UNIVERSITY OF PRETORIA

A Generalized Theoretical Deterministic Particle Swarm Model

by

Christopher Wesley Cleghorn

Submitted in partial fulfillment of the requirements for the degree
Master of Science

in the Faculty of
Engineering, Built Environment and Information Technology

July 2013

A Generalized Theoretical Deterministic Particle Swarm Model
Abstract

Master of Computer Science

Written by Christopher Wesley Cleghorn
Supervised by Prof. Andries Engelbrecht,

Particle swarm optimization (PSO) is a well known population-based search algorithm, originally developed by Kennedy and Eberhart in 1995. The PSO has been utilized in a variety of application domains, providing a wealth of empirical evidence for its effectiveness as an optimizer. The PSO itself has undergone many alterations subsequent to its inception, some of which are fundamental to the PSO's core behavior, others have been more application specific. The fundamental alterations to the PSO have to a large extent been a result of theoretical analysis of the PSO's particle's long term trajectory. The most obvious example, is the need for velocity clamping in the original PSO. While there were empirical findings that suggested that each particle's velocity was increasing at a rapid rate, it was only once a solid theoretical study was performed that the reason for the velocity explosion was understood. There has been a large amount of theoretical research done on the PSO, both for the deterministic model, and more recently for the stochastic model.

This thesis presents an extension to the theoretical deterministic PSO model. Under the extended model, conditions for particle convergence to a point are derived. At present all theoretical PSO research is done under the stagnation assumption, in some form or another. The analysis done under the stagnation assumption is one where the personal best and neighborhood best are assumed to be non-changing. While analysis under the stagnation assumption is very informative, it could never provide a complete description of a PSO's behavior. Furthermore, the assumption implicitly removes the notion of a social network structure from the analysis. The model used in this thesis greatly weakens the stagnation assumption, by instead assuming that each particle's personal best and neighborhood best can occupy an arbitrarily large number of unique positions. Empirical results are presented to support the theoretical findings.

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.

- June Clarence, for her life long encouragement of my mathematical curiosity.

- My friends, for there motivation and for adding a healthy dose of fun to the research process.

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

Contents

Chapter 1

Introduction

1.1   Motivation

In the world of optimization there are many important factors to consider when deciding on the best approach to use. Some of the primary factors are the computational efficiency, the accuracy and the predictability of an approach. At present there are so many optimization approaches available, varying from classical approaches to the more modern computational intelligence approaches, that making an informed decision can be difficult. It is for this reason that having a complete understanding about how an optimization technique behaves is imperative, as it provides the user of the approach with as much information as possible. While this fact may seem obvious, there are certain situation were if this information is not available, the approach might never be considered, for example in a situation where human life is at risk. It is for this reason that theoretical work on optimization techniques are of great importance, as they provide the certainty that empirical findings will never be able to.

The particle swarm optimization (PSO) model has undergone a fair amount theoretical analysis, however, at present all theoretical PSO research is done under the stagnation assumption, in some form or another. An analysis done under the stagnation assumption is one where the personal best and neighbourhood best are assumed to be non-changing. While analysis under the stagnation assumption is very informative, it could never provide a complete description of a PSO's behaviour. It is for this reason that a generalization of the model is required. This generalization, and its theoretical implications, is the focus of this thesis.

1

1.2   Objectives

The primary objectives of this thesis can be summarized as follows:

- To justify why a generalization of the theoretical deterministic PSO model is needed.

- To provide a generalization of the theoretical deterministic PSO model that mitigates the old theoretical models' weaknesses, in particular a generalization that caters for the influence of the social network structure.

- To theoretically derive the conditions necessary for particle convergence to a point, under the new generalized model.

- To empirically support the findings of the theoretical derivation.

1.3   Thesis Outline

Chapter 2 begins with a detailed discussion of the particle swarm optimizer in section 2.1. The section begins with a description of the original PSO as developed by Kennedy and Eberhart [13], followed by explanations of the some of the most frequently used alterations to the original PSO in sections 2.1.2, 2.1.3, 2.1.4, and 2.1.5. A thorough discussion of the state of theoretical PSO research is made in section 2.2, split into two main categories, i.e. the deterministic PSO research in section 2.2.1 and the stochastic PSO research in section 2.2.2.

Chapter 3 presents a generalization of the theoretical deterministic PSO model. The chapter begins with a justification as to why generalization of the current theoretical model is needed in section 3.1. Conditions for particle convergence to a point are derived in section 3.3 under the generalized model. The derived conditions are discussed in section 3.4.

Chapter 4 presents an experiment designed to illustrate the convergence properties of the deterministic PSO under various social network structures in relation to the derived condition of chapter 3. The experimental set up is discussed in section 4.1, with the experimental result presented and discussed in section 4.2.

Chapter 5 presents a summary of the findings of this thesis, with an overview given in section 5.1. Topics for future research are also discussed in section 5.2.

Chapter 2

Background

This chapter begins with a compressive explanation of the particle swarm optimizer (PSO) in section 2.1. This is followed by a complete discussion of the state of theoretical research in section 2.2, broken down into two subsections: Subsection 2.2.1 discusses the deterministic PSO theoretical results, and subsection 2.2.2 discusses the stochastic PSO theoretical results.

2.1 Particle Swarm Optimization

This section begins with an explanation of the original PSO in subsection 2.1.1, in addition to a discussion in subsections 2.1.2, 2.1.3, 2.1.4, and 2.1.5 of the alterations the PSO algorithm has undergone since its inception.

2.1.1 Original Particle Swarm Optimization Algorithm

Particle swarm optimization (PSO) is a stochastic population-based search algorithm, which was developed by Kennedy and Eberhart [13]. The algorithm was originally developed to simulate the complex movement of bird blocks. The algorithm has however evolved into a simple and efficient optimization algorithm. The abstract premise of the algorithm is to have particles move through a potentially multi-dimensional search space in an attempt to find an optimal region. Each particle position represents a candidate solution. Each particle's movement is guided by its own experience in conjunction with that of its neighbours.

More technically, let $f : \mathbb{R}^k \to \mathbb{R}$ be the fitness function that the PSO aims to find an optimal solution for. For the sake of simplicity, a minimization problem is assumed from

3

this point forward. Let $\Omega(t)$ be a set of N particles in $\mathbb{R}^k$, at a discrete time step t. Then $\Omega(t)$ is said to be the particle swarm at time t. The position $\mathbf{x}_i$ of particle i, is updated using the following,

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

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

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

where $\boldsymbol{r}_1(t), \boldsymbol{r}_2(t) \in U(0,1)^k$ for all t; these random values introduce the stochastic aspect to PSO. The position $\boldsymbol{y}_i(t)$ represents the "best" position that particle i has visited, where "best" means the location where the particle has obtained the lowest fitness evaluation. Specifically,

$$\boldsymbol{y}_i(t) = \boldsymbol{x}_i(\hat{t}), \tag{2.3}$$

where

$$f(\boldsymbol{x}_i(\hat{t})) = \min\{f(\boldsymbol{x}_i(0)), f(\boldsymbol{x}_i(1)), \ldots, f(\boldsymbol{x}_i(t))\}. \tag{2.4}$$

The position $\hat{\boldsymbol{y}}_i(t)$ represents the "best" position that the particles in $\mathcal{N}_i$ have visited, where $\mathcal{N}_i$ is the neighbourhood of particle i. Specifically,

$$\hat{\boldsymbol{y}}_i(t) = \boldsymbol{y}_{\hat{i}}(t), \tag{2.5}$$

where

$$f(\boldsymbol{y}_{\hat{i}}(t)) = \min\{f(\boldsymbol{y}_1(t)), f(\boldsymbol{y}_2(t)), \ldots, f(\boldsymbol{y}_N(t))\}. \tag{2.6}$$

A detailed discussion of neighbourhoods is given in section 2.1.5.

The velocity update consists of three components:

- **Inertia component**, $\boldsymbol{v}_i(t)$: The inertia component serves as the memory of the direction of travel in the particle's immediate past. The component prevents a particle's trajectory from wildly changing, much like a bird would be unable to preform a perfect ninety degree turn mid flight. The inertia often assists in pushing a particle out of a local optimum.

- **Cognitive component**, $c_1\boldsymbol{r}_1(t)(\boldsymbol{y}_i(t) - \boldsymbol{x}_i(t))$: The cognitive component serves as the memory of a past position of success. Success implies the best fitness evaluation at the position in question. The cognitive component causes particles to be drawn back to positions of previous success, much like a human's tendency to revisit

places where a positive association exists. Kennedy and Eberhart referred to the cognitive component as the "nostalgia" of the particle [13]. The coefficient c_1 is referred to as the cognitive importance.

- **Social component**, $c_2 \boldsymbol{r}_2(t)(\hat{\boldsymbol{y}}_i(t) - \boldsymbol{x}_i(t))$: The social component serves as the memory of a past position of success visited by neighbouring particles (including itself). The component causes the particle to be drawn to a position where the most success was achieved within the neighbourhood. The behaviour is akin to individuals seeking to fit the group norm of their neighbourhood, Kennedy and Eberhart refer to the social component as particle "envy".

The PSO algorithm using synchronous updates is summarized in algorithm 1.

Algorithm 1 PSO Algorithm

Create and initialize a k-dimensional swarm, $\Omega(0)$, of N particles uniformly within a predefined hypercube.
Let f be the fitness 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 neighbourhood best of position of particle i, initialized to $\boldsymbol{x}_i(0)$.
Initialize $\boldsymbol{v}_i(0)$ to $\boldsymbol{0}$.
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
        if $f(\boldsymbol{y}_i) < f(\hat{\boldsymbol{y}}_i)$ **then**
            $\hat{\boldsymbol{y}}_i = \boldsymbol{y}_i$
        end if
    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

2.1.2 Velocity Clamping Particle Swarm Algorithm

While the PSO as discussed in section 2.1.1 was able to obtain good results [13], it exhibited a rather undesirable behaviour. The magnitude of the velocity of each particle would quickly explode to surprisingly large values. This meant that the particle were prone to not only leave the search space but continue to migrate away from it in some cases.

In an attempt to prevent the velocity explosion, velocity clamping was proposed [7]. Specifically, let $V_{\max,j}$ denote the maximum magnitude allowed for each component j of

the velocity of each particle. Then,

$$v_{i,j}(t+1) = \begin{cases} \hat{v}_{i,j}(t+1) & \text{if} \quad |\hat{v}_{i,j}(t+1)| < V_{\max,j} \\ sign(\hat{v}_{i,j}(t+1))V_{\max,j} & \text{if} \quad |\hat{v}_{i,j}(t+1)| \geq V_{\max,j} \end{cases} \tag{2.7}$$

where $\hat{v}_{i,j}$ is calculated using equation (2.2).

The best value for each $V_{\max,j}$ is unfortunately problem dependent. An even worse feature of velocity clamping is that the higher the dimension of the search space, the more user defined parameters the PSO will require, as a maximum velocity magnitude for each j will need to be set. These undesirable features of velocity clamping lead to the development of constriction coefficients as discussed in section 2.1.3. Other variations of the standard velocity clamping of equation (2.7) exist. The reader is referred to [8, 28] for more detail.

2.1.3 Constriction Coefficients

In an attempt to remove the need for velocity clamping, Clerc [2] proposed the inclusion of a constriction coefficient, the details for which were completely fleshed out later by Clerc and Kennedy [4]. The constriction coefficient is defined as

$$\chi_{i,j} = \begin{cases} \frac{2\kappa}{|2-\theta_{1,j}-\theta_{2,j}-\sqrt{(\theta_{1,j}+\theta_{2,j})(\theta_{1,j}+\theta_{2,j}-4)}|} & \text{if} \quad \theta_{1,j}+\theta_{2,j} > 4 \\ \kappa & \text{if} \quad \theta_{1,j}+\theta_{2,j} \leq 4 \end{cases} \tag{2.8}$$

where $\kappa \in (0,1)$, $\theta_{1,j} = r_{1,j}c_1$, and $\theta_{2,j} = r_{2,j}c_2$. Update equations (2.1) and (2.2) are then changed to

$$x_{i,j}(t+1) = \chi_{i,j}x_{i,j}(t) + \chi_{i,j}v_i(t+1) - (1-\chi_{i,j})\left(\frac{\theta_{1,j}y_{i,j}(t)+\theta_{2,j}\hat{y}_{i,j}}{\theta_{1,j}+\theta_{2,j}}\right) \tag{2.9}$$

$$v_{i,j}(t+1) = v_{i,j}(t) + \theta_{1,j}(t)(y_{i,j}(t) - x_{i,j}(t)) + \theta_{2,j}(t)(\hat{y}_{i,j}(t) - x_{i,j}(t)). \tag{2.10}$$

However, in practice [6] the constriction coefficient is only applied to the velocity component of the position update

$$x_{i,j}(t+1) = x_{i,j}(t) + \chi_{i,j}v_i(t+1) \tag{2.11}$$

The constriction coefficients can be interpreted as an imposed recommendation on the PSO to take smaller steps. Furthermore, the application of the constriction coefficients as in equation (2.9) will insure convergence of each particle to a point in the search space [4]. The value of κ has a direct impact on the rate of convergence of particles. If $\kappa >= 0 + \epsilon$, where $\epsilon > 0$ is "small", then the convergence of the particles is fast, with

local exploitation. On the other extreme, if $\kappa <= 1 - \epsilon$, where $\epsilon > 0$ is "small", the convergence of the particles is slow, with a high degree of exploration.

2.1.4   Inertia weight

The introduction of the inertia weight to the PSO algorithm was proposed by Shi and Eberhart [29]. The aim of the inclusion was two fold: Firstly, as a means of controlling the exploration and exploitation ability of the swarm, and secondly as a simple way of removing the need for velocity clamping. The inertia coefficient w is applied to the previous velocity of a particle, changing velocity update equation (2.2) to

$$\boldsymbol{v}_i\left(t+1\right) = w\boldsymbol{v}_i\left(t\right) + c_1\boldsymbol{r}_1(t)(\boldsymbol{y}_i(t) - \boldsymbol{x}_i\left(t\right)) + c_2\boldsymbol{r}_2(t)(\hat{\boldsymbol{y}}_i(t) - \boldsymbol{x}_i\left(t\right)). \qquad (2.12)$$

where $w \in \mathbb{R}$; usually a positive value is utilized.

The inclusion of the inertia weight successfully provide a simple control mechanism for the exploration and exploitation ability of the swarm. However, its secondary goal was not completely met. The need for velocity clamping was unfortunately still required in some cases [32–34]. It was only later when theoretical studies were done by Van den Berg and Engelbrecht [33, 34] and Trelea [32] that it was ascertained that velocity clamping was not required under specified coefficient choices. The details of this analysis are discussed in section 2.2. There exist many variations of the concept of an inertia coefficient, many of which alter the value of w dynamically as the time t changes [3, 19, 23, 24, 27, 30, 35, 38], the details of which are beyond the scope of this thesis. The use of velocity update equation (2.12) has become the standard PSO.

2.1.5   Social Network Structure

The driving feature 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, for example the work of Mohais [18].
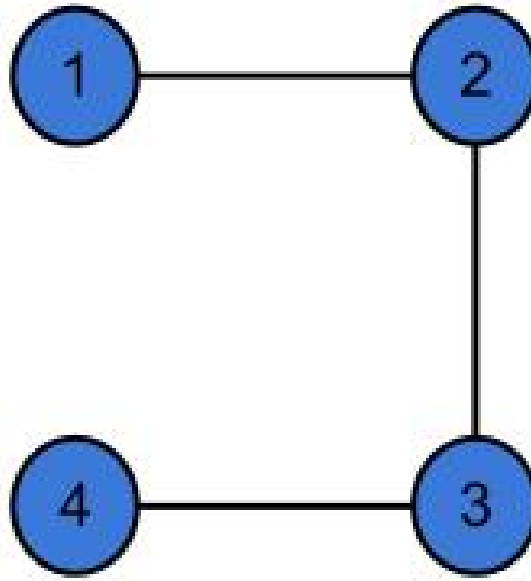
FIGURE 2.1: Simple Social Network.

The social network structure chosen has a direct impact on the behaviour of the swarm as a whole [12, 14, 22]. 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 [37].

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

- **Star**: The star network structure is one where all the particles in the swarm are interconnected as illustrated in figure 2.2. 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 2.3. 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 2.4, and the 3-D variant is illustrated in figure 2.5.
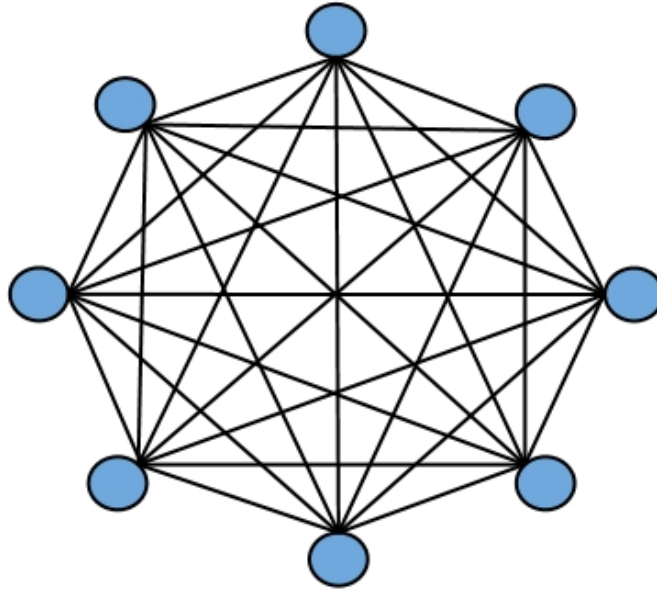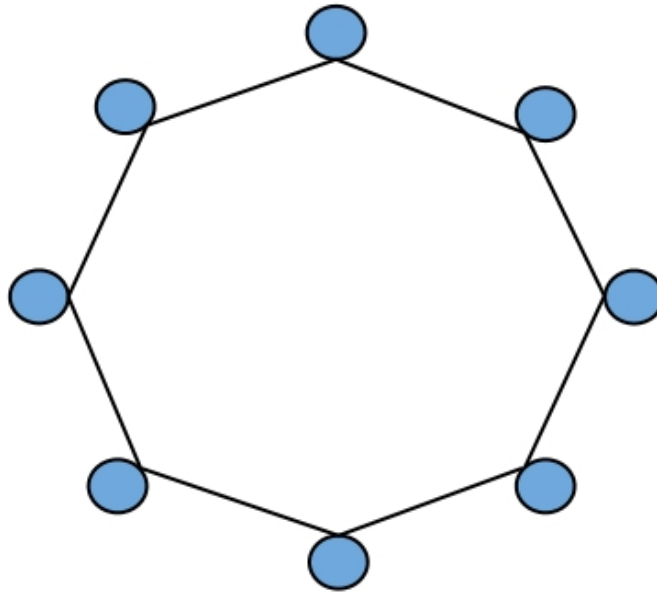
FIGURE 2.2: Star Network Structure.

FIGURE 2.3: Ring Network Structure.

There are many other available social network structures, such as the wheel, pyramid, four cluster, random, amongst others [14]. 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. 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 [14, 22].

It is worth noting that, in general, the neighbourhood structure is not constructed considering the proximity of particles to each other, but rather based on the indices of the particles in the swarm. There are, however, exceptions, for example the work
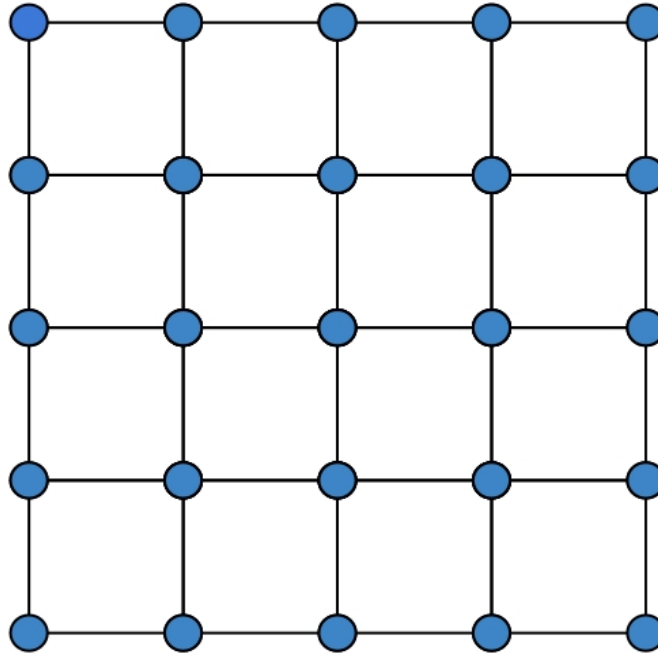
FIGURE 2.4: 2-D Von Neumann Network Structure.
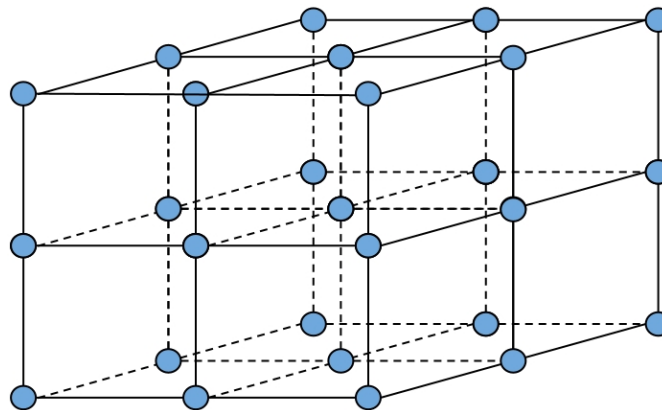
FIGURE 2.5: 3-D Von Neumann Network Structure.

of Suganthan [31] where the Euclidean distances between particles is utilized in the construction of neighbourhoods. This method of neighbourhood construction has the advantage of more accurately modeling social behaviour, as an individual is more likely to communicate with a closer individual. The disadvantage of this approach is the large amount of extra computation that is required to construct the neighbourhood, particularly if the neighbourhood changes on each iteration.

2.2 The State of Current Theoretical Research on Particle Swarm Optimization

Theoretical research of the PSO has evolved since the original study done by Ozcan and Mohan [20]. The evolution has been a process of modelling ever more accurate simplifications of the actual PSO.

The idea of theoretically modelling an optimization algorithm is a relatively new phenomenon, in contrast to modelling a natural system. Historically, a model of a complex natural system is made such that the model seems to fit the data that the system produces. The model of a natural system is almost by definition an approximation of the system, largely due to the fact that exact underlying behaviour is somewhat unknown. However, in the case of modelling an algorithm, the exact model is already known (i.e. the algorithm itself). The problem, however, is that the actual model is too complex to effectively derive meaningful results from. In the case of PSO, there are two primary complexities in the model: The inherent coupling between successive updates as a result of the non-static nature of $\boldsymbol{y}_i$ and $\hat{\boldsymbol{y}}_i$, and the influence of the stochastic component.

There are many aspects of PSO particle behaviour that are of interest from a theoretical and practical perspective. However, the focus of this discussion is on convergence results, specifically, under what conditions will convergence of each particle in the swarm to a point be guaranteed. To clarify, convergence of a particle is defined as follows:

Definition 2.2.1. **Particle convergence**: Let the dimension of the search space be k. Then, particle i is said to converge if there exists a $\boldsymbol{p} \in \mathbb{R}^k$ such that for every $\epsilon > 0$ there exists a $\hat{t}(\epsilon)$ such that if $t > \hat{t}(\epsilon)$ then

$$\|\boldsymbol{x}_i(\hat{t}(\epsilon)) - \boldsymbol{p}\| \leq \epsilon, \tag{2.13}$$

where $\|\bullet\|$ is an arbitrary norm on $\mathbb{R}^k$.

The definition of convergence in definition 2.2.1 should not be confused with the convergence of a particle to an optimum, but simply that the particle has come to rest at a point in the search space.

The current theoretical research on PSO can be split into two sub-branches: The first is where the stochastic component is assumed to be fixed, resulting in a deterministic PSO, as discussed in section 2.2.1. The second is where the assumption is dropped, as discussed in section 2.2.2.

2.2.1 Deterministic Particle Swarm Model

The first analysis of the PSO was performed by Ozcan and Mohan [20] on a very simplified version of the PSO excluding inertia weights. The analysis was performed in 1-dimension on the "best" particle's trajectory. This "best" particle i is one where the position of the personal best $\boldsymbol{y}_i(t)$ and that of the neighbourhood best position $\hat{\boldsymbol{y}}_i(t)$ are assumed to be equal for all t. There were two additional assumptions made in the analysis: The stochastic components of the update equations were assumed to be fixed, specifically:

Assumption 2.2.1. **Deterministic assumption**: It is assumed that $\boldsymbol{\theta_1} = \boldsymbol{\theta_1}(t) = c_1\boldsymbol{r}_1(t)$, and $\boldsymbol{\theta_2} = \boldsymbol{\theta_2}(t) = c_2\boldsymbol{r}_2(t)$, for all t.

The second assumption is that the personal best position remains fixed for all t, specifically:

Assumption 2.2.2. **Best particle stagnation assumption**: It is assumed that $\boldsymbol{y}_i(t) = \hat{\boldsymbol{y}}_i(t) = \boldsymbol{y}_i$, for all t.

Giving these assumptions and dropping the particle index the update equations are reformulated as

$$x\left(t+1\right) = x\left(t\right) + v\left(t+1\right), \tag{2.14}$$

$$v\left(t+1\right) = v\left(t\right) - \theta x\left(t\right) + \theta y, \tag{2.15}$$

where $\theta = \theta_1 + \theta_2$. This was then reformulated into the recurrence relation,

$$x\left(t+1\right) = (2-\theta)x(t) - x(t-1) + \theta y, \tag{2.16}$$

with initial conditions $x\left(0\right) = x_0$, $v\left(0\right) = v_0$ and $x\left(1\right) = x_0(1-\theta) + v_0 + \theta y$. From this recurrence relation the closed form of $x(t)$ is obtained, as

$$x(t) = \alpha \left(\frac{2-\theta+\delta}{2}\right)^t + \beta \left(\frac{2-\theta-\delta}{2}\right)^t + y, \tag{2.17}$$

where

$$\delta = \sqrt{\theta^2 - 4\theta}, \tag{2.18}$$

$$\beta = (x_0 - y)(\delta + \theta)2\delta - \frac{v_0}{\delta}, \tag{2.19}$$

$$\alpha = x_0 - y - \beta. \tag{2.20}$$

The analysis of the long term trajectory of a particle under equation (2.17), was split into two groups of cases: One where δ is real, the other where it is complex. This analysis was the first to give an idea of the trajectories of a particle in a PSO, and provided a

theoretical bases for intelligent parameter selection. The exact cases are discussed in
[20].

The fundamental issue with the original research was that the "best" particle is not of
practical interest, as the selection of which particle is "best" is very unlikely to remain
constant through a whole run. This results in a poor approximation of the actual PSO,
as the model effectively provides a prediction on the trajectory of a particle that is
unlikely to exist in practice. This issue was subsequently addressed again by Ozcan and
Mohan in a later paper [21], where the trajectory of a "general" particle is considered.
The assumptions remained the same as in the 1998 paper [20], however $\boldsymbol{y}_i(t)$ and $\hat{\boldsymbol{y}}_i(t)$
were not assumed equal for all t, instead the best particle stagnation assumption was
replaced with:

Assumption 2.2.3. **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.

The analysis was also done in $\mathbb{R}^k$. Though it is worth noting that giving the assumption
that $\boldsymbol{y}_i(t)$ and $\hat{\boldsymbol{y}}_i(t)$ are assumed fixed, an analysis in $\mathbb{R}^k$ and $\mathbb{R}$ are equivalent as the
assumption inherently removes the coupling between dimensions. The paper derives a
new recurrence relation,

$$x(t) = 2\left(2 - \theta_1 - \theta_2\right)x(t-1) - x(t-2) + \theta_1 y + \theta_2 \hat{y}, \tag{2.21}$$

where $y(t) = y$, and $\hat{y}(t) = \hat{y}$ for all t, with the initial conditions $x(0) = x_0$, $v(0) = v_0$,
and $x(1) = x_0\left(1 - \theta_1 - \theta_2\right) + v_0 + \theta_1 y + \theta_2 \hat{y}$. Then, using methods similar to that of
[20], the following closed form is derived:

$$x(t) = \eta \alpha^t + \iota \beta^t + \zeta, \tag{2.22}$$

where

$$\delta = \sqrt{\left(2 - \theta_1 - \theta_2\right)^2 - 4}, \tag{2.23}$$

$$\alpha = \frac{\left(2 - \theta_1 - \theta_2 + \delta\right)}{2}, \tag{2.24}$$

$$\beta = \frac{\left(2 - \theta_1 - \theta_2 - \delta\right)}{2}, \tag{2.25}$$

$$\eta = \left(\frac{1}{2} - \frac{\theta_1 + \theta_2}{2\delta}\right)x_0 + \frac{v_0}{\delta} + \frac{\theta_1 y + \theta_2 \hat{y}}{2\delta} - \frac{\theta_1 y + \theta_2 \hat{y}}{2\theta_1 + 2\theta_2}, \tag{2.26}$$

$$\iota = \left(\frac{1}{2} + \frac{\theta_1 + \theta_2}{2\delta}\right)x_0 - \frac{v_0}{\delta} - \frac{\theta_1 y + \theta_2 \hat{y}}{2\delta} + \frac{\theta_1 y + \theta_2 \hat{y}}{2\theta_1 + 2\theta_2}, \tag{2.27}$$

$$\zeta = \frac{\theta_1 y + \theta_2 \hat{y}}{\theta_1 + \theta_2}. \tag{2.28}$$

The analysis of particle trajectory is once again split into two groups of cases, one for δ real, and the other for δ complex. The exact cases are discussed in [21].

While both [20] and [21] provided a good overview of the behaviour of the PSO trajectory, they did not provide the necessary or sufficient conditions on the PSO's coefficients which would ensure convergence. This is not surprising, as it was later theoretically proven by Clerc and Kennedy [4] that the PSO using update equations (2.1) and (2.2) could not actually be guaranteed to converge under any coefficient choice, even if velocity clamping is used. This result lead to the development of the constriction coefficients, as discussed in section 2.1.3, in an attempt to guarantee particle convergence. Even though the application of constriction coefficients have become less prevalent, in favour of simply including the inertia weight as discussed in section 2.1.4, its development marks the first practical use of theoretical results with regards to PSO. The analysis done by Clerc and Kennedy [4] utilized the same assumptions as the 1999 paper by Ozcan and Mohan [21].

The theoretical research focus shifted to PSO models that included the inertia weight. There are three primary papers that address the deterministic PSO model with the inclusion of the inertia weight [32, 34, 38]. All of this research was done utilizing the same assumptions as the 1999 paper of Ozcan and Mohan [21]. The work done by Zheng *et al.* [38] was one of first to study the convergence properties of the PSO with inertia weights. The study looked at a multitude of possible cases where convergence would or would not occur, however the results did not provide a general requirement on the coefficients which would guarantee convergence. The work of Van den Bergh and Engelbrecht [33, 34], and Trelea [32] provided a more complete analysis than the findings of Zheng *et al.* [38] by ascertaining general conditions on the PSO's coefficients under which particle convergence would be guaranteed.

The approach utilized by Van den Bergh and Engelbrecht [33, 34] mirrors that of Ozcan and Mohan [21] in a more complex setting. Utilizing update equations (2.1) and (2.12), the second order recurrence relation,

$$x\left(t+1\right) = \left(1 - w - \theta_1 - \theta_2\right)x\left(t\right) - wx\left(t-1\right) + \theta_1 y + \theta_2 \hat{y} \qquad (2.29)$$

is derived, with initial conditions defined as $x\left(0\right) = x_0$, $x\left(1\right) = x_1$, and $x\left(2\right) = \left(1 + w - \theta_1 - \theta_2\right)x_1 - wx_0 + \theta_1 y + \theta_2 \hat{y}$. The following closed form is then derived:

$$x\left(t\right) = k_1 + k_2 \tau_1^t + k_3 \tau_2^2, \qquad (2.30)$$

where

$$\tau_1 = \frac{1 + w - \theta_1 - \theta_2 + \gamma}{2}, \qquad (2.31)$$

$$\tau_1 = \frac{1 + w - \theta_1 - \theta_2 - \gamma}{2}, \tag{2.32}$$

$$\gamma = \sqrt{(1 + w - \theta_1 - \theta_2)^2 - 4w}, \tag{2.33}$$

$$k_1 = \frac{\theta_1 y + \theta_2 \hat{y}}{\theta_1 + \theta_2}, \tag{2.34}$$

$$k_2 = \frac{\tau_2 (x_0 - x_1) - x_1 + x_2}{\gamma (\tau_1 - 1)}, \tag{2.35}$$

$$k_3 = \frac{\tau_1 (x_1 - x_0) + x_1 - x_2}{\gamma (\tau_2 - 1)}. \tag{2.36}$$

An analysis of the long term behaviour of the closed form in equation (2.30) yielded the conditions required to guarantee convergence, specifically,

$$0 < w < 1, \qquad c_1 > 0, \qquad c_2 > 0, \tag{2.37}$$

$$w > \frac{c_1 + c_2}{2} - 1. \tag{2.38}$$

The approach utilized by Trelea [32] deviates from the approach used in earlier theoretical PSO research, in that no closed form is derived. The analysis begins with fixing

$$r_1 = r_2 = \frac{1}{2}, \tag{2.39}$$

and by defining

$$\hat{\theta} = \frac{\theta_1 + \theta_2}{2}, \tag{2.40}$$

$$p = \frac{\theta_1}{\theta_1 + \theta_2} y + \frac{\theta_2}{\theta_1 + \theta_2} \hat{y}. \tag{2.41}$$

This results in the update equations,

$$v(t+1) = wv(t) + \hat{\theta}(p - x(t)), \tag{2.42}$$

and

$$x(t+1) = x(t) + v(t+1). \tag{2.43}$$

Equations (2.43) and (2.42) can be combined into the compact form,

$$\begin{vmatrix} x(t+1) \\ v(t+1) \end{vmatrix} = M \begin{vmatrix} x(t) \\ v(t) \end{vmatrix} + \begin{vmatrix} \frac{c_1 + c_2}{2} \\ \frac{c_1 + c_2}{2} \end{vmatrix} p \tag{2.44}$$

where

$$M = \begin{vmatrix} 1 - \hat{\theta} & w \\ -\hat{\theta} & w \end{vmatrix}. \tag{2.45}$$

Trelea realized that the eigenvalues of the matrix M are of fundamental importance when considering the convergence behaviour of a particle in a PSO. Specifically, the magnitude of the eigenvalues of M must be less than 1 in order to guarantee convergence. This restriction lead to the following conditions for convergence:

$$0 < w < 1, \qquad c_1 > 0, \qquad c_2 > 0, \tag{2.46}$$

$$w > \frac{c_1 + c_2}{4} - 1. \tag{2.47}$$

At first glance there is a discrepancy between the results of Trelea [32] and that of Van den Bergh and Engelbrecht [34]. The difference between equations (2.38) and (2.47) results from a difference in the treatment of the stochastic component. Specifically, the conditions of Van den Berg and Engelbrecht are more restrictive than that of Trelea's, this is seen in figure 2.6 where $c_1 + c_2$ is plotted against w, with the coloured region implying convergence of particles to a point. The reason for the difference is that the conditions of Van den Bergh and Engelbrecht cater for the fact that r_1 and r_2 could in fact equal 1 despite their expected value being 0.5, where Trelea's conditions are strictly derived with $r_1 = r_2 = 0.5$. What this means is that the conditions in equation (2.38) are a guarantee that every iteration will encourage convergence, where as the conditions in equation (2.47) could potential only guarantee that there is at worst a 50 percent chance that any given iteration will encourage convergence.
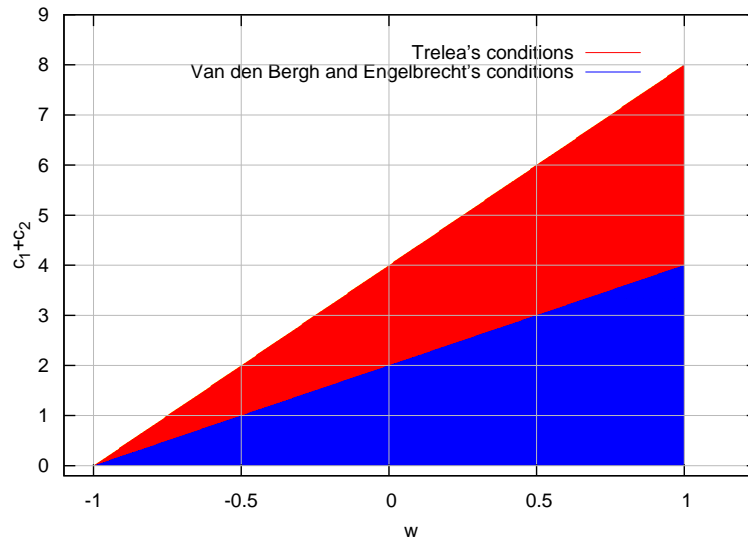

FIGURE 2.6: Parameter Regions in which Each Particle Converges to a Point, Under Van den Bergh and Engelbrecht, and Trelea's Conditions

2.2.2 Stochastic Particle Swarm Model

The theoretical research has at present shifted to almost exclusively analysing PSO convergence with the re-inclusion of the stochastic component.

The first analysis that included the stochastic component was done by Kadirkamanathan *et al.* [11], where the analysis is performed on a PSO model in 1-dimension on the "best" particles with the inclusion of the inertia weight. The best particle stagnation assumption is made. The analysis of particle trajectory is done by modelling particles as a nonlinear feedback controlled system as formulated by Lure, discussed in [5, 36]. Utilizing the passivity theorem [36] and Lyapunov stability [15], the following conditions were derived to guarantee asymptotic stability:

$$|w| < 1, \quad w \neq 0, \tag{2.48}$$

$$c_1 + c_2 < \frac{2\left(1 - 2|w| + w^2\right)}{1 + w}. \tag{2.49}$$

While the conditions of equations (2.48) and (2.49) do guarantee convergence, the parameter restriction is unfortunately very conservative. The conservativeness is a direct result of utilizing Lyapunov stability, which is inherently a conservative approach, implying that there may in fact be a less strict parameter restriction that will guarantee convergence.

Shortly, after the work of Kadirkamanathan *et al.* [11], Jiang *et al.* [10] independently provided a more extensive analysis of the stochastic PSO. The analysis is done in both 1-dimension and k-dimensions on an arbitrary particle under a partial stagnation assumption. The partial stagnation being that $\hat{y}$ is held fixed, and y is allowed to vary in a restricted manner. The conditions derived for convergence are as follows:

$$c_1 \geq 0, \quad c_2 \geq 0, \quad 0 \leq w < 1, \tag{2.50}$$

$$c_1 + c_2 > 0, \quad 0 < h(1) < \frac{c_2^2\left(1 + w\right)}{6}, \tag{2.51}$$

where

$$h(1) = -\left(c_1 + c_2\right)w^2 + \left(\frac{1}{6}c_1^2 + \frac{1}{6}c_2^2 + \frac{1}{2}c_1 c_2\right)w + c_1 + c_2 - \frac{1}{3}c_1^2 - \frac{1}{3}c_2^2 - \frac{1}{2}c_1 c_2 \tag{2.52}$$

It is worth noting that these conditions are not necessary for convergence but only sufficient.

The subsequent research by Martinez *et al.* [16] utilized a novel approach by modelling the particle dynamics as being analogous to that of a damped mass-spring system. The

analysis is done in 1-dimension under the stagnation assumption. Both a continuous and a discrete version of the mass spring system were considered, and the following condition for convergence was derived:

$$|w| < 1, \quad 0 < r_1(t)c_1 + r_2(t)c_2 < 2\left(w + 1\right), \quad \forall t. \tag{2.53}$$

The conditions of equation (2.53) correspond with those of Van den Bergh and Engelbrecht [33, 34] in equations (2.37) and (2.38), as well as the condition of Trelea [32] in equations (2.46), and (2.47). However, the range of coefficients allowed by the conditions of equation (2.53) are wider, due to the inclusion of possible negative c_1, c_2, and w.

There has been further research done under alterations of the definitions of stability, e.g. first order, and second order stability (details can be found in [26]). Most notably is the work of Poli [25, 26]. However no change to the conditions of equation (2.53) was needed.

The most recent contribution to the theoretical PSO research was by Gazi [9], where an analysis in k-dimensions is made on the stochastic PSO dynamics of a best and non-best particle under the stagnation assumption. The work extends the method of Kadirkamanathan *et al.* [11] by finding conditions less conservative than those of the standard Lyapunov stability. The following conditions were derived:

$$|w| < 1, \quad 0 < r_1(t)c_1 + r_2(t)c_2 < \frac{24\left(1 - 2|w| + w^2\right)}{7\left(1 + w\right)}, \quad \forall t \tag{2.54}$$

2.3   Summary

This chapter provided a detailed description of the original PSO, as well as some of the most common PSO alterations in section 2.1. The theoretical research done on PSOs was discussed in great detail in section 2.2, the assumptions made in the theoretical models as well as the outcomes of the theoretical analysis were highlighted.

In the next chapter, issues with the current theoretical models' assumptions are discussed in detail. A new generalized theoretical deterministic PSO model is proposed, under which condition necessary for particles to converge to a point are derived and discussed.

Chapter 3

Theoretical Analysis

This chapter provides a detailed theoretical analysis of the PSO, with a justification for the theoretical model generalization. Conditions for particle convergence to a point are provided under this generalization.

This chapter begins in section 3.1 with a justification for why the theoretical model needs an extension. This is followed by section 3.2 with a brief overview of some elementary mathematical definitions and theorems that will be used in section 3.3. Section 3.2 may be used as a reference and need not be read first. Section 3.3 contains a detailed proof of the PSO particle convergence under the weak chaotic assumption, and an arbitrary social network structure. This is followed by a discussion of the coefficient space where particle convergence will occur in section 3.4, as derived from the results of section 3.3.

3.1 Generalization of the Current Theoretical Particle Swam Model

The theoretical studies of PSO convergence have evolved greatly over the last decade, and is discussed in great detail in section 2.2. Despite the great advances made, a truly accurate analysis of the non-simplified PSO has not yet been made from a theoretical standpoint. This is not surprising as the true PSO contains a large amount of coupling between particles and iterations, making a true model incredibly difficult to work with. All of the theoretical results available to date were derived under the stagnation assumption, with the closest to an exception being the work of Jiang *et al.* [10] as discussed in section 2.2.2, where a slightly weakened version of stagnation is considered.

19

The stagnation assumption is the last remaining simplification still present in the current theoretical PSO models. In many regards it is the most simplifying of all the assumptions. If the "neighbourhood" best $\hat{y}$ of particle i is assumed fixed, it actually decouples all particles in the swarm from each other. This decoupling actually removes the notion of a swarm from the analysis. It has occasionally been stated that theoretical analysis thus far were done in the setting of a star topology [10, 33, 34] (that is using a Gbest PSO). However, the notion of a social network structure is completely meaningless if $\hat{y}_i$ is fixed for each particle. Furthermore, the stagnation assumption removes the true memory that each particle has, specifically if the personal best y of particle i is assumed to be fixed; the "memory" is then fixed. It is clear that a fixed "memory" is a very poor approximation of an actual PSO, where the memory could potentially change after each iteration.

The stagnation assumption in many ways is unrealistic; most fundamentally, a stagnated PSO is no longer an optimizer. What this implies is that even if convergence of a PSO is proved in a stagnated PSO, the convergence of the particles has little practical implication, since if a PSO was in a true stagnant state subsequent iterations of a PSO algorithm would not even alter the best located solution of the swarm, or even that of an individual particle, making the iterations pointless. Now, if the PSO was guaranteed to enter a stagnant state, the analysis would be more meaningful. However, this has never been proved. Realistically, it might not even be possible to prove that a stagnant state will always occur, other than the case where stagnation and convergence are the same (each particle's position has stagnated, not only the personal and neighbourhood best position). It is not hard to imagine a situation where stagnation will not occur, other than the case of convergence.

For example, consider figure 3.1 where the fitness function is simply a parabolic curve in $\mathbb{R}$. If a PSO was used to search the parabolic landscape it would not be surprising for the particles to oscillate around the optimum. It is also very conceivable that on each iteration a particle's current position change causes an improvement in fitness evaluation, but stagnation of the particle never occurs, unless the particle has completely stopped moving (i.e. converged to a point).

It is clear that, given the implications of the stagnation assumption, a theoretical study of the PSO without the assumption is of great importance. This is the primary issue that this thesis addresses. Specifically, the focus is that of the deterministic PSO with the stagnation assumption removed. The assumption is however replaced with a substantially weaker one. Instead of assuming that $y_i(t)$ and $\hat{y}_i(t)$ are fixed, the following assumption is rather made:
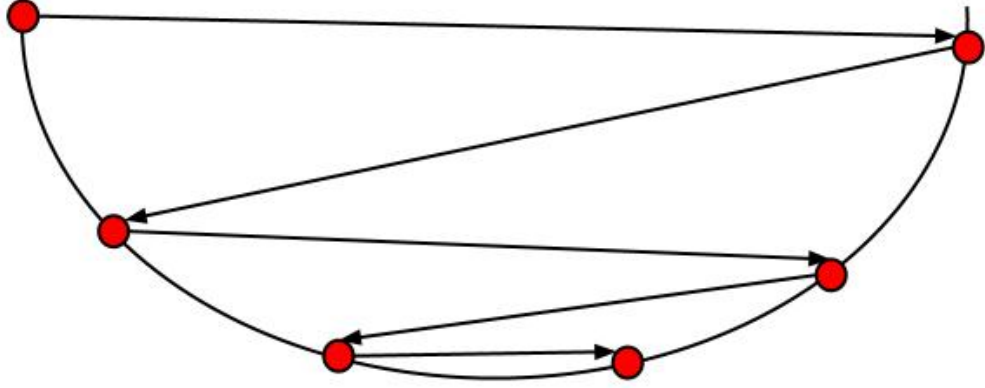
FIGURE 3.1: Particle search on a parabola.

Assumption 3.1.1. **Weak chaotic assumption**: It is assumed that both $y_i(t)$ and $\hat{y}_i(t)$ will occupy an arbitrarily large finite number of unique positions, ψ_i and $\hat{\psi}_i$, respectively.

The weak chaotic assumption provides a closer approximation to an actual PSO, as it allows $y_i(t)$ and $\hat{y}_i(t)$ to change over time, an aspect of the PSO that the stagnation assumption completely removes. The term "chaotic" is used to emphasizes that $y_i(t)$ and $\hat{y}_i(t)$ can change in a seemingling chaotic manner. The term "weak" is used to emphasizes that while the positions of $y_i(t)$ and $\hat{y}_i(t)$ can change in a seemingling chaotic manner, it is not a completely chaotic, as in reality $y_i(t)$ and $\hat{y}_i(t)$ could potentially occupy an infinite number of unique positions.

Working under this assumption, the theoretical analysis provides a meaningful step towards the "grand" PSO model, with "grand" meaning no assumptions. The situation is summarized in figure 3.2. The detailed convergence analysis done under the weak chaotic assumption is presented in section 3.3.

3.2 Preliminary Concepts

For the sake of completeness, a few elementary definitions and theorems are given in this section.

Definition 3.2.1. **Metric space**.

A metric space is a pair (X, d), where X is a set and d is a metric on X. The metric is defined on $X \times X$ such that $\forall x, y, z \in X$ the following hold:

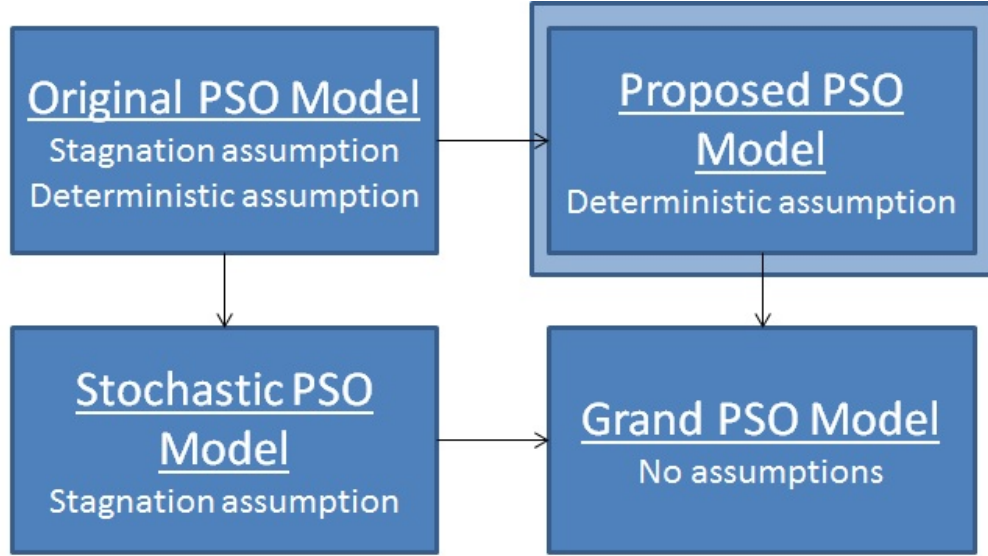1. $d(x, y) \geq 0$, d is real valued and finite.

2. $d(x, y) = 0 \iff x = y$.

FIGURE 3.2: Theoretical PSO Models

3. $d(x, y) = d(y, x)$.

4. $d(x, y) \leq d(x, z) + d(z, y)$.

Definition 3.2.2. **Cauchy sequence**.

A sequence (x_n) in a metric space (X, d) is Cauchy if $\forall \epsilon > 0 \; \exists$ a $G(\epsilon) \in \mathbb{N}$, such that $d(x_m, x_n) < \epsilon \; \forall m, n > G(\epsilon)$.

Definition 3.2.3. **Completeness**.

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

Definition 3.2.4. **Normed Space**.

A normed space $(X, \|\bullet\|)$ is a vector space with a norm defined on it. A norm on a real valued vector space X is a real valued function defined such that $\forall x, y \in X$ and $\forall \alpha \in \mathbb{R}$ the following hold:

1. $\|x\| \geq 0$

2. $\|x\| = 0 \iff x = 0$

3. $\|\alpha x\| = |\alpha| \|x\|$

4. $\|x + y\| \leq \|x\| + \|y\|$

Also, a norm on X defines the metric d on X given by $d(x, y) = \|x - y\|$.

Definition 3.2.5. **Banach Space**.

A Banach space $(X, \|\bullet\|)$ is a complete normed space, where completeness is defined in definition 3.2.3.

Definition 3.2.6. **Spectral radius**.

The spectral radius of a matrix A is defined as $\rho(A) = max_{i=1,\cdots,dim(A)}|\lambda_i|$, where λ_i is the ith eigenvalue of A.

Definition 3.2.7. **Convergent Matrix**.

An $n \times n$ matrix A is said to be convergent if $\lim_{\kappa\to\infty} A^\kappa = \Theta$, where Θ is the zero matrix.

Theorem 3.2.1. **Convergence Of Matrix**.

A is convergent if and only if $\rho(A) < 1$ [1].

Definition 3.2.8. **Affine Mapping**.

The mapping $F : \mathbb{R}^m \to \mathbb{R}^n$ is affine if there is a linear function $L : \mathbb{R}^m \to \mathbb{R}^n$ and a vector $\boldsymbol{b} \in \mathbb{R}^n$ such that $F(\boldsymbol{x}) = L(\boldsymbol{x}) + \boldsymbol{b} \ \forall \boldsymbol{x} \in \mathbb{R}^m$.

Definition 3.2.9. **Contraction**.

Let $X = (X, d)$ be a metric space. A mapping $T : X \to X$ is called a contraction on X if there is a positive real number $\alpha < 1$ such that $\forall \boldsymbol{x}, \boldsymbol{y} \in X$, $d(T\boldsymbol{x}, T\boldsymbol{y}) \le \alpha d(\boldsymbol{x}, \boldsymbol{y})$.

Theorem 3.2.2. **Infinite composition of affine mapping**.

Let $\{G_h : h = 1, \cdots, U\}$ be affine mappings on $\mathbb{R}^k$, where $U \in \mathbb{N}$. If there exists an $s \ge 1$ such that for every $\sigma_1, \cdots, \sigma_s \in \{1, \cdots, U\}$ the composition $G_{\sigma_1} \circ \cdots \circ G_{\sigma_s}$ is a contraction, then for any infinite sequence $\sigma_1, \sigma_2, \cdots \in \{1, \cdots, U\}$ and any $\boldsymbol{z} \in \mathbb{R}^k$, $\lim_{\xi\to\infty} G_{\sigma_1} \circ \cdots \circ G_{\sigma_\xi}(\boldsymbol{z})$ is convergent [17].

3.3 Convergence of the Particle Swarm

It is assumed without loss of generality that the PSO is applied on $\mathbb{R}$, instead of $\mathbb{R}^k$ (this is a reasonable assumption, as the proof that follows is valid regardless of potential dimension dependency, as will become clear during the proof). The original update equations become:

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{3.1}$$

$$v_i(t+1) = wv_i(t) + c_1r_1(t)(y_i(t) - x_i(t)) + c_2r_2(t)(\hat{y}_i(t) - x_i(t)) \tag{3.2}$$

These update equations can be reformulated as the following second order non-homogeneous non-autonomous recurrence relation:

$$x_i(t+1) = (1 + w - (\theta_1(t) + \theta_2(t)))x_t(t) - wx_i(t-1) + \theta_1(t)y_i(t) + \theta_2(t)\hat{y}_i \tag{3.3}$$

where $\theta_1(t) = c_1r_1(t)$ and $\theta_2(t) = c_2r_2(t)$. In the standard PSO model the initial position $x_i(0)$ and velocity $v_i(0)$ are given for $1 < i < N$. In the reformulation of equation (3.3) this is equivalent to the values of $x_i(0)$ and $x_i(1)$ being given for $1 < i < N$.

To simplify the discussion, the following definition of the PSO's state is used: The state S_t of the PSO at time step t is defined as containing all particle positions $x_i(t)$, as well as all the best personal positions $y_i(t)$, in addition to the best position so far within each particle's neighbourhood $\hat{y}_i(t)$ for each $1 < i < N$. More briefly, $S_t = \bigcup_{i=0}^{N}\{(x_i(t), y_i(t), \hat{y}_i(t))\}$.

Under the assumption that $x_i(0)$ and $x_i(1)$ for $1 < i < N$ are given, the initial states of the PSO, S_0 and S_1 are completely known.

Update equation (3.3) can then be reformulated as

$$\begin{vmatrix} x_i(t+1) \\ x_i(t) \end{vmatrix} = \begin{vmatrix} 1 + w - (\theta_1(t) + \theta_2(t)) & -w \\ 1 & 0 \end{vmatrix} \begin{vmatrix} x_i(t) \\ x_i(t-1) \end{vmatrix} + \begin{vmatrix} \theta_1(t)y_i(t) + \theta_2(t)\hat{y}_i \\ 0 \end{vmatrix}. \quad (3.4)$$

It is at this point that the stochastic component is removed by assuming that $\theta_1(t) = \theta_1$, and $\theta_2(t) = \theta_2$ (in effect r_1 and r_2 are fixed), producing the following deterministic form of equation (3.4):

$$\begin{vmatrix} x_i(t+1) \\ x_i(t) \end{vmatrix} = \begin{vmatrix} 1 + w - (\theta_1 + \theta_2) & -w \\ 1 & 0 \end{vmatrix} \begin{vmatrix} x_i(t) \\ x_i(t-1) \end{vmatrix} + \begin{vmatrix} \theta_1 y_i(t) + \theta_2 \hat{y}_i \\ 0 \end{vmatrix}. \quad (3.5)$$

This reformulation allows the calculation of the elements of $S(t+1)$ from those of $S(t)$ and $S(t-1)$. Specifically, the state can be calculated by first calculating particle positions using equation (3.5), and then the remaining information is readily calculated. At this point it is worth noting that the mapping is defined only for a specific particle i, and for a specific $S(t)$ and $S(t-1)$. To make this more clear, equation (3.5) is rather defined as an operator (it is customary not to use parentheses when working with operators) of the following form:

$$\begin{vmatrix} x_i(t+1) \\ x_i(t) \end{vmatrix} = T^i_{(t+1)} \begin{vmatrix} x_i(t) \\ x_i(t-1) \end{vmatrix} = \begin{vmatrix} 1 + w - (\theta_1 + \theta_2) & -w \\ 1 & 0 \end{vmatrix} \begin{vmatrix} x_i(t) \\ x_i(t-1) \end{vmatrix} + \begin{vmatrix} \theta_1 y_i(t) + \theta_2 \hat{y}_i \\ 0 \end{vmatrix}, \quad (3.6)$$

where $T^i_{(t+1)}$ is an operator applied to $[x_i(t), x_i(t-1)]^T$.

Equation (3.6) is shortened to

$$T^i_{(t+1)} \begin{vmatrix} x_i(t) \\ x_i(t-1) \end{vmatrix} = A \begin{vmatrix} x_i(t) \\ x_i(t-1) \end{vmatrix} + \boldsymbol{b}^i_t, \quad (3.7)$$

where

$$A = \begin{vmatrix} 1 + w - (\theta_1 + \theta_2) & -w \\ 1 & 0 \end{vmatrix} \quad (3.8)$$

and

$$b_t^i = \begin{vmatrix} \theta_1 y_i(t) + \theta_2 \hat{y}_i \\ 0 \end{vmatrix}. \tag{3.9}$$

This means that, in order to calculate the position of $x_i(t)$, the operator $T_{(t)}^i \circ T_{(t-1)}^i \circ \ldots \circ T_{(3)}^i \circ T_{(2)}^i$ must be applied to $|x_i(1), x_i(0)|^T$. Let $\|\bullet\|_\infty$ be the max norm on $\mathbb{R}^2$ and let d be the induced metric. Focusing on the actual operators, it is noted, that for arbitrary $\boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^2$,

$$d(T_{(t+1)}^i \boldsymbol{y}, T_{(t+1)}^i \boldsymbol{z}) = \|T_{(t+1)}^i \boldsymbol{y} - T_{(t+1)}^i \boldsymbol{z}\|_\infty = \|A(\boldsymbol{y} - \boldsymbol{z})\|_\infty \leq \|A\|_\infty \|\boldsymbol{y} - \boldsymbol{z}\|_\infty \tag{3.10}$$

If $\|A\|_\infty < 1$, $T_{(t+1)}^i$ is a contractive mapping, then theorem 3.2.2 could be utilized with $s = 1$ (how this fact could have been used will become apparent towards the end of this proof). Unfortunately A cannot be a contractive mapping given its form, as

$$\|A\|_\infty = \max\{|1 + w - (\theta_1 + \theta_2)| + |-w|, |1| + |0|\} \tag{3.11}$$

which clearly cannot be less than 1.

However, since the operator $T_{(t+1)}^i$ is a mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$, there is still a way to utilize theorem 3.2.2 if the spectral radius $\rho(A)$ is less than 1. In order for $\rho(A) < 1$, $\max\{|\lambda_1|, |\lambda_2|\}$ must be less than 1, where $|\lambda_1|$ and $|\lambda_2|$ are the modulus of the eigenvalues values of A. Simple calculation, as given in appendix B, yields the following characteristic polynomial of A,

$$\theta_1 \lambda + \theta_2 \lambda - w\lambda + w + \lambda^2 - \lambda \tag{3.12}$$

from which the following, possibly complex valued eigenvalues are obtained:

$$\lambda_1 = \frac{1}{2}(-\sqrt{(\theta_1 + \theta_2 - w - 1)^2 - 4w} - \theta_1 - \theta_2 + w + 1), \tag{3.13}$$

$$\lambda_2 = \frac{1}{2}(\sqrt{(\theta_1 + \theta_2 - w - 1)^2 - 4w} - \theta_1 - \theta_2 + w + 1). \tag{3.14}$$

If PSO parameters are chosen such that $\max\{|\lambda_1|, |\lambda_2|\} < 1$, applying theorem 3.2.1, the following is known:

$$\lim_{\kappa \to \infty} A^\kappa = \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix}, \tag{3.15}$$

which implies the existence of a K such if $A^K = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ then

$$\max\{|a_{11}| + |a_{12}|, |a_{21}| + |a_{22}|\} < 1. \tag{3.16}$$

Let $\{\xi_1, \xi_2 \ldots, \xi_K\}$ be an arbitrary set, where $\xi_j \in \{1, 2, \ldots, U_i\}$ and $U_i = \psi_i \hat{\psi}_i$ is the number of unique vectors, $\boldsymbol{b_t^i}$. Then, given an arbitrary $\boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^2$,

$$
\begin{aligned}
&d(T_{(\xi_1)}^i \circ T_{(\xi_2)}^i \circ \ldots \circ T_{(\xi_{K-1})}^i \circ T_{(\xi_K)}^i \boldsymbol{y}, T_{(\xi_1)}^i \circ T_{(\xi_2)}^i \circ \ldots \circ T_{(\xi_{K-1})}^i \circ T_{(\xi_K)}^i \boldsymbol{z}) \\
&= \|A^K \boldsymbol{y} + A^{K-1} \boldsymbol{b}_{\xi_K - 1}^i + \ldots + \boldsymbol{b}_{\xi_1 - 1}^i - (A^K \boldsymbol{z} + A^{K-1} \boldsymbol{b}_{\xi_K - 1}^i + \ldots + \boldsymbol{b}_{\xi_1 - 1}^i)\|_\infty \\
&= \|A^K (\boldsymbol{y} - \boldsymbol{z})\|_\infty \\
&= \max\{|a_{11}(y_1 - z_1) + a_{12}(y_2 - z_2)|, |a_{21}(y_1 - z_1) + a_{22}(y_2 - z_2)|\} \\
&\leq \max\{|y_1 - z_1|, |y_2 - z_2|\} \max\{|a_{11} + a_{12}|, |a_{21} + a_{22}|\} \\
&= \|\boldsymbol{y} - \boldsymbol{z}\|_\infty \max\{|a_{11} + a_{12}|, |a_{21} + a_{22}|\} \\
&= d(\boldsymbol{y}, \boldsymbol{z}) \max\{|a_{11} + a_{12}|, |a_{21} + a_{22}|\}.
\end{aligned}
\tag{3.17}
$$

From equation (3.16) and (3.17) it follows that the operator

$$
T_{(\xi_1)}^i \circ T_{(\xi_2)}^i \circ \ldots \circ T_{(\xi_{K-1})}^i \circ T_{(\xi_K)}^i,
\tag{3.18}
$$

is a contractive mapping based on definition 3.2.9, where $\alpha = \max\{|a_{11} + a_{12}|, |a_{21} + a_{22}|\}$. Using theorem 3.2.2, there exists an operator $\hat{T}_i$ such that

$$
\lim_{t \to \infty} T_{(t)}^i \circ T_{(t-1)}^i \circ \ldots T_{(3)}^i \circ T_{(2)}^i \boldsymbol{y} = \hat{T}_i \boldsymbol{y}.
\tag{3.19}
$$

Given equation (3.19) it follows that $\forall \epsilon > 0$ there exists an $M(\epsilon) \in \mathbb{N}$ such that

$$
\|T_{(m)}^i \circ T_{(m-1)}^i \circ \ldots \circ T_{(3)}^i \circ T_{(2)}^i \boldsymbol{y} - T_{(n)}^i \circ T_{(n-1)}^i \circ \ldots \circ T_{(3)}^i \circ T_{(2)}^i \boldsymbol{y}\| \leq \epsilon \ \forall n, m > M(\epsilon). \tag{3.20}
$$

Shifting focus back to the PSO model and observing that for each i the following sequence of particle i's positions are defined,

$$
\begin{aligned}
\hat{x}_i(1) &= \begin{vmatrix} x_i(1) \\ x_i(0) \end{vmatrix}, \quad \hat{x}_i(2) = \begin{vmatrix} x_i(2) \\ x_i(1) \end{vmatrix} = T_{(2)}^i \begin{vmatrix} x_i(1) \\ x_i(0) \end{vmatrix}, \quad \hat{x}_i(3) = \begin{vmatrix} x_i(3) \\ x_i(2) \end{vmatrix} = T_{(3)}^i \begin{vmatrix} x_i(2) \\ x_i(1) \end{vmatrix}, \\
\ldots \quad \hat{x}_i(n) &= \begin{vmatrix} x_i(n) \\ x_i(n-1) \end{vmatrix} = T_{(n)}^i \begin{vmatrix} x_i(n-1) \\ x_i(n-2) \end{vmatrix}
\end{aligned}
\tag{3.21}
$$

using equation (3.20), the sequence $\{\hat{x}_i(n)\}$ is Cauchy, since

$$
\|\hat{x}_i(m) - \hat{x}_i(n)\| = \|T_{(m)}^i \circ T_{(m-1)}^i \circ \ldots \circ T_{(3)}^i \circ T_{(2)}^i \hat{x}_i(0) - T_{(n)}^i \circ T_{(n-1)}^i \circ \ldots \circ T_{(3)}^i \circ T_{(2)}^i \hat{x}_i(0)\| < \epsilon \tag{3.22}
$$

for all $m, n > M(\epsilon)$. Since $\{\hat{x}_i(n)\}$ is Cauchy and $\mathbb{R}^2$ is a Banach space, $\{\hat{x}_i(n)\}$ converges to a point in $\mathbb{R}^2$, say $\hat{x}_i \in \mathbb{R}^2$. Therefore, the sequence $\{x_i(n)\}$ converges to a point say $x_i \in \mathbb{R}$ (as x_i is the first component of $\hat{x}_i$). Since i was arbitrary in $1 \leq i \leq N$, each particle's position converges to a point in $\mathbb{R}$.

3.4 Coefficient Region Guaranteeing Convergence

The results of section 3.3 can be significantly simplified for practical application. As was proved in section 3.3, if $\max\{|\lambda_1|, |\lambda_2|\} < 1$, with λ_1 and λ_2 as defined in equations (3.13) and (3.14), convergence of each particle to a point is guaranteed. If the conditions are violated, the deterministic PSO will not converge. This condition can be simplified by overlaying the inequality plots of $|\lambda_1| < 1$ and $|\lambda_2| < 1$ to the following conditions:

$$|w| < 1, \qquad 0 < \theta_1 + \theta_2 < 4, \qquad w > \frac{(\theta_1 + \theta_2)}{2} - 1. \tag{3.23}$$

The region of parameters that would ensure particle convergence to a point is illustrated in figure 3.3, where $\theta = \theta_1 + \theta_2$.
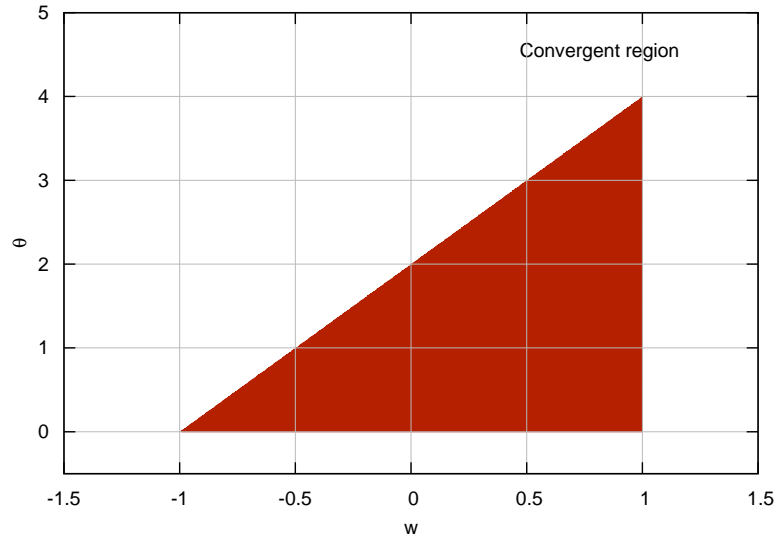

FIGURE 3.3: Parameter Region in which Each Particle Converges to a Point

The conditions presented in equation (3.23) are very positive as they match the conditions of the theoretical research on deterministic PSO done by Van den Bergh and Engelbrecht [33, 34], as well as the work of Trelea [32], just with the inclusion of a possible negative inertia weight. In both of the theoretical studies, r_1 and r_2 are fixed to a specific value in the derivation of the conditions needed for particles to converge to a point. However, a slightly different approach to handling the stochastic component is used in this thesis, namely the stochastic component is left unfixed throughout the analysis. The relation between the research can be seen by replacing θ_1 and θ_2 with

$\frac{1}{2}c_1$ and $\frac{1}{2}c_2$ in equation (3.23) to match Trelea's conditions in equation (2.47). Similarly, by replacing θ_1 and θ_2 with $1c_1$ and $1c_2$ in equation (3.23), Van den Bergh and Engelbrechts' conditions in equation (2.38) are matched.

The conditions of equation (3.23) are left in their most general form, as it provides the most clarity with regards to the influence of the stochastic component. Specifically, it is clear how likely it is for an iteration to encourage convergence (follow the condition of equation (3.23)), for example if $r_1 = r_2 = 0.2$, then an iteration that encourages convergence would at worst only have a 20 percent chance of occurring (if the coefficients used are on the edge of the allowable convergent parameter region). This is obviously not a comment on the long term convergence of particles to a point in the stochastic PSO. It is, however, useful when choosing parameters to have knowledge about the impact that the choice has on the individual iterations.

The similarity between the derived conditions of equation (3.23) and the condition of Van den Bergh and Engelbrecht [33, 34], as well as Trelea [32], is not very surprising. This is because, if the assumption is made that each particle's personal best and neighbourhood's best position are in fact fixed (the stagnation assumption), then in section 3.3 U_i is 1 for all particles, and the same conditions as in equation (3.23) can be derived. This is a clear indication that the proof utilized in section 3.3 is in fact a generalization of a proof needed to prove convergence of particles to a point in the deterministic PSO under the stagnation assumption.

One aspect of the proof utilized in section 3.3, which is also a generalization, is that at no point is a specific fitness function assumed, nor is it assumed that the fitness function is static. What this implies is that even if the fitness function is time dependent (effectively a dynamic environment), the conditions for particle convergence to a point in equation (3.23) remain valid.

3.5   Summary

This chapter provided a detailed justification for the generalization of the theoretical deterministic PSO model in section 3.1. Under the deterministic PSO model generalization, conditions necessary for each particle in the swarm to converge to a point were derived in section 3.3. The derived conditions necessary for convergence to a point were contrasted to the conditions obtained in previous theoretical deterministic PSO research, and were shown to be a generalization of previous findings in section 3.4.

The conditions of equation (3.23) are very useful as they provide guarantees about particle convergence derived under a more complete model. The more complete the

model, the more accurate the conclusions derived from the model will be. The next chapter provides empirical results, to support the theoretical conclusion of this chapter.

Chapter 4

Empirical Analysis

This chapter contains the details of an experiment designed to illustrate the convergence/divergent properties of the PSO under certain parameter choices in relation to the theoretical findings of the previous chapter. The chapter begins with the experimental setup in section 4.1, followed by the obtained results in section 4.2. The chapter ends with a discussion of the results and how they relate to the conditions derived in section 3.4.

4.1 Experimental Set Up

In theoretical research on deterministic PSOs, if any experimental results are presented, they are often generated from the theoretical model itself. The model is however based on a situation where the stagnation assumption is forced. While the results obtained in section 3.3 were derived under the weak chaotic assumption, the experiment performed in this chapter are on an actual PSO with the only simplification being the deterministic assumption, where $\boldsymbol{\theta}_1 = \mathbf{1}c_1$, and $\boldsymbol{\theta}_1 = \mathbf{1}c_2$. This approach provides a better real world approximation, as it includes the impact of the more complicated interactions between particles in the analysis.

There is an inherent difficulty in empirically analysing the convergent behaviour of PSO particles, specifically with regards to understanding the influence of the underlying fitness landscape on the PSO. In an attempt to try and mitigate this issue, a specific fitness function that will make it "hard" for the PSO to become stagnant is used. This type of function is ideal, as it makes analysis of the PSO's convergent properties more difficult, as a non-stagnant PSO is by definition more chaotic. The function used is

$$f(\boldsymbol{x}) = f(\boldsymbol{x}, t) \in U\left(-1000, 1000\right), \tag{4.1}$$

30

where t is the iterative time stamp, and $U(-1000, 1000)$ are uniformly sampled real values between -1000 and 1000. What the fitness function in equation (4.1) provides is an environment that constantly changing in inherently unpredictable manner, as well as being rife with discontinuities (actually, it is discontinuous almost everywhere), resulting in a search space where finding the global optimum is very difficult.

The experiment utilizes the following static parameters:

- **Population Size:** 64.

- **Iterations:** 2000.

- **Search Space:** Dimension $k = 100$.

- **Population initialisation:** Particle's positions are instantiated within the hypercube $(-1000, 1000)^k$. Particle's velocities are instantiated to $\mathbf{0}$.

- **Fitness function:** Equation (4.1) is utilized as the fitness function.

The measure of convergence is as follows:

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

Convergence of each particle to $\mathbf{0}$ will occur if and only if Δ approaches zero.

Four social network structures were utilized in the experiment: star, ring, 2-D Von Neumann, and 3-D Von Neumann social network structures. The choice of a population of size 64 was made to ensure that both the 2-D Von Neumann and the 3-D Von Neumann social network structures are completely formed, i.e. a complete grid and cube respectively. The convergence of the PSO is tested under each of the parameter sets in table 4.1.

TABLE 4.1: Parameter Sets

No	w	c_1	c_2
1	1	2	2
2	0.7298	1.49618	1.49618
3	0.9	1.85	1.85
4	-0.5	0.4	0.4
5	1	0.5	0.5
6	0.9	-0.1	-0.1
7	0.9	2.01	2.01
8	0.5	1.5	1.5

Each of the parameter sets were chosen in such a way as to illustrate the impact that certain parameter choices have on the convergent (to a point) or divergent long term behaviour of the PSO, or to utilize common or historically chosen parameters. The justification for each parameter set is given in section 4.2, along with the detailed results of this experimentation. Each reported experiment is the result of averaging over 50 independent runs (50 random swarm initializations).

4.2 Experimental Results and Discussion

This section presents the results of the experiment as described in section 4.1. For each parameter set in table 4.1, a figure illustrating the variation of Δ as t increases is provided. To ensure that the figures are as informative as possible, two aspects of the figures vary. Firstly, the number of iterations displayed has been chosen to best illustrate the more volatile parts of the convergence behaviour. Secondly, in the cases where radical divergence occurred, Δ has been log scaled in order to avoid later iterations obscuring the results of earlier ones. If a figure contains a log scaled Δ it has been made explicit in the figure's legend.

Parameter set 1 in table 4.1 are the parameters used in the original PSO [13]. This parameter set has known issues as mentioned in section 2.2.1. The parameter set clearly does not satisfy the conditions of equation (3.23); specifically, it fails to satisfy all three of the conditions. The results of running the PSO under this parameter set are illustrated in figure 4.1. The PSO was clearly divergent, with Δ rapidly increasing to the extent where log scaling was needed to even analyse the divergence. While there was a slight difference in the rate of divergence between different social network structures, the divergent property was apparent throughout.

Parameter set 2 was originally proposed by Eberhart and Shi [6]. Checking the conditions of equation (3.23), it is seen that $|0.7298| < 1$, $0 < 2.99236 < 0$, and $0.7298 > 1.49618 - 1 = 0.49618$, so all the conditions for convergence are satisfied. Figure 4.2 shows that convergence did in fact occur, with the PSO using the star social network structure converging the fastest. The long term convergent property of the PSO under parameter set 2 was, however, not effected by the social network structure as is clear from figure 4.2, once again only the rate of convergence seems to be affected.

Parameter set 3 was chosen as it comes very close to violating every condition in equation (3.23), but does not actually violate any. The parameter set results in the evaluation $|0.9| < 1$, $0 < 3.6 < 4$, and $0.9 > 1.8 - 1 = 0.8$, which should ensure convergence. In figure 4.3 convergence of Δ to 0 occurred under all social network structures with the
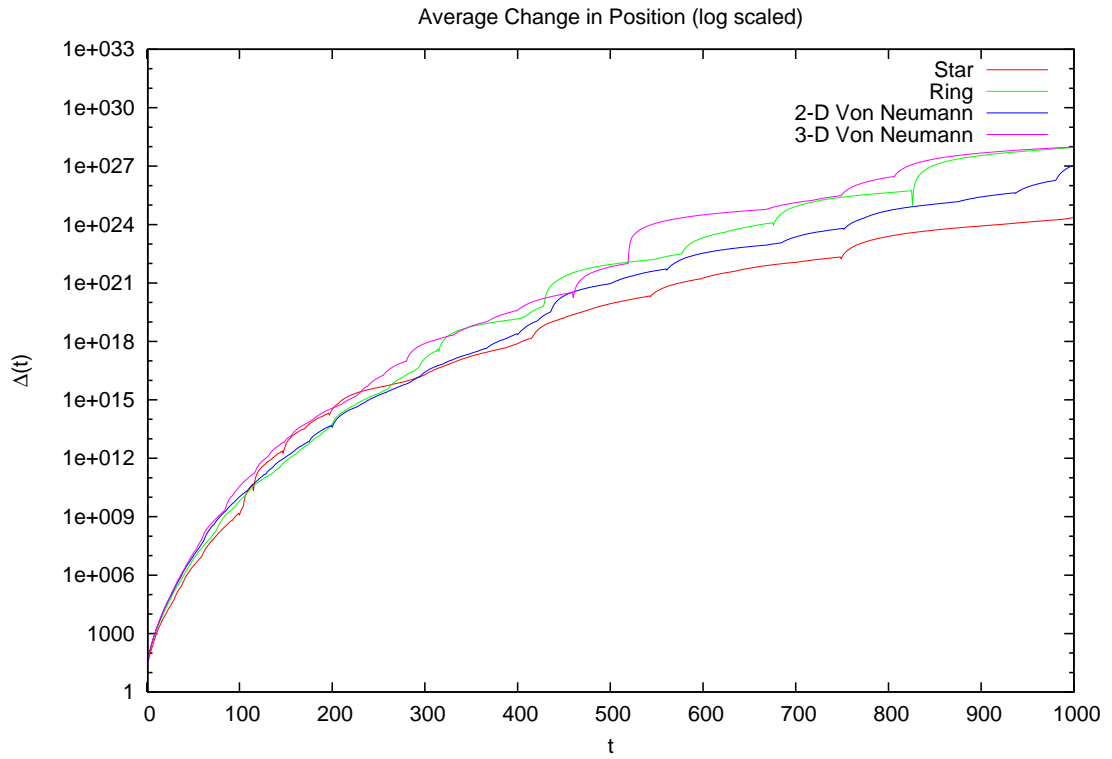
FIGURE 4.1: Parameter set $w = 1$, $c_1 = 2$, and $c_2 = 2$

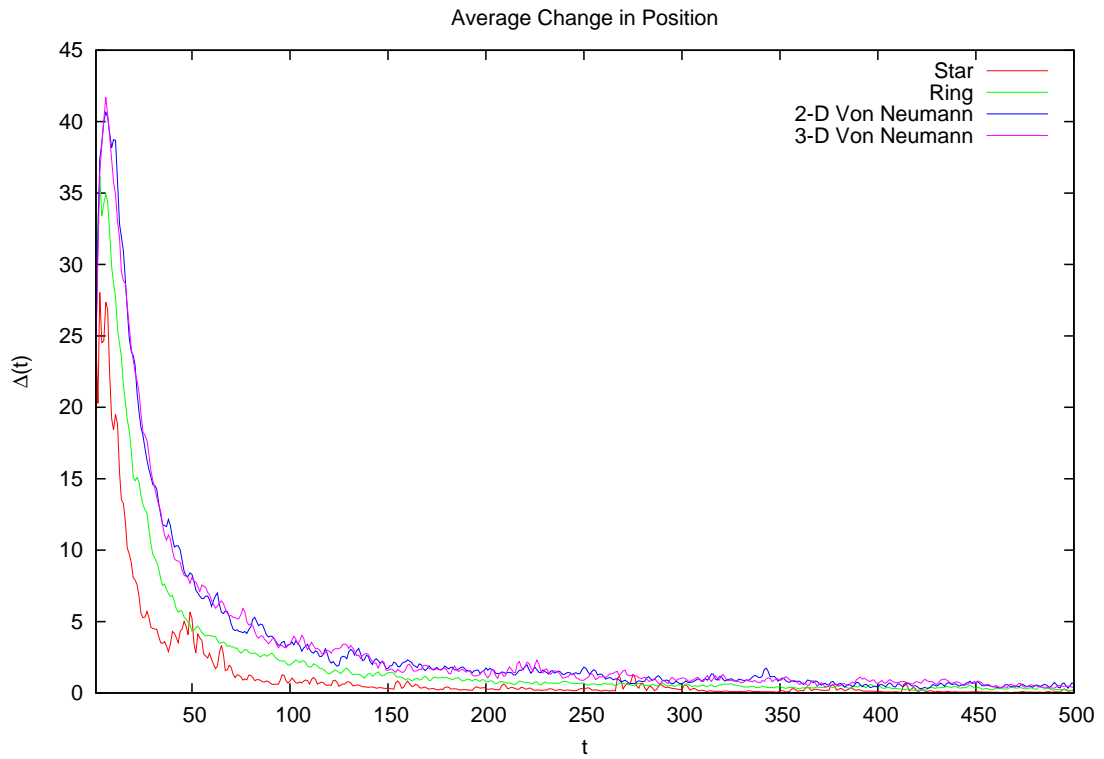FIGURE 4.2: Parameter set $w = 0.7298$, $c_1 = 1.49618$, and $c_2 = 1.49618$

star social network structure providing the quickest and most stable (smallest variations between iterations) convergence to 0, and the 3-D Von Neumann social network structure being the slowest to converge in addition to being the least stable.
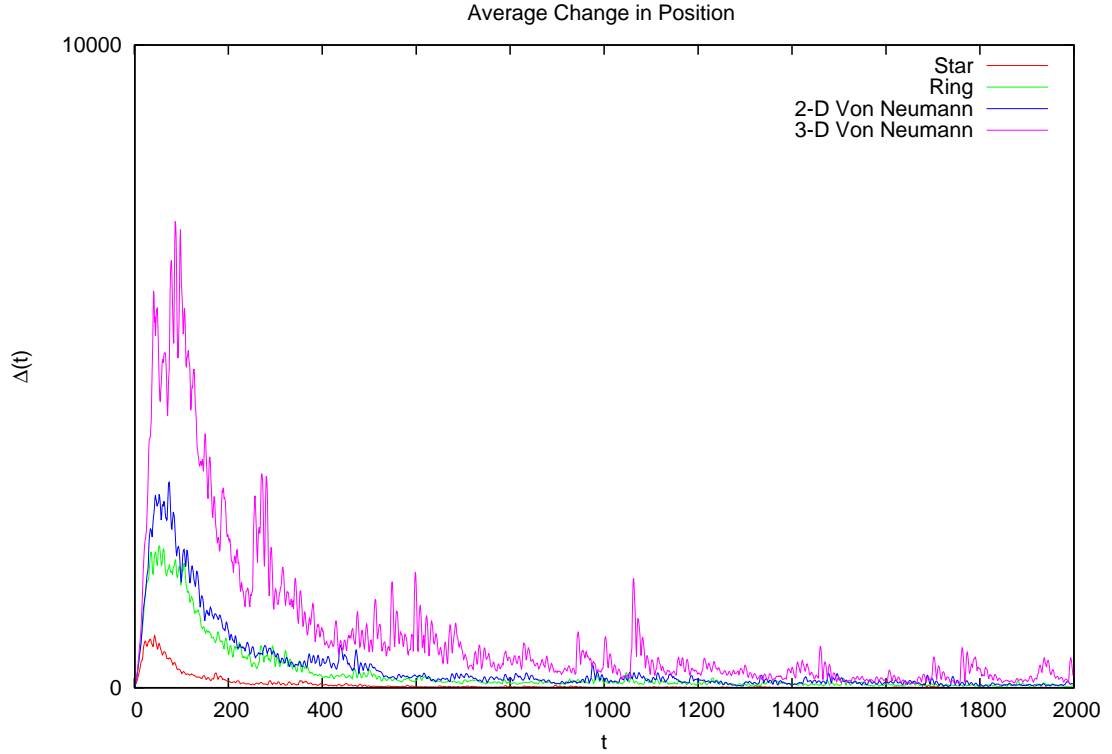

FIGURE 4.3: Parameter set $w = 0.9$, $c_1 = 1.85$, and $c_2 = 1.85$

Parameter set 4 was chosen as it utilizes a negative inertia weight, a case not often considered. The parameter set results in the evaluation $|-0.5| < 1$, $0 < 0.8 < 4$, and $-0.5 > 0.4 - 1 = -0.6$, which should ensure convergence. Once again, convergence of Δ to 0 occurred under all social network structures as can be seen in figure 4.4. Parameter set 4 was the first set under which the star social network structure does not converge in the quickest and most stable manner, in fact the opposite was true. The convergent behaviour of the remaining social network structures were virtually identical under parameter set 4.

Parameter set 5 was chosen as it violates only the first condition of equation (3.23), and is on the boundary of the convergent region of figure 3.4. The parameter set results in the evaluation $|1| \nless 1$, $0 < 1 < 4$, and $1 > 0$, so the theory does not guarantee convergence. In figure 4.5 the value of Δ is clearly not converging to 0 under any social network choice. However, Δ was not necessarily diverging to ∞, as was the case with parameter set 1. The long term behaviour of Δ more closely resembles convergence of Δ to a non-zero constant. While this is a more stable behaviour, it is not necessarily a
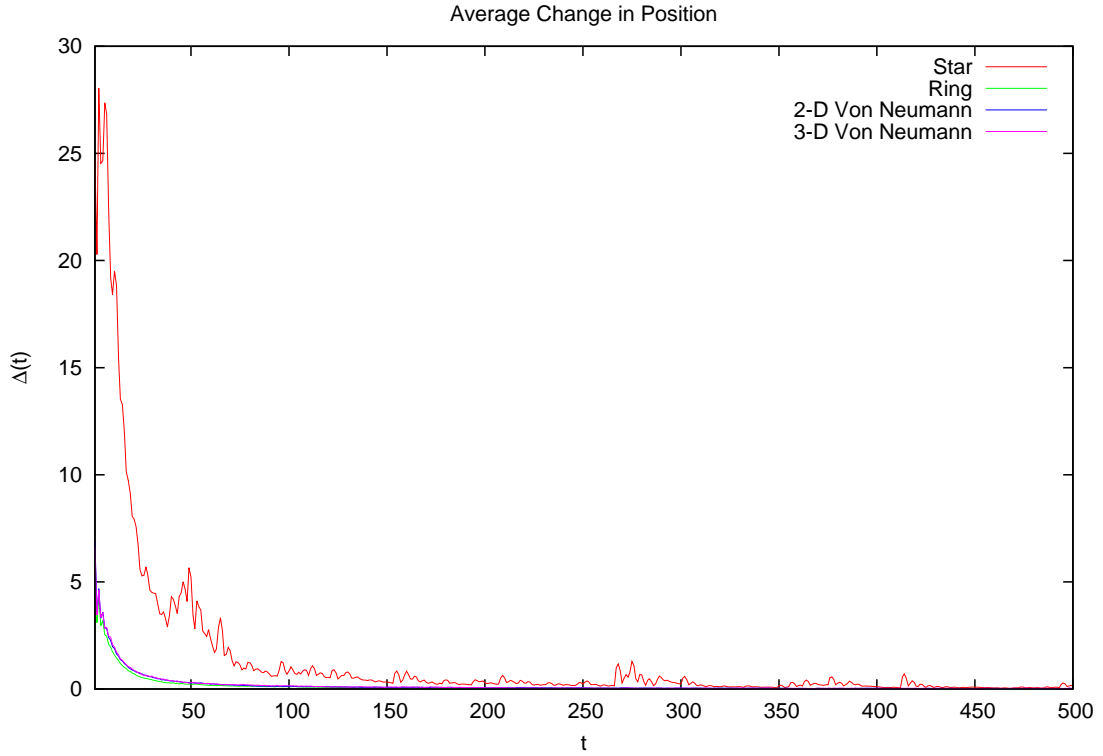
FIGURE 4.4: Parameter set $w = -0.5$, $c_1 = 0.4$, and $c_2 = 0.4$

desirable one, as a constant Δ could still result in a particle travelling infinitely far from the desired search space.

Parameter set 6 was chosen as it violates only condition two. The parameter set results in the evaluations $|0.9| < 1$, $0 \not< -0.2 < 4$, and $0.9 > -1.1$, so the theory does not guarantee convergence. Also, parameter set 7 was chosen as it violates condition two from above, and by implication condition three. The parameter set results in the evaluation $|0.9| < 1$, $0 < 4.02 \not< 4$, and $0.9 \not> 2.1 - 1 = 1.1$, so once again the theory does not guarantee convergence. Under the utilization of both parameter sets 6 and 7, rapid divergence occurred, with Δ causing overflow before even 500 iterations have occurred. This divergent behaviour of Δ occurred regardless of social network choice. The choice of social network structure only appears to slightly affect the rate at which Δ diverges, as is seen in figure 4.6 and figure 4.7.

Parameter set 8 was chosen as it violates only condition three. The parameter set results in the evaluations $|0.5| < 1$, $0 < 3 < 4$, and $0.5 \not> 1.5 - 1 = 0.5$, so the theory does not guarantee convergence. The long term behaviour of Δ can be seen in figure 4.8. Parameter set 8 resulted in one of the more interesting behaviours. All social network structures except the star network structure clearly exhibited near linear divergence. The star social network structure was in fact diverging. The divergence was however, exceptionally slow in contrast to the other social network structures. Parameter set 8
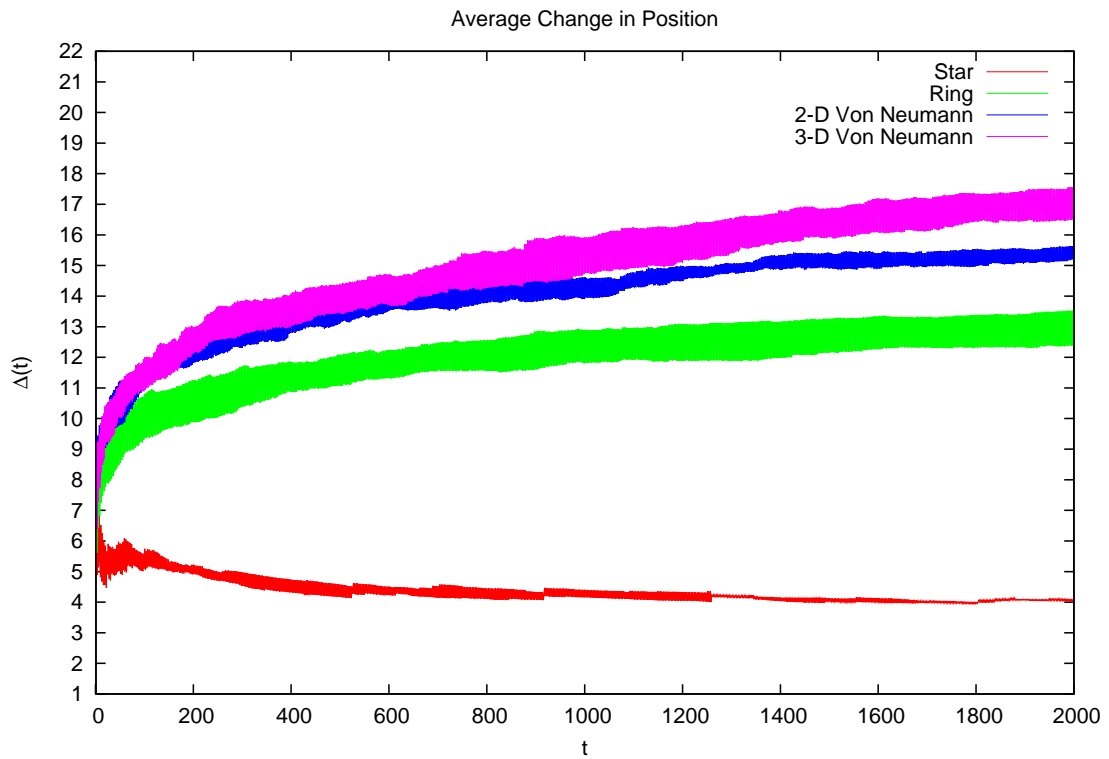
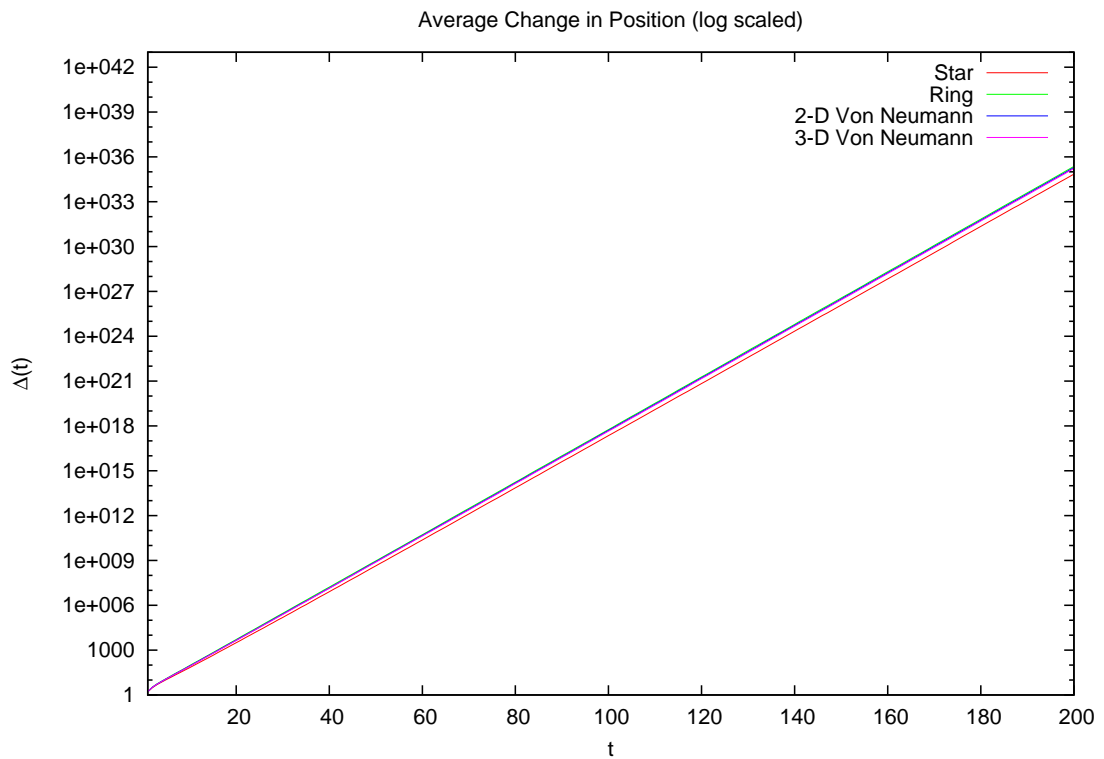FIGURE 4.5: Parameter set $w = 1$, $c_1 = 0.5$,and $c_2 = 0.5$

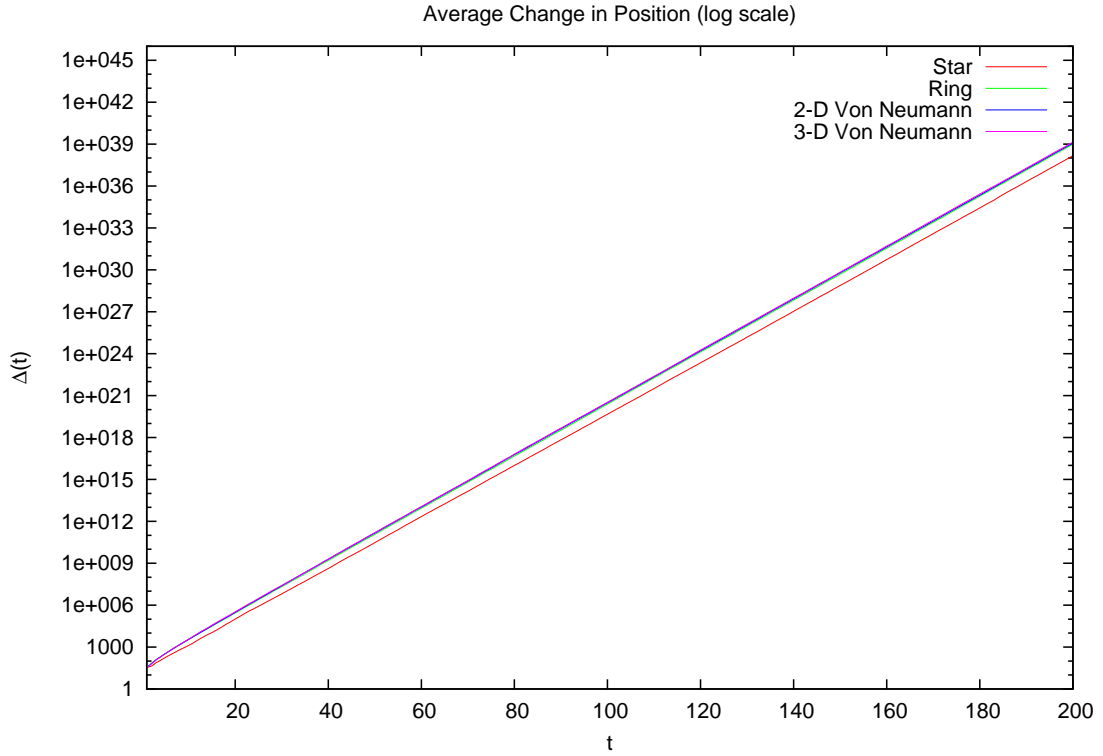FIGURE 4.6: Parameter set $w = 0.9$, $c_1 = -0.1$,and $c_2 = -0.1$

FIGURE 4.7: Parameter set $w = 0.9$, $c_1 = 2.01$, and $c_2 = 2.01$

provided the best indications that the social network choice does in fact have an effect on the rate on divergence. However, the social network choice did not have a large enough effect to cause a convergent Δ.

4.3 Summary

In this chapter the details of an experiment designed to illustrate the convergent/divergent properties of the PSO under certain parameter choices in relation to the theoretical findings of the previous chapter was given. It was found that the PSO behaved in the theoretically expected manner with regards to long term behaviour under all the parameter sets. The results also illustrated that the choice of social network structure does in fact affect the particles' convergence. However the effect is limited to the rate at which convergence or divergence occurred. This conclusion matches the results of the theoretical analysis done in chapter 3, namely that long term convergence will occur if the parameters used satisfy the conditions of equation (3.23) regardless of the social network structure used. This empirical study on the PSO, differs from other studies done, as it directly analyses the effect that the social network structure choice has on the swarms ability to have each particle converge to a point.
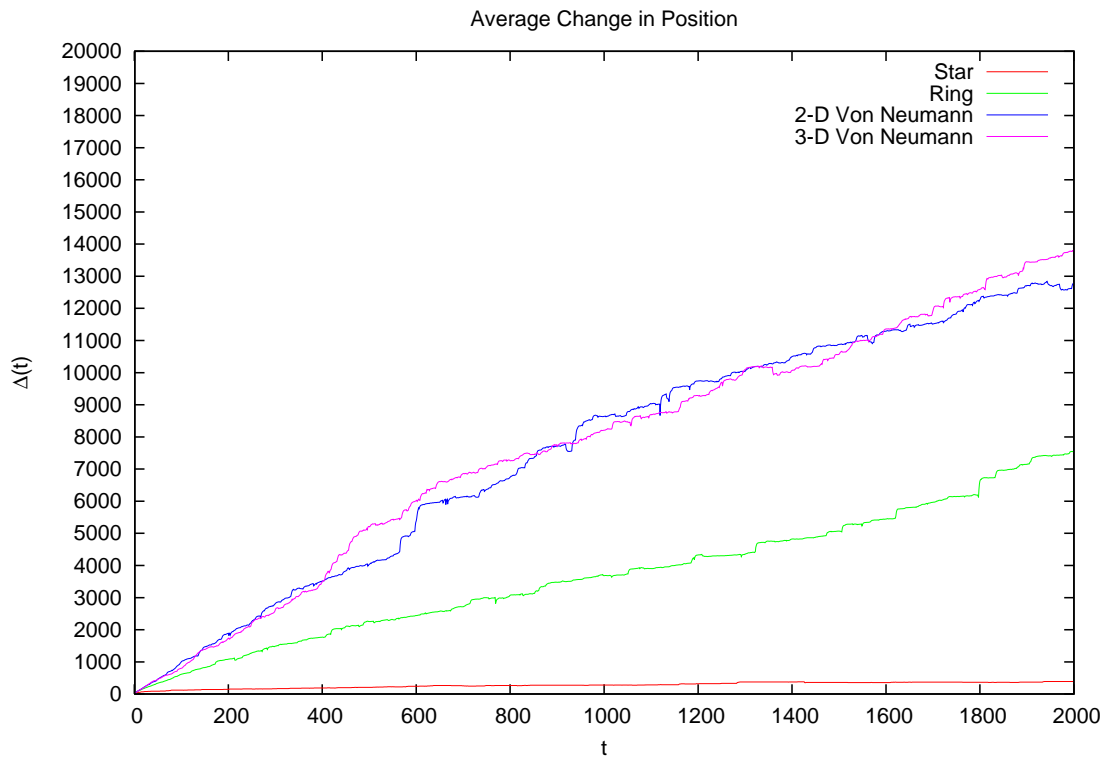
FIGURE 4.8: Parameter set $w = 0.5$, $c_1 = 1.5$, and $c_2 = 1.5$

The next chapter presents an overview of the findings of this thesis, in addition to a discussion of potential future work.

Chapter 5

Conclusion

This chapter begins with a summary of the findings and contributions of this thesis in section 5.1, followed by a discussion of potential future work in section 5.2.

5.1 Overview

This thesis analysed the current theoretical research on PSO, in both the stochastic and deterministic context. The primary focus was on the derivation of conditions necessary for each particle in a swarm to converge to a point. It was found that all theoretical models used in the research had one common shortcoming: all of the models derived their conditions under the stagnation assumption as discussed in chapter 2. While the theoretical PSO models utilized in the research have provided a wealth of insight into the PSO's underlying behaviour, the presence of the stagnation assumption has removed many aspects of the PSO's behaviour.

Chapter 3 begun with a detailed explanation of the aspects of the PSO that the presence of the stagnation assumption had inherently removed from the previous theoretical models. A new theoretical model for the deterministic PSO was proposed, under the weak chaotic assumption, a substantially weaker assumption than the stagnation assumption. Under the proposed model, the conditions necessary for each particle in a swarm to converge to a point were derived. The conditions were derived under an arbitrary social network structure, and an arbitrary swarm size, and under an arbitrary fitness fitness function.

Chapter 4 provided an experiment designed to test the theoretically derived condition for particle convergence to a point. The experiment was run on an actual PSO under multiple social network structures, with only the stochastic component of the update

equations fixed. The empirical findings of the experiment fully supported the theoretical results. Convergence of all particles to a point occurred when the convergence conditions were met, and a failure to converge to a point occurred when the conditions were violated.

In summation, a generalization of the theoretical deterministic PSO was made which better reflected the actual PSO algorithm. Under the new model, conditions to guarantee particle convergence to a point were derived and the influence of parameter choices were illustrated by empirical findings on an array of parameter values.

5.2 Future Work

There are two very clear directions for future work which are directly related to the work of this thesis. The first would be to derive the conditions necessary for each particle in a swarm to converge to a point under the theoretical stochastic PSO model with the weak chaotic assumption, instead of the usual stagnation assumption. The second would be to derive the conditions necessary for each particle in a swarm to converge to a point, without any assumptions on the nature of the personal best and neighbourhood best particles at all. This might not actually be an achievable result, never the less it should be investigated.

One of the more interesting, but less direct areas of further research, is the potential utilization of the more general deterministic PSO model of this thesis to extend the guaranteed convergence PSO (GCPSO)[33], and prove that the GCPSO (or a slight variant of the GCPSO) will converge to local optima under an arbitrary social network structure (at maximum, one unique local optimum for each neighbourhood) instead of the just the one local optimum under the star social network structures.

Appendix A

Symbols

Ω	Particle swarm population.
i	Index of the particle in its population.
$\boldsymbol{x}_i$	Position vector of particle i.
$\boldsymbol{v}_i$	Velocity vector of particle i.
j	Component index of a particle.
N	Population size.
k	Particle dimension.
c_1	Cognitive importance coefficient.
c_2	Social importance coefficient.
w	Inertia weight.
$\boldsymbol{y}_i$	Best position found by particle i.
$\hat{\boldsymbol{y}}_i$	Best position found in particle i's neighbourhood.
$\boldsymbol{r}_1, \boldsymbol{r}_2$	Stochastic variables from the uniformly disturbed hyper cube $U\left(0,1\right)^k$.
$\boldsymbol{\theta}_1$	Shorthand for $\boldsymbol{r}_1 c_1$.
$\boldsymbol{\theta}_2$	Shorthand for $\boldsymbol{r}_2 c_2$.
$\mathcal{N}_i$	Neighbourhood of particle i.
$\chi_{i,j}$	Constriction coefficient of component j of particle i.
κ	A real valued coefficient within the open interval $(0,1)$.

41

Appendix B

Extra Calculations

In section 3.3, the characteristic polynomial of the matrix,

$$A = \begin{vmatrix} 1 + w - (\theta_1 + \theta_2) & -w \\ 1 & 0 \end{vmatrix} \tag{B.1}$$

was needed. The calculation to obtain the characteristic polynomial of matrix A are as follows:

$$
\begin{aligned}
0 &= det(A - \lambda I) & \text{(B.2)} \\
&= det \begin{vmatrix} 1 + w - (\theta_1 + \theta_2) - \lambda & -w \\ 1 & \lambda \end{vmatrix} & \text{(B.3)} \\
&= (1 + w - (\theta_1 + \theta_2) - \lambda)(-\lambda) - (-w)(1) & \text{(B.4)} \\
&= \theta_1 \lambda + \theta_2 \lambda - w\lambda + w + \lambda^2 - \lambda & \text{(B.5)}
\end{aligned}
$$

Bibliography

[1] R.L. Burden and J.D. Faires. *Numberical Analysis, 9th edition*. Cengage Learning, 2010.

[2] M. Clerc. The swarm and the queen: Towards a deterministic and adaptive particle swarm optimization. *In proceedings of the IEEE Congress on Evolutionary Computation*, 3:1951–1957, 1999.

[3] M. Clerc. Think locally, act locally: The way of life of cheap-PSO. *Technical report, http://clerc.maurice.free.fr/pso/*, 2001. Accessed 01/2013.

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

[5] C.A. Dosoer and M. Vidyasagar. *Feedback Systems: Input Output Properties*. Academic Press, 1975.

[6] R.C Eberhart and Y. Shi. Comparing inertia weights and constriction factors in particle swarm optimization. *In Proceedings of the IEEE Congress on Evolutionary Computation*, 1:84–88, 2000.

[7] R.C Eberhart, P.K. Simpson, and R.W. Dobbins. Computational intelligence pc tools. *Academic Press Professional, first edition*, 1996.

[8] H-Y. Fan. A modification to particle swarm optimization algorithm. *Engineering Computations*, 19(8):970–989, 2002.

[9] 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, 2012.

[10] 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*, 102:8–16, 2007.

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

[12] J. Kennedy. Small worlds and mega-minds: effects of neighborhood topology on particle swarm performance. *In Proceedings of the IEEE Congress on Evolutionary Computation*, 3:1931–1938, 1999.

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

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

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

[16] F.J.L. Martinez, G.E. Gonzalo, and F.J.P Alvarez. Theoretical analysis of particle swarmtrajectories through a mechanical analogy. *International Journal of Computational Intelligence Research*, 4(2):93–104, 2008.

[17] L. Mate. On infinite composition of affine mappings. *Fundamenta Mathematicae*, 159:85–90, 1999.

[18] A.S. Mohais. Randomized directed neighborhoods with edge migration in particle swarm optimization. *Congress on Evolutionary Computation*, 1:548–555, 2004.

[19] S. Naka, T. Genji, T. Yura, and Y. Fukuyama. Practical distribution state estimation using hybrid particle swarm optimization. *In IEEE Power Engineering Society Winter Meeting*, 2:815–820, 2001.

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

[21] E. Ozcan and C.K. Mohan. Particle swarm optimization: Surfing the waves. *In Proceedings of the IEEE Congress on Evolutionary Computation*, volume 3, July 1999.

[22] E.S. Peer, F. van den Bergh, and A.P. Engelbrecht. Using neighborhoods with the guaranteed convergence PSO. *In Proceedings of the IEEE Swarm IntelligenceSymposium*, pages 235–242, 2003.

[23] J. Peng, Y. Chen, and R.C. Eberhart. Battery pack state of charge estimator design using computational intelligence approaches. *In Proceedings of the Annual Battery Conference on Applications and Advances*, pages 173–177, 2000.

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

[25] 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.

[26] R. Poli and D. Broomhead. Exact analysis of the sampling distribution for the canonical particle swarm optimiser and its convergence during stagnation. *In Genetic and Evolutionary Computation Conference, London,New York: ACM*, pages 134–141, 2007.

[27] 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, 2003.

[28] J.F. Schutte and A.A Groenwold. Sizing design of truss structures using particle swarms. *Structural and Multidisciplinary Optimization*, 24(4):261–269, 2003.

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

[30] Y. Shi and R.C. Eberhart. Fuzzy adaptive particle swarm optimization. *In Proceedings of the IEEE Congress on Evolutionary Computation*, 1:101–106, 2001.

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

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

[33] F. van den Bergh. An analysis of particle swarm optimizers. *PhD thesis, Department of Computer Science, University of Pretoria, Pretoria, South Africa*, 2002.

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

[35] G. Venter and J. Sobieszczanski-Sobieski. Particle swarm optimization. *Journal for the American Institute of Aeronautics and Astronautics*, 41(8):1583–1589, 2003.

[36] M. Vidyasagar. *Nonlinear Systems Analysis*. SIAM, 2002.

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

[38] Y. Zheng, L. Ma, L. Zhang, and J. Qian. On the convergence analysis and parameter selection in particle swarm optimization. *In the International Conference on Machine Learning and Cybernetics*, 3:1802–1807, 2003.