Mean-Field Techniques in Particle Swarm Optimization

Erin Scratcherd
Supervisor: Prof. Lorenzo Pareschi

Masters Project
MACS
Heriot Watt University
Summer, 2024

Abstract

Particle Swarm Optimization (PSO), is an optimization algorithm famously inspired by bird flocks and fish schools. Recent developments in simpler (CBO) models have shown that with appropriate regularization and Mean-field techniques that Vlasov-Fokker-Planck type equations can be derived which model the dynamics of the CBO model. This is dissertation looks to reproduce some recent applications of these techniques to PSO methods to also obtain deterministic PDEs analogous to the behaviour of the particle system. We then consider how analogous CBO equations can be derived in the small inertia limit and finally consider discretization techniques and perform a numerical simulation to compare the SD-PSO particle model to the MF-PSO PDE.

I would like to thank Lorenzo Pareschi and Sara Grassi for writing a very interesting paper which I have enjoyed reading this summer. Particular thanks to Lorenzo Pareschi for his support and guidance in helping me produce this project.

Contents

Chapter 1

Introduction

The origional Partical Swarm Optimization (PSO) method was inspired by the desire to simulate social behaviour. Building on simulations of synchronized movements in bird flocks and fish schools, Kennedy and Eberhart introduced an update rule which was also inspired by modeling the more abstract notion of human social behaviour. The PSO method belongs to a broader category of meta heuristic algorithms which are inspired by diverse subject areas including physical, biological, ecological and chemical processes as well as principles of swarm intelligence as in our case. Examples of such algorithms include, simulated annealing, genetic algorithms and ant colony optimization [8].

A similar model to the PSO model is the Consensus-Based-Optimization (CBO) method, which is a gradient free model based on an interacting system of particles [3]. This model which simpler than the PSO model is effective at solving complex optimization problems [7][3]. Furthermore, it has been shown in the aforementioned papers that with appropriate regularization, mean-field techniques can be employed to recover deterministic Vlaov-Fokker-Planck type systems, which characterise the dynamics of these CBO systems [11]. One aspect of this project will be to demonstrate how these techniques can be applied to the PSO case to also recover VFP equations as has been previously done [10].

Then we will will endeavour to show how the analogous CBO equations can be obtained, by taking the small inertia limit of the Mean-Field deterministic PDEs [10]. The final aspect of this project is disretize the Mean-Field PDEs and perform numerical experiments in python to compare the dynamics of the Stochastic-Differential system to the Mean-Field system.

The dissertation is organized into the following key sections. Firstly we consider the prototypical PSO algorithms, the original case and it's subsequent improvement in the form of the canonical case. Then we employ regularization and mean-field techniques as have been used in CBO models to the PSO method with and without memory. This allows for the derivation of both the stochastic differential and Mean-Field PSO equations to be obtained as will later be considered numerically. Then we undertake to derive the analogous CBO equations for the Mean-Field Vlasov-Fokker-Plank type Mean-Field equations via the small inertia limit. Finally we state discretization techniques which can be used to model the Mean-Field VFP equations and subsequently show a simulation of the simpler memory-less case.

Chapter 2

Original and Canonical PSO

2.0.1 Introducing the PSO Algorithm

The original PSO method as seen below consists in a spatial update and a velocity update. The velocity update is defined in terms of a combination of a global update and a local update. The global update is comprehensive and finds the best position found by any particle up to the current iteration. The Local best however, is the best position found by an individual particle during it's search up to the current iteration. The significance of the local best is that it allows the particles to explore their local neighborhood around their personal best position, this helps to avoid premature convergence. What specifically encourages exploration in this algorithm are the random matrices R_1^n and R_1^n, which introduce stochasticity into the update rule for both global and local best.

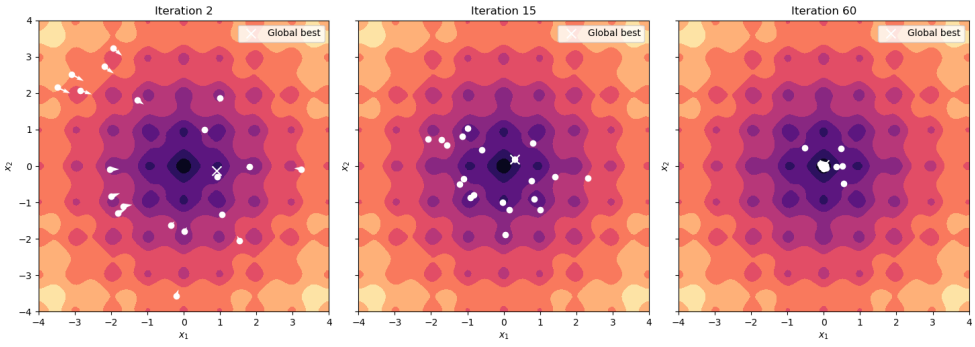

Figure 2.1: Implementation of the PSO-Algorithm (with inertia), where the Ackley function is being minimized by twenty particles. Here are snapshots at three iterations of the dynamic simulation conducted in python.

2.0.2 Original PSO

This iterative update rule below defines the original PSO method. This method initializes N particles with positions $\mathbf{x}_i$ and velocities $\mathbf{v}_i$ in $\mathbb{R}^d$ for $i = 1, \ldots, N$. So for initial positions and velocities x_i^0 and v_i^0 the update rule is as follows,

$$\begin{aligned} x_i^{n+1} &= x_i^n + v_i^{n+1} \\ v_i^{n+1} &= v_i^n + c_1 R_1^n (y_i^n - x_i^n) + c_2 R_2^n (\bar{y}^n - x_i^n) \end{aligned} \tag{2.1}$$

Where y_i^n and $\bar{y}^n$ represent the local and global updates terms respectively as discussed above. c_1 and c_2 are acceleration coefficients, which allow for a weighted choice of the global and local updates to be included in the overall update. The terms R_1^n and R_2^n refer to two d-dimensional diagonal matrices where where the diagonal entries are random numbers uniformly distributed on the interval $[0, 1]$ where the random numbers are generated anew each iteration.

In the original PSO method it is also necessary to set hard bounds for the individual particles to achieve reasonable convergence so that the velocity does not become unbounded. These usually take the form of $[x_{min}, x_{max}]$ and $[-v_{max}, v_{max}]$ [17].

One way the global update can be defined is as follows,

$$\begin{aligned} y_i^0 &= x_i^0, \\ y_i^{n+1} &= \begin{cases} y_i^n & \text{if } \mathcal{F}(x_i^{n+1}) \geq \mathcal{F}(x_i^n), \\ x_i^{n+1} & \text{if } \mathcal{F}(x_i^{n+1}) < \mathcal{F}(x_i^n), \end{cases} \\ \bar{y}^0 &= \arg\min \left\{ \mathcal{F}(x_1^0), \mathcal{F}(x_2^0), \ldots, \mathcal{F}(x_N^0) \right\}, \\ \bar{y}^{n+1} &= \arg\min \left\{ \mathcal{F}(x_1^{n+1}), \mathcal{F}(x_2^{n+1}), \ldots, \mathcal{F}(x_N^{n+1}), \mathcal{F}(\bar{y}^n) \right\}. \end{aligned}$$

2.0.3 Stochastic-Differential PSO

The time-continuous version of the discrete update rule for the PSO model can be found by translating the discrete updates into differential equations that describe the time continuous evolution of the position and velocity of the particles from the original method. To achieve this we re-write the discrete random matrices R_1^n, R_2^n with Brownian motions and the discrete time steps with continuous time increments.

Let $\bar{R}_k^n = 2R_k^n - 1$, for $k = 1, 2$. Then the previous result can be transformed into the following:

$$\begin{aligned} x_i^{n+1} &= x_i^n + v_i^{n+1} \\ v_i^{n+1} &= v_i^n + \frac{c_1}{2}(y_i^n - x_i^n) + \frac{c_2}{2}(\bar{y}^n - x_i^n) \\ &\quad + \frac{c_1}{2}\bar{R}_1(y_i^n - x_i^n) + \frac{c_2}{2}\bar{R}_2(\bar{y}^n - x_i^n) \end{aligned}$$

Where here the velocity update rule is modelled using an SDE approach written in Euler-Maruyama form and the spatial update is the implicit Euler scheme with $\Delta t = 1$.

The Euler-Maruyama Scheme:

$$v_{n+1} = v_n + \Delta t \cdot f(v_n) + g(v_n)\Delta W_n,$$

With drift and diffusion terms,

$$\Delta t \cdot f(v_i^n) = \Delta t \left(\frac{c_1}{2}(y_i^n - x_i^n) + \frac{c_2}{2}(\bar{y}^n - x_i^n) \right)$$
$$g(v_i^n) \cdot \Delta W_n = \frac{c_1}{2}\bar{R}_1(y_i^n - x_i^n) + \frac{c_2}{2}\bar{R}_2(\bar{y}^n - x_i^n)$$

The random noise in this model should have the same mean and variance as has been calculated below to preserve the dynamics of the system. The choice of distribution does not matter here so so long as this condition is met [10].

Since $R_{i,j} \sim \text{Uniform}(0,1)$, for a given diagonal element, the mean is 0 and the variance is $\frac{1}{3}$.

Since $\bar{R}_k^n = 2R_k^n - 1$, for $k = 1, 2$, the statistics for this system become:

$$\mathbb{E}[\bar{R}_k] = \mathbb{E}[2R_k - 1] = 2\mathbb{E}[R_k] - 1 = 0$$

$$\text{Var}(\bar{R}_k) = \text{Var}(2R_k - 1) = 2^2\text{Var}(R_k) = \frac{1}{3}$$

This can be written as a second order system of Stochastic Differential Equations (SDEs) in Ito form:

$$\begin{aligned} dX_i^{n+1} &= V_i^n \, dt \\ dV_i^{n+1} &= \lambda_1(Y_t^i - X_t^i) \, dt + \lambda_2(\bar{Y}_t - X_t^i) \, dt \\ &\quad + \sigma_1 D(Y_t^i - X_t^i) \, dB_t^{1,i} + \sigma_2 D(\bar{Y}_t - X_t^i) \, dB_t^{2,i} \end{aligned} \tag{2.2}$$

Where

$$\lambda_k = \frac{c_k}{2}, \quad \sigma_k = \frac{c_k}{2\sqrt{3}}, \quad k = 1, 2.$$

And

$$D(X_t) = \text{diag}\{(X_t)_1, (X_t)_2, \cdots, (X_t)_d\}.$$

Consider a d-dimensional matrix. The vectors B_t^k are defined as follows:

$$B_t^k = \left((B_t^k)_1, (B_t^k)_2, \cdots, (B_t^k)_d\right)^T \quad \text{for } k = 1, 2.$$

These vectors represent d independent 1-dimensional Brownian motions. Specifically, each $B_{tk}^{(i)}$ for $i = 1, 2, \ldots, d$ is a Brownian motion associated with the i-th particle.

It has been noted that the local best position can be considered a state of the system because it has a memory effect. The evolution of the local best the following is suggested,

$$\dot{X}_{lb} = a(X - X_{lb}) \left[I_n + \text{diag}\left(\text{sign}\left(F(X_{lb}) - F(X)\right)\right) \right]$$

Where this equation approximates the update process of the memory term by considering the current position and the objective function values. It is defined in such a way that if the current position X_i is worse than the previous position then the local position remains unchanged. The equation in this case has a zero derivative. If the current particle position is better ($F(X_i) < F(X_{LB,i})$)

then the local best moves towards the current position X_i when updated. This approximation searches for significant improvements to the local best to minimise unnecessary updates [6].

The Particle Swarm Optimization (PSO) memory update can be written as:

$$y_i^{n+1} = y_i^n + \frac{1}{2}(x_i^{n+1} - y_i^n)S(x_i^{n+1}, y_i^n),$$

where the function $S(x, y)$ is defined as:

$$S(x, y) = 1 + \text{sign}(\mathcal{F}(y) - \mathcal{F}(x)).$$

In the context of a differential system, the memory variable update is represented by:

$$dY_t^i = \upsilon(X_t^i - Y_t^i)S(X_t^i, Y_t^i)\, dt,$$

where the global best position is given by:

$$\bar{Y}_t = \text{argmin}\left\{\mathcal{F}(Y_t^1), \mathcal{F}(Y_t^2), \ldots, \mathcal{F}(Y_t^N)\right\}.$$

The initial condition is:

$$Y_0^i = X_0^i.$$

[10]

2.0.4 SD-PSO Model in 'Canonical form'

It was later observed by Shi and Eberhart that the 'canonical PSO' algorithm can be further improved by the inclusion of an inertial weight in the velocity update [18]. The interia weight plays the important role of balancing the global and local search to allow for improved convergence of the method. Without the use of an inertia term, the choice of $\lambda_k = 1$ and $\sigma_k = \frac{1}{\sqrt{3}}$ can often lead to unstable dynamics in particle velocity [10]. The inertia way is more reliable way to modulate the dynamics than use of bound such as $[-V_{max}, V_{max}]^d$ which can be challenging to determine [17].

A high inertia weight encourages particles to explore the search space preventing them from becoming prematurely stuck at local optima. A low intertia weight however makes the particles toward their current velocities and best known positions both globally and locally. For example, it has been noted in the literature that it is beneficial to shift from a high to a low inertia weight when implementing the PSO method since the dynamic weight change allows for better convergence [17].

Let $m \in (0, 1]$ be the inertia weight. Then, we can incorporate this inertia term into the original PSO method,

$$\begin{aligned} x_i^{n+1} &= x_i^n + v_i^{n+1} \\ v_i^{n+1} &= \mathbf{m}v_i^n + c_1 R_1^n(y_i^n - x_i^n) + c_2 R_2^n(\bar{y}^n - x_i^n) \end{aligned} \tag{2.3}$$

Then we can simply rewrite the previous expression in the following manner,

$$\begin{aligned} x_i^{n+1} &= x_i^n + v_i^{n+1} \\ mv_i^{n+1} &= mv_i^n + c_1 R_1^n (y_i^n - x_i^n) + c_2 R_2^n (\bar{y}^n - x_i^n) - (1-m) v_i^{n+1} \end{aligned}$$

Now we can obtain the the second order Stochastic Differential PSO model with inertia. Where $\gamma = (1-m) \geq 0$. is known as the inertia term.

Thus, we can finally obtain the SDE system,

$$\begin{aligned} dX_i^{n+1} &= V_i^n \, dt \\ mdV_i^{n+1} &= -\gamma V_i^n \, dt + \lambda_1 (Y_i^n - X_i^n) \, dt + \lambda_2 (\bar{Y}_i - X_i^n) \, dt \\ &\quad + \sigma_1 D (Y_i^n - X_i^n) \, dB_t^{1,i} + \sigma_2 D (\bar{Y}_i - X_i^n) \, dB_t^{2,i} \end{aligned} \tag{2.4}$$

Chapter 3

Deriving the Stochastic Differential PSO and Mean-Field PSO equations

3.0.1 Motivation

A similar model to the PSO model is the Consensus-Based-Optimization (CBO) method, which is a gradient free model based on an interacting system of particles. This model has a deterministic and stochastic component which drives particles towards their weighted average and to roam the search space due to the effects of multiplicative noise [3].

For N particles with position X_i for $i = 1, \cdots N$ and Heaviside regularization H^ϵ, the system takes the following form,

$$dX_i = -\lambda(X_i - \bar{X}^\alpha)H^\epsilon(L(X_i) - L(\bar{X}^\alpha))dt + \sigma|X_i - \bar{X}^\alpha|dW_i \tag{3.1}$$

For a weighted average $\bar{X}^\alpha$ defined,

$$\bar{X}^\alpha = \frac{1}{\sum_{i=1}^N e^{-\beta L(X_i)}} \sum_{i=1}^N X_i e^{-\beta L(X_i)} \tag{3.2}$$

In the previous SD-PSO system, we implemented a global update which iterativley adjusted the particle velocities according to the best known position of the swarm up to that iteration . However we would now like to be able characterise the dynamics of the Stochastic Differential PSO model in the form of a deterministic PDE. In order to do this, we need to regularize the global best update in a specific way. Here we can adopt the same regularization process as is being used in CBO methods as shown in (3.2) so that the mean-field limit can be passed and a VFP equation can be arrived at [1][3][7] [16].

To begin with, we can effectively 'remove' the memory term so we have a simplified instantaneous dynamic of particles for which we can later pass the mean-field limit. Below we define the regularised stochastic differential system.

$$\begin{aligned} dX_t^i &= V_t^i\, dt, \\ mdV_t^i &= -\gamma V_t^i\, dt + \lambda(\bar{X}_t^\alpha - X_t^i)\, dt + \sigma D(\bar{X}_t^\alpha - X_t^i)\, dB_t^i. \end{aligned} \tag{3.3}$$

To motivate this regularization, which effectively acts as a new global best we consider the weights as are defined in the CBO case. For large enough α, the weights of the particle becomes concentrated around points which minimise $F(X_t^i)$ to the greatest extent.

To illustrate this point, consider the weight defined: $\omega_\alpha(X_i^t) = e^{-\alpha F(X_i^t)}$, then the concentration of particles can be seen in the example below

Example 3.0.1. Consider 5 particles with positions X_i^t:

$$\begin{aligned} X_1^t = 1, \quad X_2^t = 1.5, \quad X_3^t = 2, \\ X_4^t = 2.5, \quad X_5^t = 3. \end{aligned}$$

Where the function to be minimised is: $F(x) = (x-2)^2$.
Compute $F(X_i^t)$:

$$\begin{aligned} F(X_1^t) &= (1-2)^2 = 1, \\ F(X_2^t) &= (1.5-2)^2 = 0.25, \\ F(X_3^t) &= (2-2)^2 = 0, \\ F(X_4^t) &= (2.5-2)^2 = 0.25, \\ F(X_5^t) &= (3-2)^2 = 1. \end{aligned}$$

Compute weights $\omega_\alpha(X_i^t) = e^{-\alpha F(X_i^t)}$:

- For $\alpha = 1$:

$$\begin{aligned} \omega_1(X_1^t) &= e^{-1} \approx 0.368, \\ \omega_1(X_2^t) &= e^{-0.25} \approx 0.779, \\ \omega_1(X_3^t) &= 1, \\ \omega_1(X_4^t) &= e^{-0.25} \approx 0.779, \\ \omega_1(X_5^t) &= e^{-1} \approx 0.368. \end{aligned}$$

- For $\alpha = 10$:

$$\begin{aligned} \omega_{10}(X_1^t) &= e^{-10} \approx 4.5 \times 10^{-5}, \\ \omega_{10}(X_2^t) &= e^{-2.5} \approx 0.082, \\ \omega_{10}(X_3^t) &= 1, \\ \omega_{10}(X_4^t) &= e^{-2.5} \approx 0.082, \\ \omega_{10}(X_5^t) &= e^{-10} \approx 4.5 \times 10^{-5}. \end{aligned}$$

As α increases, the weights for the particles which do not minimize the function becomes asymptotically smaller so that there is a clustering around the particles which minimize the function to the greatest extent.

3.0.2 Discrete Weighted Average

To make a reasonable estimate for the weighted average we need to consider the Laplace Principle.

Definition 3.1. *The Laplace Principle:*
In general, for a Lebesgue-measurable subset A of a Euclidean space $\mathbb{R}^d$ with a measurable function $F(x) : \mathbb{R}^d \to \mathbb{R}$, we have:

$$\int_A e^{-F(x)}\, dx < \infty.$$

We get:

$$\lim_{\alpha\to\infty} \frac{1}{\alpha} \log \int_A e^{-\alpha F(x)}\, dx = -ess \inf_{x\in A} F(x),$$

As in [5].
In our context this becomes,

$$\lim_{\alpha\to\infty} (-\frac{1}{\alpha} \log(\int_{\mathbb{R}^d} e^{-\alpha F(x)} d\rho(x))) = \inf_{x\in \sup(\rho)} F(x). \tag{3.4}$$

From the Laplace principle, we get that the following integral decays rapidly unless $F(x)$ is minimized:

$$I(\alpha) = \int_{\mathbb{R}^d} e^{-\alpha \mathcal{F}(X)}\, d\rho(x)$$

Based on this we would like to describe a weighted empirical distribution, which is a discrete form of this integral. The discrete weighted average below describes a new way to estimate the minimiser of a given function $F(x)$ in the context of PSO methods. It is defined as follows,

$$\bar{X}_t^\alpha = \frac{1}{N^\alpha} \sum_{i=1}^{N} X_t^i \omega_\alpha(X_t^i), \quad N^\alpha = \sum_{i=1}^{N} \omega_\alpha(X_t^i), \quad \omega_\alpha(X_t^i) = e^{-\alpha F(X_t)}. \tag{3.5}$$

We have now reformulated the Stochastic Differential PSO method into a new form based the global best defined in terms of the weighted average,

$$\begin{aligned} dX_t^i &= V_t^i\, dt, \\ m dV_t^i &= -\gamma V_t^i\, dt + \lambda(\bar{\boldsymbol{X}}_t^\alpha - X_t^i)\, dt + \sigma D(\bar{\boldsymbol{X}}_t^\alpha - X_t^i)\, dB_t^i. \end{aligned} \tag{3.6}$$

3.0.3 Definitions

We will first discuss some concepts and definitions necessary for the derivation of the deterministic PDE.

To obtain a continuous approximation of the collective distribution of the particles we want model the evolution of the particle density. Therefore, it is necessary to consider N particle probability density for a system of particles.

The N particle probability density describes the probability the ith particle will be in position x_i with velocity v_i at time t.

$$f^{(N)}(x_1, \cdots, x_N, v_1, \cdots, v_N, t)$$

We next consider the dynamics of the first marginal, which gives the probability density of finding a given particle with position x_1 with velocity v_1 at a given time and integrate out the other remaining terms.

$$f_1^{(N)}(x_1, v_1, t) = \int f^{(N)}(x_1, \cdots, x_N, v_1, \cdots, v_N, t) d\Omega_1,$$

where $d\Omega_1 = dx_2 \cdots dx_N dv_2 \cdots dv_N$.

Definition 3.2. *(Propagation of chaos assumption).* For an interacting system of particles, if we take $N \gg 1$ sufficiently large then the system approximatley behaves as if it were a system of i.i.d particles [12].

The name *'Propagation of chaos assumption'*, comes from the fact that starting from a set of initial conditions, the system evolves to a state of effective independence among particles in the large N limit, i.e the particles behave 'chaotically'.

For large N we have that

$$f^{(N)}(x_1, \cdots, x_N, v_1, \cdots, v_N, t) \approx \prod_{i=1}^{N} f^{(1)}(x_i, v_i, t)$$

That is the distributions of each of the random pairs are approximately i.i.d with distribution $f(x, v, t)$ for sufficiently large N.

Definition 3.3. *(Convergence in Law of Empirical Measures)*

An empirical measure μ_N converges to a probability measure P in law if the following holds,

$$\mu_N \xrightarrow{d} P$$

.

Where P is a probability distribution that describes the limiting behaviour of the empirical distribution for large N [11].

Example 3.0.2. For example if we chose $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, then for large N we get,

$$Y_n \xrightarrow{d} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Definition 3.4. *(Convergence of the first moment)*

Provided we have convergence in law, then we also have convergence of the first moment of $\mu_N(t)$ for a bounded continuous function ϕ and a particle in space and time $Z_i(t)$,

$$\frac{1}{N}\sum_{i=1}^{N} \phi(Z_i(t)) \to \int_E \phi(z) f_t(Z) dz$$

[11]

Where in this case the Empirical measure is,

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{Z_i}(t) \tag{3.7}$$

Where $Z_i(t) = (X_i(t), V_i(t))$ denotes the position and velocity of the i-th particle at time t [11].

3.0.4 Modelling Particle Density Dynamics

In order to derive the mean-field equations for the microscopic system we make the propagation of chaos assumption on the first marginal so that $f^{(N)} \approx f^{\otimes N}$. Then in this system $N \gg 1$ is sufficiently large and the particles share the same distribution $f(x, v, t)$ [10].

Then we arrive at the following,

$$\frac{1}{N^\alpha}\sum_{i=1}^{N}\omega_\alpha(X_t^i) \approx \int_{\mathbb{R}^d}\omega_\alpha p(x,t)\,dx, \quad \frac{1}{N^\alpha}\sum_{i=1}^{N}X_t^i\omega_\alpha(X_t^i) \approx \int_{\mathbb{R}^d}X_t^i\omega_\alpha p(x,t)\,dx. \tag{3.8}$$

As a result of the law of large numbers we can now consider the weighted average in it's continuous form where, $\bar{X}_t^\alpha \approx \bar{X}^\alpha[\rho]$. Now that we have independence of the random pairs we obtain the Mc-Kean nonlinear process where now,

$$\begin{aligned} d\bar{X}_t &= \bar{V}_t\,dt, \\ m d\bar{V}_t^i &= -\gamma\bar{V}_t\,dt + \lambda(\bar{X}^\alpha[\rho] - \bar{X}_t)\,dt + \sigma D(\bar{X}^\alpha[\rho] - X_t)\,dB_t[16]. \end{aligned} \tag{3.9}$$

Definition 3.5. (Macroscopic System):

For $N >> 1$ sufficiently large we get an approximatly $f(x, v, t)$ distributed system of random pairs $(x_1, v_1), \cdots (x_N, v_N)$. Now we get a Mc-Kean non-linear process.

$$\begin{aligned} d\bar{X}_t &= \bar{V}_t\,dt, \\ m d\bar{V}_t^i &= -\gamma\bar{V}_t\,dt + \lambda(\bar{X}^\alpha[\rho] - \bar{X})\,dt + \sigma D(\bar{X}^\alpha[\rho] - X)\,dB_t. \end{aligned} \tag{3.10}$$

Where (3.10) can be written in Vlasov-Fokker-Plank form with,

$$\kappa[\rho_t](x) = -\frac{\gamma}{m}v + \frac{\lambda}{m}(\bar{X}^\alpha - x), \quad \mu[\rho_t](x) = \frac{\sigma^2}{m^2}D(\bar{X}^\alpha - X)^2[16].$$

So that upon rearrangement we get Vlasov-Fokker-Plank equation,

$$\partial_t f + v\cdot\nabla_x f = \nabla_v\cdot(\frac{\gamma}{m}vf + \frac{\lambda}{m}(X - \bar{X}\alpha)f + \frac{\sigma^2}{2m^2}D(X - \bar{X}^\alpha)^2)\nabla_v f) \tag{3.11}$$

Definition 3.6. *(Vlasov-Fokker-Planck equation).*

$$\partial_t f + v \cdot \nabla_x f + (\kappa[\rho_t]) \cdot \nabla_v f = \mu[\rho_t] \Delta_v f, \tag{3.12}$$

Where f, is the distribution of particles and

$$\rho(t, x) = \int_{\mathbb{R}^d} f(t, x, \rho)\, d\rho,$$

is the spatial density of the macroscopic system [**?**].

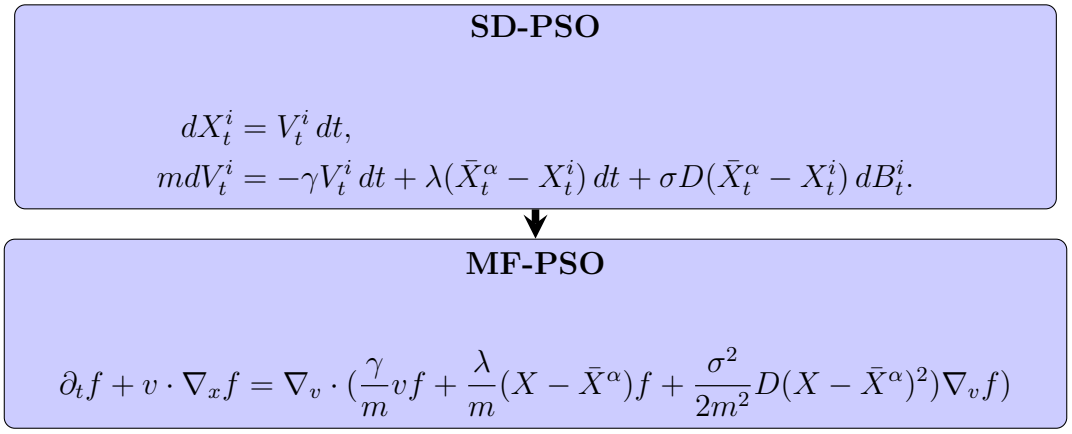

Figure 3.1: From SD-PSO to MF-PSO indicating that the dynamics of the SD-PSO system can be approximated by the dynamics of the MF-PSO PDE

3.1 Regularized PSO dynamic with memory

Below we have the second order system of SDEs for the canonical PSO method where now we take into account the global and local best, which have been regularized.

$$\begin{aligned} dX_t^i &= V_t^i\, dt, \\ dY_t^i &= v(X_t^i - Y_t^i) S(X_t^i, Y_t^i)\, dt, \\ dV_t^i &= -\gamma V_t^i\, dt + \lambda_1(Y_t^i - X_t^i)\, dt + \lambda_2(\bar{Y}_t^\alpha - X_t^i)\, dt \\ &\quad + \sigma_1 D(Y_t^i - X_t^i)\, dB_t^{1,i} + \sigma_2 D(\bar{Y}_t^\alpha - X_t^i)\, dB_t^{2,t} \end{aligned} \tag{3.13}$$

Where the weighted average is defined similarly to the case without memory,

$$\bar{Y}_t^\alpha = \frac{\sum_{i=1}^N Y_t^i \omega_\alpha(Y_t^i)}{\sum_{i=1}^N \omega_\alpha(Y_t^i)}, \quad \omega_\alpha(Y_t) = e^{-\alpha \mathcal{F}(Y_t)}. \tag{3.14}$$

Again, starting with the N particle probability density, we have a probability density which describes the probability the ith particle will be in position x_i, y_i with velocity v_i at time t.

$$f^{(N)}(x_1, \cdots, x_N, y_1, \cdots, y_N, v_1, \cdots, v_N, t)$$

We can then take the first marginal which gives the probability density of finding a given particle with position x_1, y_1 with velocity v_1 at a given time and integrate out the other remaining terms.

$$f_1^{(N)}(x_1, y_1, v_1, t) = \int_{\mathbb{R}^d} f^{(N)}(x_1, \cdots, x_N, y_1, \cdots, y_N, v_1, \cdots, v_N, t) d\Omega_1,$$

where $d\Omega_1 = dx_2 \cdots dx_N dy_2 \cdots dy_N dv_2 \cdots dv_N$.

In order to derive the mean-field equations for the microscopic system for the regularized system with memory, we again make the propogation of chaos assumption on the first marginal. That is in this system $N \gg 1$ is sufficiently large and $f^{(N)} \approx f^{\otimes N}$ and the particles are random triples (X_t^i, Y_t^i, V_t^i) sharing the same distribution $f(x, v, t)$.

Then,

$$\overline{Y_t^\alpha} \approx \overline{Y^\alpha}(\overline{\rho}) = \frac{\int_{\mathbb{R}^2} y\omega_\alpha(y)\overline{\rho}(y,t)\,dy}{\int_{\mathbb{R}^d} \omega_\alpha(y)\overline{\rho}(x,y)\,dy}, \quad \overline{\rho}(y,t) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x,y,v,t)\,dx\,dv. \tag{3.15}$$

The Mc-Kean non-linear process this time becomes,

$$\begin{aligned} d\bar{X}_t &= \bar{V}_t\,dt, \\ d\bar{Y}_t &= \upsilon(\bar{X}_t - \bar{Y}_t)S(\bar{X}_t, \bar{Y}_t)\,dt, \\ d\bar{V}_t &= -\gamma\bar{V}_t\,dt + \lambda_1(\bar{Y}_t - \bar{X}_t)\,dt + \lambda_2(\bar{Y}_t^\alpha - \bar{X}_t)\,dt \\ &\quad + \sigma_1 D(\bar{Y}_t - \bar{X}_t)\,dB_t^{1,i} + \sigma_2 D(\bar{Y}_t^\alpha - \bar{X}_t)\,dB_t^{2,t} \end{aligned} \tag{3.16}$$

Where (6.3) can be written in Vlasov-Fokker-Plank form with,

$$\kappa[\rho_t](x) = -\frac{\gamma}{m}v + \frac{\lambda_1}{m}(Y - x) + \frac{\lambda_2}{m}(\bar{Y}^\alpha - x),$$

and

$$\mu[\rho_t](x) = \frac{\sigma_1^2}{2m^2}D^2(Y - x)^2 + \frac{\sigma_2^2}{2m^2}D^2(\bar{Y}^\alpha - x)[16].$$

Then the distribution $f(x, v, t)$ takes satisfies the VFP-equation with memory,

$$\begin{aligned} &\partial_t f + v \cdot \nabla_x f + \nabla_y \cdot \big(\upsilon(x-y)S^\beta(x,y)f\big) \\ &\qquad = \nabla_v \cdot \left(\frac{\gamma}{m}vf + \frac{\lambda_1}{m}(x-y)f + \frac{\lambda_2}{m}\big(x - \bar{Y}^\alpha(\bar{\rho})\big)f \right. \\ &\qquad\quad \left. + \left(\frac{\sigma_2^2}{2m^2}D(x - \bar{Y}^\alpha(\bar{\rho}))^2 + \frac{\sigma_1^2}{2m^2}D(x-y)^2 \right) \nabla_v f \right) \end{aligned} \tag{3.17}$$

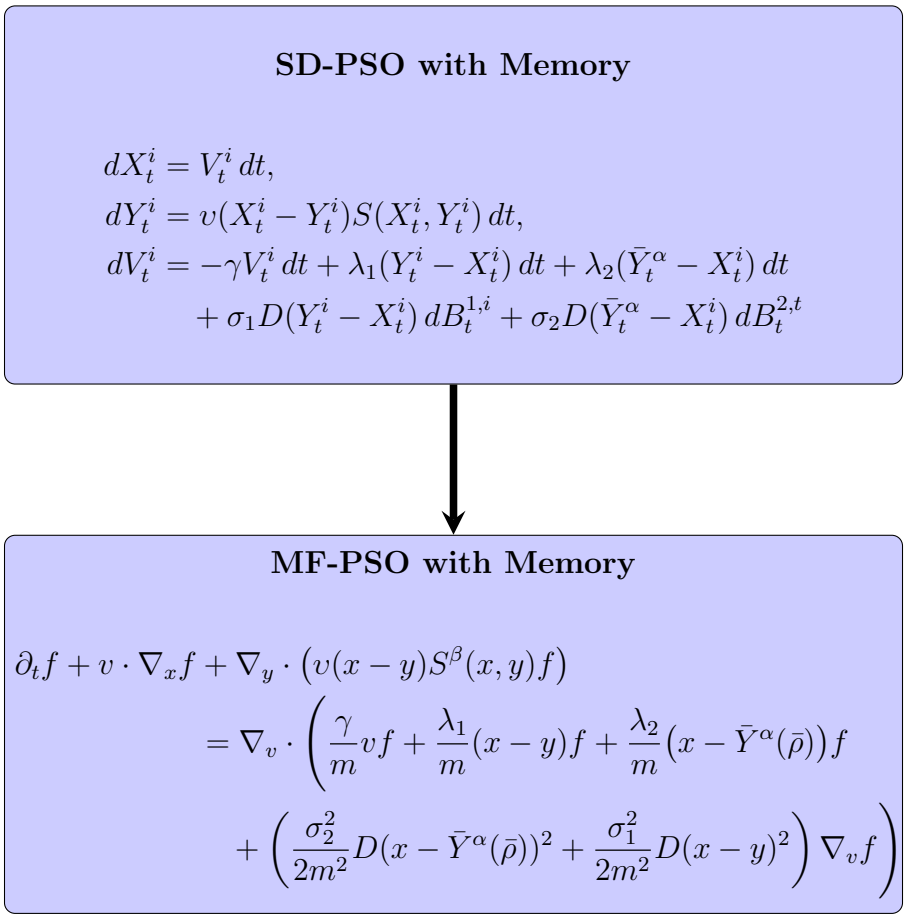

Figure 3.2: From SD-PSO to MF-PSO indicating that the dynamics of the SD-PSO system can be approximated by the dynamics of the MF-PSO PDE

Chapter 4

Small Inertia Limit - The CBO Equations

In the previous chapter we showed that in the mean-field limit that we were able to arrive a macroscopic system as represented by a Vlasov-Fokker-Planck type equation for the SDE form of the canonical PSO method.

Now we will consider that in the small inertia limit $m \to 0$, that we can derive a continuous mean-field CBO system. Since, in the small interia limit, the system of equations takes the form of the compressible Euler equations and exhibits properties of a hydrodynamic system it will first be necessary to consider some relevant ideas from kinetic theory.

(In what follows close reference has been made to [13].)

4.1 Kinetic theory

We can consider an interacting particle system, where for N particles with $x_i(t)$ and $v_i(t)$ such that $i = 1, \cdots, N$, Newton's equations are obtained

$$\dot{x_i} = v_i, \quad \dot{v_i} = F_i(x_1, \cdots, x_N)$$

where the force on the i-th particle can be considered a combination of an interaction force between particles and an external force as follows,

$$F_i(x_1, \cdots, x_N) = \frac{1}{N} \sum_{j=1}^{N} F_{int}(x_i - x_j) + F_{ext}(x_i)$$

For the case with non-interacting particles ($F_{\text{int}} = -\nabla \Phi_{\text{int}} = 0$), we get a divergence-free vector field ($\dot{X}, \dot{V} = F_{\text{ext}}(X, t)$) where the phase space volume remains constant over time under the dynamics of $\dot{X}$ and F_{ext}.

The field equations are:

$$\dot{X} = V, \quad \dot{V} = F_{\text{ext}}$$

Since the volume element $dx\, dv$ does not change over time, $f(X(t), V(t), t)$

must satisfy the conservation law:

$$\frac{d}{dt}f(X(t),V(t),t)=0$$

This indicates that particle distribution $f(X(t),V(t),t)$ does not change with time along the trajectories specified by $X(t)$ and $V(t)$.

Definition 4.1. *Collision operators*:

The other case arises when we take into account particle collisions. Let $Q(f)$ model the rate of change of f due to collisions. Then,

$$\frac{d}{dt}f(X(t),V(t),t)=\left(\frac{\partial f}{\partial t}+v\cdot\nabla_x f\right)|_{X(t),V(t),t}=Q(f)|_{X(t),V(t),t}$$

Where $Q(f)$ is known as the collision operator.

Definition 4.2. *Kinetic moments:*

In order to describe kinetic processes we start with a distribution function intending to model a large particle system,

$$f(x,v,t)\geq 0.$$

Where $f(x,v,t)$ is a non-negative function describing the density of particles in the phase space where f quantifies how many particles exist in a given phase space volume $dxdv$ about (x,v).

$$\rho(x,t)=\int_{\mathbb{R}^3} f v\,dv,\quad \rho u(x,t)=\int_{\mathbb{R}^3} f v\,dv,\quad E(x,t)=\frac{1}{2}\int_{\mathbb{R}^3} f v^2\,dv \tag{4.1}$$

Definition 4.3. *Equation of free transport*

In this equation there is no interaction between particles and this equation simply follows directly from Newton's equations.

$$\frac{\partial f}{\partial t}+v\cdot\nabla_x f=0. \tag{4.2}$$

Furthermore, with the introduction of the collision operator we obtain the Boltzmann equation:

$$\frac{\partial f}{\partial t}+v\cdot\nabla_x f=Q(f,f). \tag{4.3}$$

Definition 4.4. *Vlasov equation*

If we now introduce a macroscopic force F(x) acting on particles in the system then we obtain the linear Vlasov equation.

$$\frac{\partial f}{\partial t}+v\cdot\nabla_x f+F(x)\cdot\nabla_v f=0. \tag{4.4}$$

With the introduction of the collision operator into the Boltzmann equation with macroscopic force we obtain the following [19],

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + F(x) \cdot \nabla_v f = Q(f, f). \tag{4.5}$$

The compressible Euler equation for conservation of mass can be obtained by integrating (4.3) against the collision invariant $\phi(v)$ i.e,

$$\frac{\partial}{\partial t} \int_{\mathbb{R}^3} f\phi(v), dv + \nabla_x \left(\int_{\mathbb{R}^3} vf\phi(v), dv \right) = 0,$$

where,

$$\int_{\mathbb{R}^3} Q(f, f)\phi(v), dv = 0, \quad \phi(v) = 1, v_{x_1}, v_{x_2}, v_{x_3}, |v|^2. \tag{4.6}$$

Since collisions should preserve mass, momentum and energy. Integrating against the the other collision invariants yields the Compressible Euler equations.

Definition 4.5. *Compressible Euler equations*

$$\begin{aligned} &\frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho u) = 0 \\ &\frac{\partial (\rho u)}{\partial t} + \nabla_x \cdot (\rho u \otimes u + pI) = 0 \\ &\frac{\partial E}{\partial t} + \nabla_x \cdot ((E + p)u) = 0, \quad p = \rho T = \frac{2}{3}E - \frac{1}{3}\rho u^2. \end{aligned} \tag{4.7}$$

Definition 4.6. *Maxwellian Distributions*

Maxwellian distributions are probability distributions of the following form,

$$M(x, v, t) = p(x, t) \frac{1}{(2\pi T(x, t))^{\frac{3}{2}}} \exp\left(-\frac{|v - U(x, t)|^2}{2T(x, t)} \right) \tag{4.8}$$

Where the local Maxwellian is a function of mass $\rho(x, t)$, $U(x, t)$, and temperature $T(x, t)$[15](p.5).

For a system in thermal equilibrium, this distribution maximizes the entropy given certain constraints (such as mass, momentum and energy.) A system reaches equilibrium precisely when entropy is maximised [19] (p.36).

Definition 4.7. *Gibbs' lemma corollary:*

Solutions to $Q(f, f) = 0$ where f is a distribution function with finite mass and energy in $\mathbb{R}$ are precisely Maxwellian distributions [19](p.36).

4.2 Small Inertia Hydrodynamic Limits

Consider the PSO dynamic without memory effects in (3.3). Here we have that for $m \ll 1$ that the second equation becomes,

$$V_t^i dt = \lambda \left(\bar{X}_t^\alpha - X_t^i \right) dt + \sigma D \left(\bar{X}_t^\alpha - X_t^i \right) dB_t^i,$$

So with the observation that $\gamma = 1 - m \approx 1$ we obtain the first order CBO dynamic.

$$dX_t^i = \lambda \left(\bar{X}_i^\alpha - X_t^i\right) dt + \sigma D \left(\bar{X}_t^\alpha - X_t^i\right) dB_t^i, \tag{4.9}$$

This can be considered as a reduced order approximation of canonical SD-PSO dynamics [10].

The small intertia value is denoted $m = \epsilon > 0$ which is subsituted into (3.11) so that we can write the scaled VFP in the following form,

$$\partial_t f + v \cdot \nabla_x f + \frac{1}{\epsilon} \nabla_v \cdot (\epsilon v f + \lambda(\bar{X}^\alpha(\rho) - x) f) = L_\epsilon(f). \tag{4.10}$$

Where the left hand side takes the form of the Vlasov mean-field equation and the right-hand side is the collision operator $L_\epsilon(f)$ such that the equation is in the form of (4.5).

Then $L_\epsilon(f)$ is defined,

$$L_\epsilon(f) = \frac{1}{\epsilon} \nabla_v \cdot (vf + \frac{\sigma^2}{2\epsilon} D(x - \bar{X}^\alpha(\rho))^2 \nabla_v f)$$

4.2.1 Derivation

Overall we would like to obtain hydrodynamic limits for this system for small values of m. To do this we will derive the Compressible Euler equations. In order to do so it is neccesary to find a Maxwellian such that when the density function equals the maxwellian the collision operator equals zero.

- We would like to find a Maxwellian $M_\epsilon(x, v, t)$ such that $L_\epsilon(f) = 0$.
- To see that $M_\epsilon(x, v, t)$ is a Maxwellian we rewrite $L_\epsilon(f) = 0$ in a more amenable form.
- Density is approximately the local Maxwellian for small ϵ
- We can now derive the Compressible Euler equations.

4.2.2 Rearrange Collision Operator

Starting with $L_\epsilon(f)$, Let,

$$A = vf + \frac{\sigma^2}{2\epsilon} D(x - \bar{X}^\alpha(\rho))^2 \nabla_v f$$

Then we obtain the collision operator for the Vlasov-Fokker-Plank system.

$$\nabla_v \cdot A = \sum_{j=1}^{d} \frac{\partial}{\partial v_j} \left(v_j f + \frac{\sigma^2}{2\epsilon} d_j (x_j - \bar{X}_j^\alpha(\rho))^2 \frac{\partial f}{\partial v_j} \right)$$

Note that when we introduce the summation term since $d_j = 1$ for all j we can drop D.

Then,

$$L_\epsilon(f) = \frac{1}{\epsilon}\sum_{j=1}^{d}\frac{\sigma^2}{2}(x_j - \bar{X}_j^\alpha(\rho))^2\frac{\partial}{\partial v_j}\left(\frac{2fv_j}{\sigma^2(x_j - \bar{X}_j^\alpha(\rho))^2} + \frac{1}{\epsilon}\frac{\partial f}{\partial v_j}\right)$$

At present the collision operator is describes the evolution of the distribution function which is described by the spatial deviation of the particles from their weighted average position as well as by drift and diffusion terms.

Definition 4.8. *(Local Maxwellian):*
For unitary mass and zero momentum we have the following as defined in [10],

$$\mathcal{M}_\epsilon(x,v,t) = \prod_{j=1}^{d}\mathcal{M}_\epsilon(x_j,v_j,t),$$
$$\mathcal{M}_\epsilon(x_j,v_j,t) = \left(\frac{\epsilon^{\frac{1}{2}}}{\pi^{\frac{1}{2}}(x_j - \bar{X}_j^\alpha(\rho))^2}\right)\exp\left(-\frac{\epsilon v_j^2}{\sigma^2(x_j - \bar{X}_j^\alpha(\rho))^2}\right).$$

Next,

$$L_\epsilon(f) = \frac{1}{\epsilon^2}\sum_{j=1}^{d}\frac{\sigma^2}{2}(x_j - \bar{X}_j^\alpha(\rho))^2\frac{\partial}{\partial v_j}\left(\frac{2\epsilon fv_j}{\sigma^2(x_j - \bar{X}_j^\alpha(\rho))^2} + \frac{\partial f}{\partial v_j}\right)$$

We would like to rewrite our expression for $L_\epsilon(f)$ so that it is in terms of the local Maxwellian. This is more easily seen as being reverse engineered from the final term in our relation and by making the observation that the partial of the log Maxwellian coefficient term vanishes with respect to the partial v_j.

Consider the following,

$$\begin{aligned}\frac{2\epsilon fv_j}{\sigma^2(x_j - \bar{X}_j^\alpha(\rho))^2} + \frac{\partial f}{\partial v_j} &= f\frac{\partial}{\partial v_j}\left(\log\left(\frac{\epsilon^{\frac{1}{2}}}{\pi^{\frac{1}{2}}(x_j - \bar{X}_j^\alpha(\rho))^2}\right) + \frac{\epsilon v_j^2}{\sigma^2(x_j - \bar{X}_j^\alpha(\rho))^2}\right) + \frac{\partial f}{\partial v_j}\\ &= f\frac{\partial}{\partial v_j}\left(\log(f) - \log(M_\epsilon(x_j,v_j,t))\right)\\ &= f\frac{\partial}{\partial v_j}\log\left(\frac{f}{M_\epsilon(x_j,v_j,t)}\right)\end{aligned}$$

Then we now have,

$$L_\epsilon(f) = \frac{1}{\epsilon^2}\sum_{j=1}^{d}\frac{\sigma^2}{2}(x_j - \bar{X}_j^\alpha(\rho))^2\frac{\partial}{\partial v_j}\left(f\frac{\partial}{\partial v_j}\log\left(\frac{f}{M_\epsilon(x_j,v_j,t)}\right)\right).$$

Which more succinctly expresses the dynamics as we can consider the deviation of the distribution function $f(x,v,t)$ from the local Maxwellian $M_\epsilon(f)$. Crucially, in this form, it is clear that when $f = M_\epsilon(f)$, $L_\epsilon(f) = 0$.

4.2.3 Equilibrium Approximation

Since $L_\epsilon(f)$ is of order $\frac{1}{\epsilon^2}$, for small values of $\epsilon \ll 1$, the influence of collisions becomes much more significant driving f towards the equilibrium when $\epsilon \to 0$ [19](p.40).

Therefore we have,

$$f(x, v, t) = \rho(x, t)\mathcal{M}_\epsilon(x, v, t) \tag{4.11}$$

4.2.4 Compressible Euler Equations

Conservation of Mass Equation

Recall the identities,

$$\rho u = \int_{\mathbb{R}^d} f(x, v, t) v \, dv \quad \rho = \int_{\mathbb{R}^d} f(x, v, t) \, dv$$

Then integrating (4.10) with respect to v provides,

$$\int_{\mathbb{R}^d} \frac{\partial f}{\partial t} + \nabla_x \cdot v f \, dv = 0.$$

So that,

$$\frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho u) = 0.$$

Since, for small $\epsilon \ll 1$ we have that $L_\epsilon(f)$ tends to zero when f approaches a Maxwellian distribution $M_\epsilon(f)$

Furthermore, since $\frac{1}{\epsilon^2}$ dominates the asymptotic regime, we can neglect $\frac{1}{\epsilon}$ terms.

Conservation of Momentum Equation

Now integrating (4.10) against the collision invariant $\phi(v) = v$ allows us to obtain the conservation of momentum equation.

$$\int_{\mathbb{R}^d} \frac{\partial f}{\partial t} v + v \left(\nabla_x \cdot v f\right) \, dv = \int_{\mathbb{R}^d} \left(\frac{(1-\epsilon)}{\epsilon} f v + \frac{1}{\epsilon} \lambda (\bar{X}^\alpha(\rho) - x) f \right) \, dv$$

$$\frac{\partial \rho u}{\partial t} + \int_{\mathbb{R}^d} v \left(v \cdot \nabla_x f\right) \, dv = \frac{(1-\epsilon)}{\epsilon} \rho u + \frac{1}{\epsilon} \lambda (\bar{X}^\alpha(\rho) - x) \rho.$$

With the previous assumption (4.11), for $\epsilon \gg 1$ the i-th compment is,

$$\begin{aligned}\int_{\mathbb{R}^d} v_i \left(v \cdot \nabla_x \left(\rho \mathcal{M}_\epsilon(x,v,t)\right)\right) dv &= \sum_{j=1}^{d} \frac{\partial}{\partial x_j}\left(\rho(x,t)\int_{\mathbb{R}^d} v_i \left(v_j \mathcal{M}_\epsilon(x,v,t)\right) dv\right) \\ &= \frac{\partial}{\partial x_j}\left(\rho(x,t)\int_{\mathbb{R}^d} v_i^2 \mathcal{M}_\epsilon(x,v,t)\, dv\right) \\ &= \frac{\sigma^2}{2\epsilon}\frac{\partial}{\partial x_i}\left(\rho(x,t)\left(c_i - \bar{X}_i^\alpha(\rho)\right)^2\right)\end{aligned}$$

Where the second equality holds under the assumption that v_i, v_j vanish for $i \neq j$ that is, mixed moments vanish,

$$\int_{\mathbb{R}^d} v_i v_j(x_j, v_j, t) dv = 0.$$

The final equality from direct integration using the well known identity,

$$\int_{-\infty}^{\infty} v_i^2 e^{-a v_i^2}\, dv_i = \frac{1}{2a}\sqrt{\frac{\pi}{a}}.$$

Overall this results in the second order macroscopic model

$$\begin{aligned}\frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho u) &= 0 \\ \frac{\partial \left(\rho u\right)_i}{\partial t} + \frac{\sigma^2}{2\epsilon}\frac{\partial}{\partial x_i}\left(\rho(x,t)\left(c_i - \bar{X}_i^\alpha(\rho)\right)^2\right) &= \frac{(1-\epsilon)}{\epsilon}\left(\rho u\right)_i + \frac{1}{\epsilon}\lambda\left(\bar{X}_i^\alpha(\rho) - x_i\right)\rho.\end{aligned} \tag{4.12}$$

The limit of which as $\epsilon \to 0$ in (4.12) is,

$$\left(\rho u\right)_i = \lambda\left(\bar{X}^\alpha(\rho) - x\right)\rho - \frac{1}{\epsilon}\lambda(\bar{X}_i^\alpha(\rho) - x_i)\rho.$$

Substituting this expression into the first equation of (4.12), we get the mean-field CBO system.

$$\frac{\partial \rho}{\partial t} + \nabla_x \cdot \lambda\left(\bar{X}^\alpha(\rho) - x\right)\rho = \frac{\sigma^2}{2}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_j^2}\left(\rho(x,t)\left(x_j - \bar{X}_j^\alpha(\rho)\right)^2\right). \tag{4.13}$$

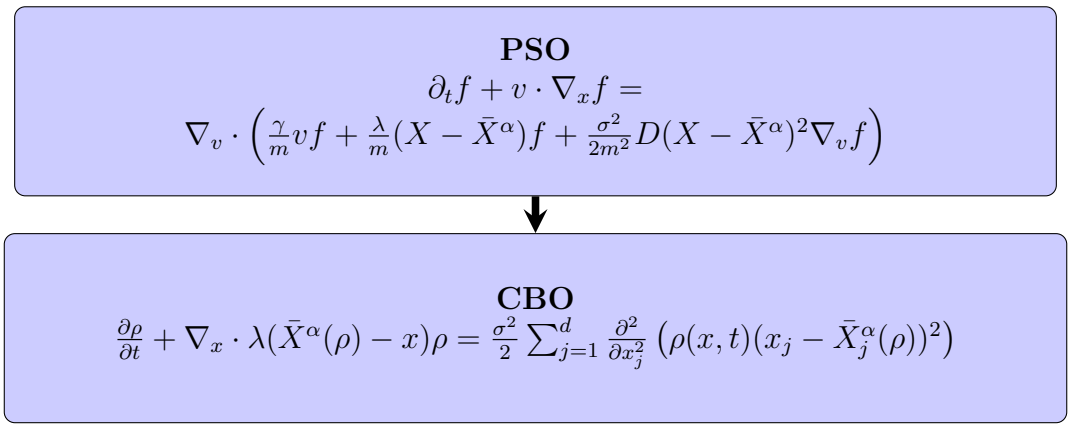

Figure 4.1: From PSO to CBO system via small inertia limit. Indicating that small inertia dynamics of the PSO system can be approximated by CBO dynamics.

4.3 Small inertia limit with memory

Starting with the Mean-Field equation which describes in deterministic form, the behaviour of the SD-PSO equation with memory (6.3) we will demonstrate the derivation of the complimentary CBO equation.

$$\begin{aligned} \partial_t f + v \cdot \nabla_x f + \nabla_y \cdot \left(\upsilon(x-y) S^\beta(x,y) f \right) \\ + \frac{1}{\epsilon} \nabla_v \cdot \left(\epsilon v f + \lambda_1 (x-y) f + \lambda_2 (\bar{Y}^\alpha(\bar{\rho}) - x) f \right) = L_\epsilon(f) \end{aligned} \tag{4.14}$$

Where,

$$L_\epsilon(f) = \frac{1}{\epsilon} \nabla_v \cdot \left(v f + \frac{\sigma_2^2}{2\epsilon} D(x - \bar{Y}^\alpha(\bar{\rho}))^2 \nabla_v f + \frac{\sigma_1^2}{2\epsilon} D(x-y)^2 \nabla_v f \right).$$

Let,

$$A = \left(v f + \frac{\sigma_2^2}{2\epsilon} D(x - \bar{Y}^\alpha(\bar{\rho}))^2 \nabla_v f + \frac{\sigma_1^2}{2\epsilon} D(x-y)^2 \nabla_v f \right).$$

Then we obtain the collision operator for the Vlasov-Fokker-Plank system with memory,

$$\nabla_v \cdot A = \sum_{d=j}^{d} \frac{\partial}{\partial v_j} \left(v_j f + \frac{\sigma_2^2}{2\epsilon} d_j (x_j - \bar{Y}_j^\alpha(\bar{\rho}))^2 \frac{\partial f}{\partial v_j} + \frac{\sigma_1^2}{2\epsilon} D(x-y)^2 \frac{\partial f}{\partial v_j} \right).$$

Where using a similar simplification as in the case with memory we get,

$$\frac{1}{2\epsilon}\sum_{j=1}^{d}\Sigma(x_j,y_j,t)^2\frac{\partial}{\partial v_j}\left(\frac{2fv_j}{\Sigma(x_j,y_j,t)^2}+\frac{1}{\epsilon}\frac{\partial f}{\partial v_j}\right),$$

Such that,

$$\Sigma(x_j,y_j,t)^2=\sigma^2(x_j-\bar{Y}_j^{\alpha}(\bar{\rho}))^2+\sigma_1^2(x_j-y_j)^2.$$

This time the local Maxwellian is the following,

$$\mathcal{M}_\epsilon(x,y,v,t)=\prod_{j=1}^{d}M_\epsilon(x_j,y_j,v_j,t),$$

$$\mathcal{M}_\epsilon(x_j,y_j,v_j,t)=\frac{\epsilon^{\frac{1}{2}}}{\pi^{\frac{1}{2}}|\Sigma(x_j,y_j,t)|}\exp\{\frac{-\epsilon v_j^2}{\Sigma(x_j,y_j,t)^2}\}$$

Then we do essentially the same steps to recover a more helpfully written collision operator once again,

$$\begin{aligned}\frac{2\epsilon fv_j}{\Sigma(x_j,y_j,t)^2}+\frac{\partial f}{\partial v_j}&=f\frac{\partial}{\partial v_j}\left(-\log\left(\frac{\epsilon^{\frac{1}{2}}}{\pi^{\frac{1}{2}}\Sigma(x_j,y_j,t)}\right)+\frac{\epsilon v_j^2}{\Sigma(x_j,y_j,t)^2}\right)+\frac{\partial f}{\partial v_j}\\&=f\frac{\partial}{\partial v_j}\left(\log(f)-\log(M_\epsilon(x_j,v_j,t))\right)\\&=f\frac{\partial}{\partial v_j}\log\left(\frac{f}{M_\epsilon(x_j,v_j,t)}\right).\end{aligned}$$

Then integrating (3.17) with respect to v provides,

$$\int_{\mathbb{R}^d}\frac{\partial f}{\partial t}+\nabla_x\cdot vf+\nabla_y\cdot(v(x-y)S^{\beta}(x,y)f)\,dv=0.$$

So that,

$$\frac{\partial\rho}{\partial t}+\nabla_x\cdot(\rho u)+\nabla_y\cdot(v(x-y)S^{\beta}(x,y)\rho)=0. \tag{4.15}$$

Since, for small $\epsilon\ll 1$ we have that $L_\epsilon(f)$ tends to zero when f approaches a Maxwellian distribution $M_\epsilon(f)$.

Integrating (4.14), with respect the collision invariant $\phi(v)=v$ allows us to obtain the conservation of momentum equation for the the case with memory.

$$\begin{aligned}\int_{\mathbb{R}^d}\left(\frac{\partial f}{\partial t}v+v\left(\nabla_x\cdot(vf)\right)+\frac{1}{\epsilon}\nabla_v\left(\epsilon vf+\lambda_1(x-y)f+\lambda_2(\bar{Y}^{\alpha}-x)f\right)v\right)dv\\=\int_{\mathbb{R}^d}\frac{1}{\epsilon}\nabla_v\left(vf\right)v\,dv\end{aligned} \tag{4.16}$$

Consider:

$$\int_{\mathbb{R}^d} v\left(\nabla_x \cdot (vf)\right)\, dv = \int_{\mathbb{R}^d} v\left(v \cdot \nabla_x f + df\right)\, dv$$

Since:

$$\int_{\mathbb{R}^d} v \cdot (df)\, dv = d \int_{\mathbb{R}^d} vf\, dv = 0$$

Since f is a Maxwellian.
Then:

$$\int_{\mathbb{R}^d} v_i v_j \frac{\partial f}{\partial x_j}\, dv = \frac{\partial}{\partial x_j}\left(\int_{\mathbb{R}^d} v_i v_j \rho \mathcal{M}_\epsilon\, dv\right)$$

Then,

$$\frac{\sigma^2}{2\epsilon}\frac{\partial}{\partial x_i}\left(\rho(x,t)\Sigma(x_i,y_i,t)^2\right)$$

Since,

$$\int_{\mathbb{R}^d} v_i v_j M_\epsilon\, dv = \mathbb{E}[v_i v_j] = \frac{\Sigma(x_i,y_i,t)^2\delta_{ij}}{\epsilon}.$$

Which in terms of the kinetic moments reads,

$$\frac{\partial(\rho u)_i}{\partial t} + \frac{\sigma^2}{2\epsilon}\frac{\partial}{\partial x_i}\left(\rho(x,t)\Sigma(x_i,y_i,t)^2\right) = \frac{1-\epsilon}{\epsilon}(\rho u)_i + \frac{1}{\epsilon}\left(\lambda_1(y_i - x_i) + \lambda_2\left(\bar{Y}_i^\alpha(\bar{\rho}) - x_i\right)\right)\rho. \tag{4.17}$$

Rearranging (4.17) and taking the limit $\epsilon \to 0$ provides,

$$(\rho u)_i = \frac{\sigma^2}{2}\frac{\partial}{\partial x_j}(\rho\Sigma(x_j,y_j,t)^2) + \lambda_1(x_i - y_i) + \lambda_2(\bar{Y}^\alpha(\bar{\rho}))\rho.$$

Substituting this into (4.15) gives the mean-field CBO system with local best,

$$\begin{aligned}\frac{\partial\rho}{\partial t} + \nabla_x \cdot \left[(\lambda_1(y-x) + \lambda_2(\bar{Y}^\alpha(\bar{\rho}) - x))\rho\right] + \nabla_y \cdot \left[v(x-y)S^\beta(x,y)\rho\right] \\ = \frac{1}{2}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_j^2}\left[\rho(x,t)\left(\sigma_1^2(x_j - y_j)^2 + \sigma_2^2(x_j - \bar{Y}_j^\alpha(\bar{\rho}))^2\right)\right].\end{aligned} \tag{4.18}$$

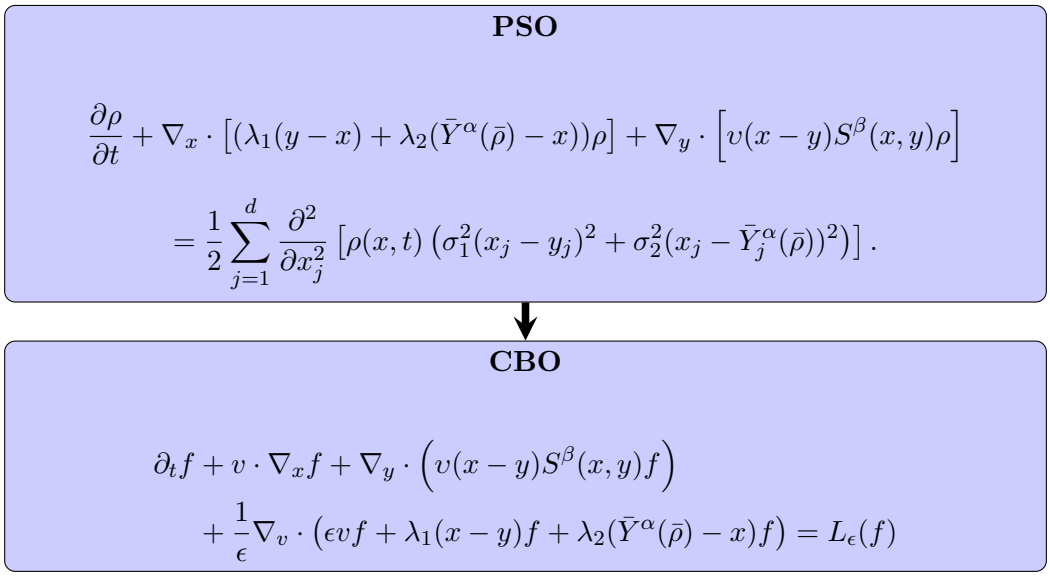

Figure 4.2: From PSO to CBO system via small inertia limit. Indicating that small inertia dynamics of the PSO system with memory can be approximated by CBO dynamics with memory.

Chapter 5

Discretization

5.0.1 Discretization without Memory

First let us consider the case without memory 3.11. To discretize this PDE, we can perform dimensional splitting where we either solve the transport part using a second-order backward semi-Lagrangian method or using an upwind method under certain conditions. The Fokker-Plank term can be solved using an implicit central scheme.

We have the nonlinear Vlasov-Fokker-Planck equation:

$$\partial_t f + v \cdot \nabla_x f = \nabla_v \cdot \left(\frac{\gamma}{m} v f + \frac{\lambda}{m}(X_t^i - \bar{X}_t^\alpha) f + \frac{\sigma^2}{2m^2} D (X_t^i - \bar{X}_t^\alpha)^2 \nabla_v f \right). \quad (5.1)$$

We would like to perform dimensional splitting so that the transport part is solved through a second-order backward semi-Lagrangian method and the remaining Fokker-Planck term is discretized using an implicit central scheme.

$$\frac{df}{dt} = \partial_t f + v \cdot \nabla_x f$$

$$\begin{aligned} \frac{f_i^{n+1} - f_i^n}{\Delta t} =& \frac{\gamma}{m} \left(f_i^{n+1} + v_i \frac{f_{i+1}^{n+1} - f_{i-1}^{n+1}}{2\Delta v} \right) \\ &+ \frac{\lambda}{m}(X_t^i - \bar{X}_t^\alpha) \left(\frac{f_{i+1}^{n+1} - f_{i-1}^{n+1}}{2\Delta v} \right) \\ &+ \frac{f_{i+1}^{n+1} - 2f_i^{n+1} + f_{i-1}^{n+1}}{\Delta v^2} \left(\frac{\sigma^2}{2m^2} D (X_i^t - \bar{X}^\alpha)^2 \right) \end{aligned} \quad (5.2)$$

5.0.2 Dimensional Splitting

From here we can introduce some new terms ν_1 and ν_2 and $a[i]dx$ to simplify the equation:

$$\nu_1 = \frac{\Delta t}{2w\Delta v}, \quad \nu_2 = \frac{\sigma^2 \Delta t}{2w^2 \Delta v^2}, \quad a[i]dx = (X_i^t - \bar{X}^\alpha)$$

So we get:

$$\begin{aligned} f_i^{n+1} - f_i^n = \frac{\Delta t \gamma}{m} f_i^{n+1} \\ + v_i \gamma \nu_1 (f_{i+1}^{n+1} - f_{i-1}^{n+1}) \\ + \nu_1 \lambda a[i] dx (f_{i+1}^{n+1} - f_{i-1}^{n+1}) \\ + \nu_2 a[i]^2 dx (f_{i+1}^{n+1} + f_{i-1}^{n+1} - 2 f_i^{n+1}) \end{aligned}$$

Then, the implicit central scheme is:

$$f_i^n = A_i f_{i-1}^{n+1} + B_i f_i^{n+1} + C_i f_{i+1}^{n+1}$$

With coefficients:

$$\begin{aligned} A_i &= -a[i]\, dx^2 \nu_2 - \nu_1 \gamma v_i - \nu_1 \lambda a[i]\, dx \\ B_i &= 1 + 2\nu_2 a[i]\, dx^2 - \frac{\gamma}{m} \Delta t \\ C_i &= -\nu_2 a[i]\, dx^2 + \gamma v_i \nu_1 + \nu_1 \lambda a[i]\, dx \end{aligned}$$

5.0.3 Semi-Lagrangian Scheme

We have that for the semi-Lagrangian scheme using linear interpolation. If you use $\frac{v\Delta t}{\Delta x} < 1$, you obtain the upwind method [14]. So provided we follow this condition, the upwind scheme can be used as detailed below:

$$f_i^{n+1} = f_i^n - \frac{\Delta t}{\Delta x} \left(\max(0, v) \cdot (f_i^n - f_{i-1}^n) + \min(0, v) \cdot (f_{i+1}^{n+1} - f_i^n) \right).$$

Alternatively, you can just use the linear semi-Lagrangian system.
The semi-Lagrangian scheme is:

$$f(x_j, v_i, t^{n+1}) = f(x_j - v\Delta t, v_i, t^n)$$

Such that:

$$f_j^{n+1} = f_{j - v\frac{\Delta t}{\Delta x}}^n = f_{j-k-\alpha}^n, \quad k + \alpha = v\frac{\Delta t}{\Delta x}, \quad k = \left\lfloor v \frac{\Delta t}{\Delta x} \right\rfloor$$

Where for linear interpolation we get:

$$f_j^{n+1} = \alpha f_{j-k-1}^n + (1 - \alpha) f_{j-k}^n [14].$$

5.0.4 Discretization with Memory

Now we consider the case with memory (3.17). We follow the same procedure as in the case with memory and use dimensional splitting, where the transport term and convection term can be solved using the upwind method. Again the Fokker-Planck term uses an implicit central scheme.

We have the nonlinear Vlasov-Fokker-Planck equation with memory:

$$\partial_t f + v \cdot \nabla_x f + \nabla_y \cdot \big(v(x-y)S^\beta(x,y)f\big) = \nabla_v \cdot \left(\frac{\gamma}{m} vf + \frac{\lambda_1}{m}(x-y)f + \frac{\lambda_2}{m}\big(x - \bar{Y}^\alpha(\bar{\rho})\big)f + \left(\frac{\sigma_2^2}{2m^2} D(x - \bar{Y}^\alpha(\bar{\rho}))^2 + \frac{\sigma_1^2}{2m^2} D(x-y)^2 \right) \nabla_v f \right)$$

5.0.5 Fokker-Planck Term

We first derive the implicit central scheme for the Fokker-Planck term.

$$\begin{aligned} \frac{f_i^{n+1} - f_i^n}{\Delta t} = \frac{\gamma}{m} & \left(f_i^{n+1} + v_i \frac{f_{i+1}^{n+1} - f_{i-1}^{n+1}}{2\Delta v} \right) \\ & + \frac{\lambda_1}{m}(x-y)\frac{f_{i+1}^{n+1} - f_{i-1}^{n+1}}{2\Delta v} \\ & + \frac{\lambda_2}{m}(x - \bar{Y}^\alpha)\frac{f_{i+1}^{n+1} - f_{i-1}^{n+1}}{2\Delta v} \\ & + \left(\frac{\sigma^2}{2m^2} D(x-y)^2 + \frac{\sigma_2^2}{2m} D(x-y)^2 \right) \frac{f_{i+1}^{n+1} - 2f_i^{n+1} + f_{i-1}^{n+1}}{\Delta v^2} \end{aligned}$$

Then,

$$\begin{aligned} f_i^{n+1} - f_i^n = & \mu_3 \gamma v \left(f_{i+1}^{n+1} - f_{i-1}^{n+1} \right) \\ & + \mu_3 \lambda_1 \left((x-y) \left(f_{i+1}^{n+1} - f_{i-1}^{n+1} \right) \right) \\ & + \mu_3 \lambda_2 \left((x - \bar{Y}^\alpha) \left(f_{i+1}^{n+1} - f_{i-1}^{n+1} \right) \right) \\ & + \mu_4 \left(\sigma_1^2 D(x-y)^2 + \sigma_2^2 D(x - \bar{Y}^\alpha)^2 \right) \left(f_{i+1}^{n+1} - 2f_i^{n+1} + f_{i-1}^{n+1} \right) \end{aligned}$$

Which can be used directly or solved using BTCS and solving a system of equations.

5.0.6 Upwind Method for Memory Term

The upwind method for the Memory term as in the upwind scheme is derived based on whether the scheme is upwind or downwind, where we accordingly implement a forward or backwards approximation.

$$\begin{aligned} f_j^{n+1} = f_j^n - \frac{\Delta t v}{\Delta y} \mathrm{adv}(j) & \left[(x - y_j) S^\beta(x, y_j) f_j^n - (x - y_{j-1}) S^\beta(x, y_{j-1}) f_{j-1}^n \right] \\ & + (1 - \mathrm{adv}(j)) \left[(x - y_{j+1}) S^\beta(x, y_{j+1}) f_{j+1}^n - (x - y_j) S^\beta(x, y_j) f_j^n \right] \end{aligned}$$

Chapter 6

Numerical Testing

6.1 Numerical Comparison

To test the accuracy of the mean-field PSO equation without memory (3.11) in modeling the dynamics of the stochastic-differential PSO system also without memory (3.3) we consider the Ackley function with global minimum in x = 0.

The results were obtained with,

$$\gamma = 0.5, \quad \lambda = \frac{1}{\sqrt{3}}, \quad \alpha = 30.$$

Where the choices for λ and α were make to match $c_{1,2} = 2$ in (2.3). The results were initialized with the uniform distribution with $N = 5 \times 10^5$ and a rectangular mesh with $(x, v) \in [-3, 3] \times [-4, 4]$.

In figure 6.3 and 6.5 the evolution of the contour plots in time can be seen for $t = 0.5, 1$ and 3 for the SD-PSO and MF-PSO cases. In figure 6.4 and 6.10 the evolution of the marginal density $\rho(x, t) = \int_{\mathbb{R}^d} f(x, v, t)dv$ is recorded for the same time points.

In the Ackley function with $x = 0$ 6.3 we see convergence towards the Dirac-delta function as time processes. We also observe increasingly good matching of the MF-system to the PSO system.

In 6.5 the case where the minimum of the Ackely function is in x = 1. We see the particles move beyond $x = 1$ initially and then settle towards $x = 1$ by $t = 3$. There is reasonable agreement in this case but we see that the MF-PSO model undershoots the stochastic model in this case. This is perhaps due to the choice of backward-semi-Lagrangian scheme for the transport part. A Lax-Wendroff scheme may perform better.

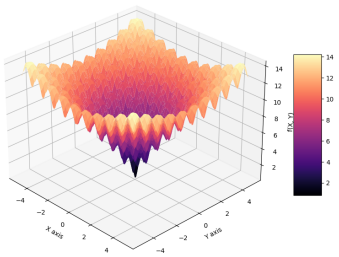

Figure 6.1: 3-D plot of the Ackley function with minimum in x = 0.

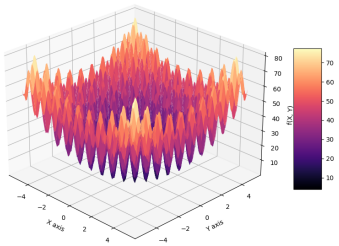

Figure 6.2: 3-D plot of the Rastrigin function with its minimum in x = 0.

Ackley Function:

$$f(\mathbf{x}) = -20\exp\left(-0.2\sqrt{\frac{x_1^2 + x_2^2}{2}}\right) - \exp\left(\frac{\cos(2\pi x_1) + \cos(2\pi x_2)}{2}\right) + 20 + e$$

Rastrigin Function:

$$f(\mathbf{x}) = 10 \cdot n + \sum_{i=1}^{n} \left[x_i^2 - 10\cos\left(2\pi x_i\right)\right]$$

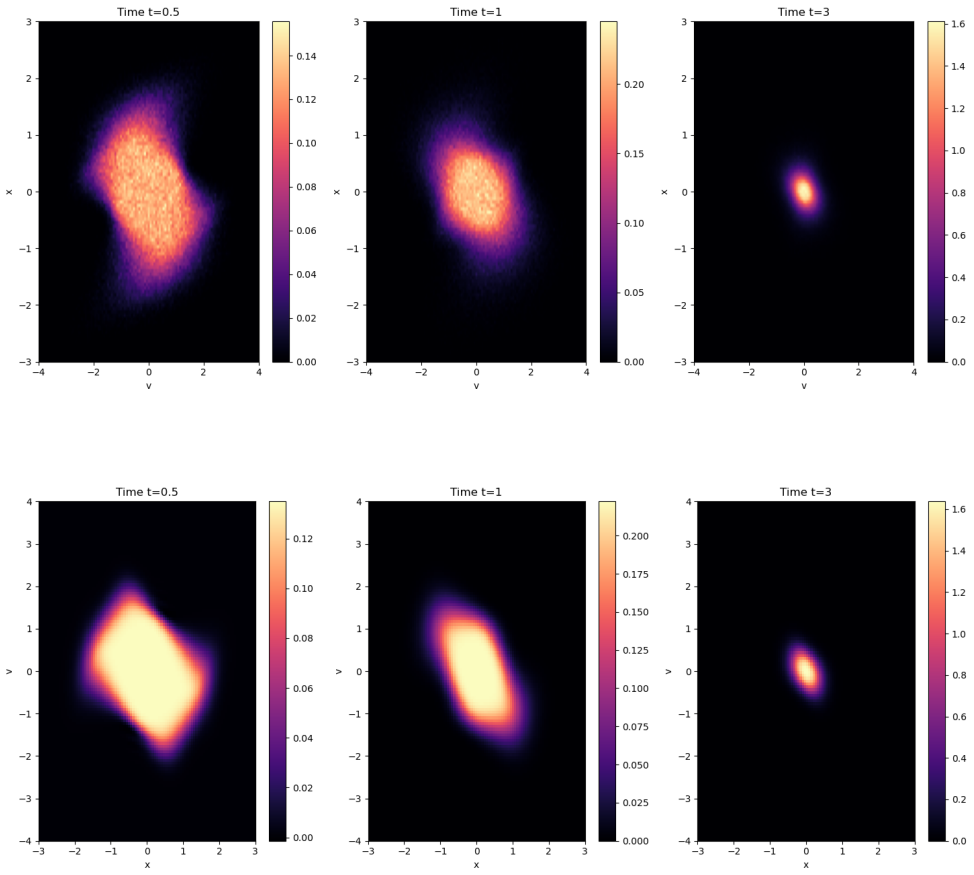

Figure 6.3: Contour plots for the time evolution of the Stochastic Differential PSO system (above) and the Mean-Field PSO equation (3.11)(below). For times $t = 0.5, 1$, and 3 where the function to be minimized is the Ackley function with a minimum at $x = 0$.

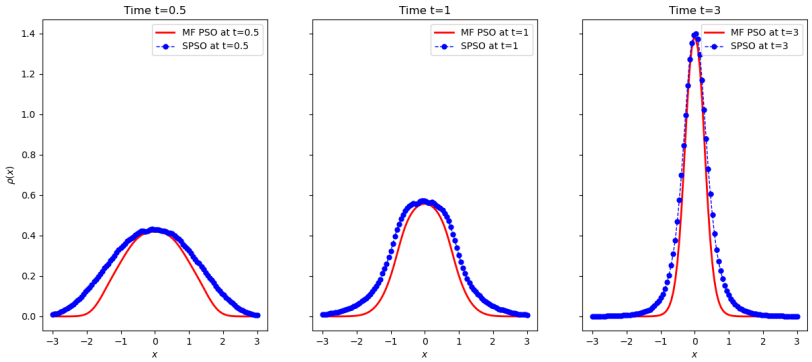

Figure 6.4: Plotting the evolution of the PDF of the particles for MF-PSO and SD-PSO at time $t = 0.5, 1$ and 3. For the Ackley function with minimum at $x = 0$.

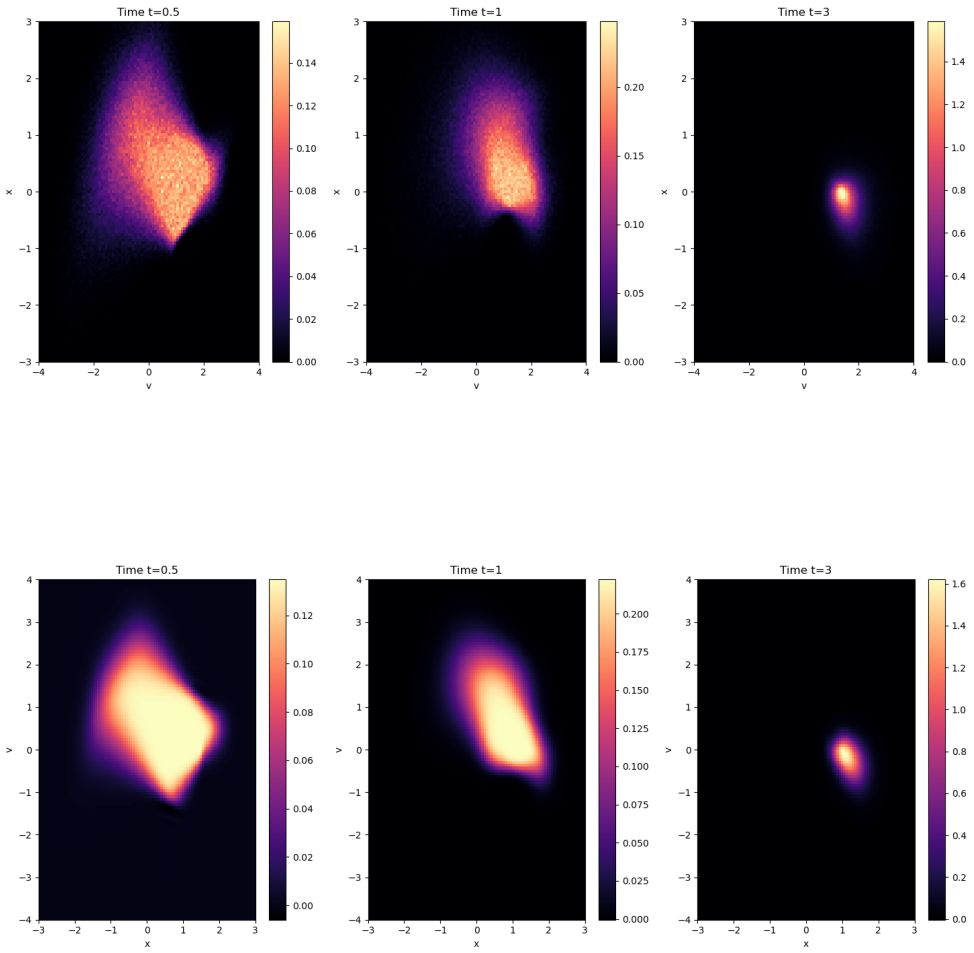

Figure 6.5: Histograms for the SD-PSO system (6.3)(above) and for the MF-PSO system (below). Dynamics at time steps $t = 0.5, 1$ and 3, to minimize the Ackley function with minimum at $x = 0$.

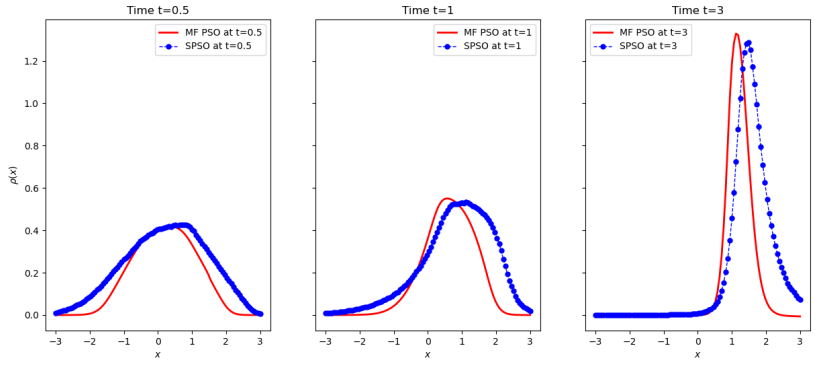

Figure 6.6: PDFs for the SD-PSO system and for the MF-PSO system. Dynamics at time steps $t = 0.5, 1$ and 3, to minimize the Ackley function with minimum at $x = 0$.

6.2 Numerical Testing with Memory

6.2.1 Case 1

In this second case, we examine the dynamics of the SD-PSO system with memory (3.13). Here we will consider the dynamics when only the local-best term is present. To isolate this term set $\lambda_2, \sigma_2 = 0$. The only additional parameters now are a choice of $\beta = 30$ and $\nu = 0.5$. In the first of the plots for this case, the Ackley function with minimum at x = 0 us being minimized, in the second the Rastrigin function also with minimum at x = 0. Starting again with uniform initial data this time we observe the dynamics from $t = 0.5, 3$ and 6. With only the 'local best' dynamics in this method, we see that both local and global minima are identified by the particles. This is because the particles are returning to their local best position.

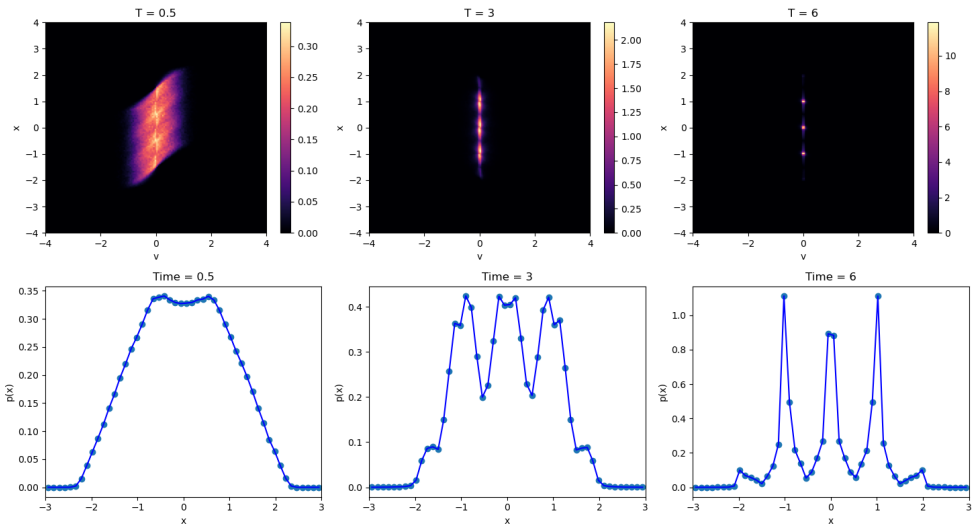

Figure 6.7: Histogram showing the evolution of the SD-PSO system with memory with only local best dynamics (3.13) (above). Evolution of the probability density function for the same method (below). In this case the Ackley function with minimum at x = 0 is being minimized and the evolution is recorded for t = 0.5, 3 and 6.

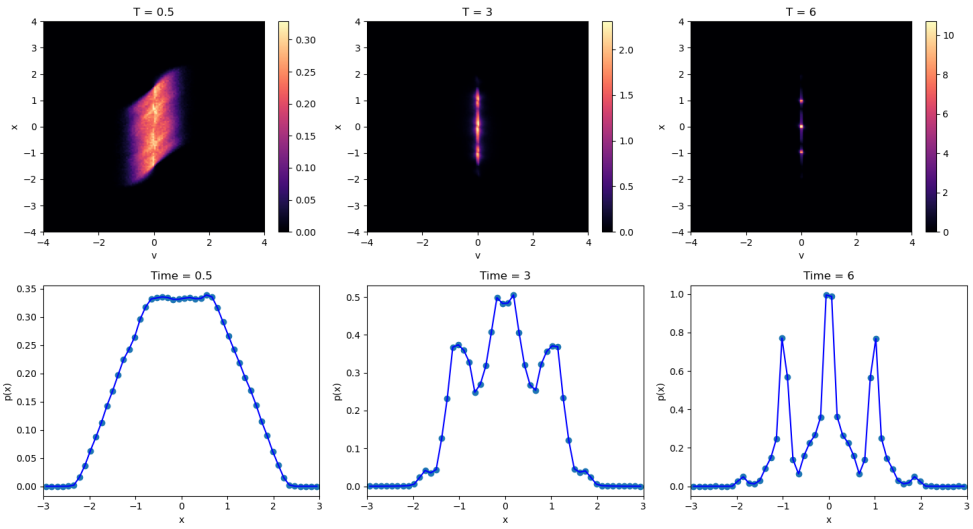

Figure 6.8: Histogram showing the evolution of the SD-PSO system with memory with only local best dynamics(3.13) (above). Evolution of the probability density function for the same method (below). In this case the Rastrigin function with minimum at x = 0 is being minimized and the evolution is recorded for t = 0.5, 3 and 6.

6.2.2 Case 2

In this third case, we examine the dynamics of the SD-PSO system with memory (3.13). Here we will consider the dynamics of both the global and local best in this system. These effects are equally weighted with the choice $\lambda_1 = \lambda_2 = 1$ and $\sigma_1, \sigma_2 = \frac{1}{\sqrt{3}}$ again with the same parameters and initial distribution as in the other cases. As in the second case, the Ackley and Rastrigin functions are being minimized with minimum at x = 0.

In this case, We no longer see the particles also identifying the local minima. In this case only the global minima are identified. Observe that in this third case we have used the time steps $t = 0.5, 1$ and 3 as in the first case and that if we compare the evolution of the density in the first case to the second for the Ackley function i.e 6.2.2 and 6.1 that we see faster convergence in the SD-PSO system with the additional state variable.

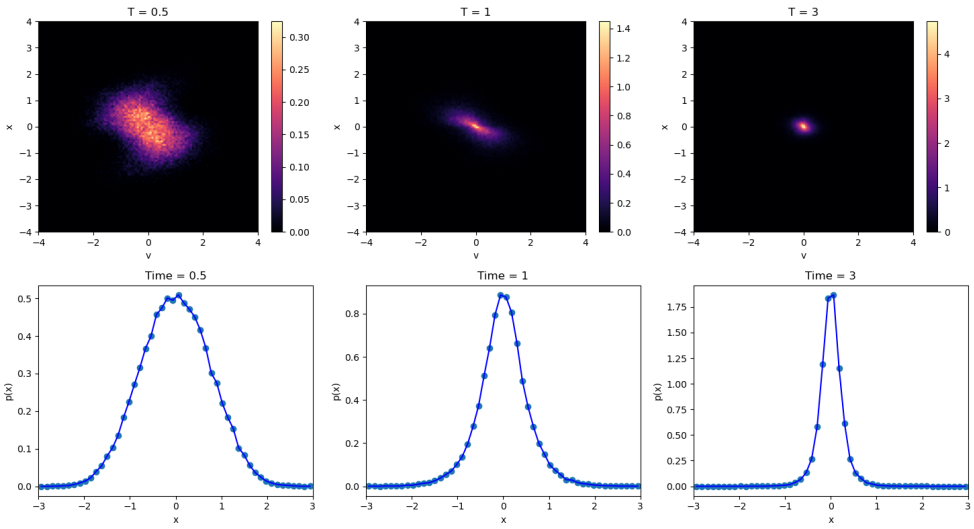

Figure 6.9: Histogram showing the evolution of the SD-PSO system with memory with global and local best dynamics (3.13) (above). Evolution of the probability density function for the same method (below). In this case the Ackley function with minimum at x = 0 is being minimized and the evolution is recorded for t = 0.5, 3 and 6.

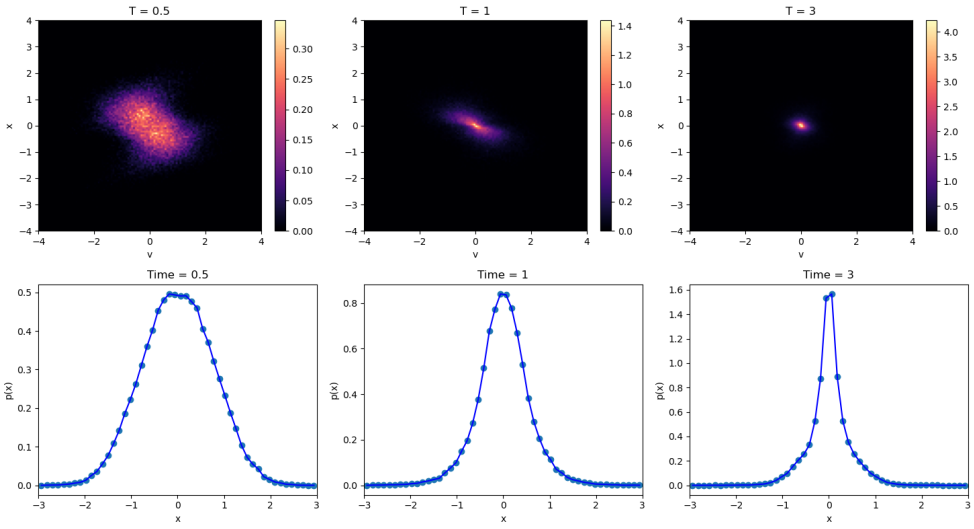

Figure 6.10: Histogram showing the evolution of the SD-PSO system with memory with global and local best dynamics (3.13) (above). Evolution of the probability density function for the same method (below). In this case the Rastrigin function with minimum at x = 0 is being minimized and the evolution is recorded for t = 0.5, 3 and 6.

6.3 Conclusion

In this dissertation, we have considered the optimization technique of Particle Swarm Optimization with a particular focus to implementing recent techniques of

regularisation, mean-field large particle limits to derive deterministic analogues of stochastic PSO systems. This was achieved by looking at the original and canonical formulations of the PSO method and demonstrating how they can be rewritten as a time continuous formulation of a second order system of SDEs. Then via regularization of the global best term, regularized PSO systems both with and without memory could be derived where in the large particle limit mean-field Vlasov-Fokker-Planck type equations are obtained. Then, with the PSO methods re-written in this new formulation, it becomes possible to make a hydrodynamic approximation of the Mean-Field VFP systems in the small inertia limit by deriving macroscopic equations from the first order CBO dynamic.

To validate the correspondence between the SD-PSO equations and the MF-PSO deterministic PDE, we looked at the simpler case without memory to make a numerical comparison of the methods. Via discritization techniques discussed in the penultimate section we were able to make this comparison by looking at evolution of the contour plots and probability density functions of the two models. It was found that indeed there was good agreement between the two models, however it should be observed that even better numerical agreement can be obtained than in the present project [10].

Furthermore we looked at some characteristics of the SD-PSO sytem with memory (3.13), to see how they compared to the case without memory. From this it was apparent that the inclusion of 'local best' dynamics only allowed for the identification of local as well as global minima and that the inclusion of local and global dynamics in this system lead to faster convergence to the global minimum of the Ackley function in particular than with a MF-PSO equation without memory.

Rigorous results concerning the derivation of the CBO system via the zero inertia limit and results quantifying the convergence rates of these methods have subsequently been addressed in [4][9]. Furthermore, related developments to this research area have been subsequently considered such as CBO with Memory Effects, (CBO-ME) with applications of the this method include image segmentation using multi-thresholding and the training of neural networks [2].

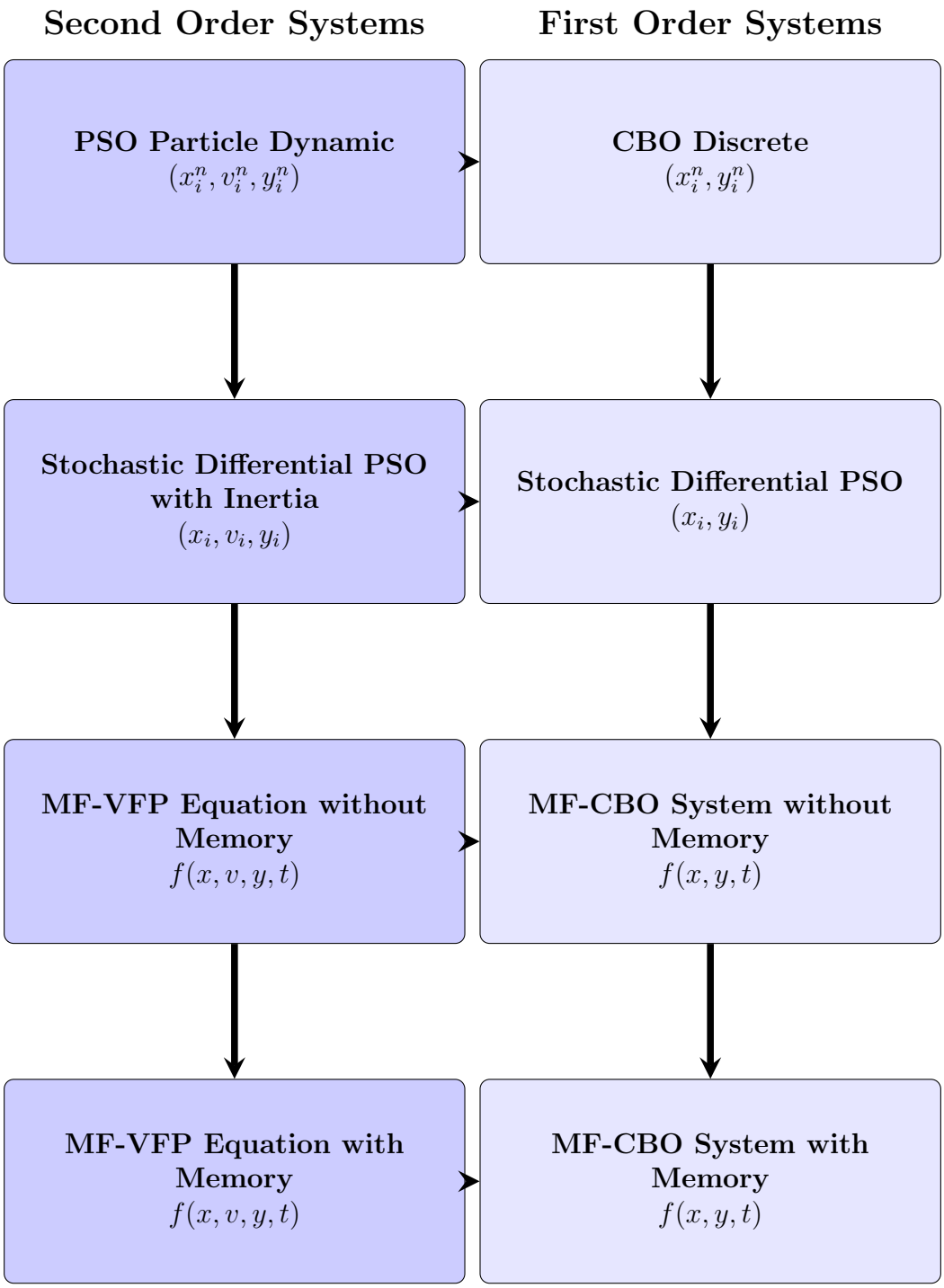

Figure 6.11: Flow chart suggesting the relationship between 'PSO' methods and 'CBO' methods.

Chapter 7

Appendix 1

7.0.1 Laplace's Principle

Proof. Assume that $F(x)$ is continuous and achieves its infimum c for some $x^* \in \text{dom}(\rho)$. Since $\nabla F(x^*) = 0$ (because x^* is a minimum), we perform a Taylor expansion around x^* to get:

$$F(x) \approx F(x^*) + \frac{1}{2}(x - x^*)^T H(x^*)(x - x^*) + \mathcal{O}(\|x - x^*\|^3),$$

where $H(x^*)$ is the Hessian matrix of F at x^*.
Now consider the function $e^{-\alpha F(x)}$. We approximate it as follows:

$$e^{-\alpha F(x)} \approx e^{-\alpha F(x^*)} e^{-\frac{\alpha}{2}(x-x^*)^T H(x^*)(x-x^*)}.$$

Next, we approximate the integral:

$$\int_{\mathbb{R}^d} e^{-\alpha F(x)}\, d\rho(x) \approx e^{-\alpha F(x^*)} \int_{\mathbb{R}^d} e^{-\frac{\alpha}{2}(x-x^*)^T H(x^*)(x-x^*)}\, d\rho(x).$$

Let $y = x - x^*$. Assuming the Hessian $H(x^*)$ is positive definite, the integral simplifies to:

$$\int_{\mathbb{R}^d} e^{-\frac{\alpha}{2} y^T H(x^*) y}\, dy = \sqrt{\frac{(2\pi)^d}{\det(H(x^*))/\alpha}}.$$

Thus,

$$\int_{\mathbb{R}^d} e^{-\alpha F(x)}\, d\rho(x) \approx e^{-\alpha F(x^*)} \sqrt{\frac{(2\pi)^d \alpha}{\det(H(x^*))}}.$$

Taking the logarithm, we have:

$$-\frac{1}{\alpha} \log \left(\int_{\mathbb{R}^d} e^{-\alpha F(x)}\, d\rho(x) \right) \approx -\frac{1}{\alpha} \log \left(e^{-\alpha F(x^*)} \sqrt{\frac{(2\pi)^d \alpha}{\det(H(x^*))}} \right).$$

This simplifies to:

$$-\frac{1}{\alpha}\left(-\alpha F(x^*)+\frac{1}{2}\log\left(\frac{(2\pi)^d\alpha}{\det(H(x^*))}\right)\right)=F(x^*)-\frac{1}{2\alpha}\log\left(\frac{(2\pi)^d\alpha}{\det(H(x^*))}\right).$$

Taking the limit as $\alpha\to\infty$, the term $\frac{1}{2\alpha}\log\left(\frac{(2\pi)^d\alpha}{\det(H(x^*))}\right)$ vanishes, yielding:

$$\lim_{\alpha\to\infty}-\frac{1}{\alpha}\log\left(\int_{\mathbb{R}^d}e^{-\alpha F(x)}\,d\rho(x)\right)=F(x^*).$$

□

Bibliography

[1]

[2] Giacomo Borghi, Sara Grassi, and Lorenzo Pareschi. Consensus based optimization with memory effects: Random selection and applications. *Chaos, Solitons amp; Fractals*, 174:113859, September 2023.

[3] José A. Carrillo, Shi Jin, Lei Li, and Yuhua Zhu. A consensus-based global optimization method for high dimensional machine learning problems, 2020.

[4] Cristina Cipriani, Hui Huang, and Jinniao Qiu. Zero-inertia limit: from particle swarm optimization to consensus based optimization, 2022.

[5] I. H. Dinwoodie. Large deviations techniques and applications (amir dembo and ofer zeitouni). *SIAM Review*, 36(2):303–304, 1994.

[6] H.M. Emara and Hossam Abdel Fattah. Continuous swarm optimization technique with stability analysis. pages 2811 – 2817 vol.3, 01 2004.

[7] Massimo Fornasier, Hui Huang, Lorenzo Pareschi, and Philippe Sünnen. Consensus-based optimization on the sphere: Convergence to global minimizers and machine learning, 2021.

[8] Michel Gendreau and Jean-Yves Potvin. *Handbook of Metaheuristics*. 01 2019.

[9] Sara Grassi, Hui Huang, Lorenzo Pareschi, and Jinniao Qiu. Mean-field particle swarm optimization, 2021.

[10] Sara Grassi and Lorenzo Pareschi. From particle swarm optimization to consensus based optimization: stochastic modeling and mean-field limit, 2020.

[11] Pierre-Emmanuel Jabin and Zhenfu Wang. *Mean Field Limit for Stochastic Particle Systems*, volume 1. 04 2017.

[12] Nikolaos Kolliopoulos, Martin Larsson, and Zeyu Zhang. Propagation of chaos for maxima of particle systems with mean-field drift interaction, 2023.

[13] Lorenzo Pareschi. Advances in numerical methods for hyperbolic and kinetic equations. lecture 5.

[14] Lorenzo Pareschi. Advances in numerical methods for hyperbolic and kinetic equations. lecture 7.

[15] Lorenzo Pareschi and Giuseppe Toscani. *Interacting multiagent systems. Kinetic equations and Monte Carlo methods*. 11 2013.

[16] René Pinnau, Claudia Totzeck, Oliver Tse, and Stephan Martin. A consensus-based model for global optimization and its mean-field limit. *Mathematical Models and Methods in Applied Sciences*, 27(01):183–204, January 2017.

[17] Riccardo Poli, James Kennedy, and Tim Blackwell. Particle swarm optimization: An overview. *Swarm Intelligence*, 1, 10 2007.

[18] Y. Shi and R. Eberhart. A modified particle swarm optimizer. In *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*, pages 69–73, 1998.

[19] Cédric Villani. A review of mathematical topics in collisional kinetic theory. *Handbook of Mathematical Fluid Dynamics*, 1, 12 2002.