Klamroth, Kathrin; Stiglmayr, Michael; Totzeck, Claudia

Article — Published Version

Consensus-based optimization for multi-objective problems: a multi-swarm approach

Journal of Global Optimization

Provided in Cooperation with:
Springer Nature

Suggested Citation: Klamroth, Kathrin; Stiglmayr, Michael; Totzeck, Claudia (2024) : Consensus-based optimization for multi-objective problems: a multi-swarm approach, Journal of Global Optimization, ISSN 1573-2916, Springer US, New York, NY, Vol. 89, Iss. 3, pp. 745-776, https://doi.org/10.1007/s10898-024-01369-1

This Version is available at:
https://hdl.handle.net/10419/315321

Journal of Global Optimization (2024) 89:745–776
https://doi.org/10.1007/s10898-024-01369-1

Consensus-based optimization for multi-objective problems: a multi-swarm approach

Kathrin Klamroth[1] · **Michael Stiglmayr**[1] · **Claudia Totzeck**[1]

Received: 28 November 2022 / Accepted: 15 January 2024 / Published online: 15 February 2024

Abstract
We propose a multi-swarm approach to approximate the Pareto front of general multi-objective optimization problems that is based on the consensus-based optimization method (CBO). The algorithm is motivated step by step beginning with a simple extension of CBO based on fixed scalarization weights. To overcome the issue of choosing the weights we propose an adaptive weight strategy in the second modeling step. The modeling process is concluded with the incorporation of a penalty strategy that avoids clusters along the Pareto front and a diffusion term that prevents collapsing swarms. Altogether the proposed K-swarm CBO algorithm is tailored for a diverse approximation of the Pareto front and, simultaneously, the efficient set of general non-convex multi-objective problems. The feasibility of the approach is justified by analytic results, including convergence proofs, and a performance comparison to the well-known non-dominated sorting genetic algorithms NSGA2 and NSGA3 as well as the recently proposed one-swarm approach for multi-objective problems involving consensus-based optimization.

Keywords Multiobjective optimization · Consensus-based methods · Global optimization

1 Introduction

Multiple conflicting objective functions occur in a variety of applications ranging from engineering design to economic and financial decisions. Economical goals are often in conflict with ecological criteria, we have to trade-off between expected return and risk, and we aim at affordable yet high quality products. We refer to the textbooks [14, 22] for a general introduction to the topic of multi-objective optimization. In this paper, we focus on *continuous multi-objective optimization problems* (MOP)

✉ Claudia Totzeck
totzeck@uni-wuppertal.de

Kathrin Klamroth
klamroth@uni-wuppertal.de

Michael Stiglmayr
stiglmayr@uni-wuppertal.de

[1] Optimization Group, School of Mathematics and Natural Sciences, University of Wuppertal, Wuppertal, Germany

$$\min_{x\in\mathcal{X}} f(x) = (f_1(x), \dots, f_p(x)), \tag{MOP}$$

with a non-empty feasible set $\mathcal{X} \subset \mathbb{R}^d, d \in \mathbb{N}$, and with $p \geq 2$ continuous objective functions $f_i \colon \mathcal{X} \to \mathbb{R}_+$, each of which has a unique global minimum on $\mathcal{X}$. We consider unconstrained problems, i.e., $\mathcal{X} = \mathbb{R}^d$, as well as box-constrained problems where $\mathcal{X} = [\ell, u]$ with lower and upper bounds $\ell, u \in \mathbb{R}^d$ with $\ell_j \leq u_j$, $j = 1, \dots, d$.

We refer to $\mathbb{R}^d$ as the *decision space* and to $\mathbb{R}^p$ as the *objective space* of (MOP). For a multiobjective optimization problem, the set $\mathcal{Y} := f(\mathcal{X}) \subset \mathbb{R}^p$ is denoted as feasible *outcome set*. Two feasible solutions $x^1, x^2 \in \mathcal{X}$ are compared based on their respective outcome vectors $z^1 = f(x^1)$ and $z^2 = f(x^2)$: We say that z^1 *dominates* z^2 (and x^1 dominates x^2) denoted by $z^1 \leq z^2$ if and only if

$$z^1_i \leq z^2_i \text{ for all } i = 1, \dots, p \text{ and } z^1 \neq z^2.$$

A feasible solution $x \in \mathcal{X}$ is called *Pareto optimal* or *efficient* if there is no other solution $\bar{x} \in \mathcal{X}$ such that $f(\bar{x}) \leq f(x)$. The corresponding image in the objective space is called *non-dominated* in this case. The set of all Pareto optimal solutions is denoted by $\mathcal{X}_P$ and referred to as the *efficient set*, and the set of all non-dominated outcome vectors, i.e., $\mathcal{Y}_P = f(\mathcal{X}_P)$, is called the *non-dominated set* or the *Pareto front* of (MOP). Note that since we assume that each objective function has a unique global minimum on $\mathcal{X}$, the Pareto front of (MOP) is non-empty since it contains these individual minima.

An important approach to generate or to approximate Pareto optimal solutions are scalarizations that transform the multi-objective problem (MOP) into a series of associated single-objective problems. Maybe the most prominent scalarization approach is the *weighted-sum scalarization* [16]: Given non-negative weights $\lambda_i \geq 0$, $i = 1, \dots, p$ (that correspond to the relative importance of the respective criteria), the *weighted sum-objective* is given by

$$f_\lambda(x) := \sum_{i=1}^{p} \lambda_i \, f_i(x). \tag{1}$$

Let $\Lambda := \{\lambda \in \mathbb{R}^p \colon \sum_{i=1}^p \lambda_i = 1 \text{ and } \lambda_i > 0,\ i = 1, \dots, p\}$ and let $\Lambda_0 := \{\lambda \in \mathbb{R}^p \colon \sum_{i=1}^p \lambda_i = 1 \text{ and } \lambda_i \geq 0,\ i = 1, \dots, p\}$. Further, we define $\mathbb{R}^p_+ := \{z \in \mathbb{R}^p \colon z_i \geq 0,\ i = 1, \dots, p\}$. The following theorem is a well-known result from the field of multi-objective optimization.

Theorem 1 (see, e.g., [14]) *If $\lambda \in \Lambda$, then an optimal solution $\bar{x}(\lambda) \in \mathcal{X}$ of (1) is efficient for (MOP). Moreover, if $\mathcal{Y} + \mathbb{R}^p_+$ is convex and $\bar{x} \in \mathcal{X}$ is efficient, then there exists a weighting vector $\bar{\lambda} \in \Lambda_0$ such that $\bar{x} = \bar{x}(\bar{\lambda})$ is optimal for (1).*

For later reference we phrase the following observations as remark.

Remark 1 *(see, e.g.,* [14]) When $\lambda \in \Lambda_0$, then an optimal solution $\bar{x}(\lambda) \in \mathcal{X}$ of (1) is at least *weakly efficient* for (MOP), i.e., there is no $\hat{x} \in \mathcal{X}$ such that $f_i(\hat{x}) < f_i(\bar{x}(\lambda))$ for all $i = 1, \dots, p$. Note also that when $\mathcal{Y} + \mathbb{R}^p_+$ is non-convex, then it can not be guaranteed in general that all efficient solutions can be obtained as an optimal solution of a weighted-sum scalarization.

The goal of this paper is to develop a provably convergent, yet efficient algorithm for high-quality representations of Pareto fronts of multi-objective optimization problems. We choose the *Consensus-based Optimization method (CBO)* as basis for the algorithm as it is a particle method for single-objective global optimization problems which is easy to implement and

allows for analytical studies. Its dynamics is governed by an interacting particle system that allows to propagate information of the individuals through a weighted mean. The weighted mean has two advantages: (a) there is no need to label individuals as current best, and (b) there is no need for pairwise interactions as the dynamic of one particle depends only on the weighted mean of the whole swarm. Advantage (a) allows for the derivation of a corresponding mean-field equation [18] which can be employed for analytical studies such as the long time behavior and the convergence to the global minimizer. Indeed, in [7, 8] it is shown that the invariant solution of the mean-field dynamics is a Dirac-delta located arbitrarily close to the global minimizer of the objective function. Before we recall the details of CBO, we want to emphasize that our focus lies on the convergence analysis of the proposed multi-objective optimization algorithm. Hence, we tailor each modeling step of the algorithm such that it is feasible for a convergence analysis. This is in contrast to many other algorithms for multi-objective optimization which are often of heuristic nature.

We recall the CBO dynamics for N particles in the single-objective case, i.e., for $p = 1$. Hence, let $f : \mathbb{R}^d \to \mathbb{R}_+$ denote the objective function. For $\alpha, \sigma > 0$ the dynamics of the j-th particle $X^j : [0, T] \to \mathbb{R}^d$, $j = 1, \dots, N$, at time $t \in [0, T]$ is given by

$$\mathrm{d}X_t^j = -(X_t^j - v_t)\,\mathrm{d}t + \sigma\ \mathrm{diag}(X_t^j - v_t)\,\mathrm{d}B_t^j, \tag{2a}$$

$$\mathrm{law}(X_0^j) = \rho_0, \qquad j = 1, \dots, N, \tag{2b}$$

where $\mathrm{law}(X)$ refers to the probability law of the random variable X and the weighted mean v_t is given by

$$v_t = \frac{\sum_{j=1}^N X_t^j\, \mathrm{e}^{-\alpha f(X_t^j)}}{\sum_{j=1}^N \mathrm{e}^{-\alpha f(X_t^j)}}, \tag{2c}$$

ρ_0 is a probability distribution on the state space, $B^j : [0, T] \to \mathbb{R}^d$, $j = 1, \dots, N$ are independent Brownian motions and $\alpha > 0$ allows to scale the difference of the local and global minima in the objective function (Laplace principle). Here and in the following we use the anisotropic noise term $\sigma\ \mathrm{diag}(X_t^j - v_t)\,\mathrm{d}B_t^j$ as proposed in [7] as it is shown to be more robust in settings with high-dimensional state space.

The idea of the multi-swarm CBO algorithm we propose in Sect. 3.1 and further develop in Sect. 3.2 is based on Theorem 1. Indeed, each swarm is associated with a weight vector that yields a scalarization of the cost function. Following the CBO dynamic the swarms globally minimize their respective scalarized problems giving us an approximation of the Pareto front. For a diverse approximation we introduce interactions between the swarms in Sect. 3.2. The multi-swarm CBO algorithm with adaptive swarms is analyzed in Sect. 3.3, where we show under appropriate assumptions that each swarm clusters at a point along the Pareto front and the approximation obtained by the points of all swarms is diverse in the sense that the clustering points admit a distance greater than a minimal distance.

As mentioned above, weighted sum scalarizations can have issues in case of non-convex problems. In particular, efficient solutions do not necessarily correspond to global minima of the weighted sum scalarization. By introducing a penalization strategy (in Sect. 4) that avoids clustering along the Pareto front we allow swarms to also converge towards efficient unsupported solutions, i.e., solutions which are not located on the convex hull of $\mathcal{Y}$. In the modeling process we focus on the weighted means of the swarms and tailor the dynamics such that the means form a high quality, i.e., representative and well-distributed approximation of the Pareto front. In Consensus-based Optimization the swarms are forced to collapse in the large time limit [8, 24]. This is disadvantageous in our setting, as we would lose local

information about the Pareto front. Hence, we prevent collapsing swarms by implementing the diffusion term from Consensus-based sampling [6].

Each modeling step is illustrated by simulation results. To not interrupt the flow, we present the details of the simulations in Sect. 5.2, where all parameters and implementation details are reported. We conduct a qualitative study comparing the proposed multi-swarm CBO with the recently proposed single-swarm CBO for multi-objective problems [4] and the well-known NSGA2 [11] and NSGA3 [10, 19] algorithms implemented in [1] in Sect. 6, before we conclude the article and give an outlook to future work.

2 Overview of known results for CBO

The CBO algorithm was tailored in order to be simple enough to allow for a rigorous analysis in terms of its mean-field approximation, and on the other hand contains the main ingredients of particle-based global optimization algorithms such as exploration and consensus dynamics.

For later reference we recall the main results from the review [28] of the convergence analysis for CBO in the following. For more details we refer to [8]. The convergence analysis is based on the mean-field approximation of the particle dynamics (2). Let us denote $\rho_0 = \mathrm{law}(X_0^j)$ for $j = 1, \dots, N$ as above, then the corresponding mean-field equation (which corresponds formally to the limit $N \to \infty$) is given by

$$\partial_t \rho_t = \frac{\sigma^2}{2} \sum_{i=1}^{d} \partial_{ii}\big((x - v[\rho_t])\, \rho_t\big) + \nabla \cdot \big((x - v[\rho_t])\, \rho_t\big), \qquad \lim_{t \to 0} \rho_t = \rho_0 \tag{3}$$

with the weighted mean given by

$$v[\rho_t] = \frac{1}{\int \mathrm{e}^{-\alpha f(x)} \, \mathrm{d}\rho_t(x)} \int x \, \mathrm{e}^{-\alpha f(x)} \, \mathrm{d}\rho_t(x)$$

and $\sigma > 0$, $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$. Under the following assumption, this mean-field equation admits a unique solution.

Assumption 1 To obtain the well-posedness results of the mean-field equation we assume that it holds:

1. The cost function $f : \mathbb{R}^d \to \mathbb{R}$ is bounded from below with $\underline{f} := \inf f$. (Note that this is satisfied with $\underline{f} = 0$ in our setting.)
2. There exist constants $L_f > 0$ and $c_u > 0$ such that

$$\begin{cases} |f(x) - f(y)| \le L_f(|x| + |y|)|x - y| & \text{for all } x, y \in \mathbb{R}^d, \\ f(x) - \underline{f} \le c_u(1 + |x|^2) & \text{for all } x \in \mathbb{R}^d. \end{cases} \tag{A1}$$

Definition 1 We say that a function has *quadratic growth* if there exist constants $M > 0$ and $c_l > 0$ such that

$$f(x) - \underline{f} \ge c_l |x|^2 \quad \text{for } |x| \ge M. \tag{A2}$$

Similar to the above, we denote by $B : [0, T] \to \mathbb{R}^d$ a d-dimensional Brownian motion.

Theorem 2 *Let f be bounded or have quadratic growth, let Assumption 1 hold and $\rho_0 \in \mathcal{P}_4(\mathbb{R}^d)$. Then there exists a unique nonlinear process $\bar{X} \in \mathcal{C}([0, T], \mathbb{R}^d)$, $T > 0$, satisfying*

$$\mathrm{d}\bar{X}_t = -(\bar{X}_t - v_f[\rho_t])\, \mathrm{d}t + \sigma |\bar{X}_t - v_f[\rho_t]| \, \mathrm{d}B_t, \qquad \rho_t = \mathrm{law}(\bar{X}_t),$$

in the strong sense, and $\rho \in \mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R}^d))$ *satisfies the Fokker–Planck equation (which, in our case, corresponds to (3)) in the weak sense with* $\lim_{t\to 0}\rho_t = \rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$.

To understand the steps of the convergence proof it is helpful to recall that vanishing variance of a crowd indicates that the crowd collapses to one point, and moreover, that the center of mass of the crowd coincides with the point where this collapse happens. Having this in mind, it is simple to understand that the convergence proof of CBO is split into two steps. The first result is concerned with the long-time behavior of the variance, in fact, it is shown that the variance vanishes in the limit $t \to \infty$ and the second result shows that the point where the crowd collapses is arbitrary close to the global minimum of the objective function f.

The result on the long-time behavior of the variance is based on the following assumption:

Assumption 2 We assume that $f \in \mathcal{C}^2(\mathbb{R}^d)$ satisfies additionally

1. $\inf f > 0$.
2. $\|\nabla^2 f\|_\infty \le c_f$ and there exist constants $c_0, c_1 > 0$, such that

$$\Delta f \le c_0 + c_1|\nabla f|^2 \quad \text{in } \mathbb{R}^d.$$

In [8] the following concentration result is proven. We emphasize that we set the drift parameter to one here. The expectation and variance of the density are defined by $\mathbb{E}(\rho_t) = \int_X x\,\mathrm{d}\rho_t$ and $V(\rho_t) = \frac{1}{2}\int_X |x-\mathbb{E}(\rho_t)|^2\,\mathrm{d}\rho_t$.

Theorem 3 *Let* f *satisfy Assumption 2 and let the parameters* α *and* σ *satisfy*

$$2\alpha\,\mathrm{e}^{-2\alpha\underline{f}}(c_0\sigma^2 + 2c_f) < \frac{3}{4}, \qquad 2b_0^2 - K - 2\,\mathrm{d}\sigma^2 b_0\,\mathrm{e}^{-\alpha\underline{f}} \ge 0,$$

with $K = V(\rho_0)$ *and* $b_0 = \|\omega_f^\alpha\|_{L^1(\rho_0)}$. *Then* $V(\rho_t) \le V(\rho_0)\,\mathrm{e}^{-qt}$ *with*

$$q = 2\big(1 - (\,\mathrm{d}\sigma^2/b_0)\,\mathrm{e}^{-\alpha\underline{f}}\big) \ge K/b_0^2.$$

Furthermore, there exists a point $\tilde{x} \in \mathbb{R}^d$ *for which* $\mathbb{E}(\rho_t) \to \tilde{x}$ *and* $v_f[\rho_t] \to \tilde{x}$ *as* $t \to \infty$.

This shows the concentration of the crowd at $\tilde{x}$. The following result exploits the Laplace principle to show that the above concentration point is located arbitrarily close to the global minimizer of f.

Theorem 4 *Let* f *satisfy Assumption 2. For any given* $0 < \varepsilon_0 \ll 1$ *arbitrarily small, there exist some* $\alpha_0 \gg 1$ *and appropriate* σ *such that uniform consensus is obtained at a point* $\tilde{x} \in B_{\varepsilon_0}(x_*) = \{x \in \mathbb{R}^d : |x - x_*| \le \varepsilon_0\}$. *More precisely, we have that* $\rho_t \to \delta_{\tilde{x}}$ *for* $t \to \infty$, *with* $\tilde{x} \in B_{\varepsilon_0}(x_*)$ *and* $\delta_{\tilde{x}}$ *the Dirac distribution positioned at* $\tilde{x}$.

This ends the section on known results for the standard CBO method for global optimization problems. We now proceed with the step-by-step modeling of the provably convergent CBO algorithms for multiobjective optimization.

3 CBO for convex multi-objective problems

We begin with the presentation of a multi-swarm approach for multi-objective optimization based on CBO (MSCBO) with fixed scalarization weights and provide analytical results. Fixed weights come with the burden of choosing appropriate weights a priori, which may result in highly unbalanced Pareto front approximations [9]. To overcome this problem, we propose an adaptive weight strategy in the second part of this section.

3.1 MSCBO with fixed scalarization weights

We begin with a simple CBO method for multi-objective problems that is gradually extended in the following: Let $K \in \mathbb{N}$ be the number of swarms involved and let $(\lambda^k)_{k=1,\dots,K} \subset \Lambda$ be their respective fixed scalarization weights. For each k we introduce the swarm size $N_k \in \mathbb{N}$ and the positions of the individuals are denoted by $X^k = (X^{k,1}, \dots, X^{k,N_k}) \colon [0,T] \to \mathbb{R}^{N_k \cdot d}$. The evolution of the swarms is given by the dynamics

$$\mathrm{d}X_t^{k,j} = -(X_t^{k,j} - v_t^k)\,\mathrm{d}t + \sigma\ \mathrm{diag}(X_t^{k,j} - v_t^k)\,\mathrm{d}B_t^{k,j}, \tag{4a}$$

$$\mathrm{law}(X_0^{k,j}) = \rho_0, \qquad j = 1, \dots, N_k,\ k = 1, \dots, K, \tag{4b}$$

where the weighted mean v_t^k is given by

$$v_t^k = \frac{\sum_{j=1}^{N_k} X_t^{k,j} \mathrm{e}^{-\alpha f_{\lambda^k}(X_t^{k,j})}}{\sum_{j=1}^{N_k} \mathrm{e}^{-\alpha f_{\lambda^k}(X_t^{k,j})}}, \tag{4c}$$

$\sigma > 0$ is a diffusive strength parameter, $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ is a probability measure with finite second moment, $B^{k,j}$ are independent Brownian motions and $\alpha > 0$ allows to scale the difference of the local and global minima in the objective function (Laplace principle). For notational convenience we define $N = \sum_{k=1}^K N_k$ and the vector $X_t^N = (X_t^1, \dots, X_t^K) \in \mathbb{R}^{N \cdot d}$. A pseudo code for the dynamics is given in Algorithm 1 in Sect. 5.

Remark 2 (*Initialization of the weights*) In order to obtain admissible initial conditions for the weight vectors λ^k, $k = 1, \dots, K$, we draw uniformly distributed random samples from the standard simplex $\mathcal{S}^p = \Lambda_0$. Thus, each weight vector has $p-1$ degrees of freedom and we can write $\lambda^k = (\lambda^{k,1}, \dots, \lambda^{k,p-1}, 1 - \sum_{i=1}^{p-1} \lambda^{k,i})$. For the numerical tests we use Algorithm 2 from [23] to sample the initial weights λ^k for $k = 1, \dots, K$.

Remark 3 In contrast to the dynamics proposed in [3, 4], where the interacting particles in the swarm converge towards different efficient solutions, here each swarm (namely its weighted mean) represents one solution. That means each swarm is supposed to find one point along the Pareto front and the swarms act independently. In Sect. 3.2 we implement interactions between the swarms through adaptive scalarization weights.

3.1.1 Analysis of the independent swarms

Note that in the dynamics proposed above the K swarms are independent and each is globally minimizing its weighted sum (1) given by λ^k. Hence, we may employ the result in [29] to prove the well-posedness of the dynamics.

Theorem 5 (Well-posedness [29]) *Let $K \in \mathbb{N}$, $(\lambda^k)_{k=1,\dots,K} \subset \Lambda_0$, $N_k \in \mathbb{N}$ fixed and f_{λ^k} locally Lipschitz continuous for every λ^k, then system (4) admits a unique strong solution $\{X_t^N : t \geq 0\}$ for each initial condition X_0^N with $\mathbb{E}[|X_0^N|^2] < \infty$.*

Proof By the independence of the swarms, we first apply Theorem 1 in [29] to every subsystem describing the evolution of each swarm and collect the corresponding solutions in X^N for the full system. □

As the convergence analysis on the particle level is highly nontrivial, we follow the steps in [7, 8] and derive statistical representations of the swarms in form of mean-field equations.

Under the standard assumption of *propagation of chaos*, i.e. that the particles decouple as $N_k \to \infty$, we denote the probability of finding a member of the k-th swarm in position x at time t by $\rho_t^k(x)$. Following the lines of [7, 8] we compute the evolution equation of these probabilities with the help of Itô-calculus. Towards this end, let for simplicity $N_k = \bar{N}$ for all $k = 1, \dots, K$. For $\bar{N} \to \infty$ the PDE system corresponding to (4) reads

$$\begin{aligned} \partial_t \rho_t^k &= \frac{\sigma^2}{2} \sum_{i=1}^{d} \partial_{ii} \left((x - v[\rho_t^k])\, \rho_t^k \right) + \nabla \cdot \left((x - v[\rho_t^k])\, \rho_t^k \right), \\ \lim_{t \to 0} \rho_t^k &= \rho_0, \quad k = 1, \dots, K, \end{aligned}$$

with the weighted mean given by

$$v[\rho_t^k] = \frac{1}{\int \mathrm{e}^{-\alpha f_{\lambda^k}(x)} \, \mathrm{d}\rho_t^k(x)} \int x \, \mathrm{e}^{-\alpha f_{\lambda^k}(x)} \, \mathrm{d}\rho_t^k(x)$$

and $\sigma > 0$, $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ as above.

Again, using the independence of the K swarms we can directly apply the result in [8] to prove the convergence of each swarm towards a limit point arbitrarily close to the efficient set. To state the result properly we define

$$\begin{aligned} E^k(t) &:= \int_{\mathbb{R}^d} x \, \mathrm{d}\rho_t^k(x), \\ V^k(t) &:= \int_{\mathbb{R}^d} |x - E^k(t)| \, \mathrm{d}\rho_t^k(x), \\ M^k(t) &:= \int_{\mathbb{R}^d} \mathrm{e}^{-\alpha f_{\lambda^k}(x)} \, \mathrm{d}\rho_t^k(x). \end{aligned}$$

The convergence theory for CBO [8] is based on the (strong) assumption that the minimizer of the objective function is unique. This can be guaranteed under a similarly strong uniqueness assumption for the minimizer of the scalarized objective functions, i.e.,

$$\bar{x}^k := \operatorname*{argmin}_{x \in \mathcal{X}} f_{\lambda^k}(x) \text{ is unique, and } \quad F^k := f_{\lambda^k}(\bar{x}^k) > 0 \quad \text{for all } \lambda^k. \tag{A1}$$

In addition, we inherit the following technical assumption from the convergence theory of CBO [8]

$$c_f^k := \max\left\{ \| \max_i |\partial_{ii} f_{\lambda^k}| \|_\infty, \ \| r(\nabla^2 f_{\lambda^k}) \|_\infty \right\} < \infty, \tag{A2}$$

where $\nabla^2 f_{\lambda^k}$ represents the Hessian of f_{λ^k}, $r(\nabla^2 f_{\lambda^k})$ is the spectral radius and $\partial_{ii} f_{\lambda^k}$ is the i-th element of the diagonal of $\nabla^2 f_{\lambda^k}$.

Remark 4 Since we assumed that all objective functions $f_i, i = 1, \dots, p$, are bounded from below, all weighted sums f_{λ^k} are bounded from below for $\lambda^k \in \Lambda_0$. Hence, $F^k > 0$ in Assumption (A1) can always be guaranteed after applying an appropriate linear transformation to the objective functions values. The uniqueness of the global minimizers of weighted sum scalarizations f_{λ^k}, however, does in general not follow from the uniqueness of the global minimizers of the individual objective functions $f_i, i = 1, \dots, p$, even if $p = 2$. As an example, suppose that $\mathcal{X} = [0, 1] \subset \mathbb{R}$, $f_1(x) = x$ and $f_2(x) = 1 - x$. Then, $\mathcal{Y} = \mathcal{Y}_P$ is the line segment connecting the two points $(0, 1)$ and $(1, 0)$ in $\mathbb{R}^2$, and while the global minimizers of f_1 and of f_2 are unique, the complete feasible set $\mathcal{X}$ is optimal for f_λ with $\lambda = (\frac{1}{2}, \frac{1}{2})$.

Theorem 6 (Convergence towards the efficient set [7]) *Let* (A1) *and* (A2) *hold. Further, let* α *and the initial condition* ρ_0 *be chosen such that for each* $k \in \{1, \dots, K\}$ *it holds*

$$\mu^k := 2 - \sigma^2 - 2\sigma^2 \frac{e^{-\alpha F^k}}{M^k(0)} > 0,$$
$$\nu^k := \frac{2\, V^k(0)}{\mu^k \cdot (M^k(0))^2} \alpha\, e^{-2\alpha F^k} c_f^k (2 + \sigma^2) \leq \frac{3}{4}.$$

Then for each $k \in \{1, \dots, K\}$ *it holds that* $V^k(t) \to 0$ *exponentially fast and there exist* $\tilde{x}^k$ *such that* $v[\rho_t^k] \to \tilde{x}^k$, $E^k(t) \to \tilde{x}^k$ *exponentially fast. Moreover, for* $\alpha \to \infty$ *the limit points* $\tilde{x}^k$ *are arbitrarily close to the efficient set. In particular, for every* $\epsilon > 0$ *there exists* $\alpha \gg 1$ *such that* $\tilde{x}^k \in B_\epsilon(\bar{x}^k)$.

Proof As the K swarms are independent the result is a direct consequence of Theorem 3.1 in [7]. □

Corollary 1 *Let* f_i *be strictly convex functions for all* $i \in \{1, \dots, p\}$ *and let Assumption* (A2) *be satisfied. Then for each point of the efficient set there exists* $\lambda^k \in \Lambda_0$ *such that the swarm corresponding to* λ^k *and following the dynamics* (4) *concentrates arbitrarily close to a point in the efficient set.*

Proof The result follows from Theorem 1 in combination with Theorem 6. □

The previous results are based on the fact that the weights are fixed. It therefore remains open to choose the weights appropriately. This is not a trivial task, as it is for example well-known that an equidistant choice of weights does not necessarily lead to an equidistant resolution of the Pareto front, even in the case of convex problems [9]. In order to circumvent the problem of choosing appropriate weights, we propose an adaptive procedure including dynamic weights in the next section.

3.2 MSCBO with dynamic weight adaption

Already in the biobjective case the choice of scalarization weights is nontrivial. In fact, assuming that f_1 and f_2 are continuously differentiable, the optimal choice of the weights depends on the ratio $\frac{\partial f_2}{\partial f_1}$ [9], which is in general unknown.

This observation motivates to incorporate adaptive weights in the multi-objective CBO method. The adaption dynamics are written as an ODE. In order to circumvent the restriction $\lambda_j^k \in [0, 1]$, we use the bijective transformation

$$\text{for } \lambda^k \in \Lambda \text{ set } \mu^k = \ln(\lambda^k) \text{ elementwise}$$

and formulate the dynamic weight adaption in terms of μ^k. The scalarized cost function then reads

$$f_{\mu^k}(x) = \frac{\sum_{i=1}^p e^{\mu_i^k} f_i(x)}{\sum_{i=1}^p e^{\mu_i^k}}.$$

Remark 5 Note that $\lambda^k \in \Lambda$ instead of Λ_0 implies that only efficient solutions with bounded trade-off can be determined. Since efficient solutions with unbounded trade-off usually correspond to extreme cases that are not very interesting from a practical point of view, this does not impose a strong restriction.

The dynamic of adaptive weights is based on a pairwise interaction of swarms k and ℓ if their respective weight vectors μ^k and μ^ℓ are close to each other, or the distance of their weighted means, v_t^k and v_t^ℓ, in the objective space is small, i.e., if the swarms map to similar points $f(v_t^k)$ and $f(v_t^\ell)$ in the objective space. Let

$$d_{k,\ell} := |\mu^k - \mu^\ell|, \qquad d_{k,\ell}^f := |f(v_t^k) - f(v_t^\ell)| < c_{\text{rep}} \leq \infty.$$

Then the repulsion is modeled by means of interaction potentials, where in the case $c_{\text{rep}} < \infty$ these potentials are assumed to have compact support.

For the numerical tests we use Morse potentials [12] which are given by

$$P(d_{k,\ell}) = R\, \mathrm{e}^{-d_{k,\ell}/r} - A\, \mathrm{e}^{-d_{k,\ell}/a}, \quad R, A \geq 0,\ r, a > 0. \tag{5}$$

The parameters R, A denote the strength of the repulsion and attraction, respectively, while r, a allow to adjust the range of the attractive and repulsive force, respectively.

Remark 6 We choose to have strong repulsive forces on a very short range and attractive forces on a medium range. Hence on larger ranges interactions can be neglected and we consider only interactions with direct neighbors in our main theorem analyzing the approximation of the Pareto front (Theorem 8).

For interacting particle systems, see for example [5], the force on μ^k that results from the interaction with $\mu^\ell \neq \mu^k$ is given by the corresponding gradient

$$\nabla_{\mu^k} P(d_{k,\ell}) = \left(\frac{A}{a}\, \mathrm{e}^{-d_{k,\ell}/a} - \frac{R}{r}\, \mathrm{e}^{-d_{k,\ell}/r} \right) \frac{\mu^k - \mu^\ell}{|\mu^k - \mu^\ell|}. \tag{6}$$

The main goal we are pursuing with the adaptive weights is to obtain a diverse approximation of the Pareto front. We therefore add another interaction strategy that depends on the distance of the weighted means in the objective space. Following the spirit of (6) we replace the distances $d_{k,\ell}$ in the prefactor by the distance of the objective function values $d_{k,\ell}^f$, leading to the second force given by

$$\left(\frac{A_f}{a_f}\, \mathrm{e}^{-d_{k,\ell}^f/a_f} - \frac{R_f}{r_f}\, \mathrm{e}^{-d_{k,\ell}^f/r_f} \right) \frac{\mu^k - \mu^\ell}{|\mu^k - \mu^\ell|}, \quad R_f, A_f \geq 0,\ r_f, a_f > 0. \tag{7}$$

Altogether, this leads to the following forces for the adaptive weight adjustment

$$\begin{aligned} \mathcal{K}(X_t^k, X_t^\ell, \mu^k, \mu^\ell) = \Bigg(& \frac{A}{a}\, \mathrm{e}^{-d_{k,\ell}/a} - \frac{R}{r}\, \mathrm{e}^{-d_{k,\ell}/r} \\ & + \frac{A_f}{a_f}\, \mathrm{e}^{-d_{k,\ell}^f/a_f} - \frac{R_f}{r_f}\, \mathrm{e}^{-d_{k,\ell}^f/r_f} \Bigg) \frac{\mu^k - \mu^\ell}{|\mu^k - \mu^\ell|}. \end{aligned}$$

More generally, one can use any force with similar properties to define an adaptive weight adjustment given by

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\, \mu^k &= -\frac{1}{\tau} \sum_{\ell=1, \ell \neq k}^{K} \mathcal{K}(X_t^k, X_t^\ell, \mu^k, \mu^\ell), \\ \mu^k(0) &= \ln(\lambda_0^k), \quad k = 1, \ldots, K. \end{aligned} \tag{8}$$

Here $\tau > 0$ is a parameter that allows us to control the time scale of the interaction dynamics. In fact, it will become important in Sect. 3.3, where we assume that the adaption of the

scalarization weights happens much faster than the interaction dynamics and concentration process of the swarms.

We recover the original weights

$$\lambda^k = \frac{\mathrm{e}^{\mu^k}}{\sum_k \mathrm{e}^{\mu^k}}.$$

Combining (4) and (8) yields a CBO dynamic for multi-objective problems with self-adaptive scalarization weights. We want to emphasize that this combination introduces interdependencies between swarms. Moreover, the dynamics of the weights lead to dynamic scalarized cost functions. A pseudo code of this method can be found in Algorithm 2 in Sect. 5, note that we do not make use of the penalization yet, hence $\beta = 0$ and f_{uni} is irrelevant.

3.2.1 Numerical comparison of static and dynamic weights

To illustrate the effect of the dynamic weight adaption, we solve the test problem Schaffer1, see Sect. 5 for more details. The weight vectors are initialized equidistantly in $(0, 1)$ as illustrated with the blue points in Fig. 1a. The corresponding front found by solving (4) is shown in Fig. 1b. The distribution of the adaptive λ^k is given by the orange points in Fig. 1a and the corresponding front in Fig. 1c. We note that equally distributed weights stress the left part of the front. The adaptive weighting counteracts this effect by shifting the weights to the right, which results in a better resolution of the front. All other details regarding the numerical results are given in Sect. 5.

The results indicate that the proposed method works fine for convex problems. However, we note that we may run into problems in non-convex settings. Indeed, as discussed in Remark 1, we can only find convex parts of the front by globally minimizing weighted sum scalarizations. For general and possible non-convex problems we propose another extension in the next section. Before we enter the part of the general problems we analyse the method with adaptive weights in terms of well-posedness, mean-field approximation and convergence.

3.2.2 Well-posedness of the SDE system with adaptive weights

To analyze the SDE system with adaptive weights we introduce the notation $Y_t = [X_t, \mu_t]$ where $X_t^N = (X_t^k)_{k=1,\dots,K}$ and $\mu_t^N = (\mu_t^k)_{k=1,\dots,K}$. Moreover, we write the SDE system in vectorized form using the notation

$$\mathrm{d}Y_t^N = -G(Y_t^N)\,\mathrm{d}t + \sigma\, D(Y_t^N)\,\mathrm{d}B_t^N, \tag{9a}$$

$$G(Y_t^N) = \left[G_X(Y_t^N), G_\mu(Y_t^N)\right], \qquad D(Y_t^N) = \operatorname{diag}(G_X(Y_t^N), \mathbf{0}), \tag{9b}$$

$$Y_0^N = [X_0^N, \mu_0^N]. \tag{9c}$$

where

$$\left(G_X(Y_t^N)\right)_k = \left((X_t^{k,i} - v_t^k)\right)_{i=1,\dots,N_k}, \text{ and}$$

$$\left(G_\mu(Y_t^N)\right)_k = \frac{1}{\tau}\left(\sum_{\ell=1,\ell\neq k}^{K} \mathcal{K}(X_t^k, X_t^\ell, \mu^k, \mu^\ell)\right).$$

To obtain a well-posedness result for the system with adaptive weights, we make the following assumptions:

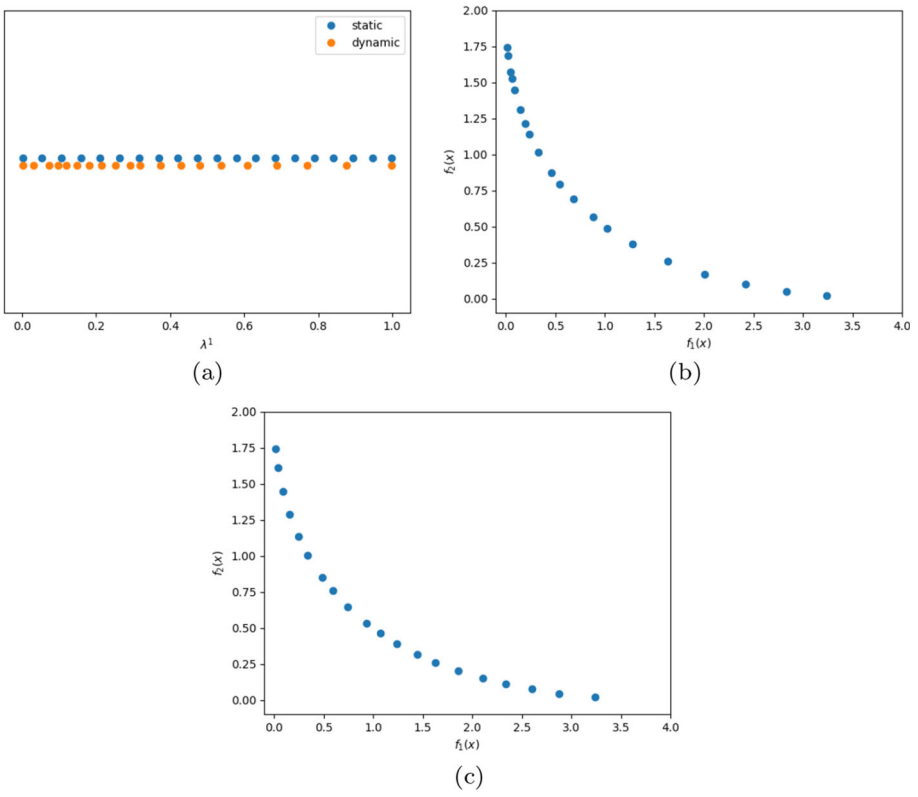

Fig. 1 Illustration of adaptive weights. The orange points in **a** show the adaptive weights at time $t = 2.5$. The corresponding front given by $f(v_t^k)_{k=1,\ldots,10}$ is shown in **c**. **b** The approximation obtained with equally distributed weights is shown

The interaction force $\mathcal{K}$ is locally Lipschitz and grows at most linearly in x, y, μ or ν. In particular, it holds

$$\left|\mathcal{K}(x, y, \mu, \nu) - \mathcal{K}(\bar{x}, \bar{y}, \bar{\mu}, \bar{\nu})\right| \leq C_L\big(|x - \bar{x}| + |y - \bar{y}| + |\mu - \bar{\mu}| + |\nu - \bar{\nu}|\big) \tag{A3}$$

and

$$\left|\mathcal{K}(x, y, \mu, \nu)\right| \leq C_b\big(|x| + |y| + |\mu| + |\nu|\big) \tag{A4}$$

for some constants $C_L, C_b > 0$ and $|(x, y, \mu, \nu) - (\bar{x}, \bar{y}, \bar{\mu}, \bar{\nu})| < M$.

Remark 7 As usual in particle interaction dynamics the forces resulting from physical potentials such as the Morse potential proposed above, do not satisfy the regularity assumption of ODE theory. In order to comply with these assumptions we have to consider a smoothed version, as for example,

$$\hat{P}(d) = R^{-d^2/r} - A\,\mathrm{e}^{-d^2/a}, \quad R, A \geq 0, \quad r, a > 0.$$

Similar to the single objective case, the proof of the well-posedness of the SDE system with adaptive weights is based on the following technical lemma.

Lemma 1 *Let $K \in \mathbb{N}$, $N_k \in \mathbb{N}$ for all $k = 1, \dots, K$, $N = \sum_{k=1}^{K} N_k$, further $\alpha, M > 0$ and (A3), (A4) hold. Then there exist constants $C_1, C_2 > 0$ depending only on M such that for any $Y_t = [X_t, \mu_t]$, $\hat{Y}_t = [\hat{X}_t, \hat{\mu}_t] \in \mathbb{R}^{dN} \times \mathbb{R}^{pK}$ with $|Y_t|, |\hat{Y}_t| \le M$ it holds*

$$\left|\left(G_X(Y_t) - G_X(\hat{Y}_t)\right)_{k,j}\right| + \left|\left(G_\mu(Y_t) - G_\mu(\hat{Y}_t)\right)_k\right| \le C_1 |Y_t^k - \hat{Y}_t^k|,$$
$$|G_X(Y_t)| + |G_\mu(Y_t)| \le C_2 |Y_t^k|.$$

Proof Let $N = \sum_{k=1}^{K} N_k$, $Y_t = [X_t, \mu_t]$, $\hat{Y}_t = [\hat{X}_t, \hat{\mu}_t] \in \mathbb{R}^{dN} \times \mathbb{R}^{pK}$. We begin with

$$\begin{aligned}
&\left(G_X(Y_t) - G_X(\hat{Y}_t)\right)_{k,j} = \\
&= \frac{\sum_{m=1}^{N_k} (X_t^{k,j} - X_t^{k,m}) \, \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,m})}}{\sum_{m=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,m})}} - \frac{\sum_{m=1}^{N_k} (\hat{X}_t^{k,j} - \hat{X}_t^{k,m}) \, \mathrm{e}^{-\alpha f_{\mu^k}(\hat{X}_t^{k,m})}}{\sum_{m=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(\hat{X}_t^{k,m})}} \\
&\quad + \frac{\sum_{m=1}^{N_k} (\hat{X}_t^{k,j} - \hat{X}_t^{k,m}) \, \mathrm{e}^{-\alpha f_{\mu^k}(\hat{X}_t^{k,m})}}{\sum_{m=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(\hat{X}_t^{k,m})}} - \frac{\sum_{m=1}^{N_k} (\hat{X}_t^{k,j} - \hat{X}_t^{k,m}) \, \mathrm{e}^{-\alpha f_{\hat{\mu}^k}(\hat{X}_t^{k,m})}}{\sum_{m=1}^{N_k} \mathrm{e}^{-\alpha f_{\hat{\mu}^k}(\hat{X}_t^{k,m})}} \\
&=: I_1 + I_2 + I_3 + I_4 + I_5
\end{aligned}$$

where

$$I_1 = \frac{\sum_{m=1}^{N_k} \left(X_t^{k,j} - \hat{X}_t^{k,j} - (X_t^{k,m} - \hat{X}_t^{k,m})\right) \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,m})}}{\sum_{m=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,m})}},$$

$$I_2 = \frac{\sum_{m=1}^{N_k} (\hat{X}_t^{k,j} - \hat{X}_t^{k,m})(\mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,m})} - \mathrm{e}^{-\alpha f_{\mu^k}(\hat{X}_t^{k,m})})}{\sum_{m=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,m})}},$$

$$I_3 = \sum_{m=1}^{N_k} (\hat{X}_t^{k,j} - \hat{X}_t^{k,m}) \mathrm{e}^{-\alpha f_{\mu^k}(\hat{X}_t^{k,m})} \frac{\sum_{m=1}^{N_k} (\mathrm{e}^{-\alpha f_{\mu^k}(\hat{X}_t^{k,m})} - \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,m})})}{\sum_{m=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,m})} \sum_{m=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(\hat{X}_t^{k,m})}},$$

$$I_4 = \frac{\sum_{m=1}^{N_k} (\hat{X}_t^{k,j} - \hat{X}_t^{k,m})(\mathrm{e}^{-\alpha f_{\mu^k}(\hat{X}_t^{k,m})} - \mathrm{e}^{-\alpha f_{\hat{\mu}^k}(\hat{X}_t^{k,m})})}{\sum_{m=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(\hat{X}_t^{k,m})}},$$

$$I_5 = \sum_{m=1}^{N_k} (\hat{X}_t^{k,j} - \hat{X}_t^{k,m}) \, \mathrm{e}^{-\alpha f_{\hat{\mu}^k}(\hat{X}_t^{k,m})} \frac{\sum_{m=1}^{N_k} (\mathrm{e}^{-\alpha f_{\mu^k}(\hat{X}_t^{k,m})} - \mathrm{e}^{-\alpha f_{\hat{\mu}^k}(\hat{X}_t^{k,m})})}{\sum_{m=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(\hat{X}_t^{k,m})} \sum_{m=1}^{N_k} \mathrm{e}^{-\alpha f_{\hat{\mu}^k}(\hat{X}_t^{k,m})}}.$$

Similar to [8] the terms can be bounded as follows:

$$|I_1| \le |X_t^{k,j} - \hat{X}_t^{k,j}| + |X_t^k - \hat{X}_t^k|,$$

$$|I_2| \le \sqrt{\frac{2}{N_k}} \, c_X^k \, |X_t^k - \hat{X}_t^k| \sqrt{N_k |\hat{X}_t^{k,j}|^2 + |\hat{X}_t^k|^2},$$

$$|I_3| \le \frac{\sqrt{2} \, c_X^k}{N_k} \, |X_t^k - \hat{X}_t^k| \sqrt{|\hat{X}_t^{k,j}|^2 + |\hat{X}_t^k|^2},$$

where $c_X^k := \alpha \|\nabla f_{\mu^k}\|_{L^\infty(B_X)} \mathrm{e}^{(\alpha \|f_{\mu^k} - \underline{f_{\mu^k}}\|_{L^\infty(B_X)})}$ for $B_X = \{x \in \mathbb{R}^d : |x| \leq M\}$ and $\underline{f_{\mu^k}} := \min_{x \in B_X} f_{\mu^k}(x)$. To bound I_4 and I_5 we first note that

$$f_{\mu^k}(x) - f_{\hat{\mu}^k}(x) = \frac{\sum_{i=1}^p (\mathrm{e}^{\mu_i^k} - \mathrm{e}^{\hat{\mu}_i^k})\, f_i(x)}{\sum_{i=1}^p \mathrm{e}^{\mu_i^k}} + \sum_{i=1}^p \mathrm{e}^{\hat{\mu}_i^k} f_i(x) \frac{\sum_{i=1}^p (\mathrm{e}^{\hat{\mu}_i^k} - \mathrm{e}^{\mu_i^k})}{\sum_{i=1}^p \mathrm{e}^{\hat{\mu}_i^k} \sum_{i=1}^p \mathrm{e}^{\mu_i^k}}$$

which leads to the estimate

$$|f_{\mu^k}(x) - f_{\hat{\mu}^k}(x)| \leq \left(1 + \frac{1}{p}\right) \|f\|_{L^\infty(B_X)} \mathrm{e}^{2M} |\mu^k - \hat{\mu}^k|$$

for $|\mu^k|, |\hat{\mu}^k| \leq M$. Using this we obtain

$$|I_4| \leq \sqrt{\frac{2}{N_k}}\, d_M^k \sqrt{N_k |\hat{X}_t^{k,j}|^2 + |\hat{X}_t^k|^2}\, \|f\|_{L^\infty(B_X)} \left(1 + \frac{1}{p}\right) \mathrm{e}^{2M} |\mu^k - \hat{\mu}^k|,$$

$$|I_5| \leq \frac{\sqrt{2}}{N_k}\, d_M^k \sqrt{|\hat{X}_t^{k,j}|^2 + |\hat{X}_t^k|^2}\, \|f\|_{L^\infty(B_X)} \left(1 + \frac{1}{p}\right) \mathrm{e}^{2M} |\mu^k - \hat{\mu}^k|,$$

where $d_M^k := \alpha\, \mathrm{e}^{(2\alpha \|f\|_{L^\infty(B_X)})}$. As these estimates are independent of $j = 1, \ldots, N_k$ and $k = 1, \ldots, K$, this proves that G_X is locally Lipschitz. Note that G_μ is locally Lipschitz by Assumption (A3).

Moreover, it is easy to see that $|(G_X(Y_t))_{k,j}| \leq |X_t^{k,j}| + |X_t^k| \leq 2|Y_t^k|$ holds which leads to the linear growth of G_X. Further, the linear growth of G_μ is ensured by Assumption (A3). □

Based on this Lemma we establish the following well-posedness and existence result.

Theorem 7 *Let the assumptions of Lemma 1 be satisfied. For each $K \in \mathbb{N}$ and $N_k \in \mathbb{N}$ for $k = 1, \ldots, K$, system (9) admits a unique strong solution $\{Y_t : t \geq 0\}$ for any initial condition Y_0 satisfying $\mathbb{E}[|Y_0|^2] < \infty$.*

Proof The proof is based on [13, Chapter 5.3, Theorem 3.1]. Note that due to Lemma 1 we only have to show that there exists $b > 0$ such that it holds

$$-2\, Y_t \cdot G(Y_t) + \sigma^2 \operatorname{trace}\big(D(Y_t) D(Y_t)^\top\big) \leq b\, |Y_t|^2.$$

For the first term we obtain, using (A3) and (A4),

$$-2\, Y_t \cdot G(Y_t) \leq 2\, C_1\, |Y_t|^2$$

for some constant $C_1 > 0$. Moreover, by the diagonal structure of $D(Y_t)$ we get

$$\operatorname{trace}\big(D(Y_t)\, D(Y_t)^\top\big) \leq 2\, C_2\, |Y_t|^2$$

for some constant $C_2 > 0$. Altogether, this proves the desired estimate and Theorem 3.1 in [13, Chapter 5.3] yields the result. □

Remark 8 Note that the regularized potential proposed in Remark 7 satisfies the regularity assumptions of Theorem 7.

3.2.3 Mean-field limit with adaptive weights

Formally passing to the mean-field limit yields a coupled PDE/ODE system given by

$$\begin{aligned}
&\partial_t \rho_t^k = \frac{\sigma^2}{2} \sum_{i=1}^{d} \partial_{ii}\big((x - v[\rho_t^k])\rho_t\big) + \nabla \cdot \big((x - v[\rho_t^k])\rho_t^k\big),\\
&\lim_{t\to 0} \rho_t^k = \rho_0, \quad k = 1, \dots, K,\\
&\frac{\mathrm{d}}{\mathrm{d}t}\mu^k = -\frac{1}{\tau} \sum_{\ell=1,\ell\neq k}^{K} \mathcal{K}(\rho_t^k, \rho_t^\ell, \mu^k, \mu^\ell), \qquad \mu^k(0) = \ln(\lambda_0^k)
\end{aligned}$$

with the weighted mean given by

$$v[\rho_t^k] = \frac{1}{\int \mathrm{e}^{-\alpha f_{\lambda^k}(x)} d\rho_t^k(x)} \int x\, \mathrm{e}^{-\alpha f_{\lambda^k}(x)}\, \mathrm{d}\rho_t^k(x), \qquad \lambda^k = \mathrm{e}^{\mu^k}$$

and $\tau, \sigma > 0$, $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ as above. Note that the coupling through the weights makes the proof of convergence towards the efficient set considerably more difficult as there is a trade-off between the concentration of the swarms and the balancing of the weight vectors. This is addressed in the next section.

3.3 Diverse approximation of the Pareto front

The aim of this section is to prove that the proposed scheme leads to a diverse approximation of the Pareto front in the case of strictly convex biobjective problems, i.e. $f\colon \mathbb{R}^d \to \mathbb{R}^2$. In fact, we prove that each swarm concentrates at one point along the Pareto front in the long-time limit, $t \to \infty$. The proof builds on results of the single-objective CBO scheme in [8, 15]. Certainly, we require additional assumptions that we motivate and formulate in the following.

For the sake of a simple presentation of the idea, we consider the smoothed version of the Morse interaction potential discussed in Remark 7. Moreover, we assume that the initial condition of the scalarization weights and the parameters of the interaction potentials are chosen such that the scalarization weights do not leave the domain $(\epsilon_\lambda, 1-\epsilon_\lambda) \times (\epsilon_\lambda, 1-\epsilon_\lambda)$ for fixed $0 < \epsilon_\lambda \ll 1$. This can be ensured by choosing A_f, R_f, a_f and r_f such that we have short-range repulsion and mid-range attraction and negligible interaction in the long-range. Note that this assumption allows us to work directly with normalized scalarization weights $\lambda^k = (\lambda_1^k, \lambda_2^k)$, $\lambda_1^k + \lambda_2^k = 1$ for all $k = 1, \dots, K$. In fact we do not require the reformulation in terms of μ that is used for the numerics.

To be more precise, we consider the dynamics

$$\frac{\mathrm{d}}{\mathrm{d}t}\lambda^k = \frac{-1}{\tau} \sum_{\ell=1}^{K} \left(\frac{A_f}{a_f}\, \mathrm{e}^{-(d_{k,\ell}^f)^2/a_f} - \frac{R_f}{r_f}\, \mathrm{e}^{-(d_{k,\ell}^f)^2/r_f} \right) (\lambda^k - \lambda^\ell) \quad \text{for } k = 1, \dots, K.$$

Note that

$$\lambda^k - \lambda^\ell = \begin{pmatrix} \lambda_1^k - \lambda_1^\ell \\ \lambda_1^\ell - \lambda_1^k \end{pmatrix} = \begin{pmatrix} \lambda_2^\ell - \lambda_2^k \\ \lambda_2^k - \lambda_2^\ell \end{pmatrix} \quad \text{for all } k, \ell.$$

As all other quantities on the right-hand side are scalar, the λ^k stay on the constraint $\lambda_1^k + \lambda_2^k = 1$ for all times $t > 0$ once they are normalized.

As a consequence, the dynamics of the vectors λ^k are uniquely defined by the dynamics of one of their components. Without loss of generality we chose the first component here and obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\lambda_1^k = -\sum_{\ell=1}^{K} u(d_{k,\ell}^f)(\lambda_1^k - \lambda_1^\ell), \qquad \frac{\mathrm{d}}{\mathrm{d}t}\lambda_2^k = -\sum_{\ell=1}^{K} u(d_{k,\ell}^f)(\lambda_1^\ell - \lambda_1^k) = 1 - \frac{\mathrm{d}}{\mathrm{d}t}\lambda_1^k,$$

where we use the notation

$$u(d_{k,\ell}^f) = \left(\frac{2A_f}{a_f}\,\mathrm{e}^{-(d_{k,\ell}^f)^2/a_f} - \frac{2R_f}{r_f}\,\mathrm{e}^{-(d_{k,\ell}^f)^2/r_f}\right).$$

Note that the unique representation by the first component allows us to sort the scalarization weights of the swarms. Without loss of generality, we assume in the following that $\lambda_1^1 < \lambda_1^2 < \cdots < \lambda_1^K$. Moreover, due to negligible long-range interaction we only consider interactions with direct neighbors.

The strategy of the proof is as follows: first, we exploit the relationship of the non-dominated points $f(\bar{x}(\lambda))$ corresponding to a given $\lambda \in \Lambda$ (i.e., $\bar{x}(\lambda)$ is the unique optimum of (1) with weight λ) and the scalarization weight λ itself. In fact, we assume that

$$\begin{pmatrix} \lambda_1^k - \lambda_1^\ell \\ \lambda_1^\ell - \lambda_1^k \end{pmatrix} = S(\lambda^k, \lambda^\ell) f(\bar{x}(\lambda^k)) - f(\bar{x}(\lambda^\ell)) \quad \text{for all } k, \ell$$

for some problem dependent matrix $S(\lambda^k, \lambda^\ell)$. Second, using the chain rule, we notice that

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\bar{x}(\lambda^k)) = \frac{\mathrm{d}}{\mathrm{d}x} f(\bar{x}(\lambda^k)) \frac{\mathrm{d}}{\mathrm{d}\lambda_1^k} \bar{x}(\lambda^k) \frac{\mathrm{d}}{\mathrm{d}t}\lambda_1^k.$$

Now, we employ a Lyapunov argument for all swarms at once, which will show that the non-dominated points $f(\bar{x}(\lambda^k))$ spread along the Pareto front with $\|\cdot\|_2$-distance $d_{\min}$ in the long time limit, where $d_{\min}$ is the unique root of $u(d)$.

Let us state the assumptions and the theorem:

(A5) The Pareto front of the biobjective problem $f\colon \mathbb{R}^d \to \mathbb{R}^2$ is strictly convex and connected.

(A6) The parameters of the interaction potential satisfy $A, R = 0$ and A_f, R_f, a_f and r_f are such that the interaction potential $\mathcal{U}$ models short-range repulsion and mid-range attractions; long-range interactions are negligible. In particular, we assume that there is a unique root $d_{\min}$ of $u(d)$ such that all K non-dominated points $f(\bar{x}(\lambda^k))$, $k = 1, \ldots K$ can spread along the front with Euclidean distance greater than $d_{\min}$. Moreover, we assume that each swarm is only interacting with its direct neighbors (in the sense of the sorted weight vectors).

(A7) The initial values of the scalarization weights λ^k satisfy $\lambda_1^k + \lambda_2^k = 1$ and $(\lambda_1^k, \lambda_2^k) \in [\epsilon_\lambda, 1 - \epsilon_\lambda]^2$ for each $k = 1, \ldots, K$ and fixed $0 < \epsilon_\lambda \ll 1$. Moreover, $\lambda_1^1 < \lambda_1^2 < \cdots < \lambda_1^K$ and the corresponding non-dominated points $f(\bar{x}(\lambda^k))$ are distributed such that $\epsilon_\lambda \le \lambda_1^1$ and $\lambda_1^K \le 1 - \epsilon_\lambda$.

(A8) For each pair λ^k, λ^ℓ with $k, \ell \in \{1, \ldots, K\}$, $k \neq \ell$, there exists a negative definite matrix $S(\lambda^k, \lambda^\ell) \in \mathbb{R}^{2\times 2}$ such that

$$\begin{pmatrix} \lambda_1^k - \lambda_1^\ell \\ \lambda_1^\ell - \lambda_1^k \end{pmatrix} = S(\lambda^k, \lambda^\ell)(f(\bar{x}(\lambda^k)) - f(\bar{x}(\lambda^\ell))) \text{ and } S(\lambda^k, \lambda^\ell) = S(\lambda^\ell, \lambda^k).$$

(A9) For each $\lambda \in [\epsilon_\lambda, 1-\epsilon_\lambda]$ there exists a negative definite matrix $T(\lambda)$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\bar{x}(\lambda)) = T(\lambda)\frac{\mathrm{d}}{\mathrm{d}t}\lambda = T(\lambda)\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\lambda_1\\ 1-\lambda_1\end{pmatrix}.$$

(A10) The product $\mathcal{T}(\lambda)\mathcal{S}(\lambda)$ with

$$\mathcal{T}(\lambda) = \begin{pmatrix} T(\lambda^1)+T(\lambda^2) & -T(\lambda^2) & & & \\ -T(\lambda^2) & T(\lambda^2)+T(\lambda^3) & -T(\lambda^3) & \ddots & \\ & -T(\lambda^3) & \ddots & \ddots & -T(\lambda^{K-1}) \\ & & \ddots & -T(\lambda^{K-1}) & T(\lambda^{K-1})+T(\lambda^K) \end{pmatrix}$$

and

$$\mathcal{S}(\lambda) = \begin{pmatrix} S(\lambda^1,\lambda^2) & & \\ & \ddots & \\ & & S(\lambda^{K-1},\lambda^K) \end{pmatrix} \tag{10}$$

is symmetric. Here $T(\lambda^k)$ are as defined in (A9) and $S(\lambda^k,\lambda^\ell)$ as defined in (A8) for each $k,\ell = 1,\dots,K, k \neq \ell$.

(A11) The dynamics of the scalarization weights and the dynamics of the swarms have different time scales. In fact, the adaption of the scalarization weights is much faster than the dynamics of the swarms.

We first state the result for the limiting case $\alpha = \infty$ as this ensures that $f(v[\rho_t^k]) = f(\bar{x}(\lambda))$ by the Laplace principle [15].

Theorem 8 *Let (A1)–(A11) hold and $\alpha = \infty$. For $K \in \mathbb{N}$ the dynamics*

$$\begin{aligned}
\partial_t \rho_t^k &= \frac{\sigma^2}{2}\sum_{i=1}^d \partial_{ii}\big((x - v[\rho_t^k])\rho_t\big) + \nabla\cdot\big((x-v[\rho_t^k])\rho_t^k\big), \quad \lim_{t\to 0}\rho_t^k = \rho_0,\\
\frac{\mathrm{d}}{\mathrm{d}t}\lambda^k &= -\frac{1}{\tau}\sum_{\ell=1}^K\left(\frac{A_f}{a_f}\,\mathrm{e}^{-(d_{k,\ell}^f)^2/a_f} - \frac{R_f}{r_f}\,\mathrm{e}^{-(d_{k,\ell}^f)^2/r_f}\right)(\lambda^k-\lambda^\ell),\\
v[\rho_t^k] &= \frac{1}{\int \mathrm{e}^{-\alpha f_{\lambda^k}(x)} d\rho_t^k(x)}\int x\,\mathrm{e}^{-\alpha f_{\lambda^k}(x)}\,\mathrm{d}\rho_t^k(x), \quad k = 1,\dots,K,
\end{aligned}$$

yields a diverse approximation of the Pareto front of f in the long time limit. In particular, each swarm concentrates at $\bar{x}(\lambda^k)$ with pairwise distance

$$\|f(\bar{x}(\lambda^k)) - f(\bar{x}(\lambda^\ell))\|_2 \geq d_{min} \text{ for all } k \neq \ell \text{ as } t \to \infty.$$

In other words, the stationary solution of the dynamics consists of K Dirac-measures located at the points $\bar{x}(\lambda^k)$ for $k = 1,\dots,K$ which have a pairwise Euclidean distance greater or equal $d_{\min}$.

Proof First note that due to $\alpha = \infty$, we have $v_f[\rho_t^k] = \bar{x}(\lambda^k)$ for $t > 0$ [15]. This allows us to focus only on the dynamics of the scalarization weights in the following. Indeed, we follow

the steps motivated above and begin with some observations on the pairwise interactions. Using (A6), (A8) and (A9) we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} f(\bar{x}(\lambda^k)) &= \frac{\mathrm{d}}{\mathrm{d}x} f(\bar{x}(\lambda^k)) \frac{\mathrm{d}}{\mathrm{d}\lambda^k} \bar{x}(\lambda^k) \frac{\mathrm{d}}{\mathrm{d}t} \lambda^k = T(\lambda^k) \frac{\mathrm{d}}{\mathrm{d}t} \lambda^k \\ &= -T(\lambda^k) \sum_{\ell=k-1}^{k+1} u(d_{k,\ell}^f)(\lambda^k - \lambda^\ell) \\ &= -T(\lambda^k) \sum_{\ell=k-1}^{k+1} S(\lambda^k, \lambda^\ell) u(d_{k,\ell}^f)(f(\bar{x}(\lambda^k)) - f(\bar{x}(\lambda^\ell)))\end{aligned}$$

for $k = 1, \ldots, K$. For simplicity we set the summand to zero whenever $\ell < 1$ or $\ell > K$.

In our strictly convex setting, the sorting of the weights λ^k induces a sorting of the points $f(\bar{x}(\lambda^1)), \ldots, f(\bar{x}(\lambda^K)$ in descending order along the f_1-axis. To reconstruct the vectors $f(\bar{x}(\lambda^k)), k = 1, \ldots, K$ at $t > 0$, it is sufficient to know their initial values and the relative positions from k to $k + 1$. This will be exploited in the following.

For $k = 1, \ldots K - 1$ we define $q_{k,k+1} = f(\bar{x}(\lambda^k)) - f(\bar{x}(\lambda^{k+1}))$. Clearly, it holds

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} q_{k,k+1} &= \frac{\mathrm{d}}{\mathrm{d}t} f(\bar{x}(\lambda^k)) - \frac{\mathrm{d}}{\mathrm{d}t} f(\bar{x}(\lambda^{k+1})) \\ &= -T(\lambda^k) \sum_{\ell=k-1}^{k+1} S(\lambda^k, \lambda^\ell) u(d_{k,\ell}^f)(f(\bar{x}(\lambda^k)) - f(\bar{x}(\lambda^\ell))) \\ &\quad + T(\lambda^{k+1}) \sum_{\ell=k}^{k+2} S(\lambda^{k+1}, \lambda^\ell) u(d_{k+1,\ell}^f)(f(\bar{x}(\lambda^{k+1})) - f(\bar{x}(\lambda^\ell))).\end{aligned}$$

Replacing the differences yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} q_{1,2} &= -\big(T(\lambda^1) + T(\lambda^2)\big) S(\lambda^1, \lambda^2) u(d_{1,2}^f) q_{1,2} \\ &\quad + T(\lambda^2) S(\lambda^2, \lambda^3) u(d_{2,3}^f) q_{2,3} \\ \frac{\mathrm{d}}{\mathrm{d}t} q_{K-1,K} &= T(\lambda^{K-1}) S(\lambda^{K-2}, \lambda^{K-1}) u(d_{K-2,K-1}^f) q_{K-2,K-1} \\ &\quad - \big(T(\lambda^{K-1}) + T(\lambda^K)\big) S(\lambda^{K-1}, \lambda^K) u(d_{K-1,K}^f) q_{K-1,K}\end{aligned}$$

and for all $k = 2, \ldots, K - 2$ it holds

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} q_{k,k+1} &= T(\lambda^k) S(\lambda^{k-1}, \lambda^k) u(d_{k-1,k}^f) q_{k-1,k} \\ &\quad - \big(T(\lambda^k) + T(\lambda^{k+1})\big) S(\lambda^k, \lambda^{k+1}) u(d_{k,k+1}^f) q_{k,k+1} \\ &\quad + T(\lambda^{k+1}) S(\lambda^{k+1}, \lambda^{k+2}) u(d_{k+1,k+2}^f) q_{k+1,k+2}\end{aligned}$$

Now, let us consider the vector $q = (q_{1,2}, \ldots, q_{K-1,K})$ and the potential

$$\mathcal{V}(q) = \sum_{k=1}^{K-1} \left(R_f \mathrm{e}^{-|q_{k,k+1}|^2/r_f} - A_f \mathrm{e}^{-|q_{k,k+1}|^2/a_f} \right)$$

with gradient

$$\nabla \mathcal{V}(q) = \left(u(d_{1,2}^f) q_{1,2}, \ldots, u(d_{K-1,K^f}) q_{K-1,K} \right)^\top.$$

The vectorized dynamics reads

$$\frac{\mathrm{d}}{\mathrm{d}t}q = -\mathcal{T}(\lambda)\mathcal{S}(\lambda)\nabla\mathcal{V}(q),$$

with $\mathcal{T}(\lambda)$ and $\mathcal{S}(\lambda)$ as in (A10).

It is easy to check using the binomial formulas and (A9) that $\mathcal{T}(\lambda)$ is negative definite. By (A10) $\mathcal{T}(\lambda)\mathcal{S}(\lambda)$ is symmetric and therefore positive definite. By (A7) we can bound the smallest eigenvalue of the product, $\sigma_0(\lambda) > 0$ by some constant depending on ϵ_λ, i.e., $\sigma_0(\lambda) > c_{\epsilon_\lambda} > 0$ for all λ.

The positive definiteness of $\mathcal{T}(\lambda)\mathcal{S}(\lambda)$ allows us to define a norm $\|\cdot\|_{\mathcal{T}\mathcal{S}}$ induced by the scalar product $\langle\cdot, T(\lambda)\mathcal{S}(\lambda)\cdot\rangle$. The time evolution of $\mathcal{V}(q)$ can now be computed as

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{V}(q) = -\|\nabla\mathcal{V}(q)\|^2_{\mathcal{T}\mathcal{S}} \leq -c_{\epsilon_\lambda}\|\nabla V(q)\|^2 < 0. \tag{11}$$

Hence $\mathcal{V}$ is a Lyapunov functional for q and hence in the long-time limit the dynamics stabilizes such that $\nabla\mathcal{V} = 0$. In particular, by (A6) that means the Euclidean distances of the non-dominated points is greater than $d_{\min}$.

Now, τ in (A11) allows us to balance the times of the stabilization of the scalarization weights and the time to collapse the swarms. For $\tau \to 0$ the scalarization weights converge very fast to a diverse approximation of the Pareto front. The scalarization weights are then stationary and the swarms concentrate at their weighted average as discussed in Theorem 6. This concludes the proof. □

We want to emphasize that in this setting each of the swarms admits a stationary solution which is the efficient point along the Pareto front where the swarm concentrates. In contrast, the particles of the dynamic proposed in [4] may move along the front for all times as Theorem 4.1 in [4] shows that the density of the particles move towards the Pareto front, but no stationarity is shown. Before presenting an example satisfying (A5)-(A10) we comment on the case $\alpha < \infty$.

Remark 9 We expect to obtain a similar result for $\alpha < \infty$ using the quantitative Laplace principle [15]. Clearly, the points $f(v[\rho_t^k]), k = 1, \ldots, K$ will only be in an ϵ-neighborhood of the Pareto front, where $0 < \epsilon \ll 1$ depends on α. The details of the proof are beyond the scope of this article.

3.3.1 Example for (A5)–(A10)

In the following we present an example with two swarms satisfying the technical assumptions (A5)-(A10). Let us consider two swarms represented by the weighting vectors $\lambda^1, \lambda^2 \in \Lambda$ in the following. Note that biobjective convex quadratic optimization problems have been extensively studied in the literature, therefore parts of the following analysis, including optimality conditions and the subsequent sensitivity analysis, can also be found, e.g., in [2, 17, 30].

Let us consider the biobjective stricly convex quadratic problem

$$f : \mathbb{R} \to \mathbb{R}^2, \qquad f(x) = \begin{pmatrix} x^2 + 1 \\ \frac{1}{2}(x-1)^2 + 1 \end{pmatrix}.$$

Let $\lambda \in [0, 1]$ and consider the scalarized objective

$$f_\lambda(x) = \lambda(x^2 + 1) + (1 - \lambda)\Big(\frac{1}{2}(x-1)^2 + 1\Big).$$

Note that this is equivalent to a weighted sum scalarization (1) with $(\lambda_1, \lambda_2) = (\lambda, 1-\lambda)$. The efficient solutions parameterized by $\lambda \in [0,1]$ can be computed using the first order optimality condition for f_λ w.r.t. the variable $x \in \mathbb{R}$:

$$\frac{\mathrm{d}}{\mathrm{d}x} f_\lambda(x) = (1+\lambda)x - (1-\lambda) \stackrel{!}{=} 0$$
$$\Longrightarrow \bar{x}(\lambda) = \frac{1-\lambda}{1+\lambda} \quad \text{and} \quad f(\bar{x}(\lambda)) = \begin{pmatrix} \left(\frac{1-\lambda}{1+\lambda}\right)^2 + 1 \\ 2\left(\frac{\lambda}{1+\lambda}\right)^2 + 1 \end{pmatrix}.$$

The function $P\colon [0,1] \to \mathbb{R}^2$ given by $P(\lambda) = f(\bar{x}(\lambda))$ is continuous, therefore (A5) holds. Moreover, we obtain for the difference of the non-dominated points of the two swarms as

$$f(\bar{x}(\lambda_1^1)) - f(\bar{x}(\lambda_1^2)) = \begin{pmatrix} \left(\frac{1-\lambda_1^1}{1+\lambda_1^1}\right)^2 - \left(\frac{1-\lambda_1^2}{1+\lambda_1^2}\right)^2 \\ 2\left(\frac{\lambda_1^1}{1+\lambda_1^1}\right)^2 - 2\left(\frac{\lambda_1^2}{1+\lambda_1^2}\right)^2 \end{pmatrix}$$
$$= \frac{-2}{(1+\lambda_1^1)(1+\lambda_1^2)} \begin{pmatrix} \frac{1-\lambda_1^1}{1+\lambda_1^1} + \frac{1-\lambda_1^2}{1+\lambda_1^2} & 0 \\ 0 & \frac{\lambda_1^1}{1+\lambda_1^1} + \frac{\lambda_1^2}{1+\lambda_1^2} \end{pmatrix} (\lambda^1 - \lambda^2).$$

As we assume that there exists $\epsilon_\lambda > 0$ sufficiently small such that all entries of λ^1, λ^2 lie in $[\epsilon_\lambda, 1-\epsilon_\lambda] \subset (0,1)$, the matrix is invertible. This allows us to define $S(\lambda^1, \lambda^2)$ through its inverse

$$S(\lambda^1, \lambda^2)^{-1} = \frac{-2}{(1+\lambda_1^1)(1+\lambda_1^2)} \begin{pmatrix} \frac{1-\lambda_1^1}{1+\lambda_1^1} + \frac{1-\lambda_1^2}{1+\lambda_1^2} & 0 \\ 0 & \frac{\lambda_1^1}{1+\lambda_1^1} + \frac{\lambda_1^2}{1+\lambda_1^2} \end{pmatrix}.$$

Concerning (A9) we compute

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\bar{x}(\lambda^1)) = \frac{-2}{(1+\lambda_1^1)^3} \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1^1 \\ 1-\lambda_1^1 \end{pmatrix} \frac{\mathrm{d}}{\mathrm{d}t} \lambda_1^1$$
$$= \frac{-2\,u(d_{1,2}^f)}{(1+\lambda_1^1)^3} \begin{pmatrix} 2(1-\lambda_1^1) & 0 \\ 0 & 2\lambda_1^1 \end{pmatrix} (\lambda^1 - \lambda^2).$$

Hence, we find

$$T(\lambda^1) = \frac{-2}{(1+\lambda_1^1)^3} \begin{pmatrix} 2(1-\lambda_1^1) & 0 \\ 0 & 2\lambda_1^1 \end{pmatrix}$$

is negative definite and analogous for $k = 2$. As $T(\lambda^1)$, $T(\lambda^2)$ and $S(\lambda^1, \lambda^2)$ are all diagonal, it is easy to check that (A10) holds.

Up to here, we mainly focused on convex objective functions since only convex parts of the efficient set can be obtained by global minimization of weighted sum scalarizations, see [9] for more details. In Sect. 4 we introduce a penalization of clusters in the objective space in order to obtain well-dispersed representations of the Pareto front and, as an intended side-effect, reach into non-convex parts of the Pareto front.

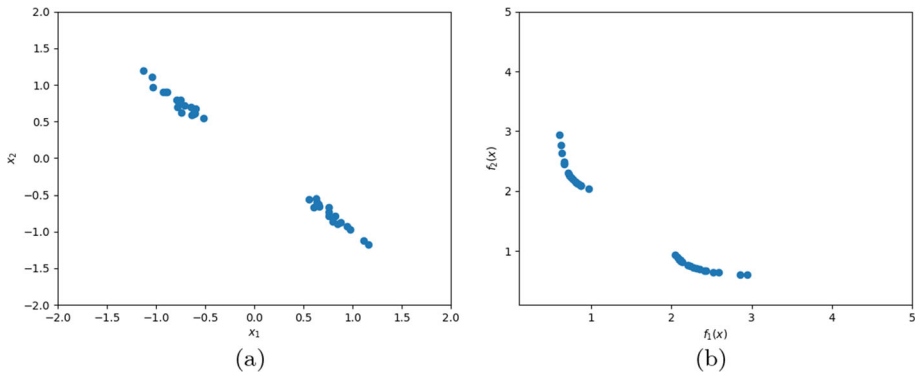

Fig. 2 Illustration of the clustering of the weighted averages in non-convex settings. For an idea of the shape of the true front see Fig. 3

4 MSCBO for general multi-objective problems

As discussed in the previous section, we cannot expect the present dynamics to work well in non-convex settings. We therefore propose a penalization strategy in order to obtain dynamics that lead to reasonable approximations for general multi-objective problems.

4.1 Penalization of clusters in the objective space

To motivate the penalization strategy we analyze the approximation obtained with the adaptive weight algorithm applied to the dent test problem, see Sect. 5.2 for more details.

Figure 2 shows the approximation of the Pareto front (b) and of the efficient set (a) that is obtained when using the algorithm with adaptive weights. We observe that the weighted means form clusters along the convex part of the Pareto front while the non-convex part is not recovered.

To overcome this issue, we propose a penalization strategy that avoids clusters along the Pareto front and tends to a uniform distribution of points along the Pareto front. Towards this end, we consider the potential leading to the interaction forces above, which is given by (5). As we want to penalize clusters in the objective space, we use the distance in the objective space leading to the penalization term

$$f_{\text{uni}}(X_t^{k,j}, v) = \sum_{\ell=1,\ell\neq k}^{K} R_c\, \mathrm{e}^{-|f(X_t^{k,j})-f(v_t^{\ell})|/r_c}, \quad R_c, r_c > 0. \tag{12}$$

By construction, the penalty term is smaller the farther away the objective of a particle is from the objective of the weighted means of the other swarms. The penalization term is added to the scalarization leading to a new cost given by

$$f_{\mu^k,p}(X_t^{k,j}) = f_{\mu^k}(X_t^{k,j}) + \frac{\beta}{\alpha} f_{\text{uni}}(X_t^{k,j}, v), \qquad \alpha > 0, \beta \geq 0.$$

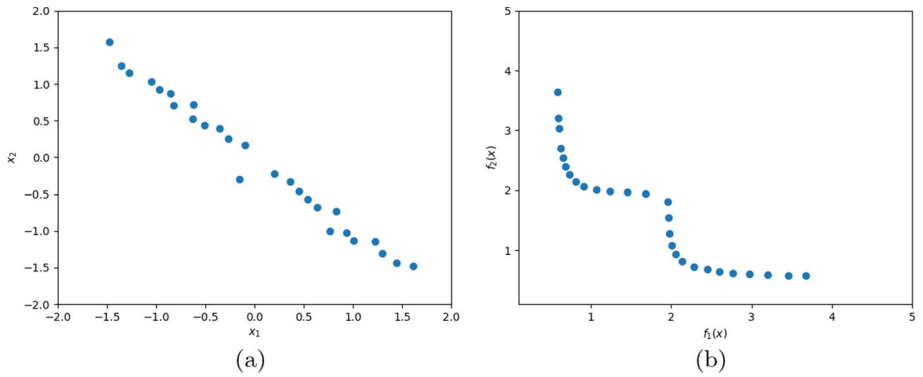

Fig. 3 Illustration of the effect of penalization. **a** We see the approximation of the efficient set by the weighted means of the swarms obtained with Algorithm 2. **b** The corresponding approximation of the Pareto front is shown

Here β allows us to balance the objective and the penalization. Using the properties of the exponential function, we can rewrite the weighted average from (4c) as

$$v_t^k = \frac{\sum_{j=1}^{N_k} X_t^{k,j} \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,j})} \mathrm{e}^{-\beta f_{\mathrm{uni}}(X_t^{k,j}, v)}}{\sum_{j=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,j})} \mathrm{e}^{-\beta f_{\mathrm{uni}}(X_t^{k,j}, v)}}. \tag{13}$$

Clearly, for $\beta = 0$ we are in the previous case without penalization of clustering in the objective space. With increasing values of $\beta > 0$ the impact of this penalization in the optimization process is increased. A pseudo code can be found in Algorithm 3 in Sect. 5.

4.2 Numerical results for adaptive weights and penalization

The following results are obtained with the adaptive weight dynamics with additional penalization given in (13). We observe that the penalization prevents the clustering and the swarms are able to approximate also the non-convex part of the Pareto front. Moreover, the swarms spread along the convex parts leading to a better result in both the decision- and the objective space, as is illustrated in Fig. 3. The improvement compared to Fig. 2 is obvious.

The influence of the penalization on the convex example of Sect. 3.2.1 is illustrated in Fig. 4. The approximation of the Pareto front obtained with penalization nicely covers the whole efficient set. In particular, the tails are better resolved compared to the approximation without penalization, cf. Fig. 1a.

We emphasize that the penalization only affects the cost function of the problem. Therefore, all the analytical results obtained in Sect. 3 still apply. The only difference is that the weighted averages of the swarms approximate the minimizers of the penalized scalarized cost functions instead of the global minima of the (original) scalarized cost functions.

The theory of CBO provides us with a convergence result showing that the invariant state of the dynamics is a consensus near the global minimizer of the cost function. We adapted the dynamics to obtain a scheme that provides a uniform approximation of the Pareto front all at once. So far, we only considered the weighed means of the swarms. Now, we want to use the additional information provided by the individuals. In fact, for our purpose it is better if the swarms do not collapse but cluster around the weighted mean, in order to obtain a local

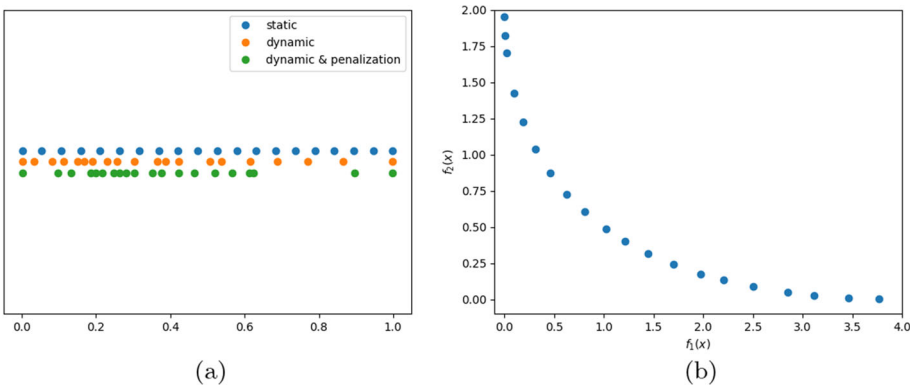

Fig. 4 Influence of penalization on the convex toy problem discussed in Sect. 3.2.1

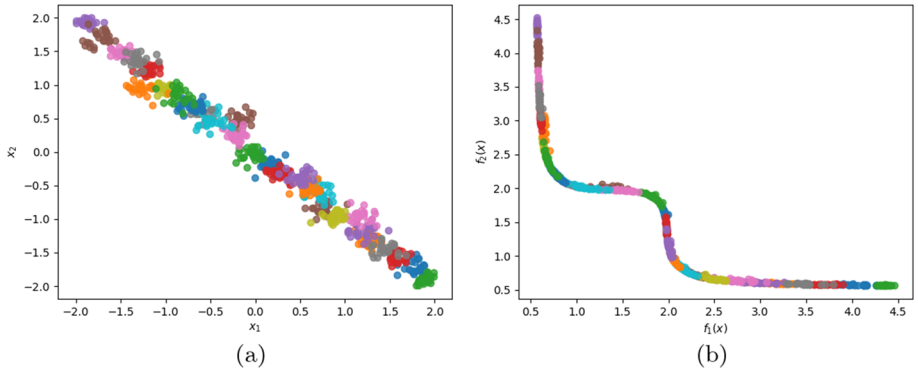

Fig. 5 Approximation of the Pareto front using the information of weighted means and all individuals. Individuals of the same swarm are plotted in the same color

approximation of the Pareto front. We achieve this by implementing an idea from Consensus-based sampling (CBS) [6]. Technically speaking, we replace the anisotropic diffusion factor in Algorithm 2 by a factor that counteracts collapse of the swarm, see Algorithm 3.

Instead of the approximation consisting of the weighted means shown in Figs. 2 and 3 we obtain now an approximation based on the weighted means and additional information of all individuals as shown in Fig. 5. The different colors of individuals illustrate the allocation to the respective swarms, plotted in the decision space (Fig. 5a) and in the objective space (Fig. 5b), respectively.

Remark 10 Let us summarize the main ideas: in contrast to many evolutionary multiobjective optimization algorithms (EMO), like e.g., NSGA2 [1], the proposed algorithms are not based on dominance tests of the swarms and/or particles. Moreover, different from common decomposition-based EMO methods, like, e.g., MOEA/D [32] and the reference point approach of NSGA3 [10, 19], the adaptive weights automatically adjust the scalarization while exploring the objective functions. Finally, the diffusion term inspired by the sampling algorithm yields non-collapsing swarms that provide local information about the Pareto front near the diversely distributed weighted means.

5 Algorithms and parameter settings

In the following we provide the implementation details such as discretization, handling of box constraints and the parameter values for the simulations underlying the illustrations above.

5.1 Discretization

The stochastic differential equations modeling the multi-swarm dynamics are solved with the Euler–Maruyama scheme [20]. In each iteration particles are projected to the feasible set $\mathcal{X}$ (if necessary). Since we study only box-shaped feasible sets, the projection can be applied component-wise. A pseudo code for the simple multi-swarm CBO with fixed scalarization weights is given in Algorithm 1.

Algorithm 1: MSCBO with fixed weights

Data: $K \in \mathbb{N}$, $N_k \in \mathbb{N}$, initial positions $\{X_0^{k,j}\}$ for $k = 1, \dots, K$ and $j = 1, \dots, N_k$, scalarization weights $\lambda^k \in \Lambda_0$, time step $\tau > 0$, diffusion coefficient $\sigma > 0$, terminal time $T > \tau$, initial time $t = 0$ and independent standard-normal distributed W_t^k, weight parameter $\alpha > 0$

Result: approximation of the Pareto front $(f_{\lambda^k}(X_T^{k,j}))_{k,j}$ and efficient set $(X_T^{k,j})_{k,j}$

while $t < T$ **do**

  for $k \leftarrow 1$ **to** K **do**

$$v_t^k \leftarrow \frac{\sum_{j=1}^{N_k} X_t^{k,j} \, \mathrm{e}^{-\alpha f_{\lambda^k}(X_t^{k,j})}}{\sum_{j=1}^{N_k} \mathrm{e}^{-\alpha f_{\lambda^k}(X_t^{k,j})}}$$

$$W_t^k \sim N(0, I)$$

$$Y_{t+\tau}^k \leftarrow X_t^k - \tau\,(X_t^k - v_t^k) + \sigma\,\sqrt{\tau}\,\operatorname{diag}(X_t^k - v_t^k)\,W_t^k$$

$$X_{t+\tau}^k \leftarrow \operatorname*{argmin}_{x \in \mathcal{X}} \|x - Y_t^k\|_2$$

  $t \leftarrow t + \tau$

The dynamic weight adaption proposed in Sect. 3.2 couples the dynamics of the different swarms. We proceed swarm by swarm and update the scalarization weight in each iteration explicitly. This leads to the pseudo code in Algorithm 2. The third modeling step implements the penalization strategy. Note that this affects only the cost function, hence leading to the adapted weighted mean (c.f. (13) above)

$$v_t^k \leftarrow \frac{\sum_{j=1}^{N_k} X_t^{k,j} \, \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,j})} \, \mathrm{e}^{-\beta f_{\mathrm{uni}}(X_t^{k,j}, v)}}{\sum_{j=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,j})} \, \mathrm{e}^{-\beta f_{\mathrm{uni}}(X_t^{k,j}, v)}}.$$

To avoid collapsing swarms we replace the diffusion term with the sampling noise leading to

$$Y_{t+\tau}^k \leftarrow X_t^k - \tau(X_t^k - v_t^k) + \sigma\sqrt{\tau |X_t^k - v_t^k|}\, W_t^k.$$

The pseudo code with all modeling features is summarized in Algorithm 3.

Algorithm 2: MSCBO with adaptive weights

Data: $K \in \mathbb{N}$, $N_k \in \mathbb{N}$, initial positions $\{X_0^{k,j}\}$ for $k = 1, \ldots, K$ and $j = 1, \ldots, N_k$, scalarization weights $\mu^k = \ln(\lambda^k)$, time step $\tau > 0$, diffusion coefficient $\sigma > 0$, terminal time $T > \tau$, initial time $t = 0$ and independent standard-normal distributed W_t^k, weight parameters $\alpha > 0$, $\beta \geq 0$

Result: approximation of the Pareto front $(f_{\mu^k}(X_T^{k,j}))_{k,j}$ and efficient set $(X_T^{k,j})_{k,j}$

while $t < T$ **do**

  for $k \leftarrow 1$ **to** K **do**

$$v_t^k \leftarrow \frac{\sum_{j=1}^{N_k} X_t^{k,j} \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,j})}}{\sum_{j=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,j})}} \text{ or } \frac{\sum_{j=1}^{N_k} X_t^{k,j} \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,j})} \mathrm{e}^{-\beta f_{\mathrm{uni}}(X_t^{k,j},v)}}{\sum_{j=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,j})} \mathrm{e}^{-\beta f_{\mathrm{uni}}(X_t^{k,j},v)}}$$

$$W_t^k \sim N(0, I)$$

$$Y_{t+\tau}^k \leftarrow X_t^k - \tau(X_t^k - v_t^k) + \sigma\sqrt{\tau}\,\mathrm{diag}(X_t^k - v_t^k) W_t^k$$

$$X_{t+\tau}^k \leftarrow \operatorname*{argmin}_{x \in \mathcal{X}} \|x - Y_t^k\|_2$$

$$\mu_{t+\tau}^k \leftarrow \mu_k^k - \tfrac{\tau}{K} \sum_{\ell=1, \ell \neq k}^{K} \mathcal{K}(X_t^k, X_t^\ell, \mu^k, \mu^\ell)$$

  $t \leftarrow t + \tau$

Algorithm 3: MSCBO with adaptive weights, penalization and sampling noise

Data: $K \in \mathbb{N}$, $N_k \in \mathbb{N}$, initial positions $\{X_0^{k,j}\}$ for $k = 1, \ldots, K$ and $j = 1, \ldots, N_k$, scalarization weights $\mu^k = \ln(\lambda^k)$, time step $\tau > 0$, diffusion coefficient $\sigma > 0$, terminal time $T > \tau$, initial time $t = 0$ and independent standard-normal distributed W_t^k, weight parameters $\alpha > 0$, $\beta \geq 0$

Result: approximation of the Pareto front $(f_{\mu^k}(X_T^{k,j}))_{k,j}$ and efficient set $(X_T^{k,j})_{k,j}$

while $t < T$ **do**

  for $k \leftarrow 1$ **to** K **do**

$$v_t^k \leftarrow \frac{\sum_{j=1}^{N_k} X_t^{k,j} \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,j})} \mathrm{e}^{-\beta f_{\mathrm{uni}}(X_t^{k,j},v)}}{\sum_{j=1}^{N_k} \mathrm{e}^{-\alpha f_{\mu^k}(X_t^{k,j})} \mathrm{e}^{-\beta f_{\mathrm{uni}}(X_t^{k,j},v)}}$$

$$W_t^k \sim N(0, I)$$

$$Y_{t+\tau}^k \leftarrow X_t^k - \tau(X_t^k - v_t^k) + \sigma\sqrt{\tau|X_t^k - v_t^k|}\, W_t^k$$

$$X_{t+\tau}^k \leftarrow \operatorname*{argmin}_{x \in \mathcal{X}} \|x - Y_t^k\|_2$$

$$\mu_{t+\tau}^k \leftarrow \mu_k^k - \tfrac{\tau}{K} \sum_{\ell=1, \ell \neq k}^{K} \mathcal{K}(X_t^k, X_t^\ell, \mu^k, \mu^\ell)$$

  $t \leftarrow t + \tau$

5.2 Test problems

The illustrative example and also the numerical comparisons in the following sections use the following test problems.

Schaffer1 [26] $f(x) = \left((x-2)^2,\ 0.5\,x^2\right)$, biobjective test problem, convex Pareto front, $\mathcal{X} = [0, 2]$, $\mathcal{Y} \subset [0, 4] \times [0, 2]$.

Dent [31] $f(x) = (f_1(x_1, x_2),\ f_2(x_1, x_2))$ with

$$f_1(x_1, x_2) = \frac{1}{2}\left(\sqrt{1 + (x_1+x_2)^2} + \sqrt{1 + (x_1-x_2)^2} + x_1 - x_2\right) + 0.85\,\mathrm{e}^{-(x_1-x_2)^2},$$

$$f_2(x_1, x_2) = \frac{1}{2}\left(\sqrt{1 + (x_1+x_2)^2} + \sqrt{1 + (x_1-x_2)^2} - x_1 + x_2\right) + 0.85\,\mathrm{e}^{-(x_1-x_2)^2}.$$

Table 1 Algorithmic parameters used in the numerical tests

$\tau = 0.1$	$T = 5$	$\lambda = 1$	$\sigma = 0.1$
$R = 0.001$	$r = 0.01$	$R_f = 0.0001$	$r_f = 1$
$R_c = 1$	$r_c = 0.1$	$\alpha = 100$	$\beta = 0$ or 10

Biobjective test problem with a dent, $\mathcal{X} = [-2, 2]^2$, $\mathcal{Y} \subset [0, 5]^2$.
Schaffer2 $f(x) = \big(f_1(x),\ f_2(x)\big)$ with

$$f_1(x) = \begin{cases} -x, & \text{if } x \le 1, \\ x - 2, & \text{if } 1 \le x \le 3, \\ 4 - x, & \text{if } 3 \le x \le 4, \\ x - 4, & \text{if } 4 \le x \end{cases}, \qquad f_2(x) = (x - 5)^2.$$

Biobjective test problem, discontinuous Pareto front, $\mathcal{X} = [-5, 10]$, $\mathcal{Y} \subset [0, 1] \times [0, 16]$.
Three [2] $f(x_1, x_2) = \big(f_1(x_1, x_2),\ f_2(x_1, x_2),\ f_3(x_1, x_2)\big)$ with

$$\begin{aligned} f_1(x_1, x_2) &= 2(x_1 - 1)^2 + 2(x_1 - 1)(x_2 - 1) + 4(x_2 - 1)^2, \\ f_2(x_1, x_2) &= (x_1 - 2)^2 + 4(x_1 - 2)(x_2 - 3) + 8(x_2 - 3)^2, \\ f_3(x_1, x_2) &= 4x_1^2 + 2x_1x_2 + x_2^2. \end{aligned}$$

Three objectives, convex Pareto front, $\mathcal{X} = [-0.5, 3.5]$, $\mathcal{Y} \subset [0, 25] \times [0, 80] \times [0, 50]$.

Parameters for illustrative examples

In the previous sections we illustrated the modeling ideas with the help of a convex optimization problem (Schaffer1) in Figs. 1 and 4, and with a non-convex problem (dent) in Figs. 2, 3 and 5. Initial data is sampled independently from the uniform distribution on $\mathcal{X}$. All potentials are chosen to be purely repulsive, therefore A and A_f are set to zero, for completeness we set a and a_f to 1. In a post-processing we eliminate dominated individuals, we therefore use a numerical offset $\epsilon_{\text{dom}} = 10^{-5}$. For the test problems with $\mathcal{Y} \subset \mathbb{R}^2$ we initialize the scalarization weights equidistantly on $[0.001, 0.999]^2$.

For the three-objective test case, initial scalarization weights are randomly generated with Algorithm 2 in [23]. For simplicity, we set $N_k = \bar{N}$ for all $k = 1, \dots K$ for all simulations. For the illustrations in Figs. 1 and 5 we use $K = 20$ swarms with $\bar{N} = 50$ individuals each. For all other biobjective test problems we set $K = 30$ and $\bar{N} = 20$. For the test problem with three objectives we set $K = 50$ and $\bar{N} = 20$. All other parameters are reported in Table 1.

6 Comparison to other population-based algorithms

This section is devoted to a comparison to other multi-objective optimization methods. In particular, we address the well-known Non-dominated Sorting Genetic Algorithm (NSGA2) [11], its variant NSGA3 [10, 19] that uses reference points to support diversification, and the recently introduced single swarm Consensus-based optimization approach for multi-objective problems [4].

6.1 Computational effort

The computational complexity of population-based multi-objective optimization algorithms lies on the one hand in the evaluation of the objective functions and on the other hand in computing the dominance relations between the generations or agents. The improvement from NSGA [27] to NSGA2 [11] (and similarly to NSGA3 [10]) is mainly obtained by an efficient computation of the dominance relationship of the best N individuals. Note that we have refrained from using the explicit multiobjective dominance information in the algorithms proposed and therefore save the complexity of $\mathcal{O}(pN^2)$ which is governed by the non-dominated sorting algorithms (see e.g. [21]).

The single swarm multi-objective algorithm (SSCBO) proposed in [4] is stabilized by a greedy approach which ensures that individuals only move if the attempted move leads to a better position. Computationally this is cheaper then computing the dominance relationship in NSGA2, still it requires N comparisons in each iteration. Our multi-swarm CBO approach MSCBO does not require these comparisons and therefore saves $\mathcal{O}(pN)$ as compared to SSCBO.

6.2 Numerical results

To obtain comparable results, we chose the parameters in the following simulations such that all algorithms use the same number of function evaluations. The time steps of SSCBO and MSCBO are chosen equally. We set the diffusion parameter of SSCBO to 10 as reported in [4] and activate the greedy strategy. Moreover, we consider the well-known methods NSGA2 and NSGA3 in order to compare the performance to well-established methods. Moreover, we make sure that NSGA2 and NSGA3 use the same number of iterations. For the following tests this will be $T/\tau = 50$. The reference directions for NSGA3 are obtained with the energy method from `pymoo`.

The comparison is based on different performance indicators: generational distance (GD), inverted generational distance (IGD) and hypervolume (HV) from the `pymoo` package are used for the comparison (see, e.g., [25, 33] for a detailed description of performance indicators). GD gives us an indication on the precision of the approximation of the Pareto front. IGD detects clustering or sparse regions along the front and HV gives us a flavor of how good the algorithms perform compared to the non-dominated reference set. The reference points for the hypervolume indicator are the upper bounds of the Cartesian intervals containing $\mathcal{Y}$ given in Sect. 5.2. For GD and IGD we need a reference data set. Our reference data set is based on the NSGA2 algorithm as this gives the largest hypervolume. However, we check if points of this data set are dominated by points obtained with the other methods. If so we replace the point in the reference data set by the nondominated point. These choices allow us to analyze the performance without knowledge about the true Pareto front of the problems. As mentioned above we do not compute dominance relations in each iteration, but we evaluate the number of non-dominated individuals (NI) at the end of the simulations.

6.2.1 Biobjective tests

We begin our tests with the convex test problem Schaffer1 with the results reported in Table 2. All individuals of SSCBO, MSCBO and NSGA2 algorithms are non-dominated, while 10 individuals of NSGA3 are dominated. The dominated hypervolumes of the three approximations are very similar. The GD values of MSCBO, SSCBO and NSGA3 are very

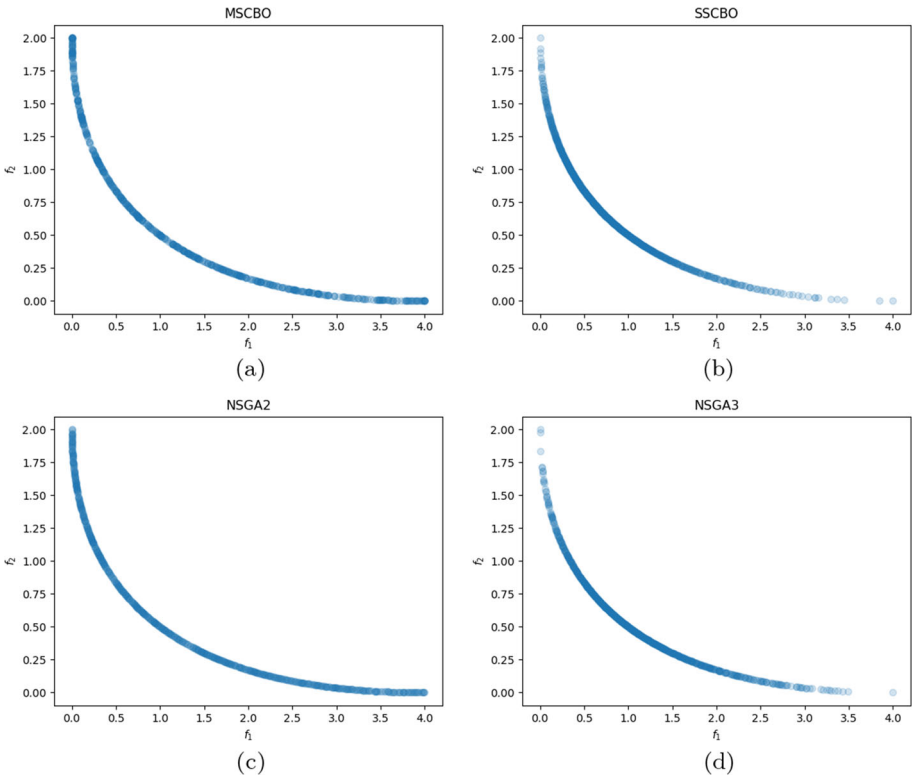

Fig. 6 Approximations of the Pareto fronts for Schaffer1: **a** MSCBO, **b** SSCBO, **c** NSGA2 and **d** NSGA3

Table 2 Simulation results for Schaffer1

$K = 30, \bar{N} = 20$	GD	IGD	HV (%)	NI
MSCBO	0.0026	0.0045	99.95	630
SSCBO	0.0023	0.0107	99.94	630
NSGA2	0.0	0.0	100	630
NSGA3	0.0024	0.0141	99.92	620

close as well, indicating that the approximations are precise. However, we see a difference in the IGD values. Further simulations and Fig. 6 suggest that the approximation points of SSCBO and NSGA3 tend to accumulate in the region of the knee of the Pareto front, therefore the distances of IGD at the tails of the front are higher leading to a higher overall value. The IGD value of NSGA3 lies in the range of SSCBO.

The second test yields approximations of the Pareto front of the test problem Dent. It turns out that many of the individuals of MSCBO are dominated, also some individuals of SSCBO and few individuals of NSGA3 are dominated. The GD values are very small for all methods. The IGD is best for SSCBO and NSGA3 lies in the middle of SSCBO and MSCBO. Further investigations suggest that the approximation of the dent region of the Pareto front is sparser for MSCBO which leads to the higher values of IGD, see Fig. 7. In contrast, we see that NSGA3 stresses the dent region. Although MSCBO and SSCBO terminate with fewer

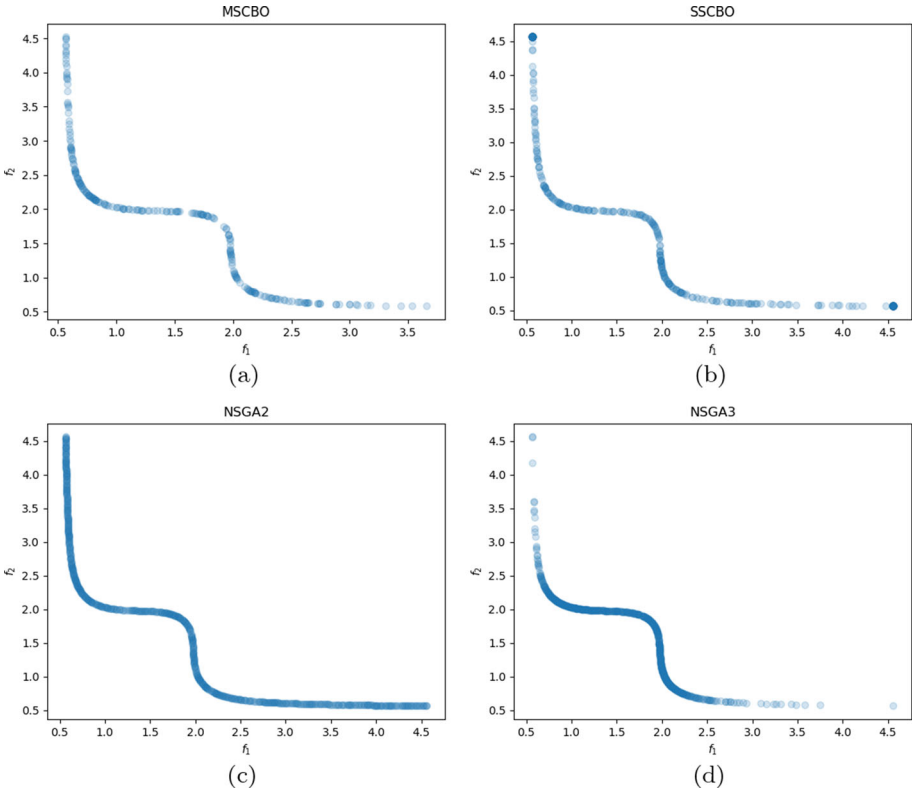

Fig. 7 Approximation of the Pareto fronts for problem Dent: **a** MSCBO, **b** SSCBO, **c** NSGA2 and **d** NSGA3

Table 3 Simulation results for problem Dent

$K = 30, \bar{N} = 20$	GD	IGD	HV (%)	NI
MSCBO	0.0037	0.0441	99.78	251
SSCBO	0.0027	0.0174	99.81	375
NSGA2	0.0003	0.0002	99.99	630
NSGA3	0.0033	0.0430	99.94	577

non-dominated individuals, the dominated hypervolumes of all four methods are in good agreement (Table 3).

The third test is concerned with the Schaffer2 problem which has a discontinuous Pareto front. The precision of the approximation of the front indicated by the GD value is again very good. The IGD values differ and simulations showed that the upper part of the Pareto front is not very well-approximated by SSCBO leading to this IGD behavior. This is visualized in Fig. 8 as well. Also in the values of the hypervolume indicator we see a difference between the SSCBO approximation. Although many of the MSCBO and NSGA3 individuals are dominated, the dominated hypervolume and the IGD values are promising. In terms of the GD, NSGA3 is similar to the other methods, in terms of IGD and HV it is very close to NSGA2 (Table 4).

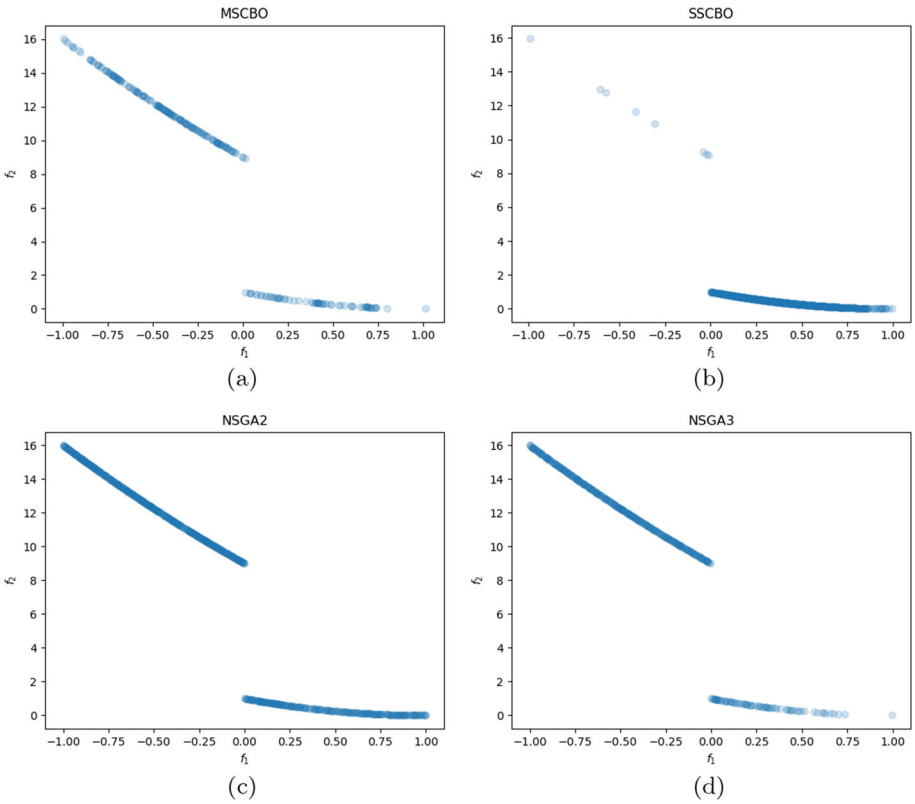

Fig. 8 Approximation of the Pareto fronts for Schaffer2: **a** MSCBO, **b** SSCBO, **c** NSGA2 and **d** NSGA3

Table 4 Simulation results for Schaffer2

$K = 30, \bar{N} = 20$	GD	IGD	HV (%)	NI
MSCBO	0.0041	0.0253	99.56	218
SSCBO	0.0024	0.2982	95.12	615
NSGA2	0.0	0.0	100	630
NSGA3	0.0049	0.0140	99.83	369

6.2.2 Three-objective test

The results of the benchmark with three objectives are shown in Table 5 and Fig. 9. We only compare MSCBO, NSGA2 and NSGA3 since the SSCBO implementation used above was only written for the biobjective case. The results show that many of the NSGA3 individuals are dominated, but the performance with respect to the indicators are in the range of the MSCBO. The results also reflect the different algorithmic strategies of the three methods: While the diversification strategy implemented in the NSGA2 method leads to a dense representation at the tails of the Pareto front (Fig. 9b), the central and structured location of the initial reference points in NSGA3 is reflected by a more structured and more centralized approximation (Fig. 9c). MSCBO compromises between these two with a slight focus on the central parts of the Pareto front while still covering all the tails reasonably well (Fig. 9a). Note that we

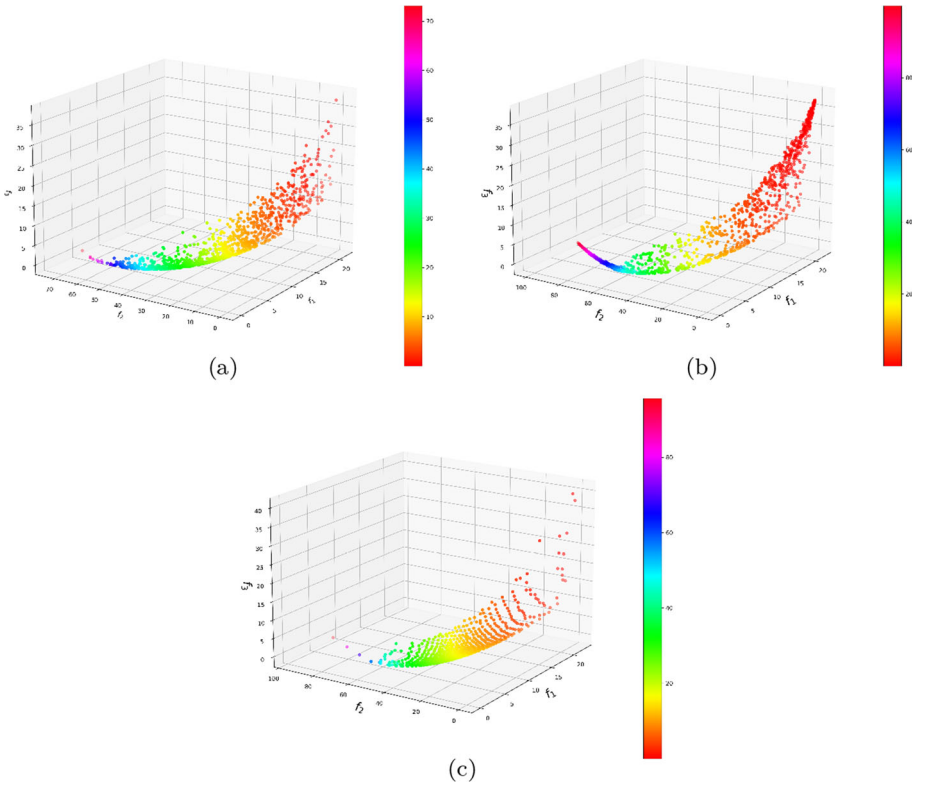

Fig. 9 Approximation of the Pareto fronts for three-objectives: **a** MSCBO, **b** NSGA2 and **c** NSGA3 with colormap wrt f_2

Table 5 Simulation results for three-objective problem

$K = 50, \bar{N} = 20$	GD	IGD	HV (%)	NI
MSCBO	0.547	2.191	99.71	1006
NSGA2	0.007	0.005	99.99	1050
NSGA3	0.602	1.474	99.70	525

do expect the magnitude of the GD and IGD indicators to be higher due to the additional objective function.

7 Conclusion and outlook

We propose a versatile population-based algorithm for the approximation of the Pareto front and of the efficient set for multi-objective optimization problems that is based on the Consensus-based optimization or sampling method, respectively. In the case of fixed scalarization weights, many analytical results of CBO can be easily generalized to this setting. For adaptive weights we proved that the methods yields a diverse approximation of the front. Moreover, with prevention of collapsing swarms, we retain local information of the

Pareto front close to the swarm means. The method is computationally cheap as no dominance relations need to be computed. The numerical results are promising and motivate for further investigations and applications of the method. In fact, the algorithm is competitive with widely-used NSGA2 and NSGA3 and the recently proposed single-swarm CBO on the considered set of classical multi-objective test instances.

Future research should be twofold: On one hand, the suggested methods can potentially be improved by incorporating additional criteria and by fine tuning the algorithmic parameters. On the other hand, the simultaneous evaluation of the decision space *and* the objective space can be further exploited, e.g., in order to incorporate decision support tools that analyze the effects of objective improvements on the properties of the corresponding solutions.

Acknowledgements We thank the authors of [4] for providing their code of the single swarm multi-objective CBO. This facilitated the comparison of the methods significantly. We acknowledge the open source code of NSGA2, NSGA3, and the performance indicators in the project `pymoo` [1].

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The data sets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Blank, J., Deb, K.: Pymoo: multi-objective optimization in python. IEEE Access **8**, 89497–89509 (2020)
2. Bolten, M., Doganay, O.T., Gottschalk, H., Klamroth, K.: Tracing locally Pareto optimal points by numerical integration. SIAM J. Control. Optim. **59**(5), 3302–3328 (2021). https://doi.org/10.1137/20M1341106
3. Borghi, G.: Repulsion dynamics for uniform Pareto front approximation in multi-objective optimization problems (2022). https://doi.org/10.48550/ARXIV.2211.03378
4. Borghi, G., Herty, M., Pareschi, L.: A consensus-based algorithm for multi-objective optimization and its mean-field description. arXiv:2203.16384 (2022)
5. Burger, M., Pinnau, R., Totzeck, C., Tse, O., Roth, A.: Instantaneous control of interacting particle systems in the mean-field limit. J. Comput. Phys. **405**, 109181 (2020)
6. Carrillo, J., Hoffmann, F., Stuart, A.M., Vaes, U.: Consensus based sampling. arXiv:2106.02519 (2021)
7. Carrillo, J., Jin, S., Li, L., Zhu, Y.: A consensus-based global optimization method for high dimensional machine learning problems. ESAIM COCV **27**, S5 (2021)
8. Choi, Y.P., Carrillo, J.A., Totzeck, C., Tse, O.: An analytical framework for consensus-based global optimization method. Math. Models Methods Appl. Sci. **28**(6), 1037–1066 (2018)
9. Das, I., Dennis, J.E.: A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. Struct. Optim. **14**, 63–69 (1997)
10. Deb, K., Jain, H.: An evolutionary many-objective optimization algorithm using reference-point-based non-dominated sorting approach, part I: solving problems with box constraints. IEEE Trans. Evol. Comput. **18**, 577–601 (2014). https://doi.org/10.1109/TEVC.2013.2281535
11. Deb, K., Pratap, A., Agarwal, A., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: Nsga-ii. IEEE Trans. Evol. Comput. **6**(2), 182 (2002)
12. D'Orsogna, M.R., Chuang, Y.L., Bertozzi, A.L., Chayes, L.S.: Self-propelled particles with soft–core interactions: patterns, stability, and collapse. Phys. Rev. Lett. **96**, 104302 (2006)
13. Durrett, R.: Stochastic Calculus: A Practical Introduction, vol. 6. CRC Press, Boca Raton (1996)
14. Ehrgott, M.: Multicriteria Optimization. Springer, Berlin (2005). https://doi.org/10.1007/3-540-27659-9

15. Fornasier, M., Klock, T., Riedl, K.: Consensus-based optimization methods converge globally. arXiv:2103.15130 (2021)
16. Gass, S., Saaty, T.: The computational algorithm for the parametric objective function. Naval Res. Logist. Q. **2**, 39 (1955)
17. Hillermeier, C.: Nonlinear Multiobjective Optimization. Birkhäuser, Basel (2001). https://doi.org/10.1007/978-3-0348-8280-4
18. Huang, H., Qiu, J.: On the mean-field limit for the consensus-based optimization. Math. Methods Appl. Sci. **45**(12), 7814–7831 (2022). https://doi.org/10.1002/mma.8279
19. Jain, H., Deb, K.: An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part II: handling constraints and extending to an adaptive approach. IEEE Trans. Evolut. Comput. **18**, 602–622 (2014)
20. Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, Berlin (1992)
21. Lang, B.: Space-partitioned ND-trees for the dynamic nondominance problem. IEEE Trans. Evolut. Comput. (2022). https://doi.org/10.1109/tevc.2022.3145631
22. Miettinen, K.: Nonlinear Multiobjective Optimization. Kluwer, Boston (1999)
23. Onn, S., Weissman, I.: Generating uniform random vectors over a simplex with implications to the volume of a certain polytope and to multivariate extremes. Ann. Oper. Res. **189**, 331 (2009)
24. 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)
25. Riquelme, N., Lucken, C.V., Baran, B.: Performance metrics in multi-objective optimization. In: 2015 Latin American Computing Conference (CLEI), IEEE (2015). https://doi.org/10.1109/clei.2015.7360024
26. Schaffer, J.D.: Multiple objective optimization with vector evaluated genetic algorithms. In: Grefensette, G., Erlbraum, J.L. (eds.), Proceedings of the First International Conference on Genetic Algorithms (1984)
27. Srinivas, N., Deb, K.: Multiobjective optimization using nondominated sorting in genetic algorithms. Evol. Comput. **2**(3), 221–248 (1994)
28. Totzeck, C.: Trends in Consensus-Based Optimization, pp. 201–226. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-93302-9_6
29. Totzeck, C., Wolfram, M.T.: Consensus-based global optimization with personal best. Math. Biosci. Eng. **17**(5), 1 (2020)
30. Toure, C., Auger, A., Brockhoff, D., Hansen, N.: On bi-objective convex-quadratic problems. In: Deb, K., Goodman, E., Coello Coello, C.A., Klamroth, K., Miettinen, K., Mostaghim, S., Reed, P. (eds.), Evolutionary Multi-criterion Optimization, pp. 3–14. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12598-1_1
31. Witting, K.: Numerical Algorithms for the Treatment of Parametric Multiobjective Optimization Problems and Applications. Dissertation, Universität Paderborn (2012)
32. Zhang, Q., Li, H.: MOEA/D: a multiobjective evolutionary algorithm based on decomposition. IEEE Trans. Evolut. Comput. **11**, 712–731 (2007). https://doi.org/10.1109/TEVC.2007.892759
33. Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Grunert da Fonseca, V.: Performance assessment of multiobjective optimizers: an analysis and review. IEEE Trans. Evolut. Comput. **7**(2), 117–132 (2003)