



An Adaptive Consensus Based Method for Multi-objective Optimization with Uniform Pareto Front Approximation

Giacomo Borghi¹ · Michael Herty¹ · Lorenzo Pareschi²

Accepted: 11 June 2023 / Published online: 10 August 2023
© The Author(s) 2023

Abstract

In this work we are interested in stochastic particle methods for multi-objective optimization. The problem is formulated via scalarization using parametrized, single-objective sub-problems which are solved simultaneously. To this end a consensus based multi-objective optimization method on the search space combined with an additional heuristic strategy to adapt parameters during the computations is proposed. The adaptive strategy aims to distribute the particles uniformly over the image space, in particular over the Pareto front, by using energy-based measures to quantify the diversity of the system. The resulting gradient-free metaheuristic algorithm is mathematically analyzed using a mean-field approximation of the algorithm iteration and convergence guarantees towards Pareto optimal points are rigorously proven. In addition, we analyze the dynamics when the Pareto front corresponds to the unit simplex, and show that the adaptive mechanism reduces to a gradient flow in this case. Several numerical experiments show the validity of the proposed stochastic particle dynamics, investigate the role of the algorithm parameters and validate the theoretical findings.

Keywords Stochastic particle methods · Consensus-based optimization · Multi-objective optimization · Gradient-free methods · Mean-field limit

Mathematics Subject Classification 65K10 · 35Q93 · 35Q90 · 90C29 · 90C56

✉ Giacomo Borghi
borghi@eddy.rwth-aachen.de

Michael Herty
herty@igpm.rwth-aachen.de

Lorenzo Pareschi
lorenzo.pareschi@unife.it

¹ Institute for Geometry and Practical Mathematics, RWTH Aachen University, Aachen, Germany

² Department of Mathematics and Computer Science & Center for Modelling, Computing and Statistics, University of Ferrara, Ferrara, Italy

1 Introduction

Motivated by the problem in which two or more objectives must be considered at the same time, in this work we are interested in the design of stochastic algorithms for multi-objective optimization. This type of problem is commonly found in everyday life, for example, in physics, engineering, social sciences, economy, biology, and many others [17, 22, 44, 45, 53]. Investing in the financial market while maximizing profit and minimizing risk, or building a vehicle while maximizing performance and minimizing fuel consumption and pollutant emissions are examples of multi-objective optimization problems, where objectives typically conflict with each other.

From a mathematical viewpoint, the problem can be formulated through a variable $x \in \mathbb{R}^d$ describing a possible decision and assuming that $g_i(x)$ is the i -th objective for $i = 1, \dots, m$, with $m \in \mathbb{N}$ being the total number of objectives. A multi-objective problem then requires to solve for a decision x

$$\min_{x \in \mathbb{R}^d} g(x) \quad (1.1)$$

where $g(x) = (g_1(x), \dots, g_m(x))^T$. A solution to (1.1) corresponds to several optimal decisions. Here, we consider optimality in the sense of Pareto [45], i.e., no objective can be improved without necessarily degrading another objective. Without additional information about subjective preferences, there may be a (possibly infinite) number of Pareto optimal solutions, all of which are considered equally good. Therefore, the optimization task consists of providing a set of optimal decisions. To this end, it is also desirable to have a *diverse* set, that is, addressing the problem not only by optimizing the objectives, but also by aiming to best describe the (possibly) broad set of optimal decisions.

Several methods have been proposed to numerically solve (1.1) and, as for single-objective optimization, they typically belong to either the class of metaheuristic algorithms or mathematical programming methods [59]. Among metaheuristics, multi-objective evolutionary algorithms [17], such as NSGA-II [18] and MOEA/D [64], have gained popularity among practitioners due to their flexibility and ease of use. At the same time, they usually lack of convergence analysis compared to mathematical programming methods. Some of these algorithms, like MOEA/D, make use of scalarization strategies [45] which allow to translate problem (1.1) into a set of parametrized single-objective problems. With this detour, all classical, non-heuristic methods can also be employed in multi-objective optimization, at the cost of solving a possibly large amount of single-objective sub-problems. Many well-known techniques such as descend methods, trust-region methods, and Newton's method have been successfully adapted to directly solve multi-objective problems avoiding scalarization and they typically guarantee to find locally Pareto optimal solutions, see e.g. [30, 37]. Among deterministic algorithms for non-convex problems we mention branch-and-bound methods [27], DIRECT methods [11], and worst-case optimal algorithms [65]. For more details on mathematical programming methods and evolutionary algorithms in multi-objective optimization we refer to the recent surveys [15, 22].

We are interested in a particular class of gradient-free stochastic particle optimization methods, called consensus-based optimization (CBO), which has recently gained popularity due to the use of techniques developed in statistical physics that can provide them with a rigorous mathematical foundation. Such methods consider interacting particle systems which combine a consensus mechanism towards the estimated minimum and random exploration of the search space [2, 12, 13, 56, 61]. From a mathematical viewpoint, this class of metaheuristic methods is inspired by the corresponding *mean-field* dynamics based on particle swarming and multi-agent social interactions, which have been widely used to study complex systems in life sciences, social sciences and economics [19, 50, 51, 54, 63]. In the context of stochastic particle optimization, mean-field techniques first approximate the heuristic iterative strategy as continuous-in-time dynamics by means of a system of Stochastic Differential Equations (SDEs). Then, the system is further approximated by a mono-particle process, the corresponding mean-field model, which is more amenable to mathematical analysis. Quantitative estimates of such approximations can eventually provide precise error estimates for the original heuristic method, see e.g. [32]. This approach has proven fruitful to demonstrate convergence towards a global minimum for single-objective problems, not only in the case of CBO methods, but also for the popular Particle Swarm Optimization (PSO) algorithm [36, 43], thus paving the way to provide a mathematical foundation for other metaheuristics.

In the same spirit, the authors proposed in [7] a multi-objective optimization algorithm (M-CBO) by prescribing a CBO-type interaction among several particles making use of a scalarization strategy. As already mentioned, scalarization allows to reduce (1.1) into many single-objective sub-problems, which can be solved simultaneously in the case of particle-based optimization methods. In this paper, we provide a convergence analysis for the method proposed in [7] based on the mean-field description of the M-CBO dynamics. Furthermore, we improve the method in order to capture with uniform accuracy the shape of the Pareto front. This is done by iteratively updating the parameters of the method to minimize specific diversity measures, based on two-body energy potentials. Mathematically, this last feature is achieved by enlarging the phase space of the particles and introducing Vlasov-type dynamics in the space of the parameters. A detailed analysis of the extended model is also presented by studying a mean-field approximation of the particle dynamics which allows to recover convergence guarantees towards optimal points. We further analyze the adaptive strategy for bi-objective problems in simplified settings where the Pareto front is the unit simplex, that is, a linear segment for $m = 2$. Under suitable regularity assumptions on the potential, the particle dynamics takes a gradient-flow structure in the space of the parameters. The latter, in particular, suggests that the heuristic strategy leads to a better quality approximation of the Pareto front, where particles are uniformly spread over it. We remark that another CBO algorithm for multi-objective optimization has been recently proposed in [47]. This variant makes use of a multi-swarm approach rather than a one-swarm approach, as it is the case of the algorithm proposed in [7] and extended in this work.

The rest of the paper is organized as follows. In Sect. 2, we formally introduce the concept of optimality for (1.1) and present the scalarization strategy. Next, in Sect. 3,

we illustrate the particle dynamics both in the search space and in the space of parameters. Section 4 is devoted to the mathematical analysis of the system evolution using a mean-field description. Finally, in Sect. 5 numerical examples on convex and non-convex as well as disjoint Pareto fronts are presented, which confirm the theoretical results and the good performance of the new method. Some concluding remarks are discussed in the last section.

2 Problem Definition and Scalarization

We will use the following notation. Let $a \in \mathbb{R}^n$, $|a|$ indicates its euclidean norm and $(a)_l$ its l -th component. For given vectors $a, b \in \mathbb{R}^d$, $a \cdot b$ denotes the scalar product. For a Borel set $A \subset \mathbb{R}^d$, $|A|$ indicates its Lebesgue measure. $\mathcal{P}(\mathbb{R}^d)$ is the set of Borel probability measures over \mathbb{R}^d and $\mathcal{P}_q(\mathbb{R}^d) = \{\mu \in \mathcal{P}(\mathbb{R}^d) \mid \int |x|^q d\mu < \infty\}$ the set of Borel probability measures with finite q -th moment. The symbols $<$ and \leq indicate the partial ordering with respect to the cone $\mathbb{R}_{>0}^m$ and $\mathbb{R}_{\geq 0}^m$ respectively.

2.1 Pareto Optimality and Diversity

When dealing with a vector-valued objective function $g : \mathbb{R}^d \rightarrow \mathbb{R}^m$

$$g(x) = (g_1(x), \dots, g_m(x))^T \quad (2.1)$$

with $m \geq 2$, the interpretation of the minimization problem (1.1) is not unique, as the image space \mathbb{R}^m is not fully ordered. We consider the notions of strong and weak Edgeworth-Pareto optimality which rely on the natural, component-wise, partial ordering on \mathbb{R}^m [45].

Definition 2.1 (*Edgeworth-Pareto optimality*) A point $\bar{x} \in \mathbb{R}^d$ is (strong) Edgeworth-Pareto (EP) optimal, or simply optimal, if $g(\bar{x})$ is a minimal element of the image set $g(\mathbb{R}^d)$ with respect to the natural partial ordering, that is if there is no $x \in \mathbb{R}^d$ such that

$$g_i(x) \leq g_i(\bar{x}) \text{ for all } i = 1, \dots, m, \quad g(x) \neq g(\bar{x}).$$

Alike, \bar{x} is weakly EP optimal, if there is no $x \in \mathbb{R}^d$ such that

$$g_i(x) < g_i(\bar{x}) \text{ for all } i = 1, \dots, m.$$

The set $F_x = \{\bar{x} \in \mathbb{R}^d \mid \bar{x} \text{ is EP optimal}\}$ constitutes the set of EP optimal points, while

$$F = \{g(\bar{x}) \in \mathbb{R}^m \mid \bar{x} \text{ is EP optimal}\}$$

is the Pareto front.

The multi-objective optimization problem (1.1) consists of finding the set of EP optimal points. Unlike single-objective problems, the set is typically uncountable and the optimization task involves finding a finite subset of optimal points. The image of those points should ideally cover F and the concept of *diversity* is introduced to distinguish between two approximations [17]. Intuitively, if points on the Pareto front are more distanced, the diversity is higher. In view of the minimization problem, having a diverse approximation is desirable as it provides at the same cost a broader variety of possible solutions.

The most diverse approximation is possibly given by a set of points which is uniformly distributed over the Pareto front. Quantifying the diversity of an optimal set is of paramount importance both to assess the performance of optimization methods and to design them. Indeed, oftentimes the heuristic of a specific method is constructed to specifically minimize, or maximize, a specific measure [15]. Without knowledge of the exact Pareto front, popular diversity measures are given by *hypervolume contribution* [66], *crowding distance* [18] and, recently, by the *Riesz s -energy* [25, 48]. Our proposed algorithm will aim to minimize the latter (or similar energy-based measures) as it can be embedded in a mean-field framework. The exact definitions are introduced later.

To sum up, the multi-objective optimization problem we consider is a two-objective task itself, as one needs to find a set of points which are both EP optimal and optimize a suitable diversity measure.

2.2 Scalarization Strategy

A popular way to approach (1.1) is to use a scalarization strategy [21, 45] which reduces the multi-objective problem to an (infinite) number of single-objective sub-problems.

Among the possible scalarization strategies, we consider the approximation sub-problems with weighted Chebyshev semi-norms [45] where the single-objectives sub-problems are given by

$$\min_{x \in \mathbb{R}^d} G(x, w) := \max_{k \in \{1, \dots, m\}} w_k |g_k(x)| \tag{2.2}$$

and are parametrized by a vector of weights w which belongs to the unit, or probability, simplex

$$\Omega := \left\{ w \in \mathbb{R}_{\geq 0}^m \mid \sum_{\ell=1}^m w_\ell = 1 \right\}.$$

The link between the scalarized problems and the original multi-objective problem is given by the following result.

Theorem 2.1 ([45, Corollaries 5.25, 11.21]) *Assume g is component-wise positive.*

- (a) *A point \bar{x} is weakly EP optimal if and only if \bar{x} is a solution to (2.2) for some $w \in \Omega$.*

- (b) Assume all sub-problems (2.2) attain a unique minimum. Then, \bar{x} is EP optimal if and only if \bar{x} is the solution to (2.2) for some $w \in \Omega$.

Theorem 2.1 shows the strength of the Chebyshev scalarization strategy which allows to find all the weakly EP optimal points, contrary to other strategies like linear scalarization [45]. We remark that the proposed algorithm can also be applied to solve any other scalarized problems of the form (2.2) where the parameters are taken from the unit simplex Ω .

In practice, it will be sufficient to solve a finite number $N \in \mathbb{N}$ of sub-problems to obtain a good approximation of the Pareto front. Even though solving N sub-problems with corresponding weights vectors $\{W^i\}_{i=1}^N \subset \Omega$ ensures to find N optimal points, we note that there is no guarantee to obtain a diverse approximation. Therefore, scalarization targets only one of the two aspects of the problem, without addressing the diversity of the solution. In the following, we introduce an algorithm where the parameters W^i dynamically change during the computation to obtain a set of EP points which is also diverse.

3 Adaptive Multi-objective Consensus Based Optimization

We propose a dynamics where $N \in \mathbb{N}$ particles interact with each other to solve N scalar sub-problems given in the form (2.2).

At the iteration step $k = 0, 1, \dots$, every particle is described by its position $X_k^i \in \mathbb{R}^d$ and its vector of weights $W_k^i \in \Omega$ which determines the optimization sub-problem the particle aims to solve. As a result, particles are described by N tuples

$$(X_k^i, W_k^i) \quad \text{for } i = 1, \dots, N \quad \text{for } k = 0, 1, \dots$$

in the augmented space $\mathbb{R}^d \times \Omega$.

The initial configuration is generated by sampling the positions X_0^i from a common distribution $\rho_0 \in \mathcal{P}(\mathbb{R}^d)$ and by taking uniformly distributed weights vectors W_0^i over Ω . The iteration is prescribed to solve the multi-objective optimization task. We recall that (1.1) not only requires to find optimal points, but also points that are diverse, that is, well-distributed over the Pareto front. To this end, the optimization process is made of two mechanisms which address these two aims separately.

3.1 A Consensus Based Particle Dynamics in the Search Space

The first mechanism prescribes the update of the positions $\{X_k^i\}_{i=1}^N$, such that they converge towards EP optimal points. As in [7], this is done by introducing a CBO-type interaction between the particles. To illustrate the CBO update rule, let us consider for the moment a fixed single-objective sub-problem parametrized by $w \in \Omega$. Similar to Particle-Swarm Optimization methods, in CBO methods at step k , the particles move towards an attracting point $Y_k^\alpha(w)$ which is given by a weighted average of their

position:

$$Y_k^\alpha(w) = \frac{\sum_{j=1}^N X_k^j \exp(-\alpha G(X_k^j, w))}{\sum_{j=1}^N \exp(-\alpha G(X_k^j, w))}. \tag{3.1}$$

Due to the coefficients used in (3.1), if $\alpha \gg 1$, $Y_k^\alpha(w)$ is closer to the particles with low values of the objective function $G(\cdot, w)$ and, in the limiting case, it holds

$$Y_k^\alpha(w) \longrightarrow \operatorname{argmin}_{X_k^j, j=1, \dots, N} G(X_k^j, w) \text{ as } \alpha \rightarrow \infty,$$

if the above minimum uniquely exists. This promotes the concentration of the particles in areas of the search space where the objective function $G(\cdot, w)$ attains low values and hence, more likely, a global minimum. We remark that the exponential coefficients correspond to the Gibbs distribution associated with the objective function and, moreover, that this choice is justified by the Laplace principle [20]. The latter is an essential result to study the convergence of CBO methods [33] and it states that for any absolutely continuous probability density $\rho \in \mathcal{P}(\mathbb{R}^d)$ we have

$$\lim_{\alpha \rightarrow \infty} \left(-\frac{1}{\alpha} \log \left(\int e^{-\alpha G(x, w)} d\rho(x) \right) \right) = \inf_{x \in \operatorname{supp}(\rho)} G(x, w).$$

Since in the multi-objective optimization dynamics each particle addresses a different sub-problem, each of them moves towards a different attracting point given by $Y_k^\alpha(W_k^i)$. The attraction strength is given by $\lambda > 0$, while another parameter $\sigma > 0$ determines the strength of an additional stochastic component. The update rule reads

$$X_{k+1}^i = X_k^i + \lambda \Delta t \left(Y_k^\alpha(W_k^i) - X_k^i \right) + \sigma \sqrt{\Delta t} D_k^i(X_k^i, W_k^i) B_k^i \tag{3.2}$$

for all $i = 1, \dots, N$ where B_k^i are multivariate independent random vectors, $B_k^i \sim \mathcal{N}(0, I_d)$, I_d being the d -dimensional identity matrix, and $\Delta t > 0$ is the step-length. The update rule (3.2) is overparametrized and in CBO optimization schemes typically $\lambda = 1$ is used. The matrices $D_k^i(X_k^i, W_k^i)$ characterize the random exploration process which might be isotropic [56]

$$D_{k, \text{iso}}^i(X_k^i, W_k^i) = |Y_k^\alpha(W_k^i) - X_k^i| I_d, \tag{3.3}$$

or anisotropic [13]

$$D_{k, \text{aniso}}^i(X_k^i, W_k^i) = \operatorname{diag} \left(Y_k^\alpha(W_k^i) - X_k^i \right). \tag{3.4}$$

where $\operatorname{diag}(a)$ is the diagonal matrix with elements of $a \in \mathbb{R}^d$ on the main diagonal.

Both explorations depend on the distance between X_k^i and the corresponding attracting point making, the variance of the stochastic component larger if the particle is far

from $Y_k^\alpha(W_k^i)$. The difference lies in the direction of the random component: while in the isotropic exploration all dimensions are equally explored, the anisotropic one explores each dimension with a different magnitude.

We note that (3.2) can be interpreted as a simple Euler-Maruyama discretization [41] of the following system of SDEs

$$dX_t^i = \lambda \left(Y_t^\alpha(W_t^i) - X_t^i \right) dt + \sigma D_t^i(X_t^i, W_t^i) dB_t^i, \quad (3.5)$$

for all $i = 1, \dots, N$, where B_t^i are d -dimensional independent Brownian processes and $Y_t^\alpha(W_t^i)$, $D_t^i(X_t^i, W_t^i)$ are defined as in (3.1), (3.3) and (3.4) depending on whether isotropic or anisotropic diffusion is used. In Sect. 4 we will use this interpretation to obtain convergence results for (3.2).

The expected outcome of the position update rule (3.2) is that every particle will find a minimizer of a sub-problem and hence, by Theorem 2.1, a weak EP optimal point. We have already mentioned that if the weights vectors are fixed to the initial, uniform, distribution, i.e. $W_k^i = W_0^i$ for all $i = 1, \dots, N$, there is no guarantee to obtain equidistant points on the front. Since it is impossible to determine beforehand the optimal distribution on Ω , we propose a heuristic strategy which updates the vector weights promoting diversity.

Remark 3.1 Assume the particle position X_k^i attains a good value for a different problem W_k^j . Given that the i -th particle aims to optimize a different objective, it will probably leave such good location in the following iterations, possibly resulting in a deterioration of the weighted average for the j -th particle. As we will see, a key role is played by the Gaussian noise which ensures that the probability of finding another particle in such good position X_k^i is non-zero throughout the whole computation. Similarly to Particle Swarm Optimization methods, one could also avoid losing valuable information by adding memory mechanisms to the CBO dynamics, see e.g [6, 58].

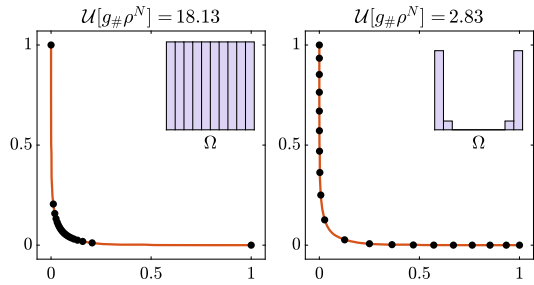
3.2 Uniform Approximation of the Pareto Front

A popular diversity metric in multi-objective optimization is the *hypervolume* contribution metric [66], which has the drawbacks of being computationally expensive [3] and, by definition, dependent on estimates of upper bounds of g . Motivated by this and by the objective of designing algorithms which perform well for any shape of the Pareto front [48], new energy-based diversity measures have recently gained popularity [15, 25]. Such measures quantify the diversity of a given empirical distribution $\rho^N \in \mathcal{P}(\mathbb{R}^d)$ by considering the pairwise interaction given by a two-body potential $U : \mathbb{R}^m \rightarrow (-\infty, \infty]$ on the image space

$$\mathcal{U}[g\#\rho^N] := \iint U(g(x) - g(y)) d\rho^N(y) d\rho^N(x), \quad (3.6)$$

$g\#\rho^N$ being the push-forward measure of ρ^N .

Fig. 1 Illustration of two particle systems over the Pareto front (in red): a high-energy configuration (left) and a low-energy one (right). The configuration Riesz energy ((3.6) with (3.7)) is also displayed, together with the histogram illustrating the corresponding particle distribution on the simplex (Color figure online)



The problem of finding well-spread points over the Pareto front is then equivalent to finding a configuration which is minimal with respect to the given energy \mathcal{U} where we recall that F_x is the set of EP optimal points:

$$\min_{\rho^N \in \mathcal{P}(F_x)} \mathcal{U}[g\#\rho^N] \rightarrow \min$$

We refer to Fig. 1 for an intuition on the difference between low and high energy configurations of particles over the Pareto front.

A distribution v^N is called diverse, if and only if

$$v^N \in \operatorname{argmin}_{\rho^N \in \mathcal{P}(F_x)} \mathcal{U}[g\#\rho^N].$$

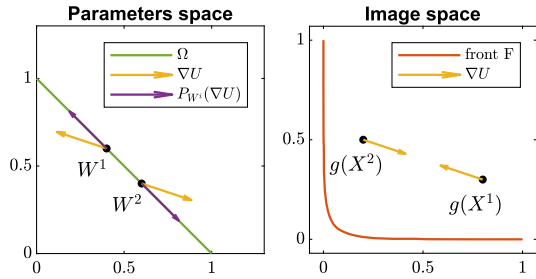
Any energy \mathcal{U} describing short range repulsion between particles, like Monge energy or repulsive-attractive power-law energy, is in principle a candidate to be a diversity measure. The Riesz s -energy, given by

$$U_R(z) = \frac{1}{|z|^s} \quad \text{with } s = m - 1, \tag{3.7}$$

is a popular choice [25] due to its theoretic guarantees of being a good measure of the uniformity of points. Indeed, if F is a $(m - 1)$ -dimensional manifold, the minimal energy configuration v^N converges to the uniform Hausdorff distribution over F as $N \rightarrow \infty$. We refer to [40] for the precise statement of the result and more details. Inspired by the electrostatic potential between charged particles, the authors in [10] used a Newtonian potential which is also empirically proven to be a suitable diversity measure [9, 10]. See [25] for a numerical comparison between two-body potentials as diversity measures in evolutionary algorithms. We will also compare different energies in Sect. 5 and consider $U \in C^1(\mathbb{R}^m \setminus \{0\})$ to be any of the above. The exact computation of minimal energy configurations of a system of N particles is a well-studied problem as it is connected to e.g. crystallization phenomenon [4]. We note that, in our settings, the configuration ρ^N is additionally mapped to the image space in (3.6), making the task even harder. Therefore, we propose a heuristic strategy that is expected to find only suboptimal configurations.

To promote diversity, we let the particles follow a vector field associated with \mathcal{U} . The movement will be only in parameter space, acting on $\{W_k^i\}_{i=1}^N$, in order not to interfere

Fig. 2 Illustration of dynamics (3.10) in case of two particles $(X_t^i, W_t^i), i = 1, 2$. Forces $\nabla U(g(X_t^i) - g(X_t^j))$ are collected in the image space and applied to the weights vectors W_t^i in the parameters space. The projection operator $P_{W_t^i}(\cdot)$ ensures that the dynamics is confined in the simplex Ω



with the CBO optimization dynamics acting on the positions $\{X_k^i\}_{i=1}^N$. Intuitively, if two particles are close to each other in the image space $g(\mathbb{R}^d)$, their weights vectors are pulled apart, see Fig. 2. This resembles a short range repulsion of U . To ensure W_k^i remains in the unit simplex Ω , a projection P_Ω is required, where

$$P_\Omega(v) = \operatorname{argmin}_{w \in \Omega} |v - w| \quad \text{for all } v \in \mathbb{R}^m,$$

see, e.g., [16] for an implementation. A parameter $\tau \geq 0$ determines the time scale of the weights adaptation process with respect to the CBO dynamics (3.2). The process can be turned off for $\tau = 0$.

In the case of bi-objective problems, where $m = 2$, we, therefore, obtain the update rule

$$W_{k+1}^i = P_\Omega \left(W_k^i + \frac{\tau}{N} \sum_{j=1}^N \nabla U \left(g(X_k^i) - g(X_k^j) \right) \Delta t \right), \quad (3.8)$$

which is well-defined as the parameters space is embedded in the image space \mathbb{R}^m . If U has a singularity in 0, we set $\nabla U(0) = 0$. We note that the lack of a minus sign in (3.8), is due to the explicit form of the relation determined by Theorem 2.1 between the Pareto front and Ω . This will become clear in the next section, as this choice gives a gradient flow structure to the parameters dynamics, under suitable assumptions.

For $m > 2$, the relation between a weight vector $w \in \Omega$ and the correspondent (weakly) EP optimal point is more involved. Indeed, (3.8) makes implicitly use of the fact that parameters can only move in one direction when $m = 2$, as the simplex is given by a segment. This is not the case for $m > 3$ and there is no straightforward way to pull back the vector field ∇U from the objective space to the simplex. Nevertheless, we prescribe a suitable heuristic dynamics as follows:

$$W_{k+1}^i = P_\Omega \left(W_k^i - \frac{\tau}{N} \sum_{j=1}^N \frac{W_k^i - W_k^j}{|W_k^i - W_k^j|} \left| \nabla U \left(g(X_k^i) - g(X_k^j) \right) \right| \Delta t \right), \quad (3.9)$$

for all $i = 1, \dots, N$. The term $|\nabla U(g(X_k^i) - g(X_k^j))|$ determines the strength of the interaction, while $(W_k^i - W_k^j)/|W_k^i - W_k^j|$ the direction of movement. We note that this second adaptive strategy can also be used in bi-objective problems, $m = 2$.

Up to our knowledge, energy-based diversity metrics have only been used as a selection criterion between candidate approximations of the Pareto front [26, 48], and this is the first time the vector field associated with \mathcal{U} is used to guide the particle dynamics in a metaheuristic multi-objective optimization method.

As before, (3.8) and (3.9) can be interpreted as an explicit discretization of continuous-in-time dynamics. In particular, to study the analytical properties of the adaptive strategy, we will consider the following Vlasov-type dynamics corresponding to (3.8)

$$dW_t^i = \tau P_{T(w_t^i, \Omega)} \left(\frac{1}{N} \sum_{j=1}^N \nabla U \left(g(X_t^i) - g(X_t^j) \right) \right) dt, \tag{3.10}$$

for all $i = 1, \dots, N$, where $P_{T(w, \Omega)}$ corresponds to the projection to the tangent cone $T(w, \Omega)$ to Ω at $w \in \Omega$, see also [14] for more details. Similarly, the time-continuous approximation of (3.9) is given by

$$dW_t^i = \tau P_{T(w_t^i, \Omega)} \left(-\frac{1}{N} \sum_{j=1}^N \frac{W_k^i - W_k^j}{|W_k^i - W_k^j|} \left| \nabla U \left(g(X_k^i) - g(X_k^j) \right) \right| \right) dt. \tag{3.11}$$

4 Convergence Analysis in Mean-Field Law

In this section, we give a statistical description of the optimization dynamics by presenting the corresponding mean-field model, which allows us to analyze the convergence of the method towards a solution to the multi-objective optimization problem. We will consider any scalarization strategy where the scalar sub-problems take the form of (2.2), with the parameters space given by the unit simplex Ω . We restrict ourselves to the case where particle positions are updated with anisotropic diffusion (3.4) for simplicity. The algorithm update rule for the positions X_k^i is given by (3.2), while weights vectors W_k^i are given by either (3.8) or (3.9).

As anticipated, we first approximate the discrete-in-time algorithm with the continuous-in-time dynamics given by (3.5) for particle locations, (3.10) or (3.11) for particle weights vectors. At the cost of introducing an approximation error which depends on Δt (see Remark 4.3 below for more details), we are able in this way to derive the correspondent mean-field dynamics. Mean-field models are obtained by taking the limit $N \rightarrow \infty$, that is, by considering an infinite number of particles. Such description allows to work with only one probability distribution of particles, rather than a (usually large) number of N different particle trajectories given by the microscopic description. Similar to [56], we make the so-called *propagation of chaos* assumption on the marginals [60]. In particular, let $F^N(t)$ be the particles probability distribution over $(\mathbb{R}^d \times \Omega)^N$ at a time $t \geq 0$. We assume $F^N(t) \approx f(t)^{\otimes N}$, that is, we assume the particles $(X_t^i, W_t^i), i = 1, \dots, N$ to be independently distributed according to $f(t) \in \mathcal{P}(\mathbb{R}^d \times \Omega)$ for some large $N \gg 1$.

In the following, we indicate with $\rho(t) \in \mathcal{P}(\mathbb{R}^d)$ the first marginal of $f(t)$ and with $\mu(t) \in \mathcal{P}(\Omega)$ the second marginal on the parameters space Ω .

As a consequence of the propagation of chaos assumption, we formally obtain that

$$Y_t^\alpha(W_t^i) = \frac{\frac{1}{N} \sum_{j=1}^N X_t^j e^{-\alpha G(X_t^j, W_t^i)}}{\frac{1}{N} \sum_{j=1}^N e^{-\alpha G(X_t^j, W_t^i)}} \approx \frac{\int x e^{-\alpha G(x, W_t^i)} d\rho(t, x)}{\int e^{-\alpha G(x, W_t^i)} d\rho(t, x)} =: y^\alpha(\rho(t), W_t^i)$$

and that

$$\frac{1}{N} \sum_{j=1}^N \nabla U(g(X_t^i) - g(X_t^j)) \approx \int \nabla U(g(X_t^i) - g(x)) d\rho(t, x),$$

where the summation with respect to the j -indices has been substituted with integration with respect to the marginal $\rho(t)$. We note that factorization of the marginals, $f(t) = \rho(t) \otimes \mu(t)$, is not required in the above derivations and it is not expected to hold true due to the coupling between the dynamics in \mathbb{R}^d and Ω .

The SDEs (3.5) and (3.10) are now independent on the index i and we obtain the process (X_t, W_t) , $t > 0$ as

$$\begin{cases} dX_t &= \lambda(y_t^\alpha(\rho(t), W_t) - X_t)dt + \sigma \text{diag}(y_t^\alpha(\rho(t), W_t) - X_t)dB_t \\ dW_t &= \tau P_{W_t} \left(\int \nabla U(g(X_t) - g(x)) d\rho(t, x) \right) dt \end{cases} \tag{4.1}$$

where we introduced $P_w := P_{T(w, \Omega)}$ and anisotropic diffusion is used. In the same way, we can derive the mean-field mono-particle process associated with the CBO evolution (3.5) and the alternative adaptive dynamics (3.11). We obtain

$$\begin{cases} dX_t &= \lambda(y_t^\alpha(\rho(t), W_t) - X_t)dt + \sigma \text{diag}(y_t^\alpha(\rho(t), W_t) - X_t)dB_t \\ dW_t &= \tau P_{W_t} \left(- \int \frac{W_t - w}{|W_t - w|} |\nabla U(g(X_t) - g(x))| df(t, x, w) \right) dt. \end{cases} \tag{4.2}$$

Processes (4.1) and (4.2) are mean-field descriptions of the microscopic dynamics generated by the optimization dynamics described in Sect. 3. We note that the rigorous mean-field limit for single-objective CBO dynamics, which are similar to (3.5), was proven in [42]. Following previous works, see e.g. [12, 33], we consider such an approximation and mathematically analyze the proposed optimization method by studying solutions to (4.1) and (4.2).

Remark 4.1 We note that the law $f(t)$ of process (4.1) weakly satisfies

$$\begin{aligned} \frac{\partial}{\partial t} f(t, x, w) &= -\lambda \nabla_x \cdot ((y^\alpha(\rho(t), w) - x) f(t, x, w)) \\ &\quad + \frac{\sigma^2}{2} \sum_{\ell=1}^d \partial_{x_\ell x_\ell}^2 \left((y^\alpha(\rho(t), w) - x)_\ell^2 f(t, x, w) \right) \end{aligned}$$

$$-\tau \nabla_w \cdot \left(P_w \left(\int \nabla U (g(x) - g(y)) d\rho(t, y) \right) f(t, x, w) \right), \tag{4.3}$$

with initial conditions $f(0, x, w) = \rho_0 \otimes \mu_0$, $\rho_0 = \text{Law}(X_0)$ and $\mu_0 = \text{Law}(W_0)$, if we take as test function any $\phi \in C_c^\infty(\mathbb{R}^d \times \Omega)$, as a consequence of Itô’s formula.

4.1 Convergence to the Pareto Front

In the following, we study solutions $(X_t, W_t)_{t \in [0, T]}$ to the mean-field approximations (4.1), (4.2) to gain insights on the method convergence properties. We assess the performance of a multi-objective algorithm by the average distance from the Pareto front F and we use the Generational Distance (GD) [62] given by

$$GD[\rho(t)] = \left(\int \text{dist}(g(x), F)^2 d\rho(t, x) \right)^{\frac{1}{2}}, \tag{4.4}$$

where $\rho(t) = \text{Law}(X_t)$. In the following, we state conditions such that $GD[\rho(t)]$ decays up to a given accuracy $\varepsilon > 0$.

Assumption 4.1 (Uniqueness) Every sub-problem (2.2) with $w \in \Omega$ admits a unique solution $\bar{x}(w) \in F$. Moreover $\bar{x} : \Omega \rightarrow F_x$ is a C^1 map satisfying $\sup_{w \in \Omega} \|D\bar{x}(w)\| < \infty$, $D\bar{x}(w)$ being the differential at w , and $\|\cdot\|$ is operator norm.

The uniqueness requirement is common in the analysis of CBO methods [33]. This is due to the difficulty of controlling the attractive term y^α , whenever there are two or more minimizers. For example, assume v is a measure concentrated in two different global minimizers of a sub-problem w , $v = (\delta_{x_1} + \delta_{x_2})/2$: the attractive term could be located in the middle between them

$$y^\alpha(v, w) = \frac{1}{2} (x_1 + x_2)$$

being obviously not(!) a minimizer. The regularity assumption on \bar{x} will be required to provide bounds on the adaptive strategies (4.1), (4.2) and may be dropped if the interaction in the weights space is not present.

The next assumption is concerned with lower and upper bounds in a neighborhood of minimizers of $G(\cdot, w)$, $w \in \Omega$. See also [35] and the references therein for more details on the following conditions.

Assumption 4.2 (Stability at the minimizer) The functions $G(\cdot, w)$, $w \in \Omega$, are continuous and, in a neighborhood of their minimizer, satisfy the following growth condition: there exists a radius $R > 0$, exponents $p_1 > 0$, $p_2 \leq p_1$ and constants $c_1, c_2 > 0$ such that for all $w \in \Omega$

$$c_1 \|x - \bar{x}(w)\|_\infty^{p_1} \leq G(x, w) - \min_{y \in \mathbb{R}^d} G(y, w) \leq c_2 \|x - \bar{x}(w)\|_\infty^{p_2}, \tag{4.5}$$

for all $x \in \mathbb{R}^d$ such that $\|x - \bar{x}(w)\|_\infty \leq R$.

Moreover, outside such a neighborhood, the function cannot be arbitrarily close to the minimum: there exists $c_3 > 0$ such that for all $w \in \Omega$

$$c_3 \leq G(x, w) - \min_{y \in \mathbb{R}^d} G(y, w) \quad \forall x : \|x - \bar{x}(w)\|_\infty \geq R.$$

We note that the lower bound in (4.5) is known in the literature as *conditioning* property [35] or also as *inverse continuity* property [33]. For a better understanding, we provide a simple example of scalarized sub-problem satisfying the above condition.

Example 1 Consider the bi-objective problem given by $g_1(x) = (x - 0.5)^2$, $g_2(x) = (x + 0.5)^2$, $x \in \mathbb{R}$, and let the scalarized sub-problems be given by $G(x, w) = w_1 g_1(x) + w_2 g_2(x)$, for any $w = (w_1, w_2)^\top \in \Omega$. By setting $\nabla_x G(x, w) = 0$, one obtains that the EP optimal points are given by $\bar{x}(w) = w_1 - 0.5$ and $\min_{y \in \mathbb{R}} G(y, w) = -w_1(w_1 - 1)$. For all $w = (w_1, w_2)^\top \in \Omega$, $x \in \mathbb{R}$, it holds

$$G(w, x) - \min_{y \in \mathbb{R}} G(y, w) = (x - w_1 + 0.5)^2, \quad |x - \bar{x}(w)| = |x - w_1 + 0.5|$$

and therefore Assumption 4.3 is satisfied for $c_1 = c_2 = 1$, $p_1 = p_2 = 2$ and any $R > 0$.

Finally, we assume the EP optimal points to be bounded. As for the single-objective case [33, 34], we also prescribe a condition on the initial data ρ_0 and μ_0 .

Assumption 4.3 (Boundedness and initial datum) The set F_x of optimal points is contained by a bounded, open set $F_x \subset H$. The initial distribution f_0 is given by $f_0 = \rho_0 \otimes \mu_0$ with any $\mu_0 \in \mathcal{P}(\Omega)$ and $\rho_0 = \text{Unif}(H)$, that is, the uniform probability distribution over H .

Boundedness of the EP optimal points will be necessary to provide estimates on the probability mass around the solutions to the sub-problems. We note that, in many applications, the prescribed search space is bounded.

Assumptions 4.1–4.3 ensure that the quantitative results on the Laplace principle [33, 34] are applicable to all the different sub-problems (2.2) with uniform choice of α . Therefore, under such assumptions, it is possible to prove convergence of the particle system towards EP optimal points by studying the large time behavior of the average ℓ_2 -error

$$\text{Err}[f(t)] := \mathbb{E} \left[|X_t - \bar{x}(W_t)|^2 \right] \tag{4.6}$$

where $f(t) = \text{Law}(X_t, W_t)$. We will do so by following the analysis of the single-objective CBO methods [33, Theorem 12], [34, Theorem 2].

Theorem 4.1 Assume scalar sub-problems and initial data $X_0 \sim \rho_0$, $W_0 \sim \mu_0$, $f_0 = \rho_0 \otimes \mu_0$ satisfy Assumptions 4.1–4.3 and $\nabla U \in L^\infty(\mathbb{R}^m)$.

For any accuracy $\varepsilon \in (0, \text{Err}[f(0)])$, let $\lambda, \sigma > 0, \tau \geq 0$ satisfy

$$\lambda > 2\sigma^2 + C \frac{\tau}{\sqrt{\varepsilon}}, \quad \text{where } C := 4 \sup_{w \in \Omega} \|D\bar{x}(w)\| \|\nabla U\|_{L^\infty(\mathbb{R}^m)} \quad (4.7)$$

and $T^* > 0$ be the time horizon given by

$$T^* = \frac{2}{\lambda - 2\sigma^2 - C\tau/\sqrt{\varepsilon}} \log \left(\frac{\text{Err}[f(0)]}{\varepsilon} \right). \quad (4.8)$$

Consider now a solution $(X_t, W_t)_{t \in [0, T^]}$ to either (4.1) or (4.2) with initial data given by (X_0, W_0) and such that $\sup_{t \in [0, T^]} \mathbb{E}[|X_t|^4] < \infty$.

Provided α is sufficiently large, it holds

$$\min_{t \in [0, T^]} \text{Err}[f(t)] \leq \varepsilon, \quad (4.9)$$

where $f(t) = \text{Law}(X_t, W_t)$. Moreover, until the desired accuracy is reached, the error decays exponentially

$$\text{Err}[f(t)] \leq \text{Err}[f(0)] \exp \left(-\frac{1}{2} \left(\lambda - 2\sigma^2 - C \frac{\tau}{\sqrt{\varepsilon}} \right) t \right). \quad (4.10)$$

We remark that the choice of α depends on the estimates given in Assumption 4.2 and, in particular, on the accuracy ε . We will provide a proof in the next section. Before, we show how to link the above convergence result to the GD performance metric.

Corollary 4.1 *Under assumptions of Theorem 4.1, consider $\rho(t) = \text{Law}(X_t)$, $f(t) = \text{Law}(X_t, W_t)$. If g is Lipschitz continuous and component-wise positive, it holds*

$$\min_{t \in [0, T^]} GD[\rho(t)] \leq \text{Lip}(g)\sqrt{\varepsilon}$$

and, until the desired accuracy is reached,

$$GD[\rho(t)] \leq \text{Lip}(g)\sqrt{\text{Err}[f(0)]} \exp \left(-\frac{\lambda - \kappa\sigma^2 - C\tau/\sqrt{\varepsilon}}{2} t \right).$$

Proof Since every sub-problem admits a unique solution (Assumption 4.1), by Theorem 2.1 every solution $\bar{x}(w)$ is EP optimal and therefore its image $g(\bar{x}(w))$ belongs to the Pareto front F . Therefore

$$\text{dist}(g(x), F) \leq |g(x) - g(\bar{x}(w))| \leq \text{Lip}(g)|x - \bar{x}(w)|$$

from which follows that $GD[\rho(t)]$ is bounded by the average ℓ_2 -error. □

Remark 4.2 In Theorem 4.1, τ must be taken of order $o(\sqrt{\varepsilon})$ suggesting that the parameters should adapt at a much slower time scale with respect to the positions, so as not to interfere with the CBO dynamics. With no weights vectors interaction, $\tau = 0$, the decay estimate (4.10) is independent of ε and the rate is larger compared to the case with $\tau > 0$.

Remark 4.3 When quantitative estimates of the mean-field limit are given, the expected ℓ_2 -error can be decomposed as

$$\mathbb{E} \left[\sum_{i=1}^N |X_k^i - \bar{x}(W_k^i)|^2 \right] \leq \mathcal{O}(\Delta t) + \mathcal{O}(N^{-1}) + \mathcal{O}(\varepsilon) \tag{4.11}$$

see, for instance, [31, 33]. The first term quantifies the distance between the algorithm iteration (3.2), (3.8) (or (3.9)) and the system of SDEs (3.5), (3.10) (or (3.11)) and it is given by classical results in numerical approximation of SDEs by means of the Euler-Maruyama scheme [57]. The second term, of order N^{-1} , is the (expected) error introduced by the mean-field approximation, see e.g. [33]. The last term comes from the convergence of the mean-field model to the sub-problems solutions and it is given by Theorem 4.1.

We note that the additional dynamics of the weight vectors in (4.1) and (4.2) adds non-linearity to the particle system. Thus, standard arguments used to prove well-posedness and mean-field convergence of the CBO dynamics, see for instance [12, 33], cannot be adapted to the present case in a straightforward manner. For this reason, we leave such investigations for future work.

4.2 Proof of Theorem 4.1

To prove Theorem 4.1, we follow the strategy introduced in [33, 34] for the analysis of single-objective CBO particle methods.

We start by studying the time evolution of $\text{Err}[f(t)]$.

Proposition 4.1 *Under assumptions of Theorem 4.1, for all $t \in [0, T^*]$ it holds*

$$\begin{aligned} \frac{d}{dt} \text{Err}[f(t)] &\leq (-\lambda + 2\sigma^2) \text{Err}[f(t)] + (\lambda + 2\sigma^2) \mathbb{E} \left[|y^\alpha(\rho(t), W_t) - \bar{x}(W_t)|^2 \right] \\ &\quad + 2\tau \sup_{w \in \Omega} \|D\bar{x}(w)\| \|\nabla U\|_{L^\infty(\mathbb{R}^m)} \sqrt{\text{Err}[f(t)]}. \end{aligned}$$

Proof By applying Itô’s formula with $\phi(x, w) = |x - \bar{x}(w)|^2$, we obtain

$$\begin{aligned} d\phi(X_t, W_t) &= \lambda \nabla_x \phi(X_t, W_t) \cdot (y^\alpha(\rho(t), W_t) - X_t) dt \\ &\quad + \sigma \sum_{\ell=1}^d \partial_{x_\ell} \phi(X_t, W_t) (y^\alpha(\rho(t), W_t) - X_t)_\ell dB_{t,\ell} \end{aligned}$$

$$\begin{aligned}
 &+ \frac{\sigma^2}{2} \sum_{\ell=1}^d \partial_{x_\ell x_\ell}^2 \phi(X_t, W_t) (y^\alpha(\rho(t), W_t) - X_t)_\ell^2 dt \\
 &+ \tau D_w \phi(X_t, W_t) dW_t.
 \end{aligned}$$

Thanks to the solution regularity $\mathbb{E}[|X_t|^4] < \infty$ for all $t \in [0, T^*]$, we have $\mathbb{E}[\int_0^t |\partial_{x_\ell} \phi(X_s, W_s) (y^\alpha(\rho(s), W_s) - X_s)_\ell|^2 ds] < \infty$ from which follows

$$\mathbb{E} \left[\int_0^t \partial_{x_\ell} \phi(X_s, W_s) (y^\alpha(\rho(s), W_s) - X_s)_\ell dB_{s,\ell} \right] = 0 \quad \forall \ell = 1, \dots, d$$

due to [52, Theorem 3.2.1 (iii)]. Therefore, by taking the expectation and applying Fubini’s theorem we obtain

$$\begin{aligned}
 \frac{d}{dt} \text{Err}[f(t)] &= \frac{d}{dt} \mathbb{E}[\phi(X_t, W_t)] \\
 &= \mathbb{E} \left[\lambda \nabla_x \phi(X_t, W_t) \cdot (y^\alpha(\rho(t), W_t) - X_t) \right] \\
 &\quad + \mathbb{E} \left[\frac{\sigma^2}{2} \sum_{\ell=1}^d \partial_{x_\ell x_\ell}^2 \phi(X_t, W_t) (y^\alpha(\rho(t), W_t) - X_t)_\ell^2 \right] \\
 &\quad + \mathbb{E} \left[\tau D_w \phi(X_t, W_t) \frac{d}{dt} W_t \right] \\
 &=: T_1 + T_2 + T_3.
 \end{aligned}$$

The first term can be bounded as follows:

$$\begin{aligned}
 T_1 &= 2\lambda \mathbb{E} \left[(X_t - \bar{x}(W_t)) \cdot (y^\alpha(\rho(t), W_t) - X_t) \right] \\
 &= 2\lambda \mathbb{E} \left[(X_t - \bar{x}(W_t)) \cdot (y^\alpha(\rho(t), W_t) - \bar{x}(W_t)) \right] - 2\lambda \mathbb{E} \left[|X_t - \bar{x}(W_t)|^2 \right] \\
 &\leq \lambda \mathbb{E} \left[|X_t - \bar{x}(W_t)|^2 \right] + \lambda \mathbb{E} \left[|y^\alpha(\rho(t), W_t) - \bar{x}(W_t)|^2 \right] \\
 &\quad - 2\lambda \mathbb{E} \left[|X_t - \bar{x}(W_t)|^2 \right],
 \end{aligned}$$

where in the last inequality we used that $a \cdot b \leq (|a|^2 + |b|^2)/2$ for any $a, b \in \mathbb{R}^d$. By employing the definition of $\text{Err}[f(t)]$, we obtain

$$T_1 \leq -\lambda \text{Err}[f(t)] + \lambda \mathbb{E} \left[|y^\alpha(\rho(t), W_t) - \bar{x}(W_t)|^2 \right]. \tag{4.12}$$

Given that $\partial_{x_\ell x_\ell}^2 \phi(x, w) = 2$, it holds

$$\begin{aligned}
 T_2 &= \sigma^2 \mathbb{E} \left[\sum_{\ell=1}^d (y^\alpha(\rho(t), W_t) - X_t)_\ell^2 \right] \\
 &= \sigma^2 \mathbb{E} \left[|y^\alpha(\rho(t), W_t) - X_t|^2 \right]
 \end{aligned}$$

$$\begin{aligned} &\leq 2\sigma^2\mathbb{E}[|y^\alpha(\rho(t), W_t) - \bar{x}(W_t)|^2] + 2\sigma^2\mathbb{E}[|\bar{x}(W_t) - X_t|^2] \\ &= 2\sigma^2\mathbb{E}[|y^\alpha(\rho(t), W_t) - \bar{x}(W_t)|^2] + 2\sigma^2\text{Err}[f(t)], \end{aligned} \tag{4.13}$$

thanks to inequality $|a - b|^2 \leq 2(|a - c|^2 + |c - b|^2)$ for any $a, b, c \in \mathbb{R}^d$. To bound the third term, we first note that $D_w\phi(x, w)(v) = 2(x - \bar{x}(w)) \cdot D\bar{x}(w)(v)$, for any $v \in \mathbb{R}^m$. Note that this requires to differentiate on the manifold and the corresponding differential is denoted by D_w .

Now, thanks to the boundedness of ∇U , it holds

$$\left| \frac{d}{dt} W_t \right| \leq \|\nabla U\|_{L^\infty(\mathbb{R}^m)},$$

no matter if the weights are adapted according to (4.1) or (4.2). Thanks to the regularity of \bar{x} , we obtain

$$\begin{aligned} T_3 &= 2\tau\mathbb{E} \left[(X_t - \bar{x}(W_t)) \cdot D\bar{x}(W_t) \frac{d}{dt} W_t \right] \\ &\leq 2\tau \sup_{w \in \Omega} \|D\bar{x}(w)\| \|\nabla U\|_{L^\infty(\mathbb{R}^m)} \mathbb{E}[|X_t - \bar{x}(W_t)|] \\ &\leq 2\tau \sup_{w \in \Omega} \|D\bar{x}(w)\| \|\nabla U\|_{L^\infty(\mathbb{R}^m)} \sqrt{\text{Err}[f(t)]} \end{aligned} \tag{4.14}$$

where we applied Jensen’s inequality in the last step. By putting together (4.12), (4.13) and (4.14), we obtain the desired estimate. \square

Lemma 4.1 (Quantitative Laplace principle) *Under assumptions of Theorem 4.1, let $\rho \in \mathcal{P}(\mathbb{R}^d)$ be any measure such that the EP optimal points belong to its support, $F_x \subset \text{supp}(\rho)$. Consider any $r \in [0, R]$, $q > 0$ such that $q + c_2r^{1/p_2} \leq c_3$, for all $w \in \Omega$ it holds*

$$|y^\alpha(\rho, w) - \bar{x}(w)| \leq \sqrt{d}c_1 (q + c_2r^{1/p_2})^{1/p_1} + \frac{\sqrt{d} \exp(-\alpha q)}{\rho(B_r^\infty(\bar{x}(w)))} \int |x - \bar{x}(w)| d\rho(x)$$

where $B_r^\infty(z)$ is the open ℓ_∞ -ball centered in $z \in \mathbb{R}^d$ with radius r .

Proof For a fixed $w \in \Omega$, we recall the analogous result from the single-objective case [34, Proposition 1]. Let $\min_{y \in \mathbb{R}^d} G(y, w) = 0$ without loss of generality, and define $G_r(w) := \sup_{y \in B_r^\infty(\bar{x}(w))} G(y, w)$. For any $r \in [0, R]$ and $q > 0$ such that $q + G_r(w) \leq c_3$, it holds

$$|y^\alpha(\rho, w) - \bar{x}(w)| \leq \sqrt{d}c_1 (q + G_r(w))^{1/p_1} + \frac{\sqrt{d} \exp(-\alpha q)}{\rho(B_r^\infty(\bar{x}(w)))} \int |x - \bar{x}(w)| d\rho(x).$$

Assumption 4.1 ensures that all sub-objectives $G(\cdot, w)$ satisfy a common growth condition around the minimizers. In particular, we have that $G_r(w) \leq c_2r^{1/p_2}$ for all $w \in \Omega$, which leads to the desired estimate. \square

Next, to apply the above result for all $t \in [0, T^*]$, we provide a lower bound on $\rho(B_r^\infty(\bar{x}(w)))$ for all $w \in \Omega$. As the lower bound will be independent on the definition of $\bar{x}(w)$, we consider in the following any point $x^* \in \mathbb{R}^d$ belonging to the initial search domain, $x^* \in \text{supp}(\rho_0) = H$ (see Assumption 4.3). To do so, we introduce the mollifier ϕ_r , for $r > 0$:

$$\phi_r(x) := \begin{cases} \prod_{k=1}^d \exp\left(1 - \frac{r^2}{r^2 - (x - x^*)_k^2}\right), & \text{if } \|x - x^*\|_\infty < r, \\ 0, & \text{else.} \end{cases}$$

Lemma 4.2 *Under the settings of Theorem 4.1, let $T \in [0, T^*]$, $r > 0$, and $x^* \in \text{supp}(\rho_0)$ be fixed. Assume*

$$\sup_{t \in [0, T]} \sup_{w \in \Omega} \|y^\alpha(\rho(t), w) - x^*\|_\infty < B \tag{4.15}$$

for some $B > 0$. It holds

$$\mathbb{P}(X_t \in B_r^\infty(x^*)) \geq \mathbb{E}[\phi_r(X_0)] \exp(-qt)$$

for $q := 2d \max \left\{ \frac{\lambda(cr + B\sqrt{c})}{(1-c)^2r} + \frac{\sigma^2(cr^2 + B^2)(2c + 1)}{(1-c)^4r^2}, \frac{2\lambda^2}{(2c - 1)\sigma^2} \right\}$

where $c \in (0.5, 1)$ can be any constant satisfying $(1 - c^2)^2 \leq (2c - 1)c$.

We refer to Appendix C for a proof. Now, we are ready to provide a proof of the main result.

Proof (Theorem 4.1) Thanks to Proposition 4.1, we have for all $t \in [0, T^*]$,

$$\begin{aligned} \frac{d}{dt} \text{Err}[f(t)] &\leq (-\lambda + 2\sigma^2) \text{Err}[f(t)] + (\lambda + 2\sigma^2) \mathbb{E} \left[|y^\alpha(\rho(t), W_t) - \bar{x}(W_t)|^2 \right] \\ &\quad + 2\tau \sup_{w \in \Omega} \|D\bar{x}(w)\| \|\nabla U\|_{L^\infty(\mathbb{R}^m)} \sqrt{\text{Err}[f(t)]}. \end{aligned}$$

We consider now

$$T := \sup \left\{ t \geq 0 \mid \text{Err}(s) > \varepsilon \text{ and } \sup_{w \in \Omega} |y_\alpha(\rho(s), w) - \bar{x}(w)| \leq \tilde{C}(s) \forall s \in [0, t] \right\}$$

where

$$\tilde{C}(t) := \frac{1}{2} \frac{\lambda - 2\sigma^2}{\lambda + 2\sigma^2} \text{Err}[f(t)].$$

For all $t \in [0, T]$ it holds $\text{Err}[f(t)] \geq \varepsilon$ and therefore we have

$$\begin{aligned}
 & 2\tau \sup_{w \in \Omega} \|D\bar{x}(w)\| \|\nabla U\|_{L^\infty(\mathbb{R}^m)} \sqrt{\text{Err}[f(t)]} \\
 & \leq 2\tau \sup_{w \in \Omega} \|D\bar{x}(w)\| \|\nabla U\|_{L^\infty(\mathbb{R}^m)} \sqrt{\text{Err}[f(t)]} \frac{\sqrt{\text{Err}[f(t)]}}{\sqrt{\varepsilon}} \\
 & =: C \frac{\tau}{2\sqrt{\varepsilon}} \text{Err}[f(t)].
 \end{aligned}$$

By using the above estimate and definition of $\tilde{C}(t)$, we then obtain

$$\begin{aligned}
 \frac{d}{dt} \text{Err}[f(t)] & \leq (-\lambda + 2\sigma^2) \text{Err}[f(t)] + (\lambda + 2\sigma^2) \frac{1}{2} \frac{\lambda - 2\sigma^2}{\lambda + 2\sigma^2} \text{Err}[f(t)] \\
 & \quad + C \frac{\tau}{2\sqrt{\varepsilon}} \text{Err}[f(t)] \\
 & = -\frac{1}{2} \left(\lambda - 2\sigma^2 - C \frac{\tau}{\sqrt{\varepsilon}} \right) \text{Err}[f(t)].
 \end{aligned}$$

By applying Grönwall’s inequality we obtain the error decay

$$\text{Err}[f(t)] \leq \text{Err}[f(0)] \exp\left(-\frac{1}{2} \left(\lambda - 2\sigma^2 - C \tau/\sqrt{\varepsilon}\right) t\right) \tag{4.16}$$

for all $t \in [0, T]$. In particular, by definition of $\tilde{C}(t)$, it holds

$$\sup_{w \in \Omega} |y_\alpha(\rho(s), w) - \bar{x}(w)| \leq \tilde{C}(s) \leq \tilde{C}(0)$$

from which follows, for any $w^* \in \Omega$,

$$\begin{aligned}
 \sup_{w \in \Omega} |y_\alpha(\rho(s), w) - \bar{x}(w^*)| & \leq \sup_{w \in \Omega} |y_\alpha(\rho(s), w) - \bar{x}(w)| + |\bar{x}(w) - \bar{x}(w^*)| \\
 & \leq \tilde{C}(0) + \text{diam}(F_x)
 \end{aligned} \tag{4.17}$$

due to the boundedness of the EP optimal points $F_x = \{\bar{x}(w) | w \in \Omega\}$. We now consider three different cases.

Case $T \geq T^*$. By definition of T^* and thanks to the error decay (4.16) over $[0, T]$, it holds $\text{Err}[f(T^*)] \leq \varepsilon$.

Case $T < T^*$ and $\text{Err}[f(T)] = \varepsilon$. There is nothing to prove here.

Case $T < T^*$, $\text{Err}[f(T)] > \varepsilon$ and $\sup_{w \in \Omega} |y_\alpha(\rho(T), w) - \bar{x}(w)| > \tilde{C}(T)$. We will show that if α is sufficiently large this case cannot occur. In particular, we prove that for large α 's, $\sup_{w \in \Omega} |y_\alpha(\rho(T), w) - \bar{x}(w)| < \tilde{C}(T)$ and therefore we have a contradiction.

Let us fix any $w \in \Omega$ and let \tilde{C}_ε be given by

$$\tilde{C}_\varepsilon := \frac{1}{2} \frac{\lambda - 2\sigma^2}{\lambda + 2\sigma^2} \varepsilon.$$

We take q, r as follows

$$q := \frac{1}{2} \min \left\{ \left(\frac{\tilde{C}_\varepsilon}{c_1 2\sqrt{d}} \right)^{p_1}, c_3 \right\},$$

$$r := \min \left\{ \left(\frac{q}{c_2} \right)^{p_2}, R \right\}.$$

This choice implies $r \leq R$ and

$$\begin{aligned} \sqrt{d}c_1 \left(q + c_2 r^{1/p_2} \right)^{1/p_1} &\leq \sqrt{d}c_1 \left(q + c_2 \frac{q}{c_2} \right)^{1/p_1} = \sqrt{d}c_1 (2q)^{1/p_1} \\ &\leq \sqrt{d}c_1 \frac{\tilde{C}_\varepsilon}{c_1 2\sqrt{d}} = \frac{\tilde{C}_\varepsilon}{2}. \end{aligned} \tag{4.18}$$

We can apply now Lemma 4.1 with $x^* = \bar{x}(w)$ and $\rho = \rho(T) = \text{Law}(X_T)$. We obtain

$$\begin{aligned} |y^\alpha(\rho(t), w) - \bar{x}(w)| &\leq \sqrt{d}c_1 \left(q + c_2 r^{1/p_2} \right)^{1/p_1} + \frac{\sqrt{d} \exp(-\alpha q)}{\mathbb{P}(X_T \in B_r^\infty(\bar{x}(w)))} \mathbb{E}[|X_T - \bar{x}(w)|] \\ &\leq \frac{\tilde{C}_\varepsilon}{2} + \frac{\sqrt{d} \exp(-\alpha q) \mathbb{E}[|X_T - \bar{x}(w)|]}{\mathbb{P}(X_T \in B_r^\infty(\bar{x}(w)))}, \end{aligned} \tag{4.19}$$

where the last inequality follows from (4.18). At the same time, thanks to the bound provided by (4.17), we can apply Lemma 4.2 to obtain a lower bound on the mass around $\bar{x}(w)$ at time T :

$$\mathbb{P}(X_T \in B_r^\infty(\bar{x}(w))) \geq \mathbb{E}[\phi_{r,w}(X_0)] \exp(-pT) \tag{4.20}$$

for some $p > 0$ independent on w, α , and where $\phi_{r,w}(x)$ is now the mollifier centered around the EP optimal point $\bar{x}(w)$. We note that $\inf_{w \in \Omega} \mathbb{E}[\phi_{r,w}(X_0)] =: m_0 > 0$, due to our choice of initial data. Moreover, we have by Jensen’s inequality

$$\begin{aligned} \mathbb{E}[|X_T - \bar{x}(w)|] &\leq \mathbb{E}[|X_T - \bar{x}(W_T)|] + \mathbb{E}[|\bar{x}(W_T) - \bar{x}(w)|] \\ &\leq \sqrt{2\text{Err}[f(0)]} + \text{diam}(F_x) \end{aligned}$$

where we used again the error decay estimate and the boundedness of F_x . By combining the above estimates, we obtain

$$\begin{aligned} \frac{\sqrt{d} \exp(-\alpha q) \mathbb{E}[|X_T - \bar{x}(w)|]}{\mathbb{P}(X_T \in B_r^\infty(\bar{x}(w)))} &\leq \frac{\sqrt{d} \exp(-\alpha q + Tp)}{m_0} \left(\sqrt{2\text{Err}[f(0)]} + \text{diam}(F_x) \right) \\ &\leq \frac{\sqrt{d} \exp(-\alpha q + T^*p)}{m_0} \left(\sqrt{2\text{Err}[f(0)]} + \text{diam}(F_x) \right) \end{aligned}$$

where the right-hand side is independent on the choice of $w \in \Omega$. We also note that the decay rate q does not depend on α . As a consequence, there exists $\alpha_0 > 0$, such that it holds

$$\frac{\sqrt{d} \exp(-\alpha q) \mathbb{E}[|X_T - \bar{x}(w)|]}{\mathbb{P}(X_T \in B_r^\infty(\bar{x}(w)))} < \frac{\tilde{C}_\varepsilon}{2} \text{ for all } w \in \Omega,$$

for any $\alpha > \alpha_0$.

Together with (4.19), this leads to $\sup_{w \in \Omega} |y_\alpha(\rho(T), w) - \bar{x}(w)| < \tilde{C}_\varepsilon$. By definition of $\tilde{C}(T)$, it holds $\tilde{C}_\varepsilon \leq \tilde{C}(T)$ and so we reached the desired contradiction. \square

Theorem 4.1 and Corollary 4.1 show that the particle distribution will converge towards a neighborhood of the Pareto front. Due to the adaptive strategy, the distribution might move along the front. In the next section, we will therefore analyze the dynamics in the parameters space Ω to investigate the diversity of the computed solution.

4.3 Decay of Diversity Measure

The aim of interactions (3.8), (3.9) is to improve the distribution of the parameters $\{W_t^i\}_{i=1}^N$ so that, in view of Theorem 2.1, the corresponding (weak) EP optimal points are well-distributed in the image space. In the following, we will investigate the special case where $m = 2$ and the Pareto front F corresponds to the unit simplex Ω . Dynamics (3.8) can then be shown to be a discretization of a gradient flow on the unit simplex Ω when Chebyshev scalarization (2.2) is used. While considering $F = \Omega$ is clearly a strong assumption, the analysis sheds some light on why (3.8) may be a suitable adaptive strategy for the parameters.

We restrict again to the case of Assumption 4.1, which ensures uniqueness of global minima and continuity of \bar{x} , where $\bar{x}(w)$ is the global solution to (2.2) for a given $w \in \Omega$. As we are interested in the relation between w and its corresponding point on the Pareto front $g(\bar{x}(w)) \in F$, let us assume

$$X_t = \bar{x}(W_t) \tag{4.21}$$

that is, every particle is assumed to be located at its corresponding EP optimal point. This is justified by the convergence result (Theorem 4.1) and by the fact that the particle interactions in the search space \mathbb{R}^d take place at a faster time scale than the adaptation of the parameters, see Remark 4.2.

The parameter evolution in the mean-field model (4.1) becomes then independent on X_t and can be written as

$$dW_t = \tau P_{W_t} \left(\int \nabla U (\bar{g}(W_t) - \bar{g}(w)) d\mu(t, w) \right) dt \tag{4.22}$$

where $\mu(t) = \text{Law}(W_t)$ and, for simplicity, we introduced $\bar{g} := g \circ \bar{x}$.

Assumption 4.4 Let (1.1) be a bi-objective problem, $m = 2$, and the scalar sub-problems be given by the Chebyshev scalarization strategy (2.2). We assume the Pareto front F is given by

$$F = \Omega = \{(s, 1 - s)^\top : s \in [0, 1]\}$$

and that the binary potential energy U is radially symmetric.

Lemma 4.3 Under Assumption 4.4, for all $w, v \in \Omega$ it holds

$$P_w(\nabla U(\bar{g}(w) - \bar{g}(v))) = -P_w(\nabla U(w - v)). \tag{4.23}$$

Proof We note that when $F = \Omega$, all weakly EP optimal points are also EP optimal and hence by Theorem 2.1 $\bar{g}(w) = g(\bar{x}(w)) \in F$ for all $w \in \Omega$. Then, for any $w \in \Omega$, there exists $s \in [0, 1]$ such that $\bar{g}(w) = (s, 1 - s)$. By definition of the sub-problem (2.2) with $w = (w_1, w_2) = (w_1, 1 - w_1)$,

$$\bar{g}(w) = \min_{y \in F} \max \{y_1 w_1, y_2 w_2\} = \min_{s \in [0,1]} \max \{s w_1, (1 - s)(1 - w_1)\},$$

due to the definition of $G(x, w)$. At the minimizer, it must hold $s w_1 = (1 - s)(1 - w_1)$ and hence $s = (1 - w_1)$. It follows that

$$\bar{g}(w) = Aw, \quad \text{where } A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{4.24}$$

Since U is radially symmetric it holds $\nabla U(Aw - Av) = A\nabla U(w - v)$.

Finally, let us consider the basis $n_1 = (1, 1)^\top, n_2 = (1, -1)^\top$ and a vector $u \in \mathbb{R}^2, u = u_1 n_1 + u_2 n_2$ for some $u_1, u_2 \in \mathbb{R}$. We note that the tangent cone at $w \in \Omega$ satisfies $T(w, \Omega) \subseteq \text{span}(n_2)$ and therefore $P_w(n_1) = 0$. Together with the fact that $Au = u_1 n_1 - u_2 n_2$, this leads to

$$P_w(Au) = P_w(-u_2 n_2) = P_w(-u)$$

and identity (4.23) follows. □

Thanks to Lemma 4.3, under Assumption 4.4 equation (4.22) can be simplified to

$$dW_t = \tau P_{W_t} \left(- \int \nabla U(W_t - w) d\mu(t, w) \right) dt. \tag{4.25}$$

We note that $\mu(t)$ weakly satisfies

$$\frac{\partial}{\partial t} \mu(t, w) = -\tau \nabla_w \cdot \left(P_w \left(- \int \nabla U(w - v) d\mu(t, v) \right) \mu(t, w) \right) \tag{4.26}$$

with initial conditions $\mu(0) = \mu_0$ over the test set $C_c^\infty(\Omega)$. Equation (4.26) describes the continuum dynamics of particles which binary interact and that are confined to the set Ω . Such aggregation model on bounded domains has been the subject of several works, see for instance [14, 28, 29, 55]. Particularly relevant to the present work is [14] where general prox-regular sets, like Ω , are considered. We recall the main result from [14] for completeness, after introducing the necessary notation.

In the following, $\text{Conv}(\Omega - \Omega)$ denotes the convex envelop of $(\Omega - \Omega) = \{w - v : w, v \in \Omega\}$ and $U \in C^1(\mathbb{R}^m)$ is said to be $\tilde{\lambda}$ -geodesically convex on a set $A \subset \mathbb{R}^m$ if for all $x, y \in A$

$$U(x) \geq U(y) + \nabla U \cdot (y - x) + \frac{\tilde{\lambda}}{w} |y - x|^2.$$

Theorem 4.2 ([14, Theorem 1.5]) *Assume $U \in C^1(\mathbb{R}^m)$ to be $\tilde{\lambda}$ -geodesically convex on $\text{Conv}(\Omega - \Omega)$ for some $\tilde{\lambda} \in \mathbb{R}$. For any initial data $\mu_0 \in \mathcal{P}_2(\Omega)$ there exists a locally absolutely continuous curve $\mu(t) \in \mathcal{P}_2(\Omega), t > 0$, such that μ is a gradient flow with respect to \mathcal{U} . Also, μ is a weak measure solution to (4.26).*

Furthermore,

$$\frac{d}{dt} \mathcal{U}[\mu(t)] \leq - \int |P_w(\nabla U * \mu(t)(w))|^2 d\mu(t, w), \tag{4.27}$$

where $*$ denotes the convolution operator.

In these specific settings, thanks (4.24) and continuity of \bar{x} we have

$$\mathcal{U}[\mu(t)] = \mathcal{U}[\bar{g}_\# \mu(t)] = \mathcal{U}[g_\# \rho(t)]$$

since $\bar{g} = g \circ \bar{x}$ and $\rho(t) = \bar{x}_\# \mu(t)$ due to (4.21). Therefore, Theorem 4.2 states that the energy over the front is decreasing. We note that the flow may converge to stationary points of (4.22) that are not minimal configurations, as observed in [28] for even simple domains.

Clearly, without ansatz (4.21), there is no guarantee that the potential decreases along the evolution of the algorithm. Quite the opposite, by Theorem 4.1 particles are expected to concentrate on the Pareto front leading to an increased potential \mathcal{U} . Nevertheless, if Theorem 4.1 applies, there exists a time $T > 0$ where

$$\mathbb{E}[|X_t - \bar{x}(W_t)|^2] < \varepsilon \quad \text{and hence} \quad X_t \approx \bar{x}(W_t)$$

making ansatz (4.21) valid. Therefore, we claim that the reduced model (4.22) describes the dynamics for $t > T$. We will numerically investigate two phases of the algorithm: the first one when concentration over the Pareto happens, and the second when the potential \mathcal{U} decays leading to an improved diversity of the solution.

5 Numerical Experiments

In this section, we numerically investigate the performance of the proposed method by testing it against several benchmark bi-objective problems. We also examine the role of the parameters – in particular of τ which controls the step size of the adaptation process of the weights vectors W_k^i .

For the sake of reproducible research, an implementation in MATLAB code of the proposed algorithm is made available in the GitHub repository <https://github.com/borghig/AM-CBO>.

We recall that the particles positions $\{X_k^i\}_{i=1}^N$ are updated according to (3.2), while the sub-problem parameters $\{W_k^i\}_{i=1}^N$ are updated according to (3.8) in the following experiments. For experiments with the second adaptive strategy (3.9) and validation against tri-objective problems, we refer to [5]. The complete optimization method is described by Algorithm 1. A remark on the algorithm computational complexity follows.

Algorithm 1 AM-CBO

```

Set parameters:  $\alpha, \lambda, \sigma, \tau, \Delta t$ 
Initialize the positions:  $X_0^i \sim \rho_0, i = 1, \dots, N$ 
Initialize the weights vectors  $\{W_0^i\}_{i=1}^N$  uniformly in  $\Omega$ 
 $k \leftarrow 0$ 
while stopping criterion is NOT satisfied do
  Compute  $g(X_k^i), i = 1, \dots, N$ 
  for  $i = 1, \dots, N$  do
    compute  $Y_k^\alpha(W_k^i)$  according to (3.1)
    sample  $B_k^i$  from  $\mathcal{N}(0, I_d)$ 
    update  $X_{k+1}^i$  according to (3.2)
    update  $W_{k+1}^i$  according to (3.8)
  end for
   $k \leftarrow k + 1$ 
end while
return  $\{X_k^i\}_{i=1}^N$ 

```

Remark 5.1 Even though in every iteration the objective function g is evaluated only N times, the overall computational complexity is $\mathcal{O}(N^2)$ because the computation of $Y_k^\alpha(w)$ requires $\mathcal{O}(N)$ computations, as well as the parameters update (3.8) which is particularly costly.

One can reduce the computation complexity by considering only a random subset $I_k^M \subset \{1, \dots, N\}$ of $M \ll N$ particles when computing (3.1) and (3.8), by substituting

$$\frac{1}{N} \sum_{j=1}^N (\cdot)^j \quad \text{with} \quad \frac{1}{M} \sum_{j \in I_k^M} (\cdot)^j,$$

whenever a sum over the different particles is performed. Inspired by Monte-Carlo particle simulations [1, 46], this mini-random batch technique allows to lower the

complexity to $\mathcal{O}(NM)$. We also note that Fast Multipole Methods (FMM) [38] may additionally be used to speed up the computation of the potential field. Then, the computational complexity of (3.8), (3.9) is further reduced.

5.1 Performance Metrics

Denote by $\{X_k^i\}_{i=1}^N$ the set of particle positions at the k -th algorithm iteration and their empirical distribution by $\rho_k^N \in \mathcal{P}(\mathbb{R}^d)$. We employ three different energies, the Riesz s -energy (3.7), Newtonian and the Morse potentials, both to measure the diversity of the solutions and to determine the dynamics of the weights vectors. The Newtonian binary potential is given by

$$U_N(z) = \begin{cases} -\log |z| & \text{if } m = 2 \\ |z|^{2-m} & \text{if } m > 2 \end{cases}, \quad (5.1)$$

while the Morse potential is given

$$U_M(z) = e^{-C|z|} \quad \text{with } C > 0. \quad (5.2)$$

All considered potentials describe short-range repulsion between the particles. While the Morse potential is λ -geodesically convex, the Newtonian and Riesz potentials are not. Since we will also employ the corresponding energies $\mathcal{U}_R, \mathcal{U}_N, \mathcal{U}_M$ to define the interaction between weights vectors, the constant C can be considered as an algorithm parameter when the Morse repulsion is used.

To show the validity of the energy-based diversity metrics, we additionally consider the hypervolume contribution metric \mathcal{S} [66]. Let $g^* \in \mathbb{R}^m$ be a maximal element with respect to the natural partial ordering

$$y_i < g_j^* \quad \text{for all } y \in F,$$

the hypervolume measure is given by the Lebesgue measure of the set of points between the computed solution and the maximal point g^* , that is

$$\mathcal{S}[\rho_k^N] = \left| \bigcup_{i=1}^N \left\{ y \in \mathbb{R}^m \mid g(X_k^i) < y < g^* \right\} \right|. \quad (5.3)$$

Maximizing \mathcal{S} has been shown to lead to a diverse approximation of the Pareto front [23].

In Sect. 4.1, the convergence of the mean-field dynamics towards the Pareto front is shown by studying the evolution of the Generation Distance GD (4.4). In the experiments, we approximate this quantity by considering a reference approximation $\{y^j\}_{j=1}^M$ of the front with $M = 100$ points $y^i \in F, i = 1, \dots, M$ for every test problem. More details on the reference solution are given in Appendix A. For simplicity, we indicate

the numerical approximation of the Generational Distance again by GD , which is defined by

$$GD[\rho_k^N] = \left(\frac{1}{N} \sum_{i=1}^N \text{dist}(g(X_k^i), F_M)^2 \right)^{\frac{1}{2}}. \tag{5.4}$$

The Inverted Generational Distance (IGD) is also considered. It consists of the average distance between the points of the reference solution $\{y^i\}_{j=1}^M$ and the computed front

$$IGD[\rho_k^N] = \left(\frac{1}{M} \sum_{j=1}^M \text{dist}(y^j, G_k)^2 \right)^{\frac{1}{2}} \quad \text{with } G_k := \{g(X_k^i)\}_{i=1}^N. \tag{5.5}$$

Contrary to GD , which only measures the distance from the Pareto front, IGD takes into account the diversity of the computed solution, too. Hence, IGD is also a suitable indicator of the overall quality of the solution.

5.2 Test Problems

Test problems with diverse Pareto front geometries are selected to show the performance of the proposed method. In the Lamé problems [24] the parameter $\gamma > 0$ controls the front curvature: we use $\gamma = 0.25, 1, 3$ to obtain convex, linear and concave fronts respectively. We also consider the DO2DK [8] problems with $k = 2, s = 1$ and $k = 4, s = 2$. Here, the Pareto fronts have more complex geometries as they are not symmetric and, in one case, the front is discontinuous. All the above problems are scalable to any dimension of the search space d and in the image space m . For presentation purposes, we restrict ourselves to bi-objective optimization problems by setting $m = 2$, but consider possibly large d . In the considered benchmark problems, the analytical description of the fronts is known, allowing us to obtain reference solutions (see Appendix A). The problems definitions are recalled in Appendix B for completeness.

In this section, we use Algorithm 1 in four different scenarios

1. No parameters interaction $\tau = 0$;
2. Riesz potential (3.7), with $\tau = 10^{-5}$;
3. Newtonian potential (5.1), with $\tau = 10^{-3}$;
4. Morse potential (5.2), with $\tau = 10^{-1}, C = 20$;

The first scenario clearly corresponds to the standard M-CBO approximation, while the others to different AM-CBO strategies. The parameter τ for each setting has been tuned to obtain the best performance. We note how the stronger the potential singularity is at the origin, the smaller the optimal value of τ is. To validate model (4.26) and Theorem 4.2, we update the parameters of the scalarized sub-problems according to (3.8). The initial weights vectors $\{W_0^i\}_{i=1}^N$ are taken (deterministically) uniformly distributed over Ω , while the particle positions are uniformly sampled over

$H = [0, 1]^d$, $d = 10$. We employ $N = 100$ particles, which evolve for a maximum of $k_{\max} = 5000$ steps. The remaining parameters are set to $\lambda = 1$, $\sigma = 4$, $\alpha = 10^6$. This parameter choice consists of a compromise between the optimal parameters of each problem. Anisotropic diffusion (3.4) is used.

Figure 3 shows the computed solutions, in the image-space, in the four different scenarios. Regardless of the interaction on Ω , the particles always converge towards EP optimal points and hence to the Pareto front. By definition of the Chebyshev sub-problems (2.2), a uniform distribution in Ω leads to a uniform distribution of the particles over the front only when F is linear (as in the Lamé problem $\gamma = 1$). Indeed, Fig. 3 shows that the particles are well distributed even when there is no weights interaction ($\tau = 0$). If the front geometry differs from this straight segment, the optimal parameters distribution on Ω differs from the uniform one. In particular, subsets of the Pareto front which are almost parallel to the axis are difficult to approximate without any interaction in the parameter space, see for instance Lamé $\gamma = 0.25$ and the DO2DK problems in Fig. 3. When using $\tau \neq 0$, solutions improve as the particles are more distributed over the entire front.

Table 1 reports the performance metrics for all the problems. For a reference, we include the results obtained with the well-known NSGA-II algorithm [18], using the implementation provided in [49]. For better comparability, we use the same initial population size $N = 100$ and number of iterations $k_{\max} = 5000$, while the remaining parameters are set to default settings (crossover probability $p_c = 0.9$, mutation probability $p_m = 0.5$, mutation strength $m_s = 0.05$). For most problems, the strategy $\tau = 0$ with no interaction in Ω allows to reach lower values of GD . This is consistent with Theorem 4.1 and Remark 4.2, which suggested that the additional dynamics may interfere with the CBO mechanism and, as a consequence, slow down the convergence towards EP optimal points. If diversity metrics \mathcal{U}_R , \mathcal{U}_N , \mathcal{U}_M and \mathcal{S} are considered, dynamics including the interaction of parameters allow to obtain more diverse solutions. Interestingly, using Morse binary potential in the interaction leads to a final lower Newtonian energy in some cases. We will investigate the role of the choice of the potential \mathcal{U} and τ in the next section. We note that NSGA-II outperforms the proposed algorithm in the Lamé problems in terms of IGD metrics, whereas better results are obtained by AM-CBO in the DO2DK benchmarks.

In Fig. 3 the IGD performance shows that letting particles interact in parameter space improves the overall quality of the solution. While the improvement is more substantial in problems with complex Pareto fronts (see for instance Lamé $\gamma = 0.25$, or DO2DK $k = 2$), we remark that the additional mechanism allows to obtain better solutions even if the parameter distribution is already optimal from the beginning (see Lamé $\gamma = 1$). We conjecture that this is due to the additional stochasticity introduced by the potential. We will also investigate this aspect in the next subsection.

Figure 4a and b show the time evolution of GD , \mathcal{U}_R , \mathcal{U}_N , \mathcal{U}_M and IGD for two of the considered test problems. As suggested by the analysis of the mean-field model, in particular Theorem 4.1, GD exponentially decays up to a maximum accuracy within the first iterations of the algorithm. This is due to the CBO dynamics driving the particles around EP optimal points. At the same time, the potential energies increase as the particles concentrate towards the front in the image-space. If the particles are concentrated at the local minimizer, then the resulting dynamics on the mean-field

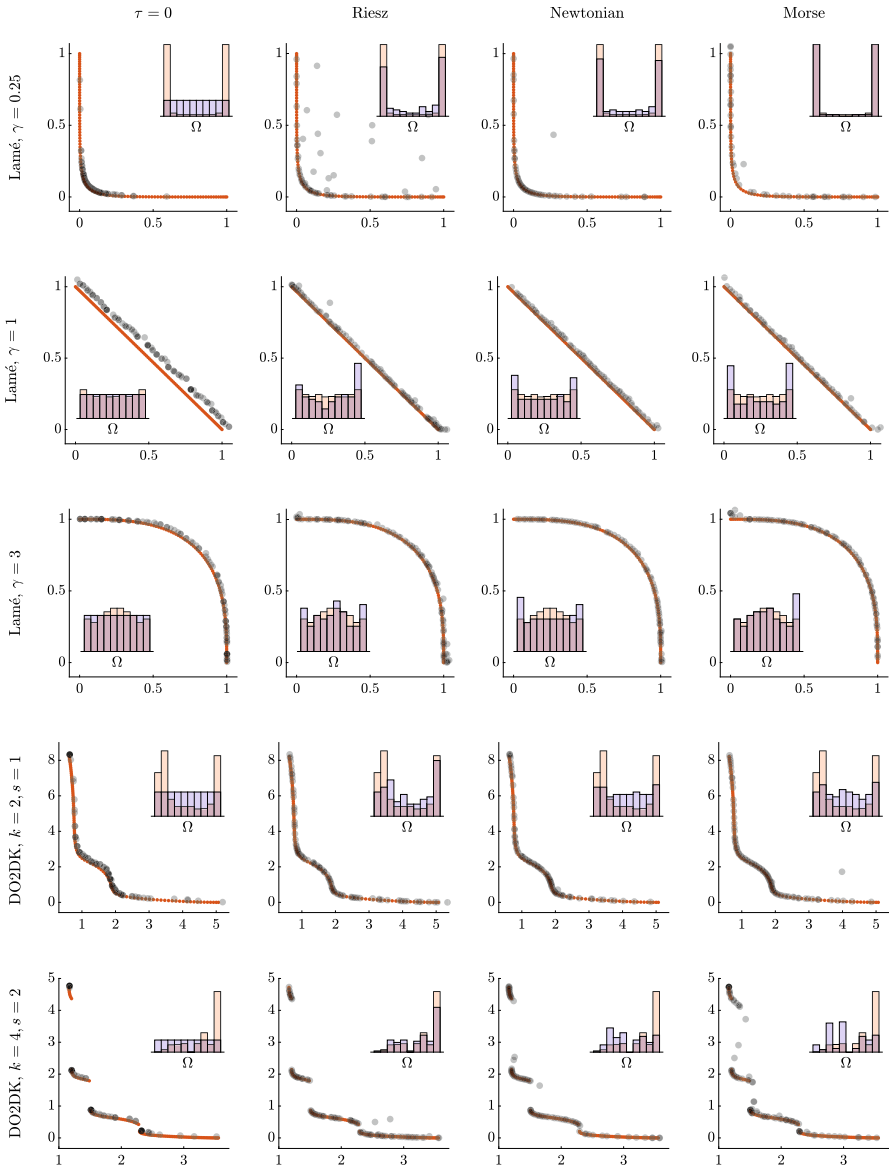


Fig. 3 Plots display, in gray, the particle positions in the image space after a single run for different benchmark problems. The reference solution is displayed in red. Four different parameters interaction strategies are used: no interaction, Riesz, Newtonian and Morse potential. Histograms show the final distribution over Ω (blue) and the optimal one (red) (Color figure online)

Table 1 Algorithm performance for the different settings and problems

Problem	Interaction	GD	U_R	U_N	U_M	S	IGD
Lamé 0.25	$\tau = 0$	2.33e-02	1.00e+10	2.41e+00	4.86e-01	9.69e-01	1.31e-01
	Riesz	8.74e+00	5.65e+00	- 1.94e-01	9.62e-02	7.77e-01	4.06e-02
	Newtonian	1.11e+01	8.23e+00	- 3.53e-01	1.14e-01	8.38e-01	4.25e-02
	Morse	1.49e+01	1.81e+04	- 1.36e+00	3.40e-02	7.45e-01	2.64e-02
Lamé 1	NSGA-II	6.99e-03	7.98+00	9.00e-01	7.45e-01	9.84e-01	4.19e-03
	$\tau = 0$	9.88e-02	9.60e+09	9.96e-01	1.26e-01	3.74e-01	8.28e-02
	Riesz	1.63e-01	6.54e+00	8.57e-01	1.23e-01	4.59e-01	1.56e-02
	Newtonian	9.81e-01	8.39e+00	5.77e-01	9.47e-02	4.62e-01	1.91e-02
Lamé 3	Morse	6.83e-01	8.97e+05	4.41e-01	7.95e-02	4.48e-01	1.78e-02
	NSGA-II	2.48e-02	5.86e+00	1.03e+00	1.06e-01	4.71e-01	4.33e-03
	$\tau = 0$	1.93e-02	8.40e+09	9.56e-01	1.30e-01	8.45e-02	2.18e-02
	Riesz	5.64e-02	7.06e+00	7.68e-01	1.14e-01	1.01e-01	1.32e-02
DO2DK k = 2, s = 1	Newtonian	2.34e-01	6.33e+00	5.74e-01	9.10e-02	1.03e-01	1.11e-02
	Morse	3.02e-01	9.57e+06	5.04e-01	7.84e-02	1.02e-01	1.29e-02
	NSGA-I	3.64e-02	4.65e+00	9.16e-01	8.10e-02	9.68e-02	5.28e-03
	$\tau = 0$	1.80e-01	1.00e+10	- 3.30e-01	6.94e-02	8.84e+01	2.82e-01
DO2DK k = 4, s = 2	Riesz	5.03e-02	1.58e+00	- 6.04e-01	4.18e-02	8.94e+01	1.18e-01
	Newtonian	6.48e-02	1.77e+01	- 5.98e-01	3.85e-02	8.94e+01	1.07e-01
	Morse	9.59e-02	7.67e+08	- 5.64e-01	3.75e-02	8.94e+01	9.33e-02
	NSGA-II	9.97e-02	9.44e-01	- 9.18e-01	1.85e-02	8.91e+01	3.42e-02
Performance of NSGA-II algorithm	$\tau = 0$	6.60e-02	1.00e+10	2.69e+00	2.64e-01	8.66e+01	1.36e-01
	Riesz	8.95e-01	3.34e+00	- 1.27e-01	7.08e-02	8.40e+01	2.61e-02
	Newtonian	1.50e+00	2.18e+01	- 2.26e-01	7.52e-02	8.44e+01	3.61e-02
	Morse	9.85e+00	1.94e+09	- 1.56e-01	9.09e-02	7.63e+01	3.45e-02
NSGA-II	7.80e-02	2.36e+00	9.85e-02	4.19e-02	8.66e+01	1.11e-01	

Performance of NSGA-II algorithm is also provided as a reference. Bold boxes show the best results among the proposed algorithm only. Results are averaged over 25 runs

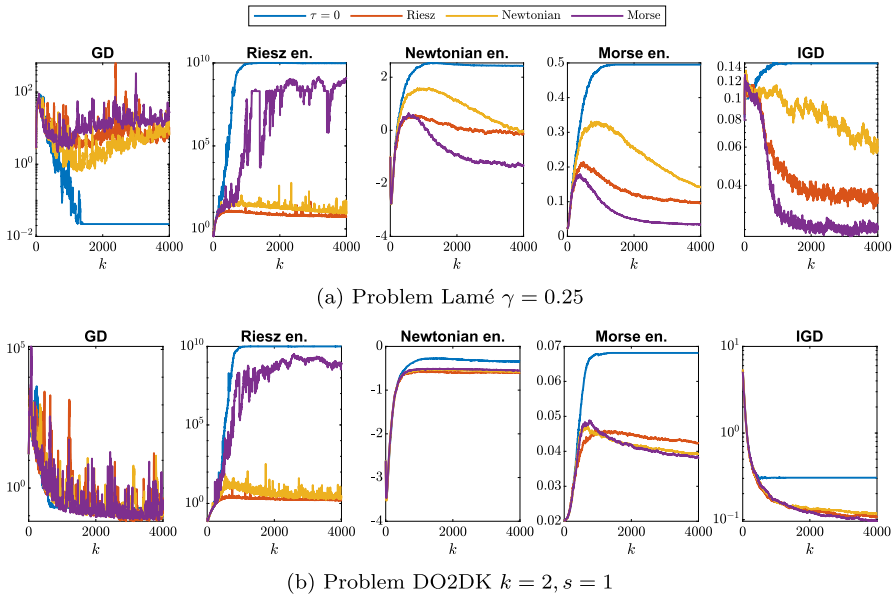


Fig. 4 Performance metric evolution, results are averaged over 25 runs

level are described by equation (4.26). This is a gradient flow for the marginal μ . While this is rigorously true only under Assumption 4.4, we nevertheless observe similar phenomena also numerically in other cases, as seen in Fig. 4a and b, where the potentials start decreasing provided that relatively low GD values are attained.

5.3 Effect of the Parameter τ and Scalability

By looking at the computational results, it becomes clear that the two phases of the algorithm, the one characterized by the CBO dynamics and the one characterized by the gradient-flows dynamics, have different scales. Typically, the former dynamics are much slower compared with the second one. This was consistent with assumptions to Theorem 4.1, where τ needs to be taken of order $o(\sqrt{\varepsilon})$. Interestingly, we note that the multi-swarm CBO approach suggested in [47] leads to the opposite behavior, where weights adapt first and particles create consensus afterwards.

To experimentally investigate the importance of parameter τ , we test the algorithm for various values of τ , keeping the remaining parameters fixed. Figure 5a and b show the final GD and IGD metrics when different binary potentials are used during the computation. As expected, relatively large values of τ lead to a strong interaction in parameter space that interferes with the CBO mechanism. As a result, the GD metric increases for large values of τ . Interestingly, the lowest GD values are not always attained for the smallest values of τ , suggesting that the additional weights vectors dynamics might help the CBO mechanism in optimizing the sub-problems.

The IGD metrics in Fig. 5b, show that the optimal value of τ is different for each test case. In particular, DO2DK problems benefit from a strong interaction in

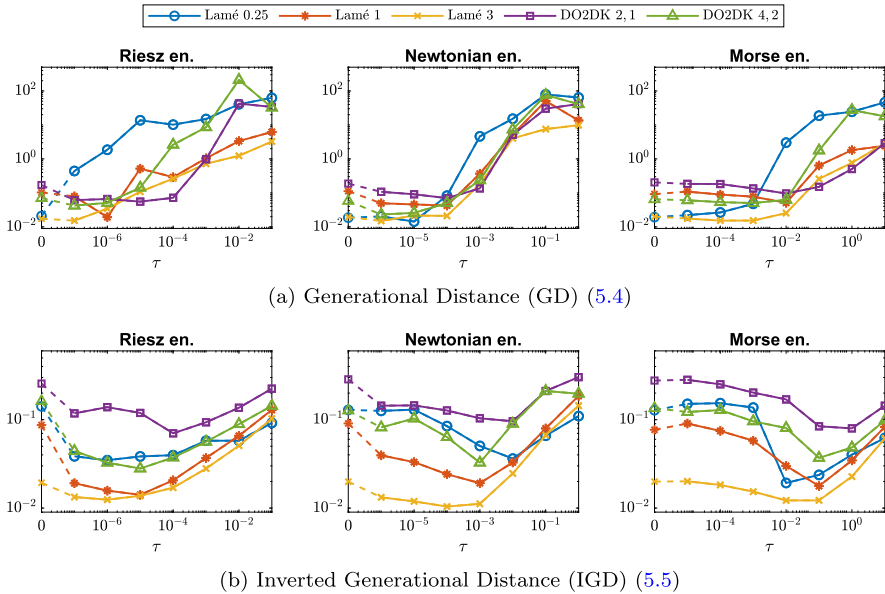


Fig. 5 Performance metrics as a function of τ for all the problems considered. Results are averaged over 10 runs

parameter space. This might be explained by the front geometry (Fig. 3): the front length is long and, as consequence, the particles tend to be further apart in the image space, making the binary potential interaction weaker. Larger values of τ mitigate this effect, leading to better algorithm performances. If the extrema of the Pareto front are known in advance, one could address this issue by estimating the front length and choosing the parameter τ accordingly. We also note that the proposed algorithm seems to perform better when the Morse potential is used during the computation.

As already mentioned, the dynamics in Ω add stochasticity to the particle evolution. Hence, the additional diffusive term $\sigma D_k^i(X_k^i, W_k^i) B_k^i$ in (3.2) might not be necessary. Yet, taking $\sigma = 0$ yields poor approximations of the Pareto front, see Fig. 6 b, suggesting that the diffusive term is still of paramount importance for the particles exploratory behavior and their statistical independence. From Fig. 6a, it is obvious that the optimal diffusion parameter σ is larger, the smaller τ is. In particular, if $\tau = 0$ the particles diverge from the EP optimal points only when $\sigma > 10$, which is consistent with other CBO methods for single-objective optimization with anisotropic diffusion, see for instance [2]. At the same time, for some problems, if σ is too small, larger values of τ improve the convergence towards optimal points.

Finally, we test the algorithm performance for different dimensions d of the search space, keeping the same parameters choice. If the same number $N = 100$ of particles is used, the *IGD* of the computed solutions increases as the space dimension d becomes larger, see Fig. 7. This effect can be simply reduced by increasing the number of particles linearly with the space dimension, see Fig. 7.

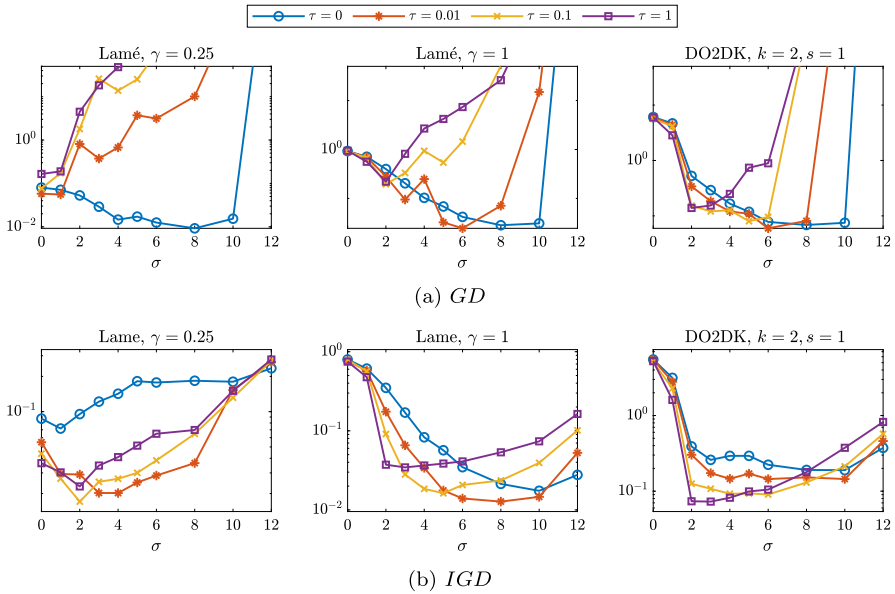


Fig. 6 Metrics as functions of σ , for different values of τ . Morse interaction is used and results are averaged over 5 runs

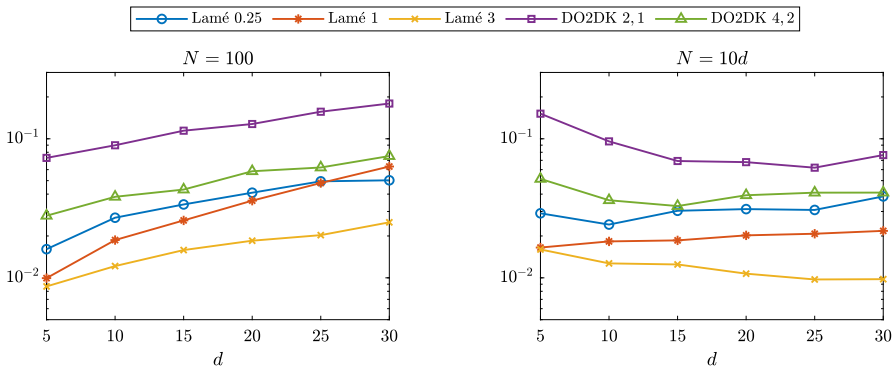


Fig. 7 *IGD* metrics as functions of the search space dimension d . Morse interaction is used and results are averaged over 20 runs

6 Conclusions

In this work, we proposed an adaptive stochastic particle dynamics based on consensus to solve multi-objective optimization problems. The method makes use of a scalarization strategy to break down the original problem into N parametrized single-objective sub-problems and can deal with non-differentiable objective functions. The proposed algorithm, AM-CBO, extends prior work on multi-objective consensus based optimization by additional adaptive dynamics in the parameter space in order to ensure

that the particles distribute uniformly over the Pareto front. This is achieved by exploiting energy-based diversity measures. A rigorous mathematical analysis is provided to validate this behavior however only in a particular case, where Assumptions 4.1–4.3 hold true. Under those conditions, we theoretically investigated the long-time behavior of the particle dynamics under the propagation of chaos assumption and establish convergence towards EP optimal points. The particles are capable of solving several single-objective problems at the same time, saving also computational cost with the respect to a naive approach. The additional dynamics on the parameter space are also analyzed based on results on non-linear aggregation equation under Assumption 4.4. Numerical experiments show that the proposed method is capable to solve multi-objective problems with very different Pareto front even those violating the previous assumptions. The algorithm scales well with the problem dimension, even when using a relatively small number of particles. Our scaling of the dynamics in the weights leads to a behavior, where first the particles concentrate along the front and then are redistributed. In future work, also the opposite may be considered as typical for classical multi-objective optimization methods.

Funding Open Access funding enabled and organized by Projekt DEAL. This work has been written within the activities of GNCS group of INdAM (National Institute of High Mathematics). L.P. acknowledges the partial support of MIUR-PRIN Project 2017, No. 2017KKJP4X “Innovative numerical methods for evolutionary partial differential equations and applications”. The work of G.B. is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projektnummer 320021702/GRK2326—Energy, Entropy, and Dissipative Dynamics (EDDy). M.H. thanks the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the financial support through 320021702/GRK2326, 333849990/IRTG-2379, CRC1481, HE5386/18-1,19-2,22-1,23-1, ERS SFDdM035 and under Germany’s Excellence Strategy EXC-2023 Internet of Production 390621612 and under the Excellence Strategy of the Federal Government and the Länder. The authors acknowledge the support of the Banff International Research Station (BIRS) for the Focused Research Group [22frg198] “Novel perspectives in kinetic equations for emerging phenomena”, July 17–24, 2022, where part of this work was done.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Construction of Reference Solutions

Even if we assume there exists an analytical representation of the Pareto front F , finding an M -approximation of F which also minimizes a given two-body potential is a computationally expensive task, which is related to the already mentioned crystallization problem in physics [4]. In [10], this was achieved by using mathematical

programming techniques, while in [26] the authors proposed the following heuristic strategy: generate $N \gg M$ points on the front and iteratively delete the point subject to the highest potential energy until only M are left. In this appendix, we propose a different heuristic strategy which not only generates low-energy approximations of the front, but also provides more insight into the choice of the proposed update strategy (3.8).

In the following, we assume F to be a $(m - 1)$ -dimensional manifold with known chart $\bar{g} \in \mathcal{C}^2(V, F)$

$$F = \{\bar{g}(z) \mid z \in V\},$$

where $V = [0, 1]^{m-1}$ or $V = \Omega$. As before, let $T(y, F)$ be the tangent cone at $y \in F$ to F . Let $\{G_t^i\}_{i=1}^M \subset F$ describe the positions at time $t \geq 0$ of M particles interacting over the front under a potential U , that is

$$\frac{d}{dt} G_t^i = P_{T(G_t^i, F)} \left(-\frac{1}{M} \sum_{j=1}^M \nabla U(G_t^i - G_t^j) \right), \tag{A.1}$$

with some given initial conditions $G_0^i = \bar{g}(Z_0^i)$, for all $i = 1, \dots, M$. Our heuristic strategy is based on the conjecture that, as $t \rightarrow \infty$, the system will eventually converge towards a low-energy configuration. Rather than solving (A.1) in \mathbb{R}^m where F is embedded, we consider the equivalent system for the coordinates $\{Z_t^i\}_{i=1}^M$,

$$\frac{d}{dt} Z_t^i = \left(D\bar{g}(Z_t^i) \right)^+ P_{T(G_t^i, F)} \left(-\frac{1}{M} \sum_{j=1}^M \nabla U(G_t^i - G_t^j) \right), \tag{A.2}$$

where $G_t^i = \bar{g}(Z_t^i)$ and $(\cdot)^+$ is the pseudo-inverse of (\cdot) . Equivalence between the systems is a consequence of the chain rule, see [39, Chapter 5]. We note that if G_t^i belongs to the extrema of F (or its ‘‘contour’’ when $m > 2$), $D\bar{g}(Z_t^i)$ might not be well-defined. In this case, though, the projection is the null map so we set $dZ_t^i/dt = 0$. System (A.2) can then be solved numerically if \bar{g} is explicitly known. The reference, low-energy, solution to (1.1) will then be given by the final configuration $\{G_T^i\}_{i=1}^M$ reached at a certain time horizon $T > 0$.

We note that dynamics (3.10) introduced in the parameters space Ω can be seen as an approximation to (A.2) when $m = 2$. Indeed, let $V = \Omega$ and the chart \bar{g} be the relation (given by Theorem 2.1) between parameters and points on F

$$\bar{g}(w) = \operatorname{argmin}_{x \in \mathbb{R}^d} G(x, w).$$

As \bar{g} and $T(y, F)$ in (A.2) are unknown during the optimization process, one could approximate them by assuming linearity on F . In particular, if no further information on the front geometry is available, let us take $F = \Omega$ as in Assumption 4.4. This leads

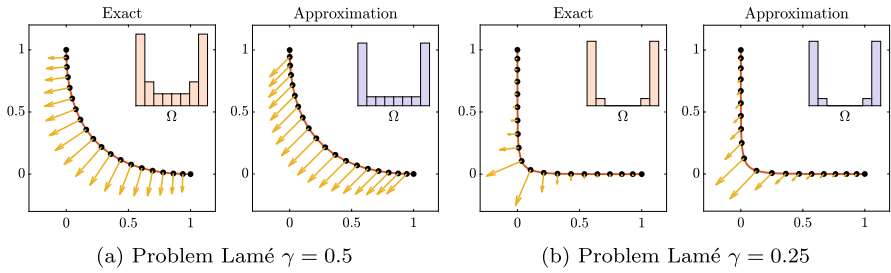


Fig. 8 Final configuration of $M = 20$ particles evolved according to (A.2) (exact dynamics) and (A.3) (approximated dynamics) under Riesz binary potential (3.7). Arrows show the total potential forces the particles are subject to. Two different front shapes are considered. The particle systems are solved with an explicit Euler scheme, $\Delta t = 10^{-8}$, $T = 0.01$. Histograms show the final distribution over the coordinate space $V = \Omega$

to

$$\bar{g}(w) \approx Aw, \quad D\bar{g}(w) \approx A \quad \text{where} \quad A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

as before in Lemma 4.3, and

$$P_{T(\bar{g}(w), F)} \approx P_{T(Aw, \Omega)} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} =: B$$

if $w > 0$ component-wise, and $P_{T(Aw, \Omega)} = 0$ otherwise. Starting from (A.2), it follows

$$\begin{aligned} \frac{d}{dt} W_t^i &= (D\bar{g}(W_t^i))^+ P_{T(G_t^i, F)} \left(-\frac{1}{M} \sum_{j=1}^M \nabla U(G_t^i - G_t^j) \right) \\ &\approx A^+ B \left(-\frac{1}{M} \sum_{j=1}^M \nabla U(G_t^i - G_t^j) \right). \end{aligned}$$

Now, since $A^+ = A$ and $AB = -B$ we obtain

$$\frac{d}{dt} W_t^i \approx -P_{T(w, \Omega)} \left(-\frac{1}{M} \sum_{j=1}^M \nabla U(G_t^i - G_t^j) \right) \tag{A.3}$$

which corresponds to the dynamics proposed in Sect. 3, provided $G_t^i \approx g(X_t^i)$.

To conclude, we remark that the above approximation has a mild impact on the final distribution over the front, even when F differs substantially from Ω , see Fig. 8.

Appendix B: Problem Definition

We report here the definition of the benchmark problems, together with the penalization strategy and known parametrization of F . The Lamé [24] and the DO2DK [8] problems are originally formulated as constrained multi-objective optimization problems where the feasible domain is given by $H = [0, 1]^d$. Moreover, the set of EP optimal points corresponds to the edge $[0, 1] \times \{0\}^{d-1}$. Adding a projection step to H has a relevant impact on the algorithm dynamics, as any point belonging to the cone $\mathbb{R} \times \mathbb{R}_{\leq 0}^{d-1}$ is projected to an EP optimal point. Therefore, we make use of an exact penalization strategy to ensure the particles remain in the feasible region adding a ℓ_1 -penalty term of the form $\beta \text{dist}(x, H)$, $\beta > 0$, to the original objective functions.

Let $x \in \mathbb{R}^d$, $x = (x_1, \dots, x_d)$ for $d \geq 1$, the objective functions are given by

- Lamé [24] with $\gamma \in \mathbb{R}_{>0}$,

$$\begin{aligned}
 g_1(x) &= \left| \cos\left(\frac{\pi}{2}x_1\right) \right|^{\frac{2}{\gamma}} (1 + r(x)) + \frac{\pi}{\gamma} \text{dist}(x, H) \\
 g_2(x) &= \left| \sin\left(\frac{\pi}{2}x_1\right) \cos\left(\frac{\pi}{2}x_2\right) \right|^{\frac{2}{\gamma}} (1 + r(x)) + \frac{\pi}{\gamma} \text{dist}(x, H)
 \end{aligned}
 \tag{B.1}$$

with $r(x) = \sqrt{\sum_{i=2}^d x_i^2}$.

- DO2DK [8] with $k \in \mathbb{N}$, $s \in \mathbb{R}_{>0}$

$$\begin{aligned}
 g_1(x) &= \sin\left(\frac{\pi}{2}x_1 + \left(1 + \frac{2^s - 1}{2^{s+2}}\right)\pi + 1\right) r_a(x)r_b(x) + 10\text{dist}(x, H) \\
 g_2(x) &= \left(\cos\left(\frac{\pi}{2}x_1 + \pi\right) + 1\right) r_a(x)r_b(x) + 10\text{dist}(x, H)
 \end{aligned}
 \tag{B.2}$$

with

$$\begin{aligned}
 r_a(x) &= 1 + \frac{9}{d-1} \sum_{i=2}^d x_i \\
 r_b(x) &= 5 + 10 \left(x_1 - \frac{1}{2}\right)^2 + \frac{2^{\frac{s}{2}} \cos(2k\pi x_1)}{k}.
 \end{aligned}$$

The parametrization used to construct reference solutions is given by

$$h : [0, 1] \rightarrow F, \quad h(r) = g((r, 0, \dots, 0)) \in F.$$

Appendix C: Proof of Lemma 4.2

Proof (Lemma 4.2) The proof closely follows the one of [34, Proposition 2] and we recall here the main steps for completeness. Given that $\text{Im}(\phi_r) \subset [0, 1]$ and

$\text{supp}(\phi_r) = B_r^\infty(x^*)$, it holds

$$\mathbb{P}(X_t \in B_r^\infty(x^*)) \geq \mathbb{E}[\phi_r(X_t)].$$

By applying Itô’s fomula as in the proof of Proposition 4.1, we study the evolution of $\mathbb{E}[\phi_r(X_t)]$

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[\phi_r(X_t)] &= \sum_{k=1}^d \mathbb{E} \left[\lambda(y^\alpha(\rho(t), W_t) - X_t)_k \partial_{x_k} \phi_r(X_t) \right] \\ &\quad + \sum_{k=1}^d \mathbb{E} \left[\frac{\sigma^2}{2} (y^\alpha(\rho(t), W_t) - X_t)_k^2 \partial_{x_k x_k}^2 \phi_r(X_t) \right] \\ &=: \sum_{k=1}^d \mathbb{E}[T_{1k}(X_t, W_t)] + \sum_{k=1}^d \mathbb{E}[T_{2k}(X_t, W_t)] \end{aligned} \tag{C.1}$$

We will show in the following that

$$\mathbb{E}[T_{1k}(X_t, W_t)] + \mathbb{E}[T_{2k}(X_t, W_t)] \geq -\frac{q}{d} \mathbb{E}[\phi(X_t)]$$

for all $k = 1, \dots, d$.

Let $\Omega_r := \{X_t \in B_r^\infty(x^*)\}$, we have $\mathbb{E}[T_{ik}(X_t, W_t) | \Omega_r^c] = 0$ for $i = 1, 2$ as $\text{supp}(\phi_r) = B_r^\infty(x^*)$. We introduce now the sets

$$K_{1k} := \{|(X_t - x^*)_k| > \sqrt{cr}\} \tag{C.2}$$

and

$$\begin{aligned} K_{2k} &:= \left\{ -\lambda(X_t - y^\alpha(\rho(t), W_t))_k (X_t - x^*)_k \left(r^2 - (X_t - x^*)_k^2 \right)^2 \right. \\ &\quad \left. > \tilde{c}r \frac{\sigma^2}{2} (X_t - y^\alpha(\rho(t), W_t))_k^2 (X_t - x^*)_k^2 \right\}, \end{aligned} \tag{C.3}$$

where $\tilde{c} = 2c - 1 \in (0, 1)$. For every $k = 1, \dots, d$, we now decompose Ω_r into three different sets

$$\Omega_r = (K_{1k}^c \cap \Omega_r) \cup (K_{1k} \cap K_{2k}^c \cap \Omega_r) \cup (K_{1k} \cap K_{2k} \cap \Omega_r). \tag{C.4}$$

Subset $K_{1k}^c \cap \Omega_r$: We have

$$\begin{aligned} &\mathbb{E}[T_{1k}(X_t, W_t) | K_{1k}^c \cap \Omega_r] \\ &= \mathbb{E} \left[2r^2 \lambda(X_t - y^\alpha(\rho(t), W_t))_k \frac{(X_t - x^*)_k}{(r^2 - (X_t - x^*)_k^2)^2} \phi_r(X_t) \mid K_{1k}^c \cap \Omega_r \right] \end{aligned}$$

$$\begin{aligned} &\geq \mathbb{E} \left[-2r^2\lambda \frac{|(X_t - y^\alpha(\rho(t), W_t))_k| |(X_t - x^*)_k|}{(r^2 - (X_t - x^*)_k^2)^2} \phi_r(X_t) \mid K_{1k}^c \cap \Omega_r \right] \\ &\geq -\frac{2\lambda(\sqrt{cr} + B)\sqrt{c}}{(1 - c)^2r} \mathbb{E} [\phi_r(X_t) \mid K_{1k}^c \cap \Omega_r] \\ &=: -q_1 \mathbb{E} [\phi_r(X_t) \mid K_{1k}^c \cap \Omega_r], \end{aligned}$$

where we used that in K_{1k}^c it holds

$$|(X_t - y^\alpha(\rho(t), W_t))_k| \leq |(X_t - x^*)_k| + \sup_{w \in \Omega} |(x^* - y^\alpha(\rho(t), w))_k| < B.$$

We now turn to T_{2k} and see that it holds

$$\begin{aligned} &\mathbb{E}[T_{2k}(X_t, W_t) \mid K_{1k}^c \cap \Omega_r] \\ &= \mathbb{E} \left[\sigma^2 r^2 (X_t - y^\alpha(\rho(t), W_t))_k^2 \cdot \right. \\ &\quad \cdot 2 \frac{(2(X_t - x^*)_k^2 - r^2)(X_t - x^*)_k^2 - (r^2 - (X_t - x^*)_k^2)^2}{(r^2 - (X_t - x^*)_k^2)^4} \phi_r(X_t) \mid K_{1k}^c \cap \Omega_r \left. \right] \\ &\geq -\frac{2\sigma^2(cr^2 + B^2)(2c + 1)}{(1 - c)^4 r^2} \mathbb{E} [\phi_r(X_t) \mid K_{1k}^c \cap \Omega_r] \\ &=: -q_2 \mathbb{E} [\phi_r(X_t) \mid K_{1k}^c \cap \Omega_r], \end{aligned}$$

where we used again $(X_t - y^\alpha(\rho(t), W_t))_k^2 \leq (\sqrt{cr} + B)^2 \leq 2(cr^2 + B^2)$ in K_{1k}^c .

Subset $K_{1k} \cap K_{2k}^c \cap \Omega_r$: Here, we have $|(X_t - x^*)_k| > \sqrt{cr}$. We note that $T_{1k}(X_t, W_t) + T_{2k}(X_t, W_t) \geq 0$ whenever

$$\begin{aligned} &\left(-\lambda(X_t - y^\alpha(\rho(t), W_t))_k (X_t - x^*)_k + \frac{\sigma^2}{2} (X_t - y^\alpha(\rho(t), W_t))_k^2 \right) (r^2 - (X_t - x^*)_k^2)^2 \\ &\leq \sigma^2 (X_t - y^\alpha(\rho(t), W_t))_k^2 (2(X_t - x^*)_k^2 - r^2) (X_t - x^*)_k^2. \end{aligned} \tag{C.5}$$

If we consider the set K_{2k}^c , the first term on the left-hand side above can be bounded as

$$\begin{aligned} &-\lambda(X_t - y^\alpha(\rho(t), W_t))_k (X_t - x^*)_k (r^2 - (X_t - x^*)_k^2)^2 \\ &\leq \tilde{c} r^2 \frac{\sigma^2}{2} (X_t - y^\alpha(\rho(t), W_t))_k^2 (X_t - x^*)_k^2 \\ &= (2c - 1) r^2 \frac{\sigma^2}{2} (X_t - y^\alpha(\rho(t), W_t))_k^2 (X_t - x^*)_k^2 \\ &\leq \left(2(X_t - x^*)_k^2 - r^2 \right) \frac{\sigma^2}{2} (X_t - y^\alpha(\rho(t), W_t))_k^2 (X_t - x^*)_k^2, \end{aligned}$$

where we used $\tilde{c} = 2c - 1$ and, in the last inequality, that we are considering events in K_{1k} . To bound the second term in (C.5) we use that $(1 - c)^2 \leq (2c - 1)c$ and, again, that we are taking events in K_{1k} :

$$\begin{aligned} & \frac{\sigma^2}{2} (X_t - y^\alpha(\rho(t), W_t))_k^2 (r^2 - (X_t - x^*)_k^2)^2 \\ & \leq \frac{\sigma^2}{2} (X_t - y^\alpha(\rho(t), W_t))_k^2 (1 - c)^2 r^4 \\ & \leq \frac{\sigma^2}{2} (X_t - y^\alpha(\rho(t), W_t))_k^2 (2c - 1)r^2 cr^2 \\ & \leq \frac{\sigma^2}{2} (X_t - y^\alpha(\rho(t), W_t))_k^2 \left(2(X_t - x^*)_k^2 - r^2 \right) (X_t - x^*)_k^2. \end{aligned}$$

Therefore, (C.5) holds in $K_{1k} \cap K_{2k}^c \cap \Omega_r$ and we have

$$\mathbb{E} [T_{1k}(X_t, W_t) + T_{2k}(X_t, W_t) \mid K_{1k} \cap K_{2k}^c \cap \Omega_r] \geq 0.$$

Subset $K_{1k} \cap K_{2k} \cap \Omega_r$: Whenever $(X_t)_k = (y^\alpha(\rho(t), W_t))_k$, $T_{1k} = 0$. Otherwise, since $\sigma > 0$ and we consider events in K_{2k} , we have

$$\begin{aligned} \frac{(X_t - y^\alpha(\rho(t), W_t))_k (X_t - x^*)_k}{(r^2 - (X_t - x^*)_k^2)^2} & \geq - \frac{|(X_t - y^\alpha(\rho(t), W_t))_k| |(X_t - x^*)_k|}{(r^2 - (X_t - x^*)_k^2)^2} \\ & > \frac{2\lambda (X_t - y^\alpha(\rho(t), W_t))_k (X_t - x^*)_k}{\tilde{c}r^2\sigma^2 |(X_t - y^\alpha(\rho(t), W_t))_k| |(X_t - x^*)_k|} \\ & \geq - \frac{2\lambda}{\tilde{c}r^2\sigma^2}. \end{aligned}$$

As a consequence, it holds

$$\begin{aligned} & \mathbb{E}[T_{1k}(X_t, W_t) \mid K_{1k} \cap K_{2k} \cap \Omega_r] \\ & = \mathbb{E} \left[2r^2\lambda (X_t - y^\alpha(\rho(t), W_t))_k \frac{(X_t - x^*)_k}{(r^2 - (X_t - x^*)_k^2)^2} \phi_r(X_t) \right. \\ & \quad \left. \mid K_{1k} \cap K_{2k} \cap \Omega_r \right] \\ & \geq - \frac{4\lambda^2}{\tilde{c}\sigma^2} \mathbb{E} [\phi_r(X_t) \mid K_{1k} \cap K_{2k} \cap \Omega_r] =: -q_3 \mathbb{E} [\phi_r(X_t) \mid K_{1k} \cap K_{2k} \cap \Omega_r]. \end{aligned}$$

By definition of K_{1k} and assumption $2(2c - 1)c \geq (1 - c)^2$ we also have

$$\mathbb{E}[T_{2k}(X_t, W_t) \mid K_{1k} \cap K_{2k} \cap \Omega_r] \geq 0.$$

Conclusion of the proof: We now plug the computed estimates in (C.1) and obtain

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[\phi_r(X_t)] & = \sum_{k=1}^d \mathbb{E}[T_{1k}(X_t, W_t)] + \sum_{k=1}^d \mathbb{E}[T_{2k}(X_t, W_t)] \\ & \geq -2d \max\{q_1 + q_2, q_3\} \mathbb{E}[\phi_r(X_t)] =: -q \mathbb{E}[\phi_r(X_t)]. \end{aligned} \tag{C.6}$$

The desired lower bound is then obtained after applying Grönwall’s inequality. □

References

- Albi, G., Pareschi, L.: Binary interaction algorithms for the simulation of flocking and swarming dynamics. *Multiscale Model. Simul.* **11**(1), 1–29 (2013)
- Benfenati, A., Borghi, G., Pareschi, L.: Binary interaction methods for high dimensional global optimization and machine learning. *Appl. Math. Optim.* **86**(1), 9 (2022)
- Beume, N., Fonseca, C.M., Lopez-Ibanez, M., Paquete, L., Vahrenhold, J.: On the complexity of computing the hypervolume indicator. *IEEE Trans. Evol. Comput.* **13**(5), 1075–1082 (2009)
- Blanc, X., Lewin, M.: The crystallization conjecture: a review. *EMS Surv. Math. Sci.* **2**(2), 255–306 (2015)
- Borghi, G.: Repulsion dynamics for uniform pareto front approximation in multi-objective optimization problems. *PAMM* **23**(1), e202200285 (2023)
- Borghi, G., Grassi, S., Pareschi, L.: Consensus based optimization with memory effects: random selection and applications. [arXiv:2301.13242](https://arxiv.org/abs/2301.13242) (2023)
- Borghi, G., Herty, M., Pareschi, L.: A consensus-based algorithm for multi-objective optimization and its mean-field description. In: 2022 IEEE 61st Conference on Decision and Control (CDC), pp. 4131–4136 (2022)
- Branke, J., Deb, K., Dierolf, H., Osswald, M.: Finding knees in multi-objective optimization. In: Yao, X., Burke, E.K., Lozano, J.A., Smith, J., Merelo-Guervós, J.J., Bullinaria, J.A., Rowe, J.E., Tiño, P., Kabán, A., Schwefel, H.-P. (eds.) *Parallel Problem Solving from Nature*, pp. 722–731. Springer, Berlin (2004)
- Braun, M.A.: Scalarized preferences in multi-objective optimization. PhD thesis, Karlsruhe Institut für Technologie (KIT) (2018)
- Braun, M.A., Shukla, P.K., Schmeck, H.: Obtaining optimal pareto front approximations using scalarized preference information. In: *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation, GECCO '15*, pp. 631–638, New York, NY, USA. Association for Computing Machinery (2015)
- Campana, E.F., Diez, M., Liuzzi, G., Lucidi, S., Pellegrini, R., Piccialli, V., Rinaldi, F., Serani, A.: A multi-objective direct algorithm for ship hull optimization. *Comput. Optim. Appl.* **71**(1), 53–72 (2018)
- Carrillo, J.A., Choi, Y.-P., Totzeck, C., Tse, O.: An analytical framework for consensus-based global optimization method. *Math. Models Methods Appl. Sci.* **28**(6), 1037–1066 (2018)
- Carrillo, J.A., Jin, S., Li, L., Zhu, Y.: A consensus-based global optimization method for high dimensional machine learning problems. *ESAIM: COCV* **27**, S5 (2021)
- Carrillo, J.A., Slepčev, D., Wu, L.: Nonlocal-interaction equations on uniformly prox-regular sets. *Discret. Contin. Dyn. Syst.* **36**(3), 1209–1247 (2016)
- Coello Coello, C.A., González Brambila, S., Figueroa Gamboa, J., Castillo Tapia, M.G., Hernández Gómez, R.: Evolutionary multiobjective optimization: open research areas and some challenges lying ahead. *Complex Intell. Syst.* **6**(2), 221–236 (2020)
- Condat, L.: Fast projection onto the simplex and the l_1 ball. *Math. Program.* **158**(1), 575–585 (2016)
- Deb, K.: *Multi-Objective Optimization using Evolutionary Algorithms*. Wiley, New York (2001)
- Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. *Trans. Evolut. Comput.* **6**(2), 182–197 (2002)
- Degond, P., Frouvelle, A., Liu, J.-G.: Phase transitions, hysteresis, and hyperbolicity for self-organized alignment dynamics. *Arch. Ration. Mech. Anal.* **216**(1), 63–115 (2015)
- Dembo, A., Zeitouni, O.: *Large Deviations Techniques and Applications*. Springer, Berlin (2010)
- Ehrgott, M.: *Multicriteria Optimization*. Springer, Berlin (2005)
- Eichfelder, G.: Twenty years of continuous multiobjective optimization in the twenty-first century. *EURO J. Comput. Optim.* **9**, 100014 (2021)
- Emmerich, M., Beume, N., Naujoks, B.: An EMO algorithm using the hypervolume measure as selection criterion. In: Coello Coello, C.A., Hernández Aguirre, A., Zitzler, E. (eds.) *Evolutionary Multi-criterion Optimization*, pp. 62–76. Springer, Berlin (2005)
- Emmerich, M.T.M., Deutz, A.H.: Test problems based on Lamé superspheres. In: *Proceedings of the 4th International Conference on Evolutionary Multi-Criterion Optimization, EMO'07*, pp. 922–936, Springer, Berlin (2007)
- Falcón-Cardona, J.G., Covantes Osuna, E., Coello Coello, C.A.: An overview of pair-potential functions for multi-objective optimization. In: Ishibuchi, H., Zhang, Q., Cheng, R., Li, K., Li, H., Wang, H., Zhou, A. (eds.) *Evolutionary Multi-criterion Optimization*, pp. 401–412. Springer, Cham (2021)

26. Falcón-Cardona, J. G., Ishibuchi, H., Coello, C. A. C.: Riesz s -energy-based reference sets for multi-objective optimization. In: 2020 IEEE Congress on Evolutionary Computation (CEC), pp. 1–8 (2020)
27. Fernández, J., Tóth, B.: Obtaining the efficient set of nonlinear biobjective optimization problems via interval branch-and-bound methods. *Comput. Optim. Appl.* **42**(3), 393–419 (2009)
28. Fetecau, R.C., Kovacic, M.: Swarm equilibria in domains with boundaries. *SIAM J. Appl. Dyn. Syst.* **16**, 1260–1308 (2017)
29. Fetecau, R.C., Kovacic, M., Topaloglu, I.: Swarming in domains with boundaries: approximation and regularization by nonlinear diffusion. *Discret. Continuous Dyn. Syst. B* **24**(4), 1815–1842 (2019)
30. Fliege, J., Drummond, L.M.G., Svaiter, B.F.: Newton’s method for multiobjective optimization. *SIAM J. Optim.* **20**(2), 602–626 (2009)
31. Fornasier, M., Huang, H., Pareschi, L., Sünnen, P.: Consensus-based optimization on the sphere: convergence to global minimizers and machine learning. *J. Mach. Learn. Res.* **22**(237), 1–55 (2021)
32. Fornasier, M., Huang, H., Pareschi, L., Sünnen, P.: Anisotropic diffusion in consensus-based optimization on the sphere. *SIAM J. Optim.* **32**(3), 1984–2012 (2022)
33. Fornasier, M., Klock, T., Riedl, K.: Consensus-based optimization methods converge globally. [arXiv:2103.15130](https://arxiv.org/abs/2103.15130) (2021)
34. Fornasier, M., Klock, T., Riedl, K.: Convergence of anisotropic consensus-based optimization in mean-field law. In: Jiménez Laredo, J.L., Hidalgo, J.I., Babaagba, K.O. (eds.) *Applications of Evolutionary Computation*, pp. 738–754. Springer, Cham (2022)
35. Garrigos, G., Rosasco, L., Villa, S.: Convergence of the forward-backward algorithm: beyond the worst-case with the help of geometry. *Math. Program.* (2022)
36. Grassi, S., Pareschi, L.: From particle swarm optimization to consensus based optimization: stochastic modeling and mean-field limit. *Math. Models Methods Appl. Sci.* **31**(08), 1625–1657 (2021)
37. Graña Drummond, L., Svaiter, B.: A steepest descent method for vector optimization. *J. Comput. Appl. Math.* **175**(2), 395–414 (2005)
38. Greengard, L., Rokhlin, V.: A fast algorithm for particle simulations. *J. Comput. Phys.* **73**(2), 325–348 (1987)
39. Hairer, E., Lubich, C., Wanner, G.: *Geometric Numerical Integration, Structure-Preserving Algorithms for Ordinary Differential Equations*, vol. 31, 2nd edn. Springer Series in Computational Mathematics. Springer, Berlin (2006)
40. Hardin, D., Saff, E.: Minimal riesz energy point configurations for rectifiable d -dimensional manifolds. *Adv. Math.* **193**(1), 174–204 (2005)
41. Higham, D.J.: An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.* **43**(3), 525–546 (2001)
42. Huang, H., Qiu, J.: On the mean-field limit for the consensus-based optimization. *Math. Methods Appl. Sci.* **45**(12), 7814–7831 (2022)
43. Huang, H., Qiu, J., Riedl, K.: On the global convergence of particle swarm optimization methods. *Appl. Math. Optim.* **88**(2), 30 (2023)
44. Hwang, C.-L., Md Masud, A.: *Multiple Objective Decision Making, Methods and Applications: A State-of-the-Art Survey*. Springer, New York (1979)
45. Jahn, J.: *Vector Optimization—Theory, Applications, and Extensions*. Springer, Berlin (2004)
46. Jin, S., Li, L., Liu, J.-G.: Random Batch Methods (RBM) for interacting particle systems. *J. Comput. Phys.* **400**, 108877 (2020)
47. Klamroth, K., Stiglmayr, M., Totzeck, C.: Consensus-based optimization for multi-objective problems: a multi-swarm approach. [arXiv:2103.15130](https://arxiv.org/abs/2103.15130) (2022)
48. Márquez-Vega, L.A., Falcón-Cardona, J.G., Covantes Osuna, E.: Towards a pareto front shape invariant multi-objective evolutionary algorithm using pair-potential functions. In: Batyrshin, I., Gelbukh, A., Sidorov, G. (eds.) *Advances in Computational Intelligence*, pp. 369–382. Springer, Cham (2021)
49. Martínez-Cagigal, V.: Non Sorting Genetic Algorithm II (NSGA-II). <https://www.mathworks.com/matlabcentral/fileexchange/65494-non-sorting-genetic-algorithm-ii-nsga-ii>. Accessed 23 March 2023 (2023)
50. Motsch, S., Tadmor, E.: Heterophilous dynamics enhances consensus. *SIAM Rev.* **56**(4), 577–621 (2014)
51. Nicolis, G., Prigogine, I.: *Self-organization in Nonequilibrium Systems*. Wiley, New York (1977)
52. Øksendal, B.: *Stochastic Differential Equations: An Introduction with Applications* (Universitext), 6th edn. Springer, New York (2014)

53. Pardalos, P.M., Žilinskas, A., Zilinskas, J.: *Non-convex Multi-objective Optimization*. Springer, Cham (2018)
54. Pareschi, L., Toscani, G.: *Interacting Multiagent Systems: Kinetic Equations and Monte Carlo methods*. Oxford University Press, Oxford (2013)
55. Patachini, F.S., Slepčev, D.: The nonlocal-interaction equation near attracting manifolds. *Discret. Continuous Dyn. Syst. Ser. A* **42**(2), 903–929 (2022)
56. Pinnau, R., Totzeck, C., Tse, O., Martin, S.: A consensus-based model for global optimization and its mean-field limit. *Math. Models Methods Appl. Sci.* **27**(1), 183–204 (2017)
57. Platen, E.: *An introduction to numerical methods for stochastic differential equations*. *Acta Numer.* **8**, 197–246 (1999)
58. Riedl, K.: Leveraging memory effects and gradient information in consensus-based optimization: On global convergence in mean-field law. [arXiv:2211.12184](https://arxiv.org/abs/2211.12184) (2022)
59. Sergeev, Y.D., Kvasov, D.E., Mukhametzhano, M.S.: On the efficiency of nature-inspired metaheuristics in expensive global optimization with limited budget. *Sci. Rep.* **8**(1), 453 (2018)
60. Sznitman, A.-S.: Topics in propagation of chaos. In: Hennequin, P.-L. (ed.) *Ecole d'Été de Probabilités de Saint-Flour XIX–1989*, pp. 165–251. Springer, Berlin (1991)
61. Totzeck, C., Wolfram, M.-T.: Consensus-based global optimization with personal best. *Math. Biosci. Eng.* **17**(5), 6026–6044 (2020)
62. Van Veldhuizen, D.A., Lamont, G.B. et al.: Evolutionary computation and convergence to a pareto front. In: *Late breaking papers at the genetic programming 1998 conference*, pp. 221–228. Citeseer (1998)
63. Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., Shochet, O.: Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.* **75**(6), 1226–1229 (1995)
64. Zhang, Q., Li, H.: MOEA/D: a multiobjective evolutionary algorithm based on decomposition. *IEEE Trans. Evolut. Comput.* **11**, 712–731 (2008)
65. Žilinskas, A.: On the worst-case optimal multi-objective global optimization. *Optim. Lett.* **7**(8), 1921–1928 (2013)
66. Zitzler, E., Thiele, L.: Multiobjective optimization using evolutionary algorithms—a comparative case study. In: *Proceedings of the 5th International Conference on Parallel Problem Solving from Nature, PPSN V*, pp. 292–304. Springer, Berlin (1998)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.