Micro-Macro Decomposition of Particle Swarm Optimization Methods

Michael Herty

RWTH Aachen University, Institute of Geometry and Applied Mathematics, Templergraben 55, 52062 Aachen, Germany

Extraordinary Professor, University of Pretoria, Department of Mathematics and Applied Mathematics, Private Bag X20, Hatfield 0028, South Africa
herty@igpm.rwth-aachen.de

Sara Veneruso*

RWTH Aachen University, Institute of Geometry and Applied Mathematics, Templergraben 55, 52062, Aachen, Germany

University of Ferrara, Department of Mathematics and Computer Science, Via Machiavelli 30, 44121 Ferrara, Italy
veneruso@igpm.rwth-aachen.de, sara.veneruso@unife.it

August 18, 2025

Abstract

Solving non-convex minimization problems using multi-particle metaheuristic derivative-free optimization methods is still an active area of research. Popular methods are Particle Swarm Optimization (PSO) methods, that iteratively update a population of particles according to dynamics inspired by social interactions between individuals. We present a modification to include constrained minimization problems using exact penalization. Additionally, we utilize the hierarchical structure of PSO to introduce a micro-macro decomposition of the algorithm. The probability density of particles is written as a convex combination of microscopic and macroscopic contributions, and both parts are propagated separately. The decomposition is dynamically updated based on heuristic considerations. Numerical examples compare the results obtained using the algorithm in the microscopic scale, in the macroscopic scale, and using the new micro-macro decomposition.

Keywords: derivative-free optimization, metaheuristics, particle swarm optimization, constrained minimization, micro-macro decomposition

MSC codes: 35Q70, 35Q84, 35Q93, 49M41, 90C56

1 Introduction

Many mathematical models nowadays study the coordinated movement of groups using multi-agent approaches [15, 45]. Metaheuristic methods are based on simple rules of interaction between individual particles, which give complex results of self-organized and swarm intelligence [2, 40, 44]. Some examples of these methods are Ant Colony Optimization [19, 20], Genetic Algorithms [25, 31], Simulated Annealing [1, 36], Consensus-Based Optimization (CBO) [8, 28, 41] and Particle Swarm

*Corresponding author

Optimization (PSO) [33–35, 43]. There are strong links between these methods, for example the CBO can be derived from the PSO via zero inertia limit [13].

In this paper, we focus on the study of Particle Swarm Optimization, proposed for solving constrained minimization problems. The goal is to find

$$\min_{x \in \mathbb{R}^d} \mathcal{F}(x) \quad \text{subject to } x \in \mathcal{M} \tag{1.1}$$

$d \geq 1$, where $\mathcal{F}(x)$ is a given continuous but possibly non convex objective function. It is assumed that $\mathcal{F}(x)$ admits an unique global minimizer $x^*_{\mathcal{M}}$ in $\mathcal{M} \subset \mathbb{R}^d$.

The PSO method solves the optimization problem (1.1) by considering a set of candidate solutions, represented by particles. Then, particles move in the search space according to certain mathematical relations on the positions and velocities of the particles, until convergence toward a neighborhood of the global minimizer. Since PSO methods are gradient-free, the function $\mathcal{F}(x)$ in (1.1) can be non-smooth.

The PSO has originally been formulated in such a way that each particle experiences a superposition of forces towards the local best position and towards the global best position found among all particles up to that iteration [34, 35]. There have been many extensions to the model; for example, in [27], a description based on a system of stochastic differential equations has been introduced, as well as a formulation modeling memory of each particle. In another version of PSO, there is an inertia weight [21, 43] to bound the possible uncontrolled speed of the particles. Additionally, PSO methods have gained prominence, drawing inspiration from CBO methods. In this variant of the PSO, the concept of a consensus point emerges, representing a weighted average of the positions of particles towards which they are attracted.

Several methods have been proposed in the literature for incorporating constraints in interacting particle consensus methods [10, 11, 22, 23]. Here, we present a method based on exact penalization. This approach has already been used to study constrained problems with PSO methods in the version that uses the local best position and the global best position [14, 39]. However, in this manuscript we use the iterative strategy introduced for the CBO [7], for which the penalty parameter is updated according to the violation of the constraint, and apply it to the version of the PSO method in which the consensus point appears.

The theoretical analysis of PSO methods has been developed on various scales [26, 27]: the microscopic scale, the mesoscopic scale, and the macroscopic scale. The microscopic scale is based on the evolution of particle trajectories using systems of stochastic differential equations. The intermediate scale, called mean-field, is the mesoscopic scale, where an equation governs the dynamics of the evolution of the probability density of the particles. These equations are a statistical description of the states of the particles. Finally, at the macroscopic scale, the particle flow is considered to behave like a fluid. At this scale, the dynamics are governed by a system of hyperbolic conservation laws.

In this manuscript, we propose an approach called micro-macro decomposition; the idea is inspired by [16, 18, 38]. This decomposition is well-known in the literature in various contexts, for example, to derive the Navier-Stokes equations from the Boltzmann equation in rarefied gas dynamics [12] or in the context of collisional kinetic equations that naturally incorporate exact space boundary conditions [16]. The decomposition is based on the decomposition of the probability density function into a contribution of the microscopic system and of the macroscopic system. The two contributions are weighted by a time-dependent function that is dynamically updated. This changes the behavior of the system to be more fluid-like or more particle-like.

This manuscript is structured as follows. In Section 2, we present the penalization approach applied to constrained minimization problems. In Section 3, the microscopic model with inertia, its mean-field limit, and finally a second-order macroscopic system of the PSO methods are presented. In Section 4, we present the micro-macro decomposition. At last, in Section 5, we present various numerical results obtained for constrained and unconstrained minimization problems, studying them at the microscopic and macroscopic scale and also using the mixed approach given by the micro-macro decomposition.

2 Particle Swarm Optimization methods for constrained problems

PSO methods are designed to solve unconstrained problems and we extend those to constrained problems of the following type:

$$\min_{x\in\mathbb{R}^d} \mathcal{F}(x) \quad \text{subject to } x \in \mathcal{M}$$

for $d \geq 1$, where $\mathcal{F}(x) : \mathbb{R}^d \to \mathbb{R}$ is a given continuous, non-convex function, with an unique global minimizer in the feasible set $\mathcal{M} \subset \mathbb{R}^d$, which is closed, bounded, possibly non-convex, and empty. The set $\mathcal{M}$ is not required to be connected. In addition, we assume $\mathcal{M}$ has a boundary $\partial\mathcal{M}$ of zero Lebesgue-measure. In this paper we will focus on the study of the constrained case using a penalized PSO method, but other approaches are possible (see [5]). We consider exact penalization, since PSO does by itself not require differentiable objective functions.

The objective function is modified by a penalty term in such a way that points outside the set $\mathcal{M}$ are reached with low probability. The penalized problem to be solved is

$$\min_{x\in\mathbb{R}^d} \{\mathcal{F}_\beta(x) := \mathcal{F}(x) + \beta r(x)\}, \tag{2.1}$$

where $\beta \in \mathbb{R}^+$ is the strength of the penalization and r is a penalty function, with the property:

$$r(x) = \begin{cases} = 0 & \text{if } x \in \mathcal{M}, \\ > 0 & \text{if } x \notin \mathcal{M}. \end{cases}$$

An example of $r(x)$ is the distance of the point $x \in \mathbb{R}^d$ from $\mathcal{M}$, e.g.,

$$r(x) = \min_{y\in\mathcal{M}} ||x - y||_p. \tag{2.2}$$

The objective function $\mathcal{F}_\beta$ coincides with the objective function $\mathcal{F}$ in the feasible set.

The goal is to obtain the same minimizer $x^*_{\mathcal{M}}$ for the penalized problem as for the constrained problem, for finite values of β. Penalization techniques that have this property are called exact penalization.

Property 1. *A function $\mathcal{F}_\beta : \Omega \to \mathbb{R}$ is called an exact penalty function at a local minimum $x^*_{\mathcal{M}}$ of* (1.1)*, if $x^*_{\mathcal{M}}$ is a local minimum of* (2.1) *[4, 5].*

F_β satisfies Property 1, if β is sufficiently large. If r is non-differentiable, then β is finite. We define this finite but a priori unknown value as $\bar{\beta}$

$$\bar{\beta} := \inf \Big\{\beta > 0 : \text{minimizers to problem (2.1) are minimizers to problem (1.1)}\Big\}.$$

In the case of the PSO, as done for the CBO in [7], we choose β by an adaptive strategy [29]. The idea is to test whether the violation value $r(x)$ is less than or equal to a fixed tolerance. If the condition is satisfied, the feasibility requirement is increased. If the condition is not fulfilled, however, the penalty parameter β is increased. The adaptive algorithm is explained in detail in Section 5.1.

To summarize, the exact penalization was employed, thereby converting the initial problem (1.1) into an unconstrained one (2.1). This approach enables the employment of the penalized PSO method. The incorporation of the penalty function introduces a non-differentiability into the objective function. However, this does not pose a significant challenge for the PSO method, as it does not depend on derivatives, thereby negating the necessity for differentiability. An additional benefit is that by implementing an adaptive update to the parameter, a convergence to the solution of constrained problem (1.1) is achieved. Furthermore, the method enables the resolution of constrained problems with unconnected constraints. In Section 5 various simulations are presented in which the feasible set $\mathcal{M}$ is not connected.

3 The PSO method for constrained problems on microscopic and macroscopic scale

3.1 Microscopic scale

In this section, we present the microscopic PSO method that describes the individual motion of particles, in the version proposed by [26].

The PSO is based on the classical second-order Newtonian dynamics of particle systems. At each time $n \geq 0$, we suppose to have N particles, each with position $X_n^i \in \mathbb{R}^d$ and velocity $V_n^i \in \mathbb{R}^d$, $i = 1, ..., N$. We assign an initial position X_0^i and an initial velocity V_0^i, then the evolution of positions and velocities is given by the canonical PSO equations

$$\begin{cases} X_{n+1}^i = X_n^i + \Delta t V_{n+1}^i, \\ V_{n+1}^i = \left(\frac{m}{m+\gamma\Delta t}\right) V_n^i + \left(\frac{\lambda\Delta t}{m+\gamma\Delta t}\right)(X_n^{\alpha,\beta} - X_n^i) + \left(\frac{\sigma\sqrt{\Delta t}}{m+\gamma\Delta t}\right) D(X_n^{\alpha,\beta} - X_n^i)\theta_n^i, \end{cases} \quad \text{for } i = 1, ..., N, \tag{3.1}$$

with inertia weight $m \in (0, 1]$, drift parameter $\lambda > 0$, noise parameter $\sigma > 0$, friction coefficient $\gamma := (1-m) \geq 0$, $\theta_n^i \sim \mathcal{N}(0, 1)$ and $D(X_n^{\alpha,\beta} - X_n^i)$ is a d-dimensional matrix. $X_n^{\alpha,\beta}$ is the weighted average, commonly called consensus point, computed using the approach introduced for the CBO methods [41]

$$X_n^{\alpha,\beta} := \frac{1}{N_n^{\alpha,\beta}} \sum_{i=1}^N X_n^i \omega^{\alpha,\beta}(X_n^i), \quad \text{with } N_n^{\alpha,\beta} := \sum_{i=1}^N \omega^{\alpha,\beta}(X_n^i). \tag{3.2}$$

The weight function $\omega^{\alpha,\beta}$ is given by

$$\omega^{\alpha,\beta}(X_n^i) := e^{-\alpha\mathcal{F}_\beta(X_n^i)}.$$

The choice of the Gibbs-type measure for $X_n^{\alpha,\beta}$ derived from the Laplace principle [17], which states that for any absolutely continuous distribution $\rho \in \mathcal{P}(\mathbb{R}^d)$ and $\mathcal{F}_\beta$ continuous, it holds

$$\lim_{\alpha\to\infty} \left(-\frac{1}{\alpha} \log\left(\int_{\mathbb{R}^d} e^{-\alpha\mathcal{F}_\beta(x)} d\rho(x)\right)\right) = \inf_{x\in supp(\rho)} \mathcal{F}_\beta. \tag{3.3}$$

The proof of convergence of (3.3) for functions with a unique global minimum is given e.g. in [41]. By the Laplace principle (3.3), and for large values of $\alpha \gg 1$, the consensus point $X_n^{\alpha,\beta}$ is an approximation of the global minimizer of the objective function $\mathcal{F}_\beta$.

Various choices for the diffusion term $D(X_n^{\alpha,\beta} - X_n^i)$ can be found in the literature. A first example is the isotropic exploration [41]

$$D^{iso}(X_n^{\alpha,\beta} - X_n^i) = ||X_n^{\alpha,\beta} - X_n^i||_2 I_d$$

where the norm used is the L^2 norm of vector in $\mathbb{R}^d$ and $I_d \in \mathbb{R}^{d\times d}$ is the identity matrix. Another possibility is the anisotropic exploration [9]

$$D^{aniso}(X_n^{\alpha,\beta} - X_n^i) = \text{diag}((X_n^{\alpha,\beta} - X_n^i)_1, ..., (X_n^{\alpha,\beta} - X_n^i)_d) \tag{3.4}$$

where diag : $\mathbb{R}^d \to \mathbb{R}^{d\times d}$ is the operator that maps a vector onto a diagonal matrix. For high-dimensional problems, it has been shown to be more efficient to use anisotropic exploration [9, 24] and we use $D(X_n^{\alpha,\beta} - X_n^i) = D^{aniso}(X_n^{\alpha,\beta} - X_n^i)$.

The change of the velocity in (3.1) is subject to three forces. The first term is the friction term. It depends only on the speed of the i-th particle in the previous time step. The second term is the exploitation term, which models the drift of the i-th particle towards the consensus point $X_n^{\alpha,\beta}$. The contribution of exploitation to acceleration depends on the distance of the particle from $X_n^{\alpha,\beta}$. The third term is the exploration term and it is the only random contribution.

The system (3.1) is an Euler-Maruyama time discretization of the system of stochastic differential equations (SDEs) in Itô form [30]

$$\begin{cases} dX_t^i = V_t^i dt, \\ dV_t^i = -\frac{\gamma}{m} V_t^i dt + \frac{\lambda}{m}(X_t^{\alpha,\beta} - X_t^i)dt + \frac{\sigma}{m} D(X_t^{\alpha,\beta} - X_t^i)dB_t^i, \end{cases} \quad \text{for } i = 1, ..., N,$$

where $X_t^{\alpha,\beta}$ is defined as (3.2) and $\{(B_t^i)_{t\geq 0}\}_{i=1}^N$ are N independent d-dimensional standard Brownian motions in $\mathbb{R}^d$.

3.2 Mean-field limit and macroscopic scale

We derive a mean-field formulation and the macroscopic system associated with the system (3.1). On the mesoscopic scale, a probabilistic approach is considered through a distribution function $f = f(t,x,v) \in \mathcal{P}(\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d)$ that represents the density of particles in the phase-space (x,v). The behavior of the particles is predicted through the study of statistical quantities of f. For the derivation, refer to [32].
The mean-field equation of the system (3.1) is

$$\partial_t f + v \cdot \nabla_x f = \nabla_v \cdot \left(\frac{\gamma}{m} v f + \frac{\lambda}{m}(x - X_{\alpha,\beta}(\rho)) f + \frac{\sigma^2}{2m^2} D(x - X_{\alpha,\beta}(\rho))^2 \nabla_v \cdot f \right) \tag{3.5}$$

where

$$X_{\alpha,\beta}(\rho) := \frac{\int_{\mathbb{R}^d} x\omega_{\alpha,\beta}(x)\rho(t,x)dx}{\int_{\mathbb{R}^d} \omega_{\alpha,\beta}(x)\rho(t,x)dx} \quad \text{and} \quad \rho = \rho(t,x) = \int_{\mathbb{R}^d} f(t,x,v)dv. \tag{3.6}$$

The mean-field equation can be written in compact form as

$$\partial_t f + v \cdot \nabla_x f + Q[f] = 0,$$

where

$$Q[f] := -\nabla_v \cdot \left(\frac{\gamma}{m} v f + \frac{\lambda}{m}(x - X_{\alpha,\beta}(\rho)) f + \frac{\sigma^2}{2m^2} D(x - X_{\alpha,\beta}(\rho))^2 \nabla_v \cdot f \right)$$

is the collisional operator. It acts as a non-local operator on f only in velocity and models how binary collisions between particles affect the evolution of the distribution function f, accounting for both the gain and loss of particles in a given velocity state due to these interactions [12]. We observe that the collisional operator Q satisfies the local conservation property for $\phi(v) := \begin{pmatrix} 1 \\ v \end{pmatrix}$, i.e.,

$$\langle \phi(v), Q(g) \rangle = 0 \qquad \forall g \in \mathcal{P}(\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d).$$

The components of the vector ϕ are the collisional invariants, namely the locally conserved quantities in (t,x). Therefore we define the vector $U = U(t,x)$ as

$$U(t,x) = \begin{pmatrix} \rho \\ \rho u \end{pmatrix}(t,x) := \int_{\mathbb{R}^d} f(t,x,v) \begin{pmatrix} 1 \\ v \end{pmatrix} dv.$$

To obtain equations for U of the macroscopic system, we integrate the mean-field equation (3.5) against the collisional invariants $\phi(v)$ and obtain

$$\partial_t \begin{pmatrix} \rho \\ \rho u \end{pmatrix} + \nabla_x \cdot \begin{pmatrix} \rho u \\ \int v^2 f(t,x,v)dv \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{\gamma}{m}(\rho u) + \frac{\lambda}{m}(x - X_{\alpha,\beta}(\rho))\rho \end{pmatrix} = 0. \tag{3.7}$$

The initial data $U_0(x) = U(0,x)$ is given by

$$U_0(x) = \begin{pmatrix} \rho_0 \\ (\rho u)_0 \end{pmatrix}(x) = \int_{\mathbb{R}^d} f(0,x,v)\phi(v)dv.$$

In its compact form, system (3.7) is written as

$$\partial_t U + \nabla_x \cdot F(U) + S(U) = 0,$$

where

$$U := \begin{pmatrix} \rho \\ \rho u \end{pmatrix} \text{ is the vector of conserved variables,} \tag{3.8}$$

$$F(U) := \begin{pmatrix} \rho u \\ \int v^2 f(t,x,v)dv \end{pmatrix} \text{ is the flux vector,}$$

$$\text{and } S(U) := \begin{pmatrix} 0 \\ \frac{\gamma}{m}\rho u + \frac{\lambda}{m}(x - X_{\alpha,\beta}(\rho))\rho \end{pmatrix} \text{ is the source term,} \tag{3.9}$$

with $X_{\alpha,\beta}(\rho)$ defined as in (3.6). The derivation of the macroscopic system is conducted by first starting from the mean-field equation and then exploiting the properties of fluid dynamics, as referenced in [3, 12].
The system (3.7) depends on the distribution f. To eliminate this dependence, a closure relation can be imposed by selecting a specific distribution, $f^* = f^*(t, x, v)$, which ensures that the conservation laws are hyperbolic or strictly hyperbolic. By expressing the system (3.7) in quasi-linear form, it is possible to determine the properties that the function f^* must satisfy.
In order to derive the quasi-linear form, it is necessary to express the spatial derivative of the flows $\nabla_x \cdot F(U)$, as a function of the derivatives of U. To accomplish this, the chain rule is applied component by component

$$\frac{\partial F_i(U)}{\partial x_i} = \frac{\partial F_i(U)}{\partial U}\partial_{x_i} U = A_i(U)\partial_{x_i} U, \qquad i = 1, ..., d.$$

For the system (3.7), we obtain the following:

$$A_i(U) := \frac{\partial}{\partial U} F_i(U) = \frac{\partial}{\partial U} \begin{pmatrix} \rho u \\ \int v^2 f(t,x,v) dv \end{pmatrix}_i = \begin{pmatrix} 0 & e_i^T \\ \partial_\rho \int v^2 f^* dv & \partial_{\rho u} \int v^2 f^* dv \end{pmatrix}. \tag{3.10}$$

Accordingly, the system (3.7) is represented in quasi-linear form as

$$\partial_t U + \sum_{i=1}^{d} A_i(U)\partial_{x_i} U + S(U) = 0 \tag{3.11}$$

where U, $S(U)$ and $A_i(U)$ are defined in (3.8), (3.9) and (3.10), respectively.
In order to obtain a well-posed initial value problem, we introduce a basic algebraic property.

Definition 1. *A system of balance laws*

$$\partial_t U + \sum_{i=1}^{d} A_i(U)\partial_{x_i} U + S(U) = 0$$

in $\mathbb{R}^n$*, with initial data* $U(x, 0) = U_0(x)$*, is strictly hyperbolic if, for every* U*, the composite Jacobian matrix in the direction* ξ

$$A(\xi) = \sum_{i=1}^{d} A_i(U)\xi_i = \sum_{i=1}^{d} \frac{\partial F_i}{\partial U}\xi_i$$

has n *real and distinct eigenvalues:* $\lambda_1(U) < ... < \lambda_n(U)$*.*

From now on, we focus on problems in dimension 1. It is possible to derive a sufficient condition for the system (3.11) to be strictly hyperbolic.

Lemma 3.1. *In dimension 1, the system is strictly hyperbolic if the following condition is satisfied*

$$\left(\partial_{\rho u} \int v^2 f^* dv\right)^2 > 4\partial_\rho \int v^2 f^* dv. \tag{3.12}$$

Proof. In order to identify a sufficient condition for strict hyperbolicity, which f^* must satisfy, we compute the eigenvalues of the matrix $A(U)$. In dimension 1 the system in quasi-linear form is

$$\partial_t U + A(U)\partial_x U + S(U) = 0, \quad \text{with} \quad A(U) = \begin{pmatrix} 0 & 1 \\ \partial_\rho \int v^2 f^* dv & \partial_{\rho u} \int v^2 f^* dv \end{pmatrix}. \tag{3.13}$$

We proceed by calculating the eigenvalues of $A(U)$

$$det(A(U) - \lambda Id) = 0,$$

$$det \begin{pmatrix} -\lambda & 1 \\ \partial_\rho \int v^2 f^* dv & \partial_{\rho u} \int v^2 f^* dv - \lambda \end{pmatrix} = 0,$$

$$\lambda^2 - \partial_{\rho u} \int v^2 f^* dv \lambda - \partial_\rho \int v^2 f^* dv = 0.$$

To obtain real and distinct eigenvalues, it is necessary that the discriminant of the last quadratic equation be strictly positive, i.e.

$$\left(\partial_{\rho u}\int v^2 f^* dv\right)^2 - 4\partial_\rho \int v^2 f^* dv > 0.$$

It thus follows that in order for the system (3.13) to be strictly hyperbolic and therefore well-posed, the distribution f^* selected for closure must satisfy the following condition:

$$\left(\partial_{\rho u}\int v^2 f^* dv\right)^2 > 4\partial_\rho \int v^2 f^* dv.$$

□

In dimension 1, a potential selection for f^* that satisfies (3.12) is the Maxwellian distribution with parameter $T \in \mathbb{R}^+$

$$f^*(t,x,v) = \frac{\rho}{\sqrt{2\pi T^2}} \exp\left(-\frac{1}{2}\frac{(v-u(t,x))^2}{T^2}\right), \tag{3.14}$$

where $u(t,x) := \frac{\rho u(t,x)}{\rho(t,x)}$ indicates the average value of the velocity along the direction v, at time t in x, and T^2 is the variance of the velocity. T is chosen in $\mathbb{R}^+$ because it represents the standard deviation.
Using the expected value and variance properties of the normal distribution, we obtain $\int v^2 f^* dv = \rho u^2 + \rho T^2$. In this case the system of differential equations for U is

$$\partial_t \begin{pmatrix} \rho \\ \rho u \end{pmatrix} + \partial_x \begin{pmatrix} \rho u \\ \rho u^2 + \rho T^2 \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{\gamma}{m}(\rho u) + \frac{\lambda}{m}(x - X_{\alpha,\beta}(\rho))\rho \end{pmatrix} = 0 \tag{3.15}$$

and with initial data U_0

$$U_0(x) = \begin{pmatrix} \rho_0 \\ (\rho u)_0 \end{pmatrix}(x) = \int_{\mathbb{R}^d} f(0,x,v)\phi(v)dv.$$

The associated quasi-linear system can be expressed as follows:

$$\partial_t U + A(U)\partial_x U + S(U) = 0 \tag{3.16}$$

with $U = \begin{pmatrix} \rho \\ \rho u \end{pmatrix}$, $A(U) = \begin{pmatrix} 0 & 1 \\ T^2 - u^2 & 2u \end{pmatrix}$ and $S(U) = \begin{pmatrix} 0 \\ \frac{\gamma}{m}(\rho u) + \frac{\lambda}{m}(x - X_{\alpha,\beta}(\rho))\rho \end{pmatrix}$.
The eigenvalues of the matrix $A(U)$ are given by

$$\lambda_1 = u + |T| \quad \text{and} \quad \lambda_2 = u - |T|.$$

The system (3.16) is strictly hyperbolic for $T \neq 0$, which implies that the system (3.15) is well posed [37].

Remark 1. *It should be noted that the monokinetic closure $f^* = \rho\delta(v-u)$ does not guarantee the strict hyperbolicity of the system, as it does not satisfy condition* (3.12). *Furthermore, it does not guarantee hyperbolicity, as the matrix $A(U) = \begin{pmatrix} 0 & 1 \\ -(\rho u)^2/\rho^2 & 2(\rho u)/\rho \end{pmatrix}$, obtained in the quasi-linear form, is not diagonalizable.*

Remark 2. *It is also possible to derive a third-order macroscopic system. In this case we will have that the vector of conserved quantities is defined as $\phi(v) := \begin{pmatrix} 1 \\ v \\ v^2 \end{pmatrix}$ and the vector U is defined as*

$$U(t,x) = \begin{pmatrix} \rho \\ \rho u \\ \rho T^2 \end{pmatrix}(t,x) := \int_{\mathbb{R}^d} f(t,x,v) \begin{pmatrix} 1 \\ v \\ v^2 \end{pmatrix} dv.$$

Again, the Maxwellian distribution (3.14) *can be chosen as the closure. Given* U_0 *initial data, we obtain the third-order macroscopic system in quasi-linear form as*

$$\partial_t U + A(U)\partial_x U + S(U) = 0,$$

$$\textit{with} \quad U = \begin{pmatrix} \rho \\ \rho u \\ \rho T^2 \end{pmatrix}, \quad A(U) = \begin{pmatrix} 0 & 1 & 0 \\ -u^2 & 2u & 1 \\ -3uT^2 & 3T^2 & u \end{pmatrix} \quad \textit{and}$$

$$S(U) = \begin{pmatrix} 0 \\ \frac{\gamma}{m}(\rho u) + \frac{\lambda}{m}(x - X_{\alpha,\beta}(\rho))\rho \\ 2\frac{\gamma}{m}\rho T^2 \end{pmatrix}.$$

The system is strictly hyperbolic since the eigenvalues are real and distinct

$$\lambda_1 = u, \quad \lambda_2 = u + \sqrt{3}T \quad \textit{and} \quad \lambda_3 = u - \sqrt{3}T.$$

In conclusion, two formulations were proposed to solve the problems presented in (1.1) and in (2.1). The initial approach was the microscopic system (3.1), which is based on the dynamics of individual particles. The evolution of the particles is given by the evolution of their positions and velocities at each time n. The velocity is updated based on the previous velocity of the particle and the distance of the particle from the consensus point, which is calculated at each time instant. The convergence of this system is governed by the relationship $X_{\alpha,\beta} \to argmin\mathcal{F}_\beta$.

The second approach was to solve the minimization problem using the macroscopic system (3.15), which is based on the evolution of U. We also derived the quasi-linear formulation of these conservation laws. In addition, we proved a criterion that the closure distributions must meet in dimension 1 to guarantee the strict hyperbolicity of the system. In the simulations in Section 5, tests obtained by implementing the conservation laws in dimension 1 and with the closure relation obtained using the Maxwellian distribution (3.14) are presented. In this case, convergence is given by the concentration of the mass of the system around the minimizer $x^*_{\mathcal{M}}$, i.e. $supp(\rho) \to argmin\mathcal{F}_\beta$.

4 Micro-macro decomposition

We propose an approach based on a formal micro-macro decomposition of the kinetic probability density to exploit the scales. The approach develops the idea of studying and comparing the behavior of the PSO method applied to the microscopic scale and the macroscopic scale, which has been presented in Section 3. Specifically, the decomposition is employed to calculate the two dynamics (microscopic and macroscopic) and couple them, with the objective of efficiently moving mass between the two scales to achieve method convergence.
The micro-macro decomposition typically leads to more efficient computational methods by reducing the number of particles to be simulated. Additionally, merging the two dynamics leverages the fact that the macroscopic model is defined across the entire spatial domain. This property enables the computation of consensus points that are a good approximation of the minimizer $x^*_{\mathcal{M}}$, even for small times.

The idea of this approach is inspired by [16, 18, 38], where micro-macro decomposition is used to decompose the spatial domain to match fluid and kinetic models. Here, we use the micro-macro approach directly on the distribution function $f = f(t, x, v) \in \mathcal{P}(\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d)$, introduced in (3.5).
The starting point is to formally split the distribution function f in (3.5) in a microscopic contribution $\hat{f} = \hat{f}(t, x, v) \in \mathcal{P}(\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d)$ and a macroscopic contribution, specifically, we write using the closure of the macroscopic system $f^* = f^*(t, x, v) \in \mathcal{P}(\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d)$ (3.14)

$$f(t, x, v) = \zeta(t, x)\hat{f}(t, x, v) + (1 - \zeta(t, x))f^*(t, x, v), \qquad \zeta(t, x) \in [0, 1],$$

where $(x, v) \in (\mathbb{R}^d \times \mathbb{R}^d)$. To compute the evolution of the distribution, we do not solve the mean-field equation but instead use particle dynamics and the macroscopic system. At the microscopic level the evolving quantities are the positions and velocities of the particles, so the microscopic contribution is given by the empirical distribution associated with the system (3.1)

$$\hat{f}(t,x,v) = \frac{1}{N}\sum_{i=1}^{N}\delta(x - X_n^i)\delta(v - V_n^i).$$

Instead, the macroscopic contribution is made by exploiting the closure relation

$$f^*(t,x,v) = \frac{\rho(t,x)}{\sqrt{2\pi T^2}}\exp\left(-\frac{1}{2}\frac{(v-u(t,x))^2}{T^2}\right),$$

which requires to propagate $(\rho, \rho u)^T$ accordingly to (3.15). The weight of the total distribution f allocated to the microscopic and macroscopic scales is given by the function $\zeta(t,x)$. Therefore, the role of $\zeta(t,x)$ is to establish a connection between the two dynamics. The coupling is determined by the movement of mass shift between the microscopic and macroscopic scales. There are several possible choices for this function. The approach employed in this study involves the partitioning of the spatial domain into cells. The j-th cell of the spatial grid is denoted by I_j. The value of the function ζ is given by

$$\zeta(t,x) := \frac{\sum_{j=1}^{K} w(t,I_j)|u(t,I_j) - \sum_{i=1}^{N}\mathbb{1}_{I_j}(X_t^i)\cdot V_t^i|}{\sum_{j=1}^{K} w(t,I_j)\max_{i=1,\dots,N}|u(t,I_j) - \mathbb{1}_{I_j}(X_t^i)\cdot V_t^i|}, \quad \forall x \in I_j, \tag{4.1}$$

where K is the number of cells, N is the number of particles, $u(t,x)$ is defined as

$$u(t,I_j) := \int I_j \frac{\rho u(t,x)}{\rho(t,x)}dx$$

and V_t^i are the velocities of the particles. The weight $w(t,I_j)$ is defined as the relative density

$$w(t,I_j) := \int_{I_j}\frac{\rho^m(t,x)}{\rho^M(t,x)+\rho^m(t,x)}dx,$$

where

$$\rho^m(t,x) = \int_{\mathbb{R}^d}\hat{f}(t,x,v)dv \quad \text{and} \quad \rho^M(t,x) = \int_{\mathbb{R}^d} f^*(t,x,v)dv, \quad \text{respectively.} \tag{4.2}$$

The definition of ζ is motivated by the following considerations. The difference $|u(t,I_j)-\sum_{i=1}^{N}\mathbb{1}_{I_j}(X_t^i)\cdot V_t^i|$ measures the distance between the macroscopic velocity defined by u (3.15) and the average velocity of the particles in the same cell (3.1). This quantity is significant since if microscopic velocities tend towards the chosen velocity of the macroscopic system, they become less relevant. Hence, we move mass to macroscopic scale. The distance is weighted by the local mass $w(t,I_j)$. Therefore, the effect of distance is determined by the microscopic mass in the cell I_j, and its contribution becomes more significant as the local mass increases. The denominator normalizes the quantity ζ.
The function ζ balances the total density $\rho(t,x) = \int_{\mathbb{R}^d} f(t,x,v)dv$ of the system. This balance is based on the comparison of the output of densities obtained at each scale at each time instant. If the microscopic system gives results that are far from those obtained by the macroscopic system, ζ will take large values. Then, the mass is shifted to the microscopic scale that is expected to be more accurate. In the opposite case, where the particle velocities have values close to those of the macroscopic velocity, the microscopic system is expected to be closer to the macroscopic system. Then, we expect a higher density at the macroscopic scale.
The detailed algorithm is given in Section 5.1.

5 Numerical algorithm and results

In this section, several simulations are presented. The numerical results obtained using the algorithm introduced at the microscopic and macroscopic scales are reported, both in the classical and in the constrained case. In addition, results are given for the micro-macro decomposition.

5.1 Constrained micro-macro PSO method

In Section 2 we introduced the concept of adaptive exact penalty in order to solve constrained minimization problems. This strategy succeeds in handling even non-differentiable functions. The idea is to minimize a modified objective function $\mathcal{F}_\beta$ over the entire space $\mathbb{R}^d$, i.e.

$$\min_{x\in\mathbb{R}^d} \mathcal{F}_\beta(x) := \mathcal{F}(x) + \beta r(x),$$

where $\beta \in \mathbb{R}^+$ and r is the penalty function, in our simulations is given by

$$r(x) = \min_{y\in\mathcal{M}} ||x-y||_p.$$

In this study, an iterative algorithm in β is considered, and the dependence on the interaction n is denoted by β_n. The evolution of the parameter β_n is based on the violation of the constraint r (2.2) by the microscopic/macroscopic system. This violation can be measured on the microscopic scale by

$$\mathcal{R}^m(X_n) := \frac{\sum_{i=1}^N r(X_n^i)\exp(-\alpha\mathcal{F}_\beta(X_n^i))}{\sum_{i=1}^N \exp(-\alpha\mathcal{F}_\beta(X_n^i))} \tag{5.1}$$

and on the macroscopic scale by

$$\mathcal{R}^M(x) := \frac{\int r(x)\exp(-\alpha\mathcal{F}_\beta(x))\rho(x)dx}{\int \exp(-\alpha\mathcal{F}_\beta(x))\rho(x)dx}, \text{ respectively.}$$

The violation is measured with a weighted expectation of r based on the Boltzmann-Gibbs distribution, the same one used for the calculation of the consensus point (3.2) (respectively (3.6)). This choice results in an adaptive mechanism that is more accurate than that obtained with other feasibility conditions, as shown in [7].
In addition, a tolerance value, denoted by κ_n, is introduced and subject to temporal variation. This value is used as a threshold for the constraint violations. In each temporal interval, it is necessary to ascertain whether the following feasibility condition is satisfied:

$$\mathcal{R}^m \le \frac{1}{\sqrt{\kappa_n}} \quad \text{or respectively} \quad \mathcal{R}^M \le \frac{1}{\sqrt{\kappa_n}}. \tag{5.2}$$

The formulation of the adaptive strategy is founded upon the evaluation of the validity of the inequality (5.2), which is tested at each time step, resulting in updated values of β_n and κ_n. If the condition (5.2) holds, it means that the violation of the constraint is contained within the imposed value $1/\sqrt{\kappa_n}$. Consequently, the variation of the parameter β_n, which governs the penalty contribution, is avoided

$$\beta_{n+1} = \beta_n.$$

However, the necessity arises to update the value of κ_n in order to test the inequality at the next time step under more restrictive conditions. The value κ_{n+1} is updated in the following way

$$\kappa_{n+1} = \eta_\kappa \kappa_n, \quad \text{with } \eta_\kappa > 1.$$

On the other hand, if the condition (5.2) is not satisfied, it means that the contribution of the violation $\mathcal{R}^M$ exceeds the established tolerance κ_n. It is therefore necessary to increase the parameter β_{n+1} in order to impose additional particles (or mass density, respectively) in the constraint

$$\beta_{n+1} = \eta_\beta \beta_n, \quad \text{with } \eta_\beta > 1.$$

The result is that the contribution of $\mathcal{R}^M$ is decreased. Consequently, the value of κ_{n+1} is reduced in order to test the inequality with a lower tolerance at the subsequent time. Therefore,

$$\kappa_{n+1} = \min\{\kappa_n/\eta_\kappa, \kappa_0\}, \quad \text{with } \eta_\kappa > 1.$$

The strategy is presented in Algorithm 1.

Algorithm 1 Algorithm for the micro-macro decomposition in the constrained case

Require: initial data $N_t, \alpha, \beta, X_0,\ \rho_0,\ \zeta_0,\ \mu_0$
Require: parameters of the penalized microscopic algorithm $\kappa_0^m, \eta_\kappa^m, \beta_0^m, \eta_\beta^m$
Require: parameters of the penalized macroscopic algorithm $\kappa_0^M, \eta_\kappa^M, \beta_0^M, \eta_\beta^M$
compute the density ρ_0^m associated to the microscopic scale (4.2)
for $n = 1, ..., N_t$ **do**
  % penalized microscopic algorithm
  apply the Euler-Maruyama scheme (3.1)
  $X_n^{\alpha,\beta} = \frac{1}{N_n^{\alpha,\beta}} \sum_{i=1}^N X_n^i \omega^{\alpha,\beta}(X_n^i)$
  if microscopic feasibility condition (5.2) holds **then**
    $\kappa_{n+1}^m = \eta_\kappa^m \kappa_n^m$
    $\beta_{n+1}^m = \beta_n^m$
  else
    $\kappa_{n+1}^m = \min\{\kappa_n^m/\eta_\kappa^m, \kappa_0^m\}$
    $\beta_{n+1}^m = \eta_\beta^m \beta_n^m$
  end if
  % penalized macroscopic algorithm
  apply the Lax-Friedrichs scheme (5.3) for the system (3.15)
  $X_{\alpha,\beta}(\rho) = \frac{\int_{\mathbb{R}^d} x\omega_{\alpha,\beta}(x)\rho(t,x)dx}{\int_{\mathbb{R}^d} \omega_{\alpha,\beta}(x)\rho(t,x)dx},$
  if macroscopic feasibility condition (5.2) holds **then**
    $\kappa_{n+1}^M = \eta_\kappa^M \kappa_n^M$
    $\beta_{n+1}^M = \beta_n^M$
  else
    $\kappa_{n+1}^M = \min\{\kappa_n^M/\eta_\kappa^M, \kappa_0^M\}$
    $\beta_{n+1}^M = \eta_\beta^M \beta_n^M$
  end if
  % micro-macro algorithm
  compute ζ_n according to (5.4)
  $\mu_n = \zeta_n \cdot \mu_0$ % update of the mass of particles (μ_0 is the initial mass of particles)
  compute the density ρ_n^m associated to the microscopic scale (4.2)
  if $\mu_n \leq \mu_{n-1}$ **then**
    $\rho_n^M = \rho_n^M + (\rho_n^m - \rho_{n-1}^m)$
  else
    $\rho_n^M = \rho_n^M - (\rho_n^m - \rho_{n-1}^m)$
  end if
end for

Remark 3. *The iterative strategy previously outlined has been implemented and its convergence is demonstrated for another meta-heuristic method for solving minimization problems, namely the Consensus-Based Optimization method, as detailed in [7].*

The penalized adaptive strategy presented for solving constrained problems can be used, as already mentioned, on both the microscopic and macroscopic scales. In the former case, the positions and velocities of the particles are calculated using the Euler-Maruyama scheme (3.1). Instead, on the macroscopic scale, the quantity that evolves is U. To compute it, we discretize the system (3.15) using the Lax-Friedrichs scheme

$$U_{n+1}^{\ell} = \frac{1}{2}(U_n^{\ell+1} + U_n^{\ell-1}) - \frac{\Delta t}{2\Delta x}(F(U_n^{\ell+1}) - F(U_n^{\ell-1})) + \Delta t S(U_n^{\ell}), \tag{5.3}$$

with $\ell = 1, ..., L$ are the grid indices on which the scheme is applied, where the total number of nodes is L, $U_n^{\ell} := \begin{pmatrix} \rho_n^{\ell} \\ \rho u_n^{\ell} \end{pmatrix}$, $F(U_n^{\ell}) := \begin{pmatrix} \rho u_n^{\ell} \\ (\rho u^2 + \rho T^2)_n^{\ell} \end{pmatrix}$, $S(U_n^{\ell}) := \begin{pmatrix} 0 \\ \frac{\gamma}{m}\rho u_n^{\ell} + \frac{\lambda}{m}(x^{\ell} - X_{\alpha,\beta}(\rho_n^{\ell}))\rho_{n+1}^{\ell} \end{pmatrix}$ and x_{ℓ} are the points of the spatial grid defined on the domain of the objective function $\mathcal{F}_{\beta}$.

We proceed to present the algorithm for micro-macro decomposition. Consider the local densities defined over a domain. Then, introduce a spatial grid over the domain, dividing it into cells. The j-th cell of the spatial grid is denoted by I_j. The value of the function ζ is computed of each cell in the grid, so that

$$\zeta(t, I_j) := \frac{\sum_{j=1}^{K}\sum_{i=1}^{N} w(t, I_j)|u(t, I_j) - \mathbb{1}_{I_j}(X_t^i) \cdot V_t^i|}{\sum_{j=1}^{K} w(t, I_j) \max_{i=1,\dots,N} |u(t, I_j) - \mathbb{1}_{I_j}(X_t^i) \cdot V_t^i|}. \tag{5.4}$$

where $u(t, I_j) := \int I_j \frac{\rho u(t,x)}{\rho(t,x)} dx$ and $w(t, I_j) := \int_{I_j} \frac{\rho^m(t,x)}{\rho^M(t,x)+\rho^m(t,x)} dx$. To compute the quantities u and w, we used a first-order quadrature rule.

The basic steps of micro-macro decomposition, introduced for the PSO method, are as follows:

- fix the initial mass of the particles $\rho(0, x) = \rho^m(0, x) + \rho^M(0, x)$;
- at each time step compute the new value of the particle mass. This implies a total mass change in the microscopic scale;
- add or remove from the macroscopic scale the density lost or gained from the microscopic scale, so as to have the conservation property of the total mass of the system.

The strategy is presented in detail in Algorithm 1.

Remark 4. *It is important to note that, to modify the mass on a microscopic scale, it is not necessary to alter the number of particles; rather, their weight can be varied. This approach ensures the conservation of the total number of particles, thus obviating the issues associated with particle sampling.*

Remark 5. *In order for ζ to provide a correct comparison of the results obtained in the two scales and thus to allow the mass to be shifted correctly, we impose a minimum threshold value of the density that must be present in each scale. This means imposing threshold values on the function ζ:*

$$\zeta_{min} \leq \zeta(t, x) \leq \zeta_{max}, \quad \forall t \in [0, T], \forall x \in \mathbb{R}^d \text{ and with } 0 < \zeta_{min} < \zeta_{max} < 1. \tag{5.5}$$

Remark 6. *A potential issue that can arise from this type of micro-macro decomposition is related to the continuous and uncontrolled movement of mass between the two scales. This phenomenon is not intrinsically associated with the properties of the analyzed systems; rather, it arises from the adaptation of the microscopic and macroscopic algorithms in the presence of a mass shift.*

Also, it may be advantageous to refrain from shifting mass for small times, thereby ensuring that the algorithm initially compute results with fixed mass. In fact, it has been empirically observed that strong mass fluctuations occur for small times.

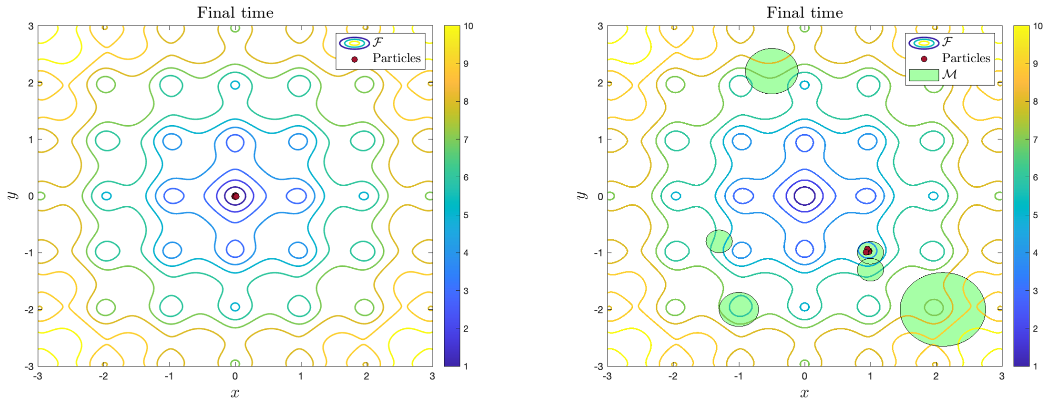
5.2 Numerical results

5.2.1 Constrained microscopic system

We started by testing the algorithm on the microscopic scale using (3.1) to compute the evolution of the position and velocity of each particle. We consider the minimization problem for the Ackley function

$$\min_{x\in\mathbb{R}^d} \mathcal{F}(x) := -20\exp\left(-0.2\sqrt{\frac{1}{d}\sum_{i=1}^{d}(x_i)^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d} cos(2\pi(x_i))\right) + 20 + e \tag{5.6}$$

that is continuous, non-differentiable and non-convex and has its global minimizer in $x^* = (0, ..., 0)$. We use $N = 480$ particles for the PSO discretization in a two-dimensional search space ($d = 2$).



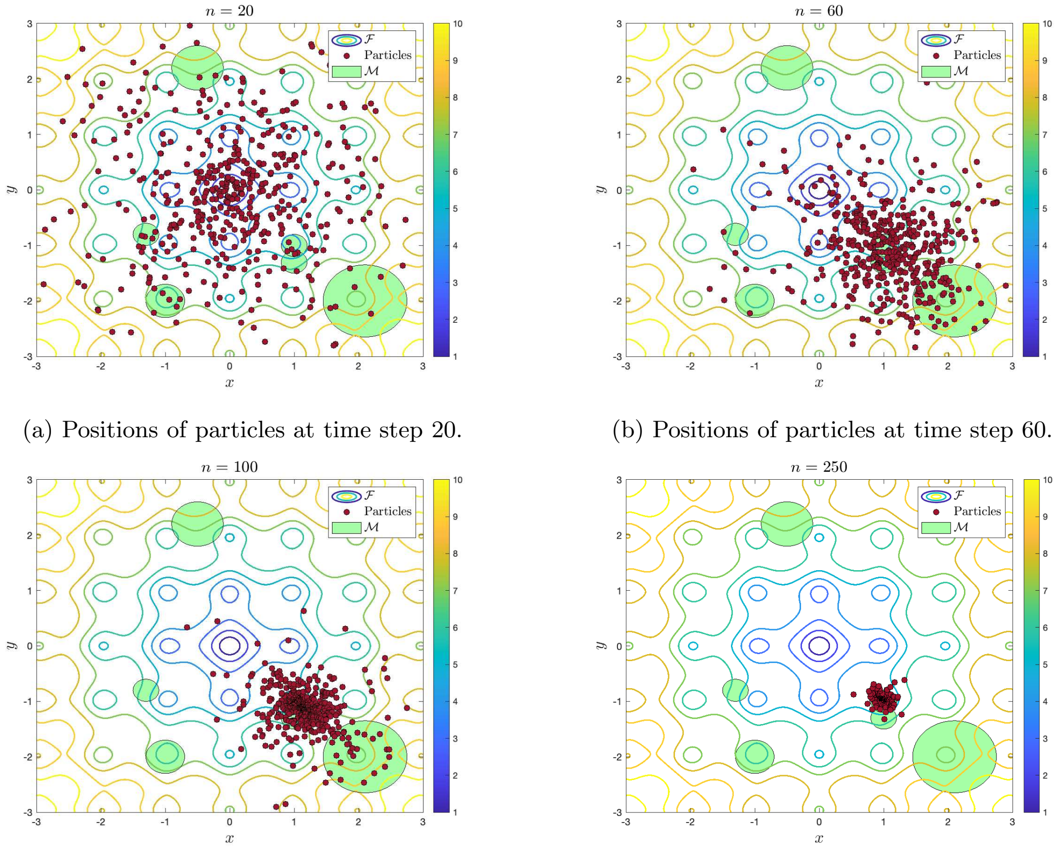
(a) The values of the objective function $\mathcal{F}(x)$ (5.6) and the final positions of the particles are shown in the unconstrained problem.

(b) The values of the objective function $\mathcal{F}(x)$ (5.6), the final positions of the particles, and the feasible set $\mathcal{M}$, which is defined as the union of the circles in green, are shown in the constrained problem.

Figure 1: Evolution of particles in the unconstrained 1(a) and constrained 1(b) case, starting from the same initial conditions. The particles in the unconstrained case 1(a) converge to the global minimizer x^*, while in the constrained case 1(b) the particles converge to the global minimizer $x^*_{\mathcal{M}}$, that is feasible for the set $\mathcal{M}$ chosen in the simulation.

We tested the classical microscopic algorithm and the constrained algorithm on the same initial data. The following results are all obtained with the following parameters: $\gamma = 0.5$, $\lambda = 1$, $\sigma = \frac{1}{\sqrt{3}}$, $\alpha = 30$. The parameter values used for the penalized algorithm are as follows: $\kappa_0 = 5$, $\eta_\kappa = 1.1$, $\beta_0 = 1$, $\eta_\beta = 1.1$. The anisotropic diffusion is used, see (3.4). In the constrained case we consider the admissible set

$$\begin{aligned}\mathcal{M} = \{x \in \mathbb{R}^d | &(x_1 + 0.5)^2 + (x_2 - 2.2)^2 \leq 0.4 \cup \\ &(x_1 - 1.3)^2 + (x_2 + 0.8)^2 \leq 0.2 \cup \\ &(x_1 - 1)^2 + (x_2 + 1.3)^2 \leq 0.1 \cup \\ &(x_1 - 1)^2 + (x_2 + 1)^2 \leq 0.1 \cup \\ &(x_1 - 2.1)^2 + (x_2 + 2)^2 \leq 0.65 \cup \\ &(x_1 + 1)^2 + (x_2 + 2)^2 \leq 0.3\}.\end{aligned}$$



(a) Positions of particles at time step 20.

(b) Positions of particles at time step 60.

(c) Positions of particles at time step 100.

(d) Positions of particles at time step 250.

Figure 2: Plots show the evolution of particle positions. For small time instants, the penalty is not large enough to direct the particles toward the feasible global minimizer $x^*_{\mathcal{M}}$. As time evolves, the particles concentrate in $x^*_{\mathcal{M}}$.

As expected, the particles at the final time $t = 400$ (shown in Figure 1(a)), in the absence of a penalty, accumulate in the global minimizer x^*. In contrast, in the penalized case, all particles are attracted to the minimizer within the feasible set, $x^*_{\mathcal{M}}$, as illustrated in Figure 1(b).
We have plotted the positions at different intermediate time instants before convergence to the minimizer inside $\mathcal{M}$ (Figure 2). It is evident that initially the particles tend to converge to the global minimizer which is outside the constraint $\mathcal{M}$. This phenomenon can be attributed to the fact that the initial value of β is substantially distant from the optimal value. Specifically, the initial value of β is set at $\beta_0 = 1$ and the final value of β attained by the algorithm is $\bar{\beta} \approx 6.1$ (see Figure 3 for a visual representation). This indicates that initially the value of the penalty term is insufficient to direct the particles within the constraint set. As time evolves, the particles reach an intermediate configuration in which they remain in proximity to the global constraint while expanding in space. Finally, as the value of β approaches the optimal value, the particles reach the feasible global minimizer $x^*_{\mathcal{M}}$.
Figure 3 shows the evolution of the value of the parameter β (in orange) and the constraint violation (in blue). The latter exhibits a decreasing trend, as particles tend to concentrate around the feasible global minimizer as time progresses. Consequently, the contributions of the penalty function in (5.1) tend to be negligible. It is evident that there is a significant decrease in the value of the feasibility condition corresponding to the instant at which the value of β increases. This marks the point in time at which particles begin to be attracted to the feasible global minimizer $x^*_{\mathcal{M}}$ rather than the global minimizer x^* as previously.

We now focus on the evolution of the consensus point. In Figure 2, we deduced that the movement of particles from the central region of the figure toward the constraint is indicative of a better approximation of the consensus point as time evolves. To further investigate this

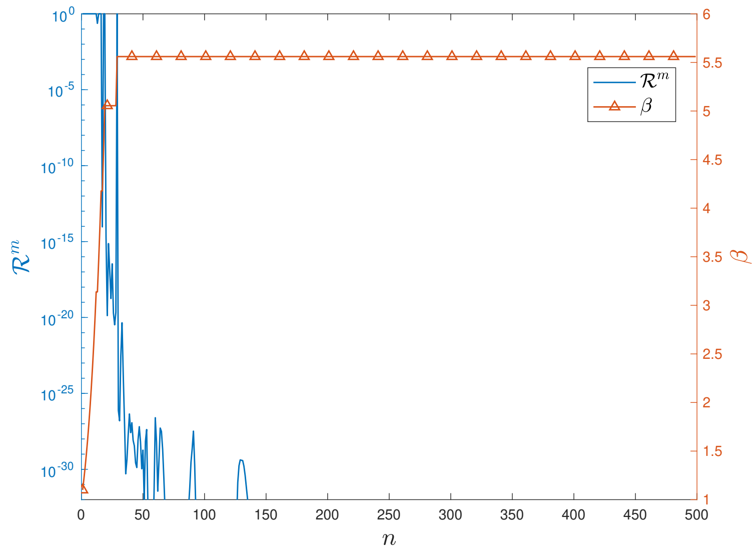


Figure 3: The evolution of two quantities is represented: the constraint violation $\mathcal{R}^m$ (5.1) and the penalty parameter β. To effectively represent the quantity $\mathcal{R}^m$, its y-axis is set to a semilogarithmic scale. As we expect, the constraint violation decreases over time, while the value of β increases.

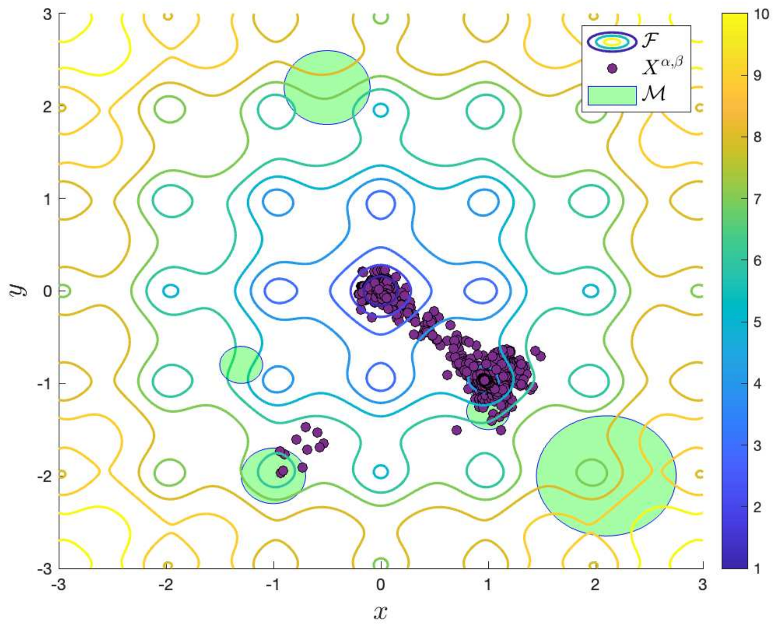


Figure 4: The plot illustrates the evolution of the consensus points obtained from 100 runs of the algorithm. It is evident in the figure, the $X^{\alpha,\beta}$ trend follows the direction of the gradient descent, whereby the algorithm initiates from the global minimizer x^* of the objective function and arrives at the feasible global minimizer $x^*_{\mathcal{M}}$.

phenomenon, we executed 100 runs of the PSO algorithm. As illustrated in Figure 4, the evolution of the consensus point exhibits characteristics reminiscent of a gradient-based method, despite the implementation of the PSO method, which is inherently gradient-free. This phenomenon has previously been examined for the CBO method (which is connected to PSO via a zero inertia limit), in the context of solving unconstrained minimization problems, as outlined in [42].

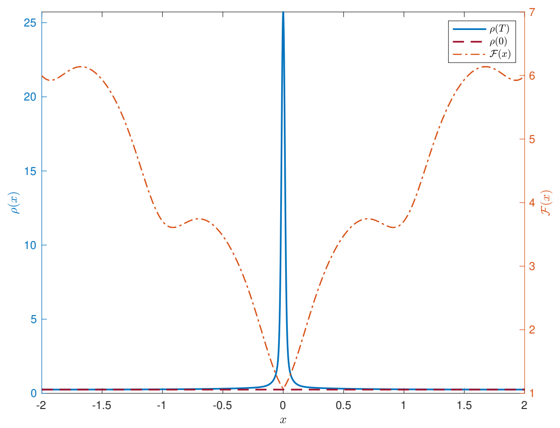
The presented method can be further evaluated based on the results already obtained for the PSO and CBO optimization methods. In [6], a comparison of the convergence rates of the two methods without penalty is provided. The PSO method has been demonstrated to be less stable and reliable when dealing with difficult objective function, in contrast to the CBO method, which has been shown to be more robust. We expect these results to remain unchanged when comparing CBO and PSO with penalization.

5.2.2 Constrained macroscopic system

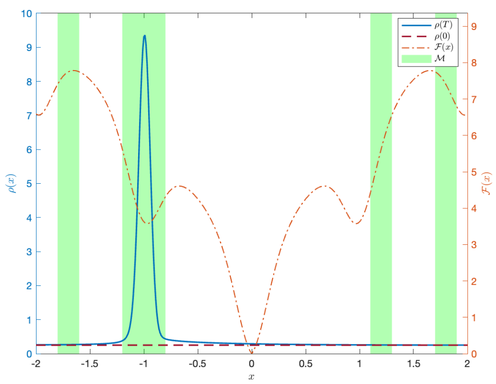
We proceed to present the results of the simulations obtained in the macroscopic case. We once again consider the minimization problem of the Ackley function (see (5.6)) on a one-dimensional space with a grid of 401 points in space and with a constraint given by

$$\mathcal{M} = \{x \in \mathbb{R} | -1.8 \leq x \leq -1.6 \cup -1.2 \leq x \leq -0.8 \cup 1.1 \leq x \leq 1.3 \cup 1.7 \leq x \leq 1.9\}.$$

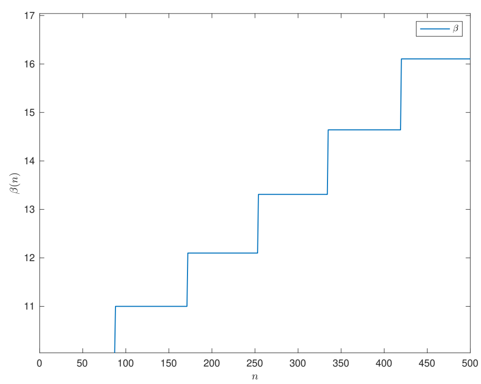
The same set of parameters employed in the microscopic case is also utilized here. The CFL condition chosen for all macroscopic simulations is equal to 0.8. In the simulations, we use the second-order macroscopic system by applying the Lax-Friedrichs numerical scheme (5.3).



(a) In the unconstrained minimization problem, we see that the density ρ converges to a Dirac delta centered in 0, which is the global minimizer x^*.



(b) In the constrained minimization problem, we see that the density ρ converges to a Dirac delta centered in -1, which is the feasible global minimizer $x^*_{\mathcal{M}}$.



(c) Evolution of the value of the penalty strength β that has an increasing trend, as expected.

Figure 5: Plots obtained by applying the PSO method to constrained and unconstrained minimization problems in which the objective function is the Ackley function.

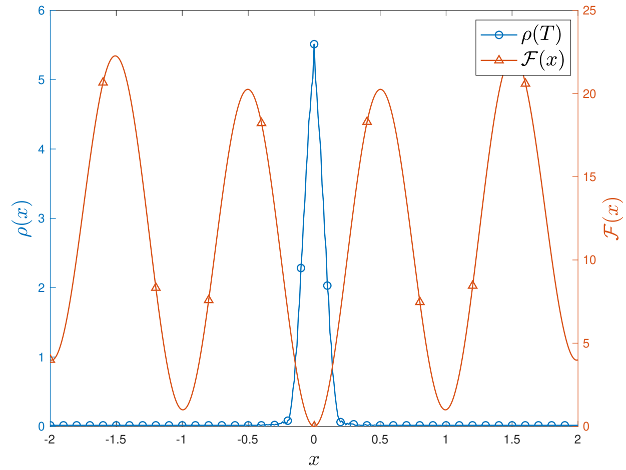
Figure 5(a) and Figure 5(b) show the initial and final densities obtained by the algorithm in the unconstrained and constrained cases. As previously highlighted, in the unconstrained scenario depicted in Figure 5(a), the mass is concentrated at the global minimizer x^*, in fact the final density tends to a Dirac delta centered at x^*. Instead in the constrained case 5(b), the final density tends to a Dirac delta centered at the minimizer within the feasible set $\mathcal{M}$.
The time evolution of the penalty parameter β is shown in Figure 5(c). The value of the parameter increases over time, which is a feature of the algorithm. This phenomenon is analogous to that observed in the microscopic algorithm (Figure 3). It is important to note that the penalized problem we are solving yields the same result as the constrained problem we are trying to solve.

5.2.3 Micro-macro decomposition

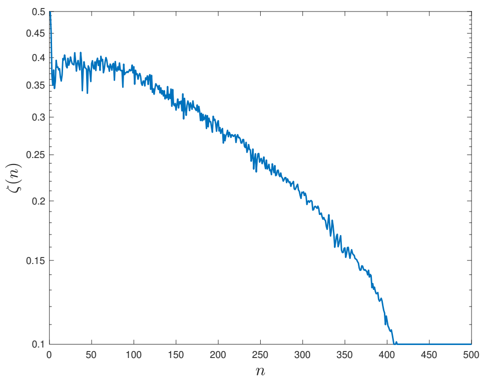
We consider the minimization problem for the Rastrigin function

$$\min_{x\in\mathbb{R}^d} \mathcal{F}(x) := 10d + \sum_{i=1}^{N} [x_i^2 - 10cos(2\pi x_i)] \tag{5.7}$$

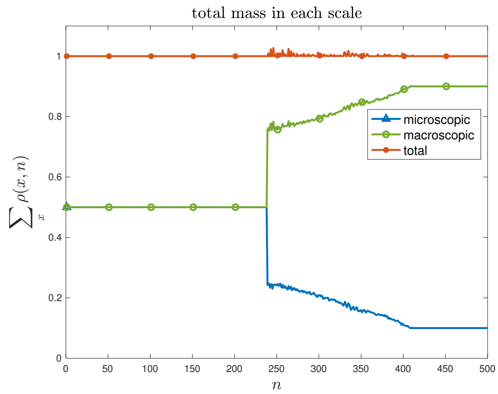
that is continuous, differentiable and non-convex, has many local minima and a global minimizer in $x^* = (0, ..., 0)$. We used $N = 480$ particles for PSO discretization for a one-dimensional search space ($d = 1$). In all the simulations following that utilize micro-macro decomposition, the threshold values (5.5) are set to $\zeta_{min} = 0.1$ and $\zeta_{max} = 0.9$. The value of ζ at the initial instant is fixed at 0.5, ensuring equivalent initial density in microscopic and macroscopic systems. As mentioned in Remark 6, to improve the stability of the algorithm, we avoid mass shifts from the first time steps. In the simulations, we set $t^* = 240$ as the first instant at which we modify the value of ζ and subsequently move the mass between the microscopic and macroscopic scales. The choice of t^* has a direct impact on the speed of convergence.
Figure 6(a) shows the final total density of the system studied. Close to the global minimizer x^*, one can distinctly see the contribution that the microscopic scale makes to the profile of the density distribution. Figure 6(b) shows the value of ζ, which, as defined, represents when the results obtained in the two scales differ from each other. It can be seen that this value tends to zero, as in the limit the results obtained from the two systems tend to the same final configuration. In fact, by definition (4.1), ζ measures the deviation of the particles from the chosen closure function for the macroscopic system, which in our work is the Maxwellian distribution (3.14). The fact that ζ tends to zero means that the particles tend to the chosen closure function.
In Figure 6(c) there is a plot of the mass present in each scale at each time instant. As previously mentioned, the initial mass on the two scales is equivalent up to the constant value, designated as $t^* = 240$. It should be noted that this choice was determined empirically to ensure a nearly monotonic movement of the mass between the two scales, thereby accelerating the achievement of convergence. Actually, a shift in mass towards the macroscopic scale is observed, accompanied by a decrease in the microscopic scale. This outcome shows that, as time progresses, particle interactions become less significant in the problem under analysis. This study may be extended to address a broader spectrum of minimization problems. In high-dimensional minimization problems, we expect that mass will not always concentrate at the macroscopic scale. Due to the inherent complexity and challenges posed by high-dimensional macroscopic behavior, mass may tend to concentrate at the microscopic scale.



(a) The density ρ converges to a Dirac delta centered in 0, which is the global minimizer.



(b) The value of ζ decreases, meaning that the microscopic contribution is dampened as time evolves.



(c) We see that up to time step 240 there is no movement of mass between the two scales. At the limit, mass tends to move toward the macroscopic scale.

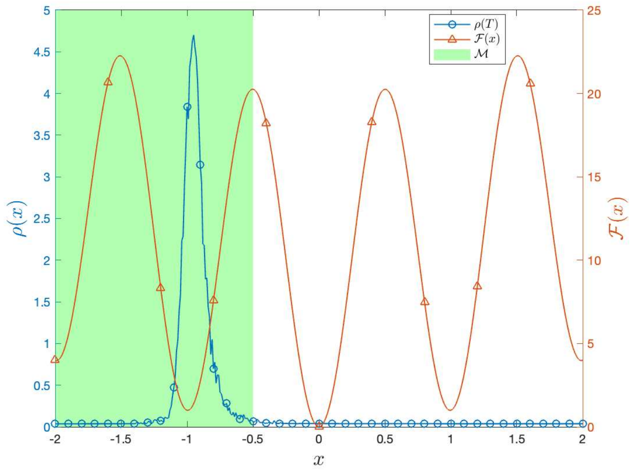
Figure 6: Plots obtained by applying micro-macro decomposition to solve an unconstrained minimization problem with objective function the Rastrigin function (5.7).

5.2.4 Constrained micro-macro decomposition

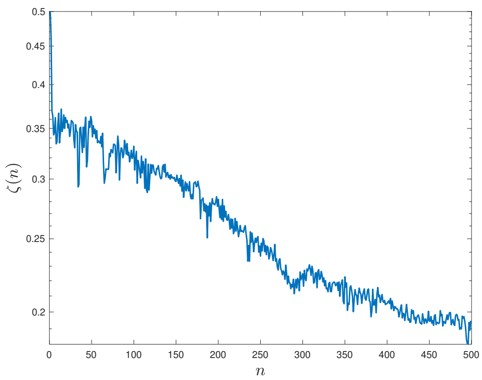
It has been shown that micro-macro decomposition can be used to address minimization problems in which the domain of the objective function is the entire space, denoted by $\mathbb{R}^d$. Furthermore, its application in constrained penalty problems, which are addressed by the adaptive penalized strategy outlined in Section 4, is also possible. At the algorithmic level, at each time step, the strategy consists of solving the penalized problem on the microscopic and macroscopic scales. Then we proceed with the comparison of the results in the two scales following Algorithm 1. Subsequently, the procedure is repeated for the next time step.
We present a simulation in which the chosen objective function is the Rastrigin function (5.7) in one-dimensional space and the chosen constraint is

$$\mathcal{M} = \{x \in \mathbb{R} | x < -0.5\}.$$

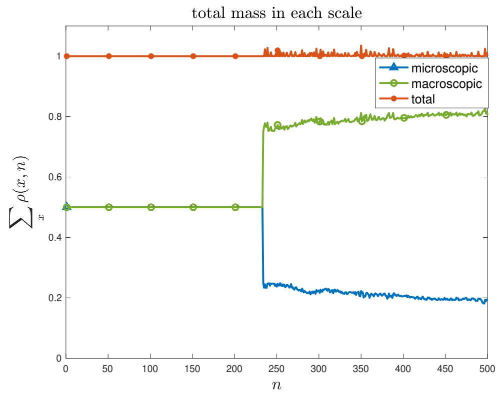
All parameters are chosen as in previous simulations.



(a) The density ρ converges to a Dirac delta centered in -1, which is the feasible global minimizer.



(b) The value of ζ decreases, meaning that the microscopic contribution is dampened as time evolves because it is multiplied by ζ.



(c) As explained in Remark 6, we see that up to time step 240 there is no mass movement between the two scales. At the limit, mass tends to move toward the macroscopic scale.

Figure 7: Plots obtained by applying micro-macro decomposition to solve a constrained minimization problem with objective function the Rastrigin function (5.7).

In Figure 7(a) we see that the final density, which is given by the microscopic and macroscopic contributions, converges to a Dirac delta centered at $x = -1$, which is exactly the feasible global minimizer. Again, the mass tends to shift from the microscopic to the macroscopic scale (see Figure 7(b)). This phenomenon indicates that, as time progresses, particle interactions become less significant for the purposes of convergence.

6 Conclusion

In this paper we conducted an analysis of the PSO method [33–35,43], which is a metaheuristic method for solving minimization problems. A novel extension to the method is presented that can be applied in the case of constrained minimization problems. This extension was introduced for the CBO model [7]. The strategy consists of working with an unconstrained minimization problem with a novel objective function that incorporates a penalty term. This penalty is in addition to the objective function of the initial minimization problem, and it is adaptively computed to reach its optimal value. The penalized adaptive algorithm was implemented at both the microscopic and macroscopic levels. A subsequent extension of this research would involve applying the algorithm

to a macroscopic system derived from an alternative choice of closure that satisfies the conditions presented in the paper.
Additionally, the micro-macro decomposition is presented in this work. It is a strategy designed to understand the different relevance of particle interactions as time evolves. Inspired by the works [16, 18, 38], this approach involves working in parallel on the microscopic and macroscopic levels. In each scale, a mass quantity is calculated at each time step by comparing the results obtained at the previous time step across all scales. We have shown that the micro-macro decomposition can also be applied to the constrained case.
Numerous results and numerical applications are presented to validate the good performance of the proposed strategies.

Acknowledgements

The work of S.V. is supported by the European Union's Horizon Europe research and innovation program under the Marie Sklodowska-Curie Doctoral Network Datahyking (Grant No. 101072546). S.V. is a member of the INdAM Research National Group of Scientific Computing (INdAM-GNCS). M.H. gratefully acknowledges support from the DFG through the projects HE5386/33-1 Control of Interacting Particle Systems, and Their Mean-Field, and Fluid-Dynamic Limits (560288187), and HE5386/34-1 Partikelmethoden für unendlich dimensionale Optimierung (561130572).
S.V. would like to thank René Pinnau and Giuseppe Visconti for the stimulating discussions during the preparation of this manuscript.

References

[1] E. H. L. Aarts and J. H. M. Korst. *Simulated annealing and Boltzmann machines : a stochastic approach to combinatorial optimization and neural computing.* Wiley-Interscience series in discrete mathematics and optimization. Wiley, United States, 1989.

[2] G. Albi and L. Pareschi. Binary interaction algorithms for the simulation of flocking and swarming dynamics. *Multiscale Modeling & Simulation*, 11(1):1–29, 2013.

[3] C. Bardos, F. Golse, and D. Levermore. Fluid dynamic limits of kinetic equations. i. formal derivations. *Journal of Statistical Physics*, 63:323–344, 1991.

[4] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, 1982.

[5] J. F. Bonnans, J. C. Gilbert, C. Lemarechal, and C. A. Sagastizábal. *Numerical Optimization: Theoretical and Practical Aspects.* Universitext. Springer Berlin Heidelberg, 2006.

[6] G. Borghi, S. Grassi, and L. Pareschi. Consensus based optimization with memory effects: Random selection and applications. *Chaos, Solitons & Fractals*, 174:113859, 2023.

[7] G. Borghi, M. Herty, and L. Pareschi. Constrained consensus-based optimization. *SIAM Journal on Optimization*, 33(1):211–236, 2023.

[8] J. A. Carrillo, Y. P. Choi, C. Totzeck, and O. Tse. An analytical framework for consensus-based global optimization method. *Mathematical Models and Methods in Applied Sciences*, 28(6):1037–1066, 2018.

[9] J. A. Carrillo, S. Jin, L. Li, and Y. Zhu. A consensus-based global optimization method for high dimensional machine learning problems. *ESAIM: Control, Optimisation and Calculus of Variations*, 27:S5, 2021.

[10] J. A. Carrillo, S. Jin, H. Zhang, and Y. Zhu. An interacting particle consensus method for constrained global optimization. Preprint *arXiv:2405.00891*, 2024.

[11] J. A. Carrillo, C. Totzeck, and U. Vaes. *Consensus-based optimization and ensemble Kalman inversion for global optimization problems with constraints*, pages 195–230. Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore. World Scientific Publishing, 1 edition, 2023.

[12] C. Cercignani. *The Boltzmann Equation and Its Applications*, volume 67 of *Applied Mathematical Sciences*. Springer, New York, 1988.

[13] C. Cipriani, H. Huang, and J. Qiu. Zero-inertia limit: From particle swarm optimization to consensus-based optimization. *SIAM Journal on Mathematical Analysis*, 54(3):3091–3121, 2022.

[14] C. A. Coello Coello. Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art. *Computer Methods in Applied Mechanics and Engineering*, 191(11):1245–1287, 2002.

[15] F. Cucker and S. Smale. Emergent behavior in flocks. *IEEE Transactions on Automatic Control*, 52(5):852–862, 2007.

[16] P. Degond, J.-G. Liu, and L. Mieussens. Macroscopic fluid models with localized kinetic upscaling effects. *Multiscale Modeling & Simulation*, 5(3):940–979, 2006.

[17] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2009.

[18] G. Dimarco and L. Pareschi. Numerical methods for kinetic equations. *Acta Numerica*, 23:369–520, 2014.

[19] M. Dorigo and C. Blum. Ant colony optimization theory: A survey. *Theoretical Computer Science*, 344(2):243–278, 2005.

[20] M. Dorigo, V. Maniezzo, and A. Colorni. Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 26(1):29–41, 1996.

[21] R. C. Eberhart and Y. Shi. Particle swarm optimization: developments, applications and resources. In *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*, volume 1, pages 81–86 vol. 1, 2001.

[22] M. Fornasier, H. Huang, L. Pareschi, and P. Sünnen. Consensus-based optimization on hypersurfaces: Well-posedness and mean-field limit. *Mathematical Models and Methods in Applied Sciences*, 30(14):2725–2751, 2020.

[23] M. Fornasier, H. Huang, L. Pareschi, and P. Sünnen. Consensus-based optimization on the sphere: Convergence to global minimizers and machine learning. *Journal of Machine Learning Research.*, 22(237):1–55, 2020.

[24] M. Fornasier, H. Huang, L. Pareschi, and P. Sünnen. Anisotropic diffusion in consensus-based optimization on the sphere. *SIAM Journal on Optimization*, 32(3):1984–2012, 2022.

[25] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., USA, 1st edition, 1989.

[26] S. Grassi, H. Huang, L. Pareschi, and J. Qiu. *Mean-field Particle Swarm Optimization*, pages 127–193. Lecture Notes Series, Institute for Mathematical Sciences. World Scientific, 2023.

[27] S. Grassi and L. Pareschi. From particle swarm optimization to consensus based optimization: Stochastic modeling and mean-field limit. *Mathematical Models and Methods in Applied Sciences*, 31(8):1625–1657, 2021.

[28] S.-Y. Ha, S. Jin, and D. Kim. Convergence of a first-order consensus-based global optimization algorithm. *Mathematical Models and Methods in Applied Sciences*, 30(12):2417–2444, 2020.

[29] M. Herty, A. Klar, A. K. Singh, and P. Spellucci. Smoothed penalty algorithms for optimization of nonlinear models. *Computational Optimization and Applications*, 37(2):157–176, 2007.

[30] D. J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546, 2001.

[31] J.H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan Press, 1975.

[32] H. Huang and J. Qiu. On the mean-field limit for the consensus-based optimization. *Mathematical Methods in the Applied Sciences*, 45(12):7814–7831, 2022.

[33] H. Huang, J. Qiu, and K. Riedl. On the global convergence of particle swarm optimization methods. *Applied Mathematics & Optimization*, 88(2):30, 2023.

[34] J. Kennedy. The particle swarm: Social adaptation of knowledge. In *Proceedings of 1997 - IEEE International Conference on Evolutionary Computation*, pages 303–308, 1997.

[35] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948, 1995.

[36] S. Kirkpatrick, C. Gelatt, and M. Vecchi. Optimization by simulated annealing. *Science (New York, N.Y.)*, 220:671–680, 1983.

[37] H.O. Kreiss and J. Lorenz. *Initial-boundary Value Problems and the Navier-Stokes Equations*. Number v. 136 in Initial-boundary value problems and the Navier-Stokes equations. Academic Press, 1989.

[38] M. Lemou and F. Méhats. Micro-macro schemes for kinetic equations including boundary layers. *SIAM Journal on Scientific Computing*, 34(6):734–760, 2012.

[39] E. Mezura-Montes and J. I. Flores-Mendoza. *Improved Particle Swarm Optimization in Constrained Numerical Search Spaces*, pages 299–332. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[40] S. Motsch and E. Tadmor. Heterophilious dynamics enhances consensus. *SIAM Review*, 56(4):577–621, 2014.

[41] R. Pinnau, C. Totzeck, O. Tse, and S. Martin. A consensus-based model for global optimization and its mean-field limit. *Mathematical Models and Methods in Applied Sciences*, 27(1):183–204, 2017.

[42] K. Riedl, T. Klock, C. Geldhauser, and M. Fornasier. Gradient is all you need? Preprint *arXiv:2306.09778*, 2023.

[43] 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.

[44] D. Sumpter. The principles of collective animal behaviour. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 361:5–22, 2005.

[45] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.*, 75:1226–1229, 1995.