F13	100.00	0.00	22.08	5.91	100.00	0.00	20.76	5.12	100.00	0.00	19.19	4.34	100.00	0.00	19.58	3.29
F14	98.21	1.79	15.77	3.68	96.63	3.37	16.43	2.50	90.08	9.52	15.77	2.10	79.17	18.65	15.77	1.58
F15	68.06	9.52	14.98	3.42	67.66	5.75	14.98	2.76	67.06	3.57	15.37	1.84	66.87	2.98	15.77	1.18
F16	100.00	0.00	25.76	6.04	100.00	0.00	24.70	5.26	100.00	0.00	22.86	5.26	98.61	1.39	21.81	2.89
F17	100.00	0.00	25.89	6.96	99.80	0.20	24.31	5.65	98.81	1.19	23.13	4.73	94.44	5.16	21.55	2.76
F18	100.00	0.00	18.79	6.70	100.00	0.00	18.40	4.86	100.00	0.00	17.48	3.81	98.61	1.39	17.08	1.84
F19	100.00	0.00	26.81	7.62	100.00	0.00	23.78	6.31	100.00	0.00	23.26	5.78	99.21	0.79	22.34	3.42
F20	100.00	0.00	29.17	10.38	100.00	0.00	27.07	6.44	100.00	0.00	26.54	5.39	100.00	0.00	23.26	4.86
F21	100.00	0.00	13.27	3.02	100.00	0.00	13.80	1.97	100.00	0.00	14.19	1.71	100.00	0.00	13.80	2.76
F22	99.60	0.40	24.18	6.31	98.61	1.39	22.86	4.73	96.03	3.97	21.81	4.20	94.05	5.75	20.37	5.26
F23	100.00	0.00	24.84	6.83	100.00	0.00	22.86	5.78	100.00	0.00	22.21	4.20	100.00	0.00	21.81	3.42
F24	69.25	23.61	11.83	2.10	66.67	21.83	12.75	1.31	62.30	14.29	13.01	1.58	53.57	13.10	13.80	1.45
F25	100.00	0.00	30.75	8.94	100.00	0.00	29.04	6.96	100.00	0.00	26.15	5.91	100.00	0.00	24.57	5.39
F26	100.00	0.00	31.27	7.36	100.00	0.00	30.22	5.52	100.00	0.00	27.07	5.78	100.00	0.00	24.44	4.99
F27	70.24	11.31	14.06	2.76	68.65	11.11	14.59	2.63	66.47	10.12	15.51	1.45	63.89	7.34	15.51	1.18
F28	100.00	0.00	28.91	8.80	100.00	0.00	26.81	6.57	100.00	0.00	24.57	6.44	100.00	0.00	22.60	4.73

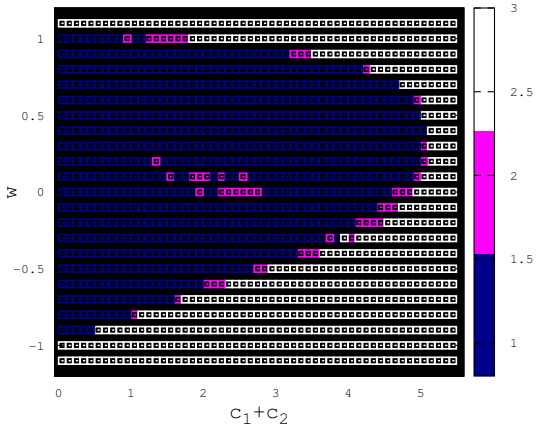
TABLE 8.4: Performance of PSO versus Random Search: 30-Dimensions

	500 Iterations				1000 Iterations				2000 Iterations				5000 Iterations			
F	S_BR	S_NDR	US_BR	US_NDR	SP_BR	SP_NDR	US_BR	US_NDR	S_BR	S_NDR	US_BR	US_NDR	S_BR	S_NDR	US_BR	US_NDR
F1	100.00	0.00	30.75	4.73	100.00	0.00	28.38	6.44	100.00	0.00	26.54	7.49	100.00	0.00	24.31	6.31
F2	100.00	0.00	28.91	5.65	100.00	0.00	27.60	5.78	100.00	0.00	26.94	6.18	100.00	0.00	24.97	5.12
F3	100.00	0.00	15.64	6.18	100.00	0.00	14.98	4.73	100.00	0.00	14.45	3.68	100.00	0.00	13.80	3.81
F4	100.00	0.00	32.59	4.20	100.00	0.00	30.49	4.86	100.00	0.00	26.94	6.57	100.00	0.00	24.70	6.31
F5	100.00	0.00	7.10	3.68	99.01	0.99	7.49	3.02	94.44	5.56	7.75	2.76	80.95	18.85	8.41	2.37
F6	52.18	7.34	2.76	0.53	50.79	7.14	3.55	0.92	48.21	6.55	4.60	1.05	41.67	9.33	6.04	0.79
F7	32.54	6.55	3.42	0.92	30.36	6.75	3.81	1.58	29.56	5.16	4.73	1.18	26.79	4.76	5.78	0.79
F8	100.00	0.00	29.43	5.78	100.00	0.00	27.46	6.04	100.00	0.00	25.36	6.04	100.00	0.00	23.13	5.65
F9	100.00	0.00	30.88	5.91	100.00	0.00	31.01	4.86	100.00	0.00	30.22	5.26	100.00	0.00	27.86	5.91
F10	67.26	10.32	4.34	2.37	66.07	10.71	5.26	2.76	48.41	25.79	3.68	4.60	24.01	46.23	2.37	6.18
F11	100.00	0.00	32.33	4.34	100.00	0.00	30.09	5.26	100.00	0.00	28.12	6.31	100.00	0.00	26.81	6.04
F12	100.00	0.00	28.25	6.57	100.00	0.00	27.99	5.78	100.00	0.00	25.36	7.10	100.00	0.00	23.26	5.26
F13	100.00	0.00	16.16	6.96	100.00	0.00	15.90	5.39	100.00	0.00	14.32	4.73	100.00	0.00	13.67	3.81
F14	100.00	0.00	5.91	2.76	100.00	0.00	6.18	2.23	100.00	0.00	7.36	1.71	100.00	0.00	7.36	1.58
F15	99.80	0.20	9.20	3.55	100.00	0.00	9.46	3.02	100.00	0.00	8.80	3.55	99.80	0.20	9.33	2.50
F16	100.00	0.00	23.39	7.62	100.00	0.00	21.42	6.83	100.00	0.00	20.63	6.96	100.00	0.00	19.05	4.73
F17	100.00	0.00	22.86	6.70	100.00	0.00	21.68	5.91	100.00	0.00	20.63	5.52	100.00	0.00	19.71	4.47
F18	100.00	0.00	18.27	6.44	100.00	0.00	17.21	5.26	100.00	0.00	16.29	5.52	100.00	0.00	15.64	4.34
F19	100.00	0.00	27.86	5.78	100.00	0.00	25.49	6.70	100.00	0.00	24.05	6.96	100.00	0.00	22.73	6.44
F20	100.00	0.00	31.27	4.99	100.00	0.00	29.96	5.12	100.00	0.00	28.12	5.91	100.00	0.00	25.62	6.31
F21	99.01	0.79	4.20	2.50	99.80	0.20	4.73	2.50	100.00	0.00	5.12	2.63	100.00	0.00	6.18	1.58
F22	100.00	0.00	22.86	5.65	100.00	0.00	20.63	6.70	100.00	0.00	19.84	6.44	100.00	0.00	18.13	4.47
F23	100.00	0.00	17.61	6.44	100.00	0.00	16.82	5.39	100.00	0.00	16.69	4.07	100.00	0.00	15.77	2.89
F24	75.20	19.64	3.42	0.66	73.21	18.45	3.94	1.05	71.23	17.86	4.99	0.92	65.08	15.48	6.57	0.53
F25	100.00	0.00	30.75	6.31	100.00	0.00	30.22	5.78	100.00	0.00	27.46	5.65	100.00	0.00	27.33	5.78
F26	100.00	0.00	31.41	5.65	100.00	0.00	29.96	5.78	100.00	0.00	28.91	4.73	100.00	0.00	27.20	5.65
F27	78.97	17.26	5.65	1.97	75.99	16.07	6.31	1.97	73.41	14.48	6.96	1.45	68.65	11.11	7.88	1.31
F28	100.00	0.00	26.28	7.23	100.00	0.00	23.26	7.23	100.00	0.00	21.81	5.52	100.00	0.00	20.76	5.52

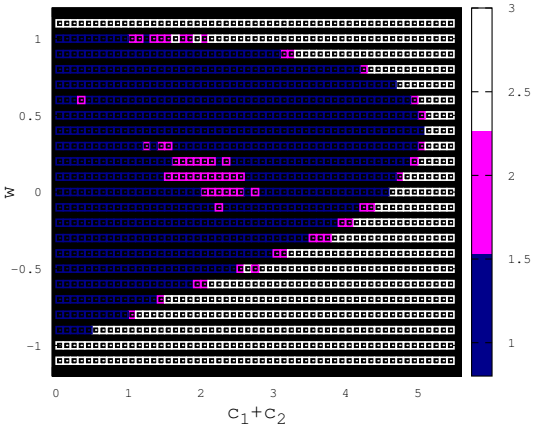
8.4 Summary

This chapter showed that the majority of parameter configurations that are theoretically unstable perform worse than random search on all objective functions tested. It was also shown that there is a degree of tolerance from which parameters can be selected just outside of the convergence region without extreme performance degradation. However, the degree to which parameter values can be selected outside of the stable region is very problem dependent. For most tested cases it was found that convergent parameter configurations drastically increased the chance of outperforming random search. Selecting parameter values within the stable region near the apex was the best strategy to ensure that PSO was always superior to random search. It was also observed that the higher the dimensionality of the problem, the more important it is to select stable parameter configurations.

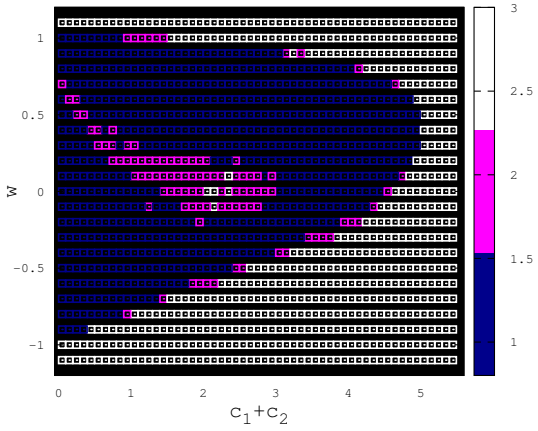
The next chapter presents an extension to the state of the art theoretical model utilized for understanding the stability of particle swarm optimization particles.



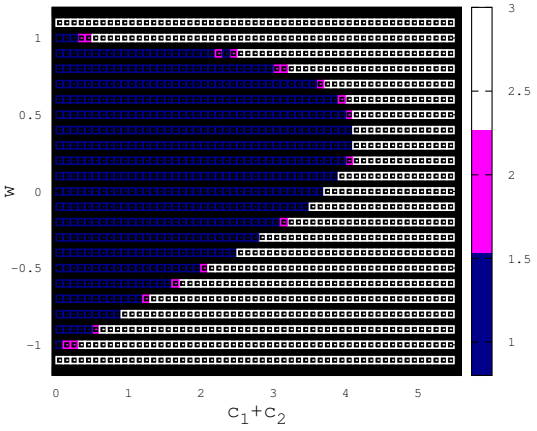
(a) Griewank, 5-dimensions, 5000 iterations.



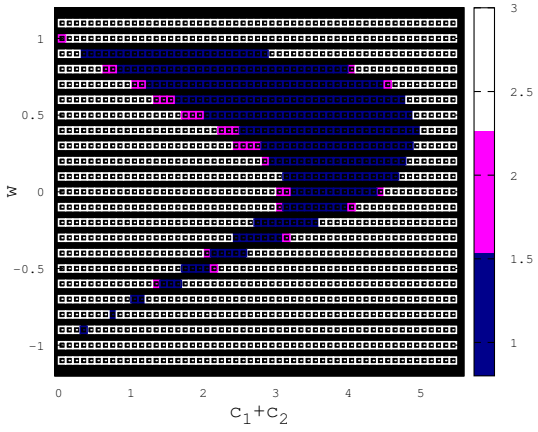
(b) Salomon, 5-dimensions, 5000 iterations.



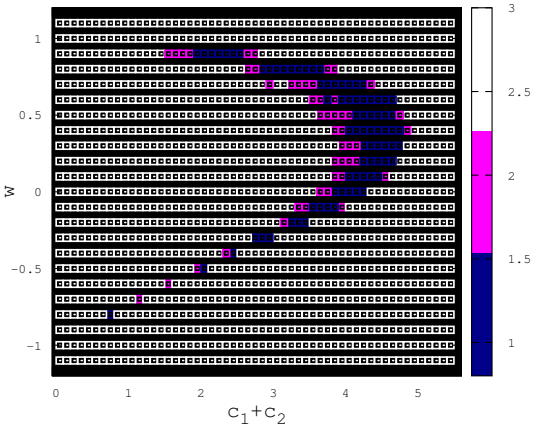
(c) Rosenbrock, 5-dimensions, 5000 iterations.



(d) Michalewicz, 30-dimensions, 1000 iterations.



(e) Elliptic, 5-dimensions, 5000 iterations.



(f) Egg Holder, 5-dimensions, 5000 iterations.

FIGURE 8.1: Performance of PSO versus random search
1 = performed better than random search, 2 = no statistical difference,
3 = random search performed better

Chapter 9

Extension of the Theoretical Particle Swarm Model

This chapter presents a meaningful extension of the state of the art theoretical model utilized for understanding the stability of particle swarm optimization's particles. 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, it is proven that the assumption that the expected value and variance of a particle's personal and neighborhood best positions are convergent sequences is in fact a necessary condition for order-1 and order-2 stability. The theoretical analysis presented in this chapter is applicable to a large class of particle swarm optimization variants.

The theoretical derivations of criteria for order-1 and order-2 stability are presented in section 9.1. Section 9.2 presents an application of the theoretical stability analysis. A summary of the chapter's findings are presented in section 9.3. Content from this chapter has been submitted for review to the Swarm Intelligence journal.

9.1 Stability Proof

This section presents a derivation of the necessary and sufficient conditions for order-1 and order-2 stability of PSO under the non-stagnant distribution assumption. The section starts with two small lemmas which are used to improve the flow of the more substantial lemma 9.3. Lemma 9.3 is an extension of a classic theorem from mathematical analysis. The main result is then presented in theorem 9.4.

Lemma 9.1. *Let* $(\boldsymbol{T}_n)$ *be a sequence of bounded linear operators [35] from* $\mathbb{R}^p$ *to* $\mathbb{R}^p$*, all with equal spectral radius, where the spectral radius is defined using definition C.11. Furthermore let* $(\boldsymbol{T}_n)$ *be uniformly operator convergent. If* $\rho(\boldsymbol{T}_1) < 1$*, then* $\lim_{n\to\infty} \boldsymbol{T}_n\boldsymbol{T}_{n-1}\cdots\boldsymbol{T}_1 = \Theta$*, 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 $\rho(\boldsymbol{T}_1) \leq ||\boldsymbol{T}_1||_\delta \leq \rho(\boldsymbol{T}_1) + \delta$ [35], 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 $n > N$,

$$\|\boldsymbol{T}_n\boldsymbol{T}_{n-1}\cdots\boldsymbol{T}_N\boldsymbol{T}_{N-1}\cdots\boldsymbol{T}_1\|_\delta \leq \|\boldsymbol{T}_n\boldsymbol{T}_{n-1}\cdots\boldsymbol{T}_{N+1}\|_\delta\|\boldsymbol{T}_N\boldsymbol{T}_{N-1}\cdots\boldsymbol{T}_1\|_\delta \tag{9.1}$$

$$\leq \sigma^{n-N}\xi^N \tag{9.2}$$

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.

Lemma 9.2. *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, where uniform operator convergence is defined using definition C.13. 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 $(T_n \cdots T_{n-j})$ 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 a 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

$$
\begin{aligned}
&\|\boldsymbol{T}_m \cdots \boldsymbol{T}_{m-j-1} - \boldsymbol{T}_n \cdots \boldsymbol{T}_{n-j-1}\| \\
&\leq \|\boldsymbol{T}_m \cdots \boldsymbol{T}_{m-j-1} - \boldsymbol{T}_m \cdots \boldsymbol{T}_{m-j}\boldsymbol{T}_{n-j-1}\| + \|\boldsymbol{T}_m \cdots \boldsymbol{T}_{m-j}\boldsymbol{T}_{n-j-1} - \boldsymbol{T}_n \cdots \boldsymbol{T}_{n-j-1}\| \\
&\leq \|\boldsymbol{T}_m \cdots \boldsymbol{T}_{m-j}\|\|\boldsymbol{T}_{m-j-1} - \boldsymbol{T}_{n-j-1}\| + \|\boldsymbol{T}_m \cdots \boldsymbol{T}_{m-j} - \boldsymbol{T}_n \cdots \boldsymbol{T}_{n-j}\|\|\boldsymbol{T}_{n-j-1}\| \\
&\leq \eta_1\|\boldsymbol{T}_{m-j-1} - \boldsymbol{T}_{n-j-1}\| + \eta_2\|\boldsymbol{T}_m \cdots \boldsymbol{T}_{m-j} - \boldsymbol{T}_n \cdots \boldsymbol{T}_{n-j}\| \\
&< \epsilon
\end{aligned} \tag{9.3}
$$

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. □

Lemma 9.3. *Let* $(\boldsymbol{x}_n)$ *be a sequence in* $\mathbb{R}^p$*, defined as*

$$\boldsymbol{x}_n = \boldsymbol{T}_n\boldsymbol{x}_{n-1} + \boldsymbol{b}_{n-1} \tag{9.4}$$

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* $\rho(\boldsymbol{T}_n) < 1$ *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 $(\boldsymbol{b}_n)$ 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 9.1 it is known that there exits 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, which is now done. Let $m, n \in \mathbb{N}$ and $m > n$.

Unwinding $\boldsymbol{x}_n$ leads to

$$\begin{aligned}\boldsymbol{x}_n &= \boldsymbol{T}_n\boldsymbol{x}_{n-1} + \boldsymbol{b}_{n-1} \\ &= \boldsymbol{T}_n(\boldsymbol{T}_{n-1}\boldsymbol{x}_{n-2} + \boldsymbol{b}_{n-2}) + \boldsymbol{b}_{n-1} \\ &= \cdots \\ &= \boldsymbol{C}_{(1,n-1)}\boldsymbol{x}_1 + \sum_{i=0}^{n-2} \boldsymbol{C}_{(n+1-i,n)}\boldsymbol{b}_{n-1-i}\end{aligned} \tag{9.5}$$

Now, using equation (9.5),

$$\begin{aligned}&||\boldsymbol{x}_m - \boldsymbol{x}_n|| \\ &= ||\boldsymbol{C}_{(1,m-1)}\boldsymbol{x}_1 + \sum_{i=0}^{m-2} \boldsymbol{C}_{(m+1-i,m)}\boldsymbol{b}_{m-1-i} - \boldsymbol{C}_{(1,n-1)}\boldsymbol{x}_1 - \sum_{i=0}^{n-2} \boldsymbol{C}_{(n+1-i,n)}\boldsymbol{b}_{n-1-i}|| \\ &\leq || \left(\boldsymbol{C}_{(1,m-1)} - \boldsymbol{C}_{(1,n-1)}\right) \boldsymbol{x}_1|| + || \sum_{i=0}^{m-2} \boldsymbol{C}_{(m+1-i,m)}\boldsymbol{b}_{m-1-i} - \sum_{i=0}^{n-2} \boldsymbol{C}_{(n+1-i,n)}\boldsymbol{b}_{n-1-i}||\end{aligned} \tag{9.6}$$

Considering the first term of the summation in equation (9.6), it is seen that

$$|| \left(\boldsymbol{C}_{(1,m-1)} - \boldsymbol{C}_{(1,n-1)}\right) \boldsymbol{x}_1|| \leq || \left(\boldsymbol{C}_{(1,m-1)} - \boldsymbol{C}_{(1,n-1)}\right) || \; ||\boldsymbol{x}_1|| \tag{9.7}$$

since $\rho(\boldsymbol{T}_n) < 1$ for all n, and using lemma 9.1, the sequence of operators $(\boldsymbol{C}_{(1,n)})$ is convergent, and therefore Cauchy. It follows that $|| \left(\boldsymbol{C}_{(1,m-1)} - \boldsymbol{C}_{(1,n-1)}\right) ||$ converges to zero, and trivially so does equation (9.7).

Focusing on the second term of the summation in equation (9.6),

$$\begin{aligned}&|| \sum_{i=0}^{m-2} \boldsymbol{C}_{(m+1-i,m)}\boldsymbol{b}_{m-1-i} - \sum_{i=0}^{n-2} \boldsymbol{C}_{(n+1-i,n)}\boldsymbol{b}_{n-1-i}|| \\ &\leq || \sum_{i=n-1}^{m-2} \boldsymbol{C}_{(m+1-i,m)}\boldsymbol{b}_{m-1-i}|| + || \sum_{i=0}^{n-2} \left(\boldsymbol{C}_{(m+1-i,m)}\boldsymbol{b}_{m-1-i} - \boldsymbol{C}_{(n+1-i,n)}\boldsymbol{b}_{n-1-i}\right) ||\end{aligned} \tag{9.8}$$

Note that, for the first term in equation (9.8), since $(\boldsymbol{b}_n)$ is convergent, there exits a ζ such that $||\boldsymbol{b}_n|| < \zeta$ for every n. It therefore follows that:

$$\begin{aligned} || \sum_{i=n-1}^{m-2} \boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{m-1-i} || &\leq \sum_{i=n-1}^{m-2} ||\boldsymbol{C}_{(m+1-i,m)}|| \; ||\boldsymbol{b}_{m-1-i}|| \\ \leq \sum_{i=n-1}^{m-2} ||\boldsymbol{C}_{(m+1-i,m)}|| \; \zeta &\leq \sum_{i=n-1}^{m-2} \prod_{j=m}^{m-1-i} ||\boldsymbol{T}_j|| \; \zeta \end{aligned} \tag{9.9}$$

It is also known that $\|\boldsymbol{T}_j\| \leq \sigma < 1$. It therefore follows from equation (9.9) that

$$\sum_{i=n-1}^{m-2} \prod_{j=m}^{m-1-i} ||\boldsymbol{T}_j|| \; \zeta \leq \sum_{i=n-1}^{m-2} \prod_{j=m}^{m-1-i} \sigma \; \zeta = \zeta \sum_{i=n-1}^{m-2} \sigma^{i+1} \tag{9.10}$$

Since $\sigma < 1$, the elementary geometric series formula can be used to transform equation (9.10) to

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

Focusing on the remaining term in equation (9.8),

$$\begin{aligned} &|| \sum_{i=0}^{n-2} \left(\boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{m-1-i} - \boldsymbol{C}_{(n+1-i,n)} \boldsymbol{b}_{n-1-i} \right) || \\ &\leq || \sum_{i=0}^{n-2} \left(\boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{m-1-i} - \boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{n-1-i} \right) \\ &+ || \sum_{i=0}^{n-2} \left(\boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{n-1-i} - \boldsymbol{C}_{(n+1-i,n)} \boldsymbol{b}_{n-1-i} \right) || \end{aligned} \tag{9.11}$$

Both terms of equation (9.11) requires a more subtle mathematical treatment, given the complexity of the internal terms. The standard epsilon-n approach from analysis will

be used. Consider the first term of equation (9.11),

$$\begin{aligned}
&||\sum_{i=0}^{n-2}\left(\boldsymbol{C}_{(m+1-i,m)}\boldsymbol{b}_{m-1-i}-\boldsymbol{C}_{(m+1-i,m)}\boldsymbol{b}_{n-1-i}\right)\\
&\leq\sum_{i=0}^{n-2}||\boldsymbol{C}_{(m+1-i,m)}||\;||\boldsymbol{b}_{m-1-i}-\boldsymbol{b}_{n-1-i}||\\
&\leq\sum_{i=0}^{n-2}\sigma^{i}\;||\boldsymbol{b}_{m-1-i}-\boldsymbol{b}_{n-1-i}||\\
&\leq\sum_{i=0}^{n-2}\sigma^{i}\;||\boldsymbol{b}_{m-1-i}-\boldsymbol{b}||+\sum_{i=0}^{n-2}\sigma^{i}\;||\boldsymbol{b}_{n-1-i}-\boldsymbol{b}||
\end{aligned}\tag{9.12}$$

Let $\boldsymbol{z}_i=\boldsymbol{b}_{i+1}-\boldsymbol{b}$. Since $\boldsymbol{b}_i\to\boldsymbol{b}$, $||\boldsymbol{z}_i||\to 0$, and 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 (9.12),

$$\begin{aligned}
\sum_{i=0}^{n-2}\sigma^{i}\;||\boldsymbol{b}_{n-1-i}-\boldsymbol{b}||&=\sum_{i=0}^{n-2}\sigma^{n-2-i}\;||\boldsymbol{b}_{i+1}-\boldsymbol{b}||\\
&\leq\tau\sum_{i=0}^{n_\epsilon-2}\sigma^{n-2-i}+\epsilon\sum_{i=n_\epsilon-1}^{n-2}\sigma^{n-2-i}\\
&\leq\frac{\sigma^{n-n_\epsilon}-\sigma^{n-1}}{1-\sigma}\tau+\frac{1-\sigma^{n-n_\epsilon}}{1-\sigma}\epsilon\\
&\leq\frac{\sigma^{n-n_\epsilon}}{1-\sigma}\tau+\frac{\epsilon}{1-\sigma}\\
&\leq\frac{\tau+1}{1-\sigma}\epsilon
\end{aligned}\tag{9.13}$$

Now, since $m>n$, the same argument can be made for the first term of equation (9.12), as was used for the second term. This implies that, for a large enough n and m, equation (9.12) can be made less than an arbitrarily small $\epsilon>0$.

The last remaining term requiring analysis is the second term of equation (9.11). Note that, since $\sigma^n\to 0$ for any $\epsilon>0$, there exits a n_ϵ such that $\sigma^n<\epsilon\,(1-\sigma)/(2\zeta)$ if

$n \geq n_\epsilon$. Equation (9.11) can then be handled as follows:

$$\begin{aligned}
&\| \sum_{i=0}^{n-2} \left(\boldsymbol{C}_{(m+1-i,m)} \boldsymbol{b}_{n-1-i} - \boldsymbol{C}_{(n+1-i,n)} \boldsymbol{b}_{n-1-i} \right) \| \\
&\leq \sum_{i=0}^{n-2} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| \| \boldsymbol{b}_{n-1-i} \| \\
&\leq \zeta \sum_{i=1}^{n_\epsilon - 1} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| + \zeta \sum_{i=n_\epsilon}^{n-2} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| \\
&\leq \zeta \sum_{i=1}^{n_\epsilon - 1} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| + 2\zeta \sum_{i=n_\epsilon}^{n-2} \sigma^i \\
&\leq \zeta \sum_{i=1}^{n_\epsilon - 1} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| + 2\zeta \frac{\sigma^{n_\epsilon}}{1-\sigma} \\
&\leq \zeta \sum_{i=1}^{n_\epsilon - 1} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| + \epsilon
\end{aligned} \tag{9.14}$$

Now that m and n have been decoupled from the summation limit of the first term in equation (9.14), the limit can be dealt with directly. It is known from lemma 9.2 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 exits 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 (9.14) becomes

$$\zeta \sum_{i=1}^{n_\epsilon - 1} \| \boldsymbol{C}_{(m+1-i,m)} - \boldsymbol{C}_{(n+1-i,n)} \| + \epsilon \leq \delta + \epsilon \tag{9.15}$$

Since ϵ and δ can be made arbitrarily small, this completes the proof. □

Now that lemma (9.3) 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

$$x_k(t+1) = x_k(t)\alpha + x_k(t-1)\beta + \gamma_t \tag{9.16}$$

are considered, where k indicates the vector component, α and β are well defined random variables, and (γ_t) is a sequence of random variables.

This class of PSOs includes PSO (with inertia), fully informed PSO [33], and unified PSO [45], though many others exist. Furthermore this class also allows arbitrary distributions

to be utilized for all internal components, provided they are dimension independent.

Theorem 9.4. *The following properties hold for all PSO variants of the form described in equation (9.16):*

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

$$A = \begin{bmatrix} E[\alpha] & E[\beta] \\ 1 & 0 \end{bmatrix} \text{ and } \boldsymbol{i}_t = \begin{bmatrix} E[\gamma_t] \\ 0 \end{bmatrix} \tag{9.17}$$

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

$$\boldsymbol{j}_t = \begin{bmatrix} E[\gamma_t] \\ 0 \\ E[\gamma_t^2] \\ 0 \\ 0 \end{bmatrix} \tag{9.18}$$

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:*

$$1 - E[\alpha] - E[\beta] \neq 0 \tag{9.19}$$

$$1 - E\left[\alpha^2\right] - E\left[\beta^2\right] - \left(\frac{2E[\alpha\beta]\,E[\alpha]}{1 - E[\beta]}\right) > 0 \tag{9.20}$$

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 (9.16) yields

$$E[x(t+1)] = E[x(t)]E[\alpha] + E[x(t-1)]E[\beta] + E[\gamma_t] \tag{9.21}$$

which is reformulated to

$$\boldsymbol{u}_t = \boldsymbol{A}\boldsymbol{u}_{t-1} + \boldsymbol{i}_t \tag{9.22}$$

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 9.3 shows that $(\boldsymbol{u}_t)$ converges if $\rho(\boldsymbol{A}) < 1$ and $(\boldsymbol{i}_t)$ is convergent, implying order-1 stability of particle positions. It is a well known theorem from analysis that if $\boldsymbol{i}_t$ is constant, then $\rho(\boldsymbol{A}) < 1$ is a necessary condition for convergence [35]. Since a constant $\boldsymbol{i}_t$ is a special case of equation (9.22), $\rho(\boldsymbol{A}) < 1$ 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 (9.16), defined as,

$$V[x(t+1)] = E[x^2(t+1)] + E[x(t+1)]^2 \tag{9.23}$$

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

$$x^2(t+1) = x^2(t)\alpha^2 + x^2(t-1)\beta^2 + \gamma_t^2 + 2x(t)\gamma_t\alpha + 2x(t)x(t-1)\alpha\beta + 2x(t-1)\beta\gamma_t \tag{9.24}$$

Application of the expectation operator produces

$$\begin{aligned} E[x^2(t+1)] = {} & E[x^2(t)]E[\alpha^2] + E[x^2(t-1)]E[\beta^2] + E[\gamma_t^2] \\ & + 2E[x(t)]E[\gamma_t\alpha] + 2E[x(t)x(t-1)]E[\alpha\beta] \\ & + 2E[x(t-1)]E[\beta\gamma_t] \end{aligned} \tag{9.25}$$

The expectation of $x(t)x(t-1)$ is obtained by multiplying equation (9.16) by $x(t)$ and applying the expectation operator to yield

$$E[x(t+1)x(t)] = E[\alpha]E[x^2(t)] + E[\beta]E[x(t)x(t-1)] + E[x(t)]E[\gamma_t] \tag{9.26}$$

Given equations (9.21), (9.25), and (9.26), 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

$$\boldsymbol{g}_t = \boldsymbol{B}_t\boldsymbol{g}_{t-1} + \boldsymbol{j}_t \tag{9.27}$$

where

$$\boldsymbol{B}_t = \begin{bmatrix} E[\alpha] & E[\beta] & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 2E[\gamma_t\alpha] & 2E[\gamma_t\beta] & E[\alpha^2] & E[\beta^2] & 2E[\alpha\beta] \\ 0 & 0 & 1 & 0 & 0 \\ E[\gamma_t] & 0 & E[\alpha] & 0 & E[\beta] \end{bmatrix} \tag{9.28}$$

(9.29)

$$\boldsymbol{g}_t = \begin{bmatrix} E[x(t)] \\ E[x(t-1)] \\ E[x^2(t)] \\ E[x^2(t-1)] \\ E[x(t)x(t-1)] \end{bmatrix}, \qquad \boldsymbol{j}_t = \begin{bmatrix} E[\gamma_t] \\ 0 \\ E[\gamma_t^2] \\ 0 \\ 0 \end{bmatrix} \tag{9.30}$$

Since the limits $(E[\gamma_t\alpha])$, $(E[\gamma_t\beta])$ and $(E[\gamma_t])$ exist by assumption, then so does the limit of $(\boldsymbol{B}_t)$. One of the conditions of lemma 9.3 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. The eigenvalues actually do not contain the terms $E[\gamma_t\alpha]$, $E[\gamma_t\beta]$, or $E[\gamma_t]$ at all (the exact eigenvalues are given in appendix D, as some are over 1000 characters). 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 9.3 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 [3]. 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 (9.19) and (9.20) are necessary conditions for order-2 stability. Note that a constant γ_t is simply a special case of equation (9.16), which implies that the necessary conditions hold equally for the class of PSOs under consideration in equation (9.16).

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 4 by contradiction. Assume that $(\boldsymbol{i}_t)$ diverges, but that $(\boldsymbol{u}_t)$ converges. Equation (9.22) can now be reformulated to

$$\boldsymbol{u}_t - \boldsymbol{A}\boldsymbol{u}_{t-1} = \boldsymbol{i}_t \tag{9.31}$$

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. □

At this point it should be made clear that the assumption that $(\boldsymbol{j}_t)$ and $(\boldsymbol{i}_t)$ are convergence sequences, is a weaker assumption than the original stagnation assumption, the weak chaotic assumption, the weak stagnation assumption, and the stagnant distribution assumption, as defined in chapter 3. Each of the existing assumptions that have been used in theoretical stability analysis of PSO are stronger special cases of the assumption that $(\boldsymbol{j}_t)$ and $(\boldsymbol{i}_t)$ converge. Therefore any stability analysis performed using theorem 9.4 is more reflective of the true PSO behavior than any previously done stability analysis.

9.2 Direct Application of Stability Theory

This section provides an illustrative example using theorem 9.4 to provide the reader with a simple procedure for using theorem 9.4 to derive stability criteria for new PSO variants that comply with the formulation given in equation (9.16).

A general version of PSO 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 [7]. It is shown that the same criteria, as in [7], can be derived for both order-1 and order-2 stability utilizing theorem 1 under the weakest allowable assumption on the personal and neighborhood best positions.

Rewriting the general version of PSO in the form of equation (9.16) is achieved by setting the following terms:

$$\begin{aligned}\alpha &= (1+w) - \theta_1 - \theta_2 \\ \beta &= -w \\ \gamma_t &= \theta_1 y_t + \theta_2 \hat{y}_t \end{aligned} \tag{9.32}$$

where y_t and $\hat{y}_t$, the personal and neighborhood best positions, are modeled as sequences of random variables which are convergent in expectation and variance. In order to use theorem 9.4 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 $(E[y_t])$ exist, that the limit of $(E[\gamma_t])$ 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

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

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

$$-1 < E[w] < 1 \qquad \text{and} \qquad 0 < \frac{E[\theta_1] + E[\theta_2]}{E[w] + 1} < 2 \tag{9.33}$$

Using the criteria for order-1 stability of equation (9.33), and property 4 of theorem 9.4, along with the assistance of Matlab's symbolic toolbox, and similar steps to that of [7],

the following necessary criteria for order-2 stability are obtained:

$$-1 < \frac{E[w]}{\sqrt{1 - V[w]}} < 1 \tag{9.34}$$

$$0 < E[\theta_1] + E[\theta_2] < \frac{-2(E[w]^2 + V[w] - 1)}{1 - E[w] + \frac{(V[\theta_1]+V[\theta_2])(1+E[w])}{(E[\theta_1]+E[\theta_2])^2}} \tag{9.35}$$

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 [7].

The next step is to verify that equations (9.34) and (9.35) are in fact sufficient conditions for order-2 stability, using properties 3 of theorem 9.4. The existence of the limits of $(E[\gamma_t\alpha])$ and $(E[\gamma_t\beta])$ must first be shown. Observe that

$$\begin{aligned} E[\gamma_t\alpha] = {} & E[\theta_1]E[y_t](1 + E[w]) + E[\theta_2]E[\hat{y}_t](1 + E[w]) - E[\theta_1^2]E[y_t] \\ & - 2E[\theta_1]E[\theta_2]E[y_t]E[\hat{y}_t] - E[\theta_2^2]E[\hat{y}_t] \end{aligned} \tag{9.36}$$

and

$$E[\gamma_t\beta] = -E[w]E[\theta_1]E[y_t] - E[w]E[\theta_2]E[\hat{y}_t] \tag{9.37}$$

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 PSO variant, the simplification of the condition $\rho(\boldsymbol{B}) < 1$ becomes intractable. However, the conditions in equations (9.34) and (9.35) can be empirically verified to be sufficient for convergence using an empirical approach similar to Bonyadi and Michalewicz [7]. 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 (9.34) and (9.35) were satisfied. It was then tested whether or not $\rho(\boldsymbol{B}) < 1$. It was found that in 100% of the cases, if equations (9.34) and (9.35) were satisfied, then the condition $\rho(\boldsymbol{B}_t) < 1$ held. This provides strong evidence that equations (9.34) and (9.35) are in fact also the sufficient conditions for order-2 stability.

It is also seen via the application of property 4 of theorem 9.4 that convergence of $(E[y_t])$, $(E[\hat{y}_t])$, $(V[y_t])$, and $(V[\hat{y}_t])$ are in fact necessary conditions for convergence. Therefore, the provided order-1 and order-2 stability regions obtained in this section are derived utilizing the weakest allowable assumption on the personal and neighborhood best positions.

It is informative to note that the criteria for order-1 and order-2 stability of the regular PSO algorithm can be directly obtained from equations (9.33), (9.34), and (9.35). Let w be a constant, and let $\theta_1 = c_1 r_1$, $\theta_2 = c_2 r_2$ as in the regular PSO 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 (9.33) leads to the following criteria for order-1 stability:

$$-1 < w < 1 \qquad \text{and} \qquad 0 < c_1 + c_2 < 4(w+1) \tag{9.38}$$

The criteria for order-1 stability in equation (9.38) agrees with the criteria for order-1 stability as derived by [50] under the more restrictive stagnation assumption. Furthermore, substituting the calculated expectations and variances into equations (9.34) and (9.35) leads to the following criteria for order-2 stability:

$$-1 < w < 1 \qquad \text{and} \qquad 0 < c_1 + c_2 < \frac{24(1-w^2)}{7-5w} \tag{9.39}$$

While this section focused on a general version of PSO, 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 (9.16).

9.3 Summary

This chapter 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 no weaker assumption is theoretically possible.

The next chapter presents a summary of the findings of this thesis, as well as a discussion on possible future research on PSO stability.

Chapter 10

Conclusion

This chapter begins with a summary of the findings and contributions of this thesis in section 10.1, followed by a discussion of potential future work in section 10.2.

10.1 Overview

This thesis proposed a method for performing assumption free empirical convergence analysis. The empirical method was successfully utilized in chapter 4 to verify which of the numerous theoretically derived criteria for PSO's particle stability were actually representative of PSO's true, unsimplified, stability behavior. It was found that the region for particle stability as derived by Poli and Broomhead [51] and Poli[50] was in fact an accurate representation of the stability criteria for PSO particles.

The proposed empirical convergence analysis method's effectiveness was tested in chapter 5. The convergence region obtained using the proposed objective function CF, as defined in equation (4.1), was compared against the intersection of the obtained convergence regions of 11 well known objective functions. It was found that just using the CF objective function was a sufficient way of obtaining the convergence region, as the found region corresponded to the intersection of the obtained convergence regions of the 11 well known objective functions. Chapter 5 showed that the social network structure used by PSO did not alter the convergence region. The empirical convergence analysis

approach was also effectively utilized to investigate the convergence region for BPSO, FIPS, SPSO 2011 in an assumption free context. One very interesting finding was that, in an assumption free context, BPSO's particles did not exhibit stable behavior, as was predicted by the work of Blackwell [30]. The discrepancy found between the theoretical and practical behavior of BPSO's particles provides a clear example of why empirical verification of the stability criteria in an assumption free context is necessary.

The stability criteria for both the FIPS and UPSO algorithms were theoretically derived under the stagnation assumption in chapters 6 and 7 respectively. Then the empirical convergence analysis approached presented in this thesis was utilized to verify the accuracy of the theoretical derivations in an assumption free context. It was found that the stability criteria that were derived for both FIPS and UPSO accurately predicted the stability criteria of the unsimplified and assumption free FIPS and UPSO.

The relation between stability criteria and PSO performance was investigated in chapter 8. It was found that the majority of parameter configurations that are theoretically unstable perform worse than random search on all objective functions tested. It was also shown that there is a degree of tolerance from which parameters can be selected just outside of the convergence region without extreme performance degradation. However, it was seen that selecting parameter values within the stable region near the apex was the best strategy to ensure that PSO was always superior to random search.

Lastly, an extension of the state of the art theoretical model utilized for understanding the stability of particle swarm optimization's particles was given in chapter 9. The necessary and sufficient conditions for order-1 and order-2 stability of particle positions were derived for a large class of PSO variants. The class of PSO variants the theory caters for includes PSO, UPSO, and FIPS to mention a few. Furthermore, the presented theoretical extension is still applicable even if the PSO variants' control parameters, such as w, c_1, and c_2, are replaced with random variables or when the existing random variables are sampled from a different distribution. The provided extension makes it easy to derive stability criteria for new PSO variants or even for minor alterations to existing PSO variants by utilizing theorem 9.4. The provided extension of PSO stability theory is all performed under what is proven to be the weakest allowable assumption on

particle informers, where the particle informers are, in the simplest case, the personal and neighborhood best particle positions.

In summation, this thesis effectively provides a complete framework for performing stability analysis on PSO variants, with the proposed empirical approach for convergence analysis allowing for assumption free verification of theoretically derived stability criteria, which can be derived in many cases by utilizing theorem 9.4. The empirical approach for convergence analysis can also be utilized for exploratory stability analysis when analytical analysis is infeasible.

10.2 Future Work

A number of directions for future work may flow from this thesis. The most obvious extension to the proposed empirical convergence analysis method is the inclusion of automated extraction of the convergence region. There exists numerous approaches that could be utilized for the extraction of the convergence region, ranging from simple curve fitting, to more sophisticated function approximation approaches such as those derived from trained neural networks.

From a theoretical perspective there are still two 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

$$x_k(t+1) = x_k(t)\alpha_t + x_k(t-1)\beta_t + \gamma_t \tag{10.1}$$

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 (10.1) includes numerous PSO variants where the inertia, cognitive or social coefficients are altered over time, as in many self-adaptive PSOs [42, 47, 55, 57, 64]. The second relatively unexplored area is to perform theoretical stability analysis on PSO variants where the particle position update equation do not operate on dimensions independently. A good example of a PSO variant

with this coupling between dimensions is SPO2011. The required mathematical techniques needed to perform this type of analysis in a tractable fashion are, unfortunately, not immediately apparent.

Appendix A

Symbols

This appendix lists the symbols used throughout this thesis, along with their corresponding meanings.

Ω	Particle swarm population.
i	Index of the particle in its swarm.
$\boldsymbol{x}_i$	Position vector of particle i.
$\boldsymbol{v}_i$	Velocity vector of particle i.
k	Component index of a particle.
N	Swarm size.
d	Particle dimension.
c_1	Cognitive weight coefficient.
c_2	Social weight coefficient.
w	Inertia weight coefficient.
$\boldsymbol{y}_i$	Best position found by particle i.
$\hat{\boldsymbol{y}}_i$	Best position found in particle i's neighbourhood.
$\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r'}_1, \boldsymbol{r'}_2$	Stochastic variables from the uniform distribution $U(0,1)^k$.
$\boldsymbol{\theta}_1$	Shorthand for $\boldsymbol{r}_1 c_1$.
$\boldsymbol{\theta}_2$	Shorthand for $\boldsymbol{r}_2 c_2$.
$\mathcal{N}_i$	Neighbourhood of particle i.
χ	Constriction coefficient.
u	Unification factor.

$\boldsymbol{g}_i$	The Gbest velocity vector in UPSO.
$\boldsymbol{l}_i$	The LBest velocity vector in UPSO.
$\boldsymbol{\zeta}_i$	The weighted average $\frac{c_1\boldsymbol{y}_i(t)+c_2\hat{\boldsymbol{y}}_i(t)}{c_1+c_2}$.
$\phi_{i,k}(t)$	The distance at time t, $\lvert y_{i,k}(t)-\hat{y}_{i,k}(t)\rvert$.
$\boldsymbol{c}_i$	The center of gravity in the SPSO2011 update equations.
$\boldsymbol{a}_i(t)$	Shorthand for $\boldsymbol{x}_i(t)+c_1\boldsymbol{r}_1\otimes(\boldsymbol{y}_i(t)-\boldsymbol{x}_i(t))$ in SPSO2011.
$\boldsymbol{b}_i(t)$	Shorthand for $\boldsymbol{x}_i(t)+c_2\boldsymbol{r}_2\otimes(\hat{\boldsymbol{y}}_i(t)-\boldsymbol{x}_i(t))$ in SPSO2011.
$\mathcal{H}(\hat{\boldsymbol{c}},r)$	A uniform distribution in the hypersphere with center $\hat{\boldsymbol{c}}$ and radius r.
γ_{fips}	The point of convergence for FIPS.
γ_{upso}	The point of convergence for UPSO.
$E[\cdot]$	The expectation operator.
$V[\cdot]$	The variance operator.

Appendix B

Objective Functions

This appendix describes the objective functions utilized within this thesis, along with the objective functions' corresponding search space. For each objective function, a reference is made to an existing source where the function is defined.

Function name	**Definition**	**Domain**
Absolute Value [21]	$f_1(\boldsymbol{x}) = \sum_{j=1}^{d} \lvert x_j \rvert$	$\boldsymbol{x} \in [-100, 100]^d$
Ackley [63]	$f_2(\boldsymbol{x}) = 20e^{-0.2\sqrt{\frac{1}{d}\sum_{j=1}^{d} x_j^2}} - e^{\frac{1}{d}\sum_{j=1}^{d} cos(2\pi x_i)} + 20 + e$	$\boldsymbol{x} \in [-32, 32]^d$
Alpine [54]	$f_3(\boldsymbol{x}) = \sum_{j=1}^{d} \lvert x_j sin(x_j) + 0.1x_j \rvert$	$\boldsymbol{x} \in [-10, 10]^d$
Bent Cigar [36]	$f_4(\boldsymbol{x}) = x_1^2 + 10^6 \sum_{j=2}^{d} x_i^2$	$\boldsymbol{x} \in [-100, 100]^d$
Discus [36]	$f_5(\boldsymbol{x}) = 10^6 x_1^2 + \sum_{j=2}^{d} x_i^2$	$\boldsymbol{x} \in [-100, 100]^d$
Egg Holder [21]	$f_5(\boldsymbol{x}) = \sum_{j=1}^{d-1} \left(-(x_{j+1} + 47) sin\left(\sqrt{\lvert x_{j+1} + x_j/2 + 47 \rvert}\right)\right)$ $+ sin\left(-x_j \sqrt{\lvert x_j - (x_{j+1} + 47) \rvert}\right)$	$\boldsymbol{x} \in [-512, 512]^d$
Elliptic [58]	$f_6(\boldsymbol{x}) = \sum_{j=1}^{d} x_j^2 (10^6)^{\frac{j-1}{d-1}}$	$\boldsymbol{x} \in [-100, 100]^d$
Expanded Griewank and Rosenbrock composite [58]	$f_6(\boldsymbol{x}) = \sum_{j=1}^{d-1} f_7\left(f_{17}\left(x_j, x_{j+1}\right)\right) + f_7\left(f_{17}\left(x_d, x_1\right)\right)$	$\boldsymbol{x} \in [-100, 100]^d$
Griewank [63]	$f_7(\boldsymbol{x}) = \frac{1}{4000} \sum_{j=1}^{d} x_j^2 - \prod_{j=1}^{d} cos\left(\frac{x_j}{\sqrt{j}}\right) + 1$	$\boldsymbol{x} \in [-600, 600]^d$
HappyCat [2]	$f_8(\boldsymbol{x}) = \lvert \sum_{j=1}^{d} x_j^2 - d \rvert^{\frac{1}{4}} + \left(0.5 \sum_{j=1}^{d} x_j^2 + \sum_{j=1}^{d} x_j\right)/d + 0.5$	$\boldsymbol{x} \in [-2, 2]^d$
HGBat [2]	$f_9(\boldsymbol{x}) = \left\lvert \left(\sum_{j=1}^{d} x_j^2\right)^2 - \left(\sum_{j=1}^{d} x_j\right)^2 \right\rvert^{\frac{1}{2}}$ $+ \left(0.5 \sum_{j=1}^{d} x_j^2 + \sum_{j=1}^{d} x_j\right)/d + 0.5$	$\boldsymbol{x} \in [-2, 2]^d$
Hyper Ellipsoid [21]	$f_{10}(\boldsymbol{x}) = \sum_{j=1}^{d} j x_j^2$	$\boldsymbol{x} \in [-5.12, 5.12]^d$

Katsuura [36]	$f_{11}(\boldsymbol{x}) = \frac{10}{d^2}\prod_{j=1}^{d}\left(1 + j\sum_{l=1}^{32}\frac{\lvert 2^l x_l - round(2^l x_l)\rvert}{2^l}\right)^{10/d^{1.2}} - \frac{10}{d^2}$	$\boldsymbol{x} \in [-5,5]^d$
Michalewicz [39]	$f_{12}(\boldsymbol{x}) = -\sum_{j=1}^{d} sin(x_j)(sin(\frac{jx_j}{\pi}))^{20}$	$\boldsymbol{x} \in [0,\pi]^d$
Norwegian [21]	$f_{13}(\boldsymbol{x}) = \prod_{j=1}^{d} cos(\pi x_j^3)\left(\frac{99+x_j}{100}\right)$	$\boldsymbol{x} \in [-1.1,1.1]^d$
Quadric [63]	$f_{14}(\boldsymbol{x}) = \sum_{j=1}^{d}\left(\sum_{l=1}^{j} x_l\right)^2$	$\boldsymbol{x} \in [-100,100]^d$
Quartic [63]	$f_{15}(\boldsymbol{x}) = \sum_{j=1}^{d} jx_j^4$	$\boldsymbol{x} \in [-1.28,1.28]^d$
Rastrigin [63]	$f_{16}(\boldsymbol{x}) = 10d + \sum_{j=1}^{d}\left(x_j^2 - 10cos(2\pi x_j)\right)$	$\boldsymbol{x} \in [-5.12,5.12]^d$
Rosenbrock [63]	$f_{17}(\boldsymbol{x}) = \sum_{j=1}^{d-1}(100(x_{j+1} - x_j^2)^2 + (x_i - 1)^2)$	$\boldsymbol{x} \in [-2.048,2.048]^d$
Salomon [53]	$f_{18}(\boldsymbol{x}) = -cos\left(2\pi\sum_{j=1}^{d} x_j^2\right) + 0.1\sqrt{\sum_{j=1}^{d} x_j^2} + 1$	$\boldsymbol{x} \in [-100,100]^d$
Schaffer 6 [21]	$f_{19}(\boldsymbol{x}) = \sum_{j=1}^{d-1}\left(0.5 + \frac{sin^2\left(x_j^2+x_{j+1}^2\right)-0.5}{\left(1+0.001\left(x_j^2+x_{j+1}^2\right)^2\right)}\right)$	$\boldsymbol{x} \in [-100,100]^d$
Schwefel 2.21 [21]	$f_{19}(\boldsymbol{x}) = max\{\lvert x_j\rvert, 1 \leq j \leq d\}$	$\boldsymbol{x} \in [-100,100]^d$
Schwefel 2.22 [63]	$f_{20}(\boldsymbol{x}) = \sum_{j=1}^{d}\lvert x_j\rvert + \prod_{j=1}^{d}\lvert x_j\rvert$	$\boldsymbol{x} \in [-10,10]^d$
Shubert [21]	$f_{21}(\boldsymbol{x}) = \prod_{j=1}^{d}\left(\sum_{l=1}^{5} lcos((l+1)x_j + l)\right)$	$\boldsymbol{x} \in [-10,10]^d$
Spherical [18]	$f_{22}(\boldsymbol{x}) = \sum_{j=1}^{d} x_j^2$	$\boldsymbol{x} \in [-5.12,5.12]^d$
Step [63]	$f_{23}(\boldsymbol{x}) = \sum_{j=1}^{d}\lfloor x_j + 0.5\rfloor^2$	$\boldsymbol{x} \in [-20,20]^d$
Vincent [21]	$f_{24}(\boldsymbol{x}) = -\left(1 + \sum_{j=1}^{d} sin\left(10\sqrt{x_j}\right)\right)$	$\boldsymbol{x} \in [0.25,10]^d$
Weierstrass [39]	$f_{25}(\boldsymbol{x}) = \sum_{j=1}^{d}\sum_{l=0}^{20} 0.5^l cos\left(2\pi 3^l(x_i + 0.5)\right)$ $-d\sum_{l=0}^{20} 0.5^l cos\left(2\pi 3^l\right)$	$\boldsymbol{x} \in [-0.5,0.5]^d$

Appendix C

Mathematical Definitions and Theorems

This appendix contains a number of well known definitions and theorems from functional and numerical analysis from [1] and [35] that are either explicitly or implicitly relevant to this thesis.

Definition C.1. Linear operator

A linear operator $\boldsymbol{T}$ is an operator with the following properties;

1. The domain $\mathcal{D}(\boldsymbol{T})$ of $\boldsymbol{T}$ is a vector space and the range $\mathcal{R}(\boldsymbol{T})$ lies in a vector space over the same field.

2. The following hold for all $\boldsymbol{z}, \boldsymbol{w} \in \mathcal{D}(\boldsymbol{T})$ and scalars ψ:

$$\boldsymbol{T}(\boldsymbol{z} + \boldsymbol{w}) = \boldsymbol{T}\boldsymbol{z} + \boldsymbol{T}\boldsymbol{w} \tag{C.1}$$

$$\boldsymbol{T}(\psi\boldsymbol{z}) = \psi\boldsymbol{T}\boldsymbol{z} \tag{C.2}$$

Definition C.2. Metric Space

A metric space is a pair (Z, d), where X is a set and d is a metric on Z. The metric is defined on $Z \times Z$ such that, for all $\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{z} \in Z$ the following properties hold:

1. $d(\boldsymbol{w}, \boldsymbol{z}) \geq 0$, and d is real-valued and finite.
2. $d(\boldsymbol{w}, \boldsymbol{z}) = 0 \iff v = w$.
3. $d(\boldsymbol{w}, \boldsymbol{z}) = d(\boldsymbol{z}, \boldsymbol{w})$.
4. $d(\boldsymbol{w}, \boldsymbol{z}) \leq d(\boldsymbol{w}, \boldsymbol{v}) + d(\boldsymbol{v}, \boldsymbol{z})$

Definition C.3. Cauchy Sequence

A sequence $(\boldsymbol{z}_n)$ in a metric space (Z, d) is Cauchy if for every $\epsilon > 0$ there exits a $G_\epsilon \in \mathbb{N}$, such that $d(\boldsymbol{z}_m, \boldsymbol{z}_n) < \epsilon$ for every $m, n > G_\epsilon$.

Definition C.4. Completeness

A metric space (Z, d) is complete if every Cauchy sequence in Z converges to a point within Z.

Definition C.5. Normed Space

A normed space $(Z, \|\cdot\|)$ is a vector space with a norm defined on it. A norm on a real-valued vector space Z is a real valued function defined such that, for all $\boldsymbol{w}, \boldsymbol{z} \in Z$ and $\alpha \in \mathbb{R}$, the following properties hold:

1. $\|\boldsymbol{z}\| \geq 0$.
2. $\|\boldsymbol{z}\| = 0 \iff z = 0$.
3. $\|\alpha \boldsymbol{z}\| = |\alpha| \|\boldsymbol{z}\|$.
4. $\|\boldsymbol{z} + \boldsymbol{w}\| \leq \|\boldsymbol{z}\| + \|\boldsymbol{w}\|$.

Also, a norm on Z defines a metric d on Z given by $d(\boldsymbol{z}, \boldsymbol{w}) = \|\boldsymbol{z} - \boldsymbol{w}\|$

Definition C.6. Banach Space

A banach space $(Z, \|\cdot\|)$ is a complete normed space, where completeness is defined in definition C.4.

Definition C.7. Bounded linear operator

Let Z and W be normed spaces and $T : \mathcal{D}(\boldsymbol{T}) \to W$ a linear operator, where $\mathcal{D}(\boldsymbol{T}) \subseteq Z$. The operator T is bounded if there exist a $h \in \mathbb{R}$ such that for all $\boldsymbol{z} \in Z$

$$\|\boldsymbol{T}\boldsymbol{z}\| \leq h\|\boldsymbol{z}\| \tag{C.3}$$

Theorem C.8.

If a normed space W is finite dimensional, then every linear operator on X is bounded. [35]

Theorem C.9. ***Matrix representation of liner operators***

Every linear operator $\boldsymbol{T}$ on finite dimensional vector spaces can be represented by a fixed matrix $\boldsymbol{M_T}$ [35].

Definition C.10. Spectrum

Let $\boldsymbol{T}$ be a linear operator on finite dimensional vector spaces. The spectrum of $\boldsymbol{T}$ is the set $\Sigma_{\boldsymbol{T}}$ of scalars λ, such that $det\,(\boldsymbol{M_T} - \lambda \boldsymbol{I}) = 0$, where det is the determinant and $\boldsymbol{I}$ is the identity matrix.

Definition C.11. Spectral radius

Let $\boldsymbol{T}$ be a linear operator on finite dimensional vector spaces with the spectrum $\Sigma_{\boldsymbol{T}}$, then the spectral radius of $\boldsymbol{T}$, is defined as

$$\rho(\boldsymbol{T}) = \max_{\lambda \in \Sigma_{\boldsymbol{T}}} |\lambda| \tag{C.4}$$

Theorem C.12.

Let $\boldsymbol{T}$ be a bounded liner operator from Z to W. For every $\delta > 0$ there exists a norm $\|\cdot\|_\delta$ such that

$$\rho(\boldsymbol{T}) \leq ||\boldsymbol{T}||_\delta \leq \rho(\boldsymbol{T}) + \delta \tag{C.5}$$

Definition C.13. Uniform Operator Convergence

Let Z and W be normed spaces. A sequence $(\boldsymbol{T}_n)$ of operators $\boldsymbol{T}_n : Z \to W$ is said to be uniformly operator convergent if

$$\|\boldsymbol{T}_n - \boldsymbol{T}\| \to 0, \tag{C.6}$$

where $\|\cdot\|$ is a operator norm.

Appendix D

Eigenvalues from Theoretical Analysis

For the sake of completeness, this appendix contains the five eigenvalues of matrix $\boldsymbol{B}_t$, from chapter 9. 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^(1/2)((E[β^2]/3 + (E[β] + E[α^2])^2/9 - (E[β]E[α^2])/3 + (2E[α]E[$\alpha\beta$])/3)/((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2 + (((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2)^2 - (E[β^2]/3 + (E[β] + E[α^2])^2/9 - (E[β]E[α^2])/3 + (2E[α]E[$\alpha\beta$])/3)^3)^(1/2))^(1/3) - ((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2 + (((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2)^2 - (E[β^2]/3 + (E[β] + E[α^2])^2/9 - (E[β]E[α^2])/3 + (2E[α]E[$\alpha\beta$])/3)^3)^(1/2))^(1/3))1i)/2 - (E[β^2]/3 + (E[β] + E[α^2])^2/9 - (E[β]E[α^2])/3 + (2E[α]E[$\alpha\beta$])/3)/(2((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2 + (((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2)^2 - (E[β^2]/3 + (E[β] + E[α^2])^2/9 - (E[β]E[α^2])/3 + (2E[α]E[$\alpha\beta$])/3)^3)^(1/2))^(1/3)) - ((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2 + (((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2)^2 - (E[β^2]/3 + (E[β] + E[α^2])^2/9 - (E[β]E[α^2])/3 + (2E[α]E[$\alpha\beta$])/3)^3)^(1/2))^(1/3)/2

Eigenvalue 5:

$E[\beta]/3 + E[\alpha^2]/3$ + (3^(1/2)((E[β^2]/3 + (E[β] + E[α^2])^2/9 - (E[β]E[α^2])/3 + (2E[α]E[$\alpha\beta$])/3)/((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2 + (((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2)^2 - (E[β^2]/3 + (E[β] + E[α^2])^2/9 - (E[β]E[α^2])/3 + (2E[α]E[$\alpha\beta$])/3)^3)^(1/2))^(1/3) - ((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2 + (((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2)^2 - (E[β^2]/3 + (E[β] + E[α^2])^2/9 - (E[β]E[α^2])/3 + (2E[α]E[$\alpha\beta$])/3)^3)^(1/2))^(1/3))1i)/2 - (E[β^2]/3 + (E[β] + E[α^2])^2/9 - (E[β]E[α^2])/3 + (2E[α]E[$\alpha\beta$])/3)/(2((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2 + (((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2)^2 - (E[β^2]/3 + (E[β] + E[α^2])^2/9 - (E[β]E[α^2])/3 + (2E[α]E[$\alpha\beta$])/3)^3)^(1/2))^(1/3)) - ((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2 + (((E[β] + E[α^2])^3/27 + ((E[β] + E[α^2])(E[β^2] - E[β]E[α^2] + 2E[α]E[$\alpha\beta$]))/6 + (E[α]E[β^2])/2 - (E[β]E[β^2])/2)^2 - (E[β^2]/3 + (E[β] + E[α^2])^2/9 - (E[β]E[α^2])/3 + (2E[α]E[$\alpha\beta$])/3)^3)^(1/2))^(1/3)/2

Bibliography

[1] K. Atkinson and W. Han. *Theoretical Numerical Analysis: A Functional Analysis Framework*. Springer, Heidelberg, Berlin, 2009.

[2] H.G. Beyer and S Fink. Happycat — a simple function class where well-known direct search algorithms do fail. In *Proceedings of the 12th International Conference on Parallel Problem Solving from Nature*, Heidelberg, Berlin, 2012. Springer.

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

[4] M.R Bonyadi and Z Michalewicz. SPSO2011- analysis of stability, local convergence, and rotation sensitivity. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 9–15, New York, NY, 2014. ACM Press.

[5] M.R Bonyadi and Z Michalewicz. Analysis of stability, local convergence, and transformation sensitivity of a variant of particle swarm optimization algorithm. *IEEE Transactions on Evolutionary Computation*, 20(3):370–385, 2016.

[6] M.R. Bonyadi and Z. Michalewicz. Particle swarm optimization for single objective continuous space problems: a review. *Evolutionary Computation*, pages 1–54, 2016.

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

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

[9] C.W. Cleghorn and A.P. Engelbrecht. Particle swarm convergence: An empirical investigation. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 2524–2530, Piscataway, NJ, 2014. IEEE Press.

[10] C.W. Cleghorn and A.P. Engelbrecht. Particle swarm convergence: Standardized analysis and topological influence. In *Proceedings of International Swarm Intelligence Conference (ANTS), Swarm Intelligence*, pages 134–145, Switzerland, 2014. Springer International Publishing.

[11] C.W. Cleghorn and A.P. Engelbrecht. Fully informed particle swarm optimizer: Convergence analysis. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 164–170, Piscataway, NJ, 2015. IEEE Press.

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

[13] C.W. Cleghorn and A.P. Engelbrecht. 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, 2016. IEEE Press.

[14] C.W. Cleghorn and A.P. Engelbrecht. Unified particle swarm optimizer: Convergence analysis. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 448–454, Piscataway, NJ, 2016. IEEE Press.

[15] M. Clerc. The swarm and the queen: Towards a deterministic and adaptive particle swarm optimization. In *Proceedings of the IEEE Congress on Evolutionary Computation*, volume 3, pages 1951–1957, Piscataway, NJ, USA, 1999. IEEE Press.

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

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

[18] K.A. De Jong. *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*. PhD thesis, University of Michigan, Ann Arbor, US, 1975.

[19] K. Du and M.N.S Swamy. *Search and Optimization by Metaheuristics: Techniques and Algorithms Inspired by Nature*. Birkhäuser, Basel, Switzerland, 2016.

[20] A.P. Engelbrecht. *Computational Intelligence, 2nd edition*. Wiley, Wiltshire, UK, 2007.

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

[22] A.P. Engelbrecht. Roaming behavior of unconstrained particles. In *Proceedings of the 1st BRICS Countries Congress on Computational Intelligence*, pages 104–111, Piscataway, NJ, 2013. IEEE Press.

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

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

[25] V. Gazi. 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, 2012. IEEE Press.

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

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

[28] M. Kendall and J. Ord. *Time Series, 3rd edition*. Edward Arnold, London, UK, 1990.

[29] J. Kennedy. 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, 1999. IEEE Press.

[30] J Kennedy. Bare bones particle swarms. In *Proceedings of the IEEE Swarm Intelligence Symposium*, pages 80–87, Piscataway, NJ, 2003. IEEE Press.

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

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

[33] J. Kennedy and R. Mendes. 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, 2003. IEEE Press.

[34] B. Kisacanin and G.C. Agarwal. *Linear Control Systems: With Solved Problems and Matlab Examples*. Springer, New York, NY, 2001.

[35] E. Kreyszig. *Introductory Functional Analysis With Applications*. Wiley, Wiltshire, UK, 1978.

[36] J.J.K. Liang, B.Y. Qu, and P.N Suganthan. Problem definitions and evaluation criteria for the CEC 2014 special session and competition on single objective real-parameter numerical optimization. Technical Report 201311, Computational Intelligence Laboratory, Zhengzhou University and Nanyang Technological University, 2013.

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

[38] V. Mirinda, H. Keko, and A.J. Duque. Stochastic star communication topology in evolutionary particle swarms (EPSO). *International Journal of Computational Intelligence Research*, 4(2):105–116, 2008.

[39] S.K. Mishra. Performance of repulsive particle swam method in global optimization of some important test functions: A fortran program. Technical report, Independant, 2006.

[40] A.S. Mohais, C. Ward, and C Posthoff. Randomized directed neighborhoods with edge migration in particle swarm optimization. *Proceedings of the IEEE Congress on Evolutionary Computation*, 1:548–555, 2004.

[41] M.A. Montes de Oca and T. Stützle. Convergence behavior of the fully informed particle swarm optimization algorithm. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 71–78, New York, NY, 2008. ACM Press.

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

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

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

[45] K.E. Parsopoulos and M.N. Vrahatis. UPSO: A unified particle swarm optimization scheme. In *Proceedings of the International Conference on Computational Methods in Sciences and Engineering*, pages 868–873, Netherlands, 2004.

[46] E.S. Peer, F. Van den Bergh, and A.P. Engelbrecht. Using neighborhoods with the guaranteed convergence PSO. In *Proceedings of the IEEE Swarm Intelligence Symposium*, pages 235–242, Piscataway, NJ, 2003. IEEE Press.

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

[48] R. Poli. On the moments of the sampling distribution of particles swarm optimisers. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 2907–141, New York, NY, 2007. ACM Press.

[49] R. Poli. Dynamics and stability of the sampling distribution of particle swarm optimisers via moment analysis. *IEEE Transactions on Evolutionary Computation*, 2008(685175):1–10, 2008.

[50] R. Poli. Mean and variance of the sampling distribution of particle swarm optimizers during stagnation. *IEEE Transactions on Evolutionary Computation*, 13(4):712–721, 2009.

[51] R. Poli and D. Broomhead. 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, 2007. ACM Press.

[52] R. Poli, J. Kennedy, and Blackwell T. Particle swarm optimization. An overview. *Swarm Intelligence Journal*, 1(1):33–57, 2007.

[53] K.V. Price, R.M. Storn, and J.A. Lampinen. *A Practical Approach to Global Optimization, Natural Computing Series*. Springer-Verlag, Berlin, 2015.

[54] S. Rahnamayan, R. Tizhoosh, and M.M.A Salama. A novel population initialization method for accelerating evolutionary algorithms. *Computers & Mathematics with Aplications*, 53(10):1605–1614, 2007.

[55] A. Ratnaweera, S. Halgamuge, and H. Watson. 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, 2003. Springer.

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

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

[58] P.N. Suganthan, N. Hanse, J.J. Liang, K. Deb, P. Chen, A. Auger, and S. Tiwart. Problem definitions and evaluation criteria for the CEC 2005 special session on real-parameter optimization. Technical report, Nanyang Technological University, Singapore, 2005.

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

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

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

[62] D.J. Watts and S.H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.

[63] X. Yoa, Y. Liu, and Lin.G. Evolutionary programming made faster. *IEEE Transactions on Evolutionary Computation*, 3(2):82–102, 1999.

[64] H. Yoshida, Y. Fukuyama, S. Takayama, and Y Nakanishi. 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, USA, 1999. IEEE Press.

[65] M. Zambrano-Bigiarini and M Clerc. Standard particle swarm optimisation 2011 at CEC-2013: A baseline for future PSO improvements. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 2337–2344, Piscataway, NJ, 2013. IEEE Press.