Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 17(5): 6026–6044.
DOI: 10.3934/mbe.2020320
Received: 15 May 2020
Accepted: 17 August 2020
Published: 11 September 2020

Research article

Consensus-based global optimization with personal best

Claudia Totzeck[1] **and Marie-Therese Wolfram**[2,3,*]

[1] School of Business Informatics and Mathematics, University of Mannheim, B 6, 68159 Mannheim, Germany

[2] Mathematics Institute, University of Warwick, Gibbet Hill Road, CV47AL Coventry, UK

[3] RICAM, Altenbergerstr. 69, 4040 Linz, Austria

* **Correspondence:** Email: m.wolfram@warwick.ac.uk; Tel: +44 (0)2476150041.

Abstract: In this paper we propose a variant of a consensus-based global optimization (CBO) method that uses personal best information in order to compute the global minimum of a non-convex, locally Lipschitz continuous function. The proposed approach is motivated by the original particle swarming algorithms, in which particles adjust their position with respect to the personal best, the current global best, and some additive noise. The personal best information along an individual trajectory is included with the help of a weighted mean. This weighted mean can be computed very efficiently due to its accumulative structure. It enters the dynamics via an additional drift term. We illustrate the performance with a toy example, analyze the respective memory-dependent stochastic system and compare the performance with the original CBO with component-wise noise for several benchmark problems. The proposed method has a higher success rate for computational experiments with a small particle number and where the initial particle distribution is disadvantageous with respect to the global minimum.

Keywords: global optimization; interacting particle systems; consensus formation; stochastic differential equations; personal best information; non-convex optimization

1. Introduction

Interacting particle systems play an important role in many applications in science - on the one hand as a modeling framework for social and biological systems, on the other as a tool for computational algorithms used in data science. In the latter case the collective behavior of interacting particle systems is used to solve high-dimensional problems, often resulting from non-convex optimization tasks in data science. Well known algorithms include particle swarm optimization (PSO) [1], ant colony optimization [2] or evolutionary [3] and genetic algorithms [4].

PSO was first introduced in [1] and has been successfully used in engineering applications [5]. Each particle in a PSO algorithm adjusts its position due to information of the global best, personal best and a noise term that allows for exploration of its neighborhood. Consensus-based optimization (CBO) [6,7] combines the idea of swarm intelligence with consensus formation techniques [8–10] to obtain a global optimization algorithm for non-convex high-dimensional problems. On the one hand particles explore the state space via an amplitude modulated random walk. On the other a drift term convects them towards the weighted global best. The method was first introduced in [6] and analyzed at the mean-field level in [7]. Recent developments of CBO include component-wise diffusion and utilize random mini-batch ideas to reduce the computational cost of calculating the weighted average [11]. Other contributions investigate a CBO dynamic that is restricted to the sphere [12,13]. Also, convergence and error estimates for time-discrete consensus-based optimization algorithms have been discussed [14]. CBO-type systems are related to large interacting particle systems, in which the dynamics are driven by weighted average quantities, see [15–17].

The model proposed in the work is based on the component-wise diffusion variant introduced in [11] and combines it with personal best information. This adjustment is motivated by the original work on PSO by Eberhart and Kennedy [1], where the particles move towards a (stochastic) linear combination of their personal best or the common global best position. The new information leads to an additional drift term in the dynamics. We investigate two types of memory effects - either using a weighted personal best over time or the personal best value in the past. The latter corresponds to record processes, see [18] for an overview. The former is used in the presented analysis and approximates the personal best of each particle. We expect by arguments similar to the Laplace principle that the weighted mean converges towards the personal best.

The proposed stochastic dynamics with weighted personal best fall into the class of stochastic functional differential equations. These equations are in general non-Markovian and their mean-field limit has been investigated in special cases only. For example, Gadat and Panloup [19] investigated a non-Markovian process with memory, which corresponds to the weighted average of the drift all along the particle's trajectory. This memory term is of a special form allowing them to rewrite the system as a 2-dimensional non-homogeneous Markovian dynamical system. Moreover, they exploit this special structure to analyze the existence and long time behavior of solutions as well as the mean-field limit. The strategy of increasing the dimension to get around the non-Markovian nature goes back to the Mori-Zwanzig formalism, see [20]. This strategy was recently adapted for non-Markovian interacting dynamics by Duong and Pavliotis in [21]. Another interesting work by Kuntzmann, see [22], investigates the ergodic behavior of self-interacting diffusions depending on the empirical mean of the process. The proposed generalization of CBO with weighted personal best does not fall into this category, hence the derivation and analysis of the respective mean-field dynamics, which often give useful insights into the dynamics, is to the best of the authors' knowledge open. This applies as well for personal best, where the update of the best function value corresponds to a record process. Hence, we focus on the well-posedness of the stochastic system as well as a detailed computational investigation of the dynamics.

This paper is organized as follows: we introduce the particle dynamics with (weighted) personal best in Section 2 and illustrate its dynamics with first toy examples. Section 3 discusses well-posedness and existence of solutions to the SDE model with weighted personal best. Section 4 presents extensive computational experiments of various benchmark optimization problems.

2. Consensus based optimization with personal best

In this section we discuss how personal best information can be included in consensus based optimization algorithms as proposed by Carrillo and co-workers in [6, 7, 11]. We start by introducing the notation before continuing with the modeling.

2.1. Notation

We refer the euclidean norm by $|x| = (x_1^2 + \cdots + x_d^2)^{1/2}$ for $x \in \mathbb{R}^d$ and $|Y| = (\sum_{i,j=1}^{dN} Y_{ij}^2)^{1/2}$ for matrices $Y \in \mathbb{R}^{dN\times dN}$. The set of natural numbers without 0 is denoted by $\mathbb{N}^* = 1, 2, 3, \ldots$ and the half-line $[0, \infty)$ by $\mathbb{R}^+$. A vector valued function or vector $x \in \mathbb{R}^{dN}$ is assumed to be of the form $x = (x^1, \ldots, x^N)$ with $x^i \in \mathbb{R}^d$. When discussing the stochastic systems we follow the notation of [23]: $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{t\geq 0})$ corresponds to the stochastic basis with sample space Ω, filtration $\mathcal{F}$ and probability function $\mathbb{P}$. Moreover, $S_d^p[0, T]$ is the space of (equivalence classes of) $\mathcal{P}$ -measurable continuous stochastic processes $X\colon \Omega \times [0, T] \to \mathbb{R}^{dN}$ such that

$$\mathbb{E} \sup_{t\in[0,T]} |X_t|^p < +\infty \quad \text{if } p > 0.$$

Two processes X, Y are called equivalent if $(X_t = Y_t \forall t \in [0, T])$ $\mathbb{P}$-almost surely ($\mathbb{P}$-a.s.). Furthermore, S_d^p is the space of (equivalence classes of) $\mathcal{P}$-measurable continuous stochastic processes $X\colon \Omega\times\mathbb{R}_+ \to \mathbb{R}^d$ such that for all $T > 0$ the restriction $X_{|[0,T]}$ of X to $[0, T]$ belongs to $S_d^p[0, T]$. Analogously, we define $\Lambda_d^p(0, T)$ as the space of (equivalent classes) of $\mathcal{P}$-measurable processes $X\colon [0, T] \to \mathbb{R}^d$ such that

$$\int_0^T |X_t|^2 dt < +\infty \quad \mathbb{P}\text{-a.s.}\omega \in \Omega \text{ if } p = 0 \quad \text{and} \quad \mathbb{E}\left(\int_0^T |X_t|^2 dt\right)^{p/2} < +\infty \quad \text{if } p > 0.$$

We refer to Λ_d^p as the space of (equivalence classes of) $\mathcal{P}$-measurable continuous stochastic processes $X\colon \Omega \times (0, +\infty) \to \mathbb{R}^d$ for which for all $T > 0$ the restriction $X_{|[0,T]}$ of X to $[0, T]$ belongs to $\Lambda_d^p(0, T)$. Moreover, for any $\phi \in C(\mathbb{R}_+, \mathbb{R}^{dN})$ we define

$$\|\phi\|_t := \sup_{0\leq s\leq t} |\phi(s)| = \sup_{0\leq s\leq t} \left(\phi_1(s)^2 + \cdots + \phi_{dN}(s)^2\right)^{1/2}.$$

2.2. The model

We wish to approximate the global minimum

$$\min_{x\in\mathbb{R}^d} f(x), \tag{2.1}$$

of a given non-negative, continuous objective function $f : \mathbb{R}^d \to \mathbb{R}$. In doing so we consider $N \in \mathbb{N}$ particles and denote the position of the i-th particle at time t by $X_t^i := X^i(t) \in \mathbb{R}^d$, $i = 1, \ldots N$. Note that we use $X_t = X(t) = (X^1(t), \ldots X^N(t)) \in \mathbb{R}^{dN}$, when referring to the positions of all particles at time t. In CBO particles compare their current function value with a weighted mean value based on the current information of the whole system. A particle moves towards the position of the weighted mean, if the function value of the weighted mean is lower. Following the ideas of [6, 7], we use the weighted average

$$v_f(t) = v_f[X_t] = \frac{\sum_{i=1}^N X^i(t) \exp(-\alpha f(X^i(t)))}{\sum_{i=1}^N \exp(-\alpha f(X^i(t)))}, \tag{2.2}$$

with $\alpha > 0$, to approximate the global best, that is, the particle with the lowest function value. Note that even though the weighted average uses only information of the current time step, it is assumed to approximate the global best over time as well, since a particle that is close to the weighed average experiences only small drift and diffusion. The parameter α scales the influence of local and global minima in the weighted mean. In fact, for $\alpha = 0$ the weights are independent of the function values, and all particles are weighted equally. For $\alpha > 0$ the particle with the best function value has the largest weight. Moreover, the Laplace principle from large deviations theory [24] assures that $v_f(t)$ converges to the global best, as $\alpha \to \infty$. For more details on the Laplace principle in the CBO context, we refer to [6, 7].

In the original version of PSO, see [25], particles compare their current position with the global best as well as their personal best value up to that time. We propose two different approaches how to include the personal best p_i of the i-th particle. First, we consider the true personal best by setting

$$P_f^i(t) = \underset{Y \in \{X^i(s) \,:\, s \in [0,t]\}}{\arg\min} f(Y). \tag{2.3}$$

Moreover, the personal best can be approximated similarly to the global best, $v_f(t)$, defined in (2.2). Hereby, we use the entire trajectory in the past and refer to this trajectory by $X = (X^1, \dots, X^N)$ with $X^i \in C(\mathbb{R}_+, \mathbb{R}^d)$ for all $i = 1, \dots, N$. Let X_0^i denote the initial position of the i-th particle at time $t = 0$, the weighted mean over time of the i-th particle is defined by

$$p_f^i(t) = \begin{cases} X_0^i, & t = 0, \\ \int_0^t X_s^i \exp(-\beta f(X_s^i))ds \Big/ \int_0^t \exp(-\beta f(X_s^i))ds, & \text{otherwise}, \end{cases} \tag{2.4}$$

with $\beta > 0$. Note that the well-posedness result presented in Section 3 holds for the weighted personal best (2.4) only. Again, by the Laplace principle, we expect that $p_f^i(t) \to P_f^i$ as $\beta \to \infty$.
We recall that particles either move towards the global or personal best state. The respective CBO dynamics for the i-th particle, $i = 1, \dots N$ are then given by the following SDE:

$$dX^i(t) = \left[-\lambda(t, X)(X^i(t) - v_f) - \mu(t, X)(X^i(t) - p_f^i)\right] dt + \sqrt{2}\sigma \mathrm{diag}(X^i(t) - v_f)\, dB_t^i, \tag{2.5}$$

where

$$\lambda(t, X) = \mathrm{H}(f(X^i(t)) - f(v_f))\, \mathrm{H}(f(p_f^i) - f(v_f)),$$
$$\mu(t, X) = \mathrm{H}(f(X^i(t)) - f(p_f^i))\, \mathrm{H}(f(v_f) - f(p_f^i)).$$

The function H corresponds to the Heaviside function and $\sigma > 0$ denotes the standard deviation. System (2.5) is supplemented with the initial condition $X_0^i = \xi_i$, $i = 1, \dots, N$. The drift and diffusion are motivated by the following considerations:

1. If the global best v_f is better than the current position X_t^i and the personal best p_f^i, the particle moves towards the current global best v_f.
2. If the personal best p_f^i is better than the current position X_t^i and the global best v_f, the particle moves towards the personal best p_f^i.
3. If none of the above holds, the particle still explores the function landscape via Brownian motion until it reaches the global best v_f.

Note that the drift coefficients depend on the past of each particle, hence system (2.5) is non-Markovian. The form of the memory does not allow us to use existing results, such as [23] to rewrite the system. Hence the existence and form of the respective mean-field model is, up to the authors' knowledge, not known.

Remark 1. *(1) The CBO version proposed in [11] can be recoverd by setting*

$$\lambda(t, X) \equiv \lambda, \qquad \mu(t, X) \equiv 0. \tag{2.6}$$

(2) Note that the personal best (2.3) *and weighted personal best* (2.4) *can be computed very efficiently due to their accumulative structure; this does not significantly increase the computational cost.*

Throughout this manuscript we will refer to the dynamics defined by (2.5) with (2.6) as CBO, and to (2.5) with (2.3) or (2.4) as personal best (PB) or weighted personal best (wPB), respectively.

2.3. Toy example: CBO vs. PB dynamics

In the following we will illustrate the differences between CBO and (w)PB using a 1D toy objective function f and 3 particles. We consider a double well-type f of the form:

$$f(x) = (x^2 - 1)^2 + 0.01x + 0.5.$$

For this function, shown in Figure 1 the global and local minimum, located at $x = -1.00125$ and $x = 0.998748$ respectively, are very close. In the following we perform 1000 Monte Carlo (MC) simulations with deterministic initial conditions ξ. We count a run as run successful, if the final position of the particles satisfy $|v_f(T) - X^i(T)| < 0.4$ for all $i = 1, \dots, N$. The final time is set to $T = 100$, the time step size $dt = 10^{-3}$ and β in (2.4) to $\beta = 30$. We study the dynamics for the following two initial conditions:

(IC1) Initialize 2 particles near the local minimizer and 1 particle near the global minimizer.
(IC2) Initialize 1 particle near the local minimizer and 2 particles near the global minimizer.

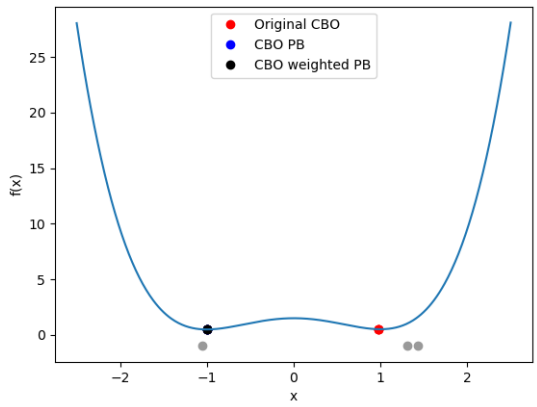

Figure 1. Corresponding to (IC1) Initial positions are depicted in gray. Points in different colors show $v_f(T)$ at $T = 100$.

Table 1. Success rates

scheme	success rate $\alpha = 10$	success rate $\alpha = 30$
CBO	30 %	60,9%
PB	100 %	100 %
wPB	100 %	100 %

The initial positions (IC1) and (IC2) of the particles correspond to the gray dots in Figure 1 and in Figure 3, respectively. We discuss (IC1) first. In this situation the weighted average, $v_f(0)$, is located near $x = 0.9$, thus, the Heaviside functions are zero and the system would be in a stationary state for $\sigma = 0$.

For $\sigma > 0$, the diffusion term drives the dynamic and the particles are exploring their neighborhood. Due to the multiplicative factor, the particle on the left is exposed to more diffusion than the particles on the right. In case of the CBO scheme, the particle on the left has a high probability of jumping out of the basin of the global minimum. Then, all particles concentrate near the local minimum. For one run, this behavior is illustrated by the positions of $v_f(T)$ shown in Figure 1 (left). This alone does not reflect the concentration which becomes apparent in Figure 2. In fact, the orange lines show fluctuations for small times but stabilize quickly indicating that no diffusion is present and thus that all particles are concentrated. This behavior changes when personal best information is included. Here, particles still explore their neighborhood, however at some point their current positions are worse than their personal best, and hence the drift starts pulling them back towards their personal best. This behavior is also illustrated by the success rates stated in Table 1. We see that (w)PB outperform CBO for large and small values of α. The 'pull-back' effect slows down the convergence of (w)PB - we observe that the respective energies decrease slower than for CBO in Figure 2. Nevertheless, they find the global minimum.

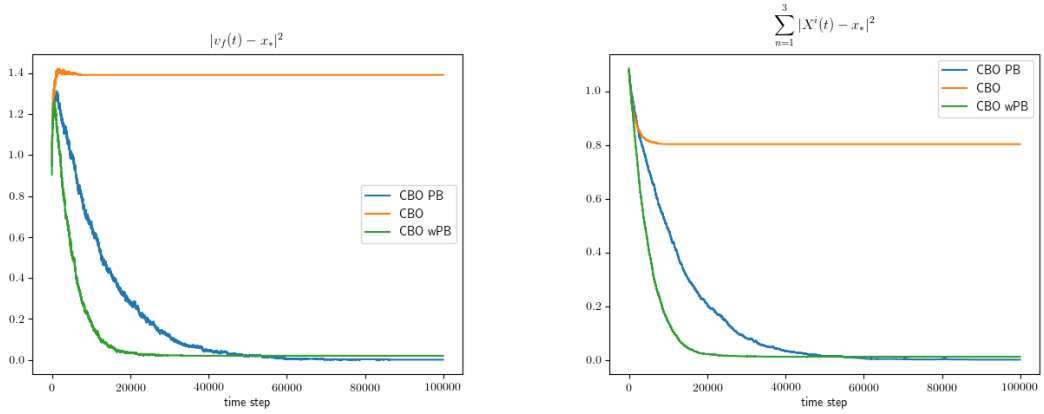

Figure 2. (IC1) CBO is not successful while CBO with personal best finds a good approximation of the global minimizer. The plot on the left shows the mean of the distances of $v_f(t)$ to the global minimizer. The plot on the right shows the mean energy $\sum_{i=1}^{3} |X^i(t) - x_*|^2$. The mean involves 1000 Monte Carlo runs.

Next we consider initial condition (IC2). Again, in the deterministic case $\sigma = 0$ the initial configuration is stationary. For $\sigma > 0$ the particles on the left are less diffusive than the particle on the right. Therefore, it is more likely that the particle on the right jumps into the basin of the global minimum. This is illustrated in Figure 3 and confirmed by the success rates in Table 2. Again, CBO converges faster than (w)PB, see Figure 4. Nevertheless, the function values at the point of concentration are smaller for (w)PB which means that the slower algorithms find better approximations. Note that the scale of the time step-axis is much smaller than in Figure 2.

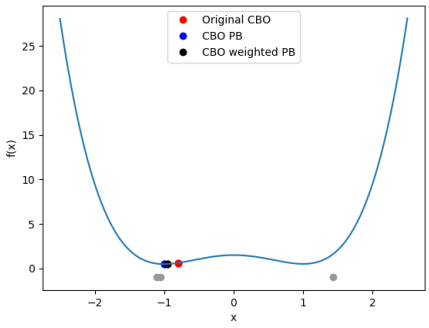

Figure 3. Corresponding to (IC2) The initial positions are depicted in gray. The points in different colors show $v_f(t)$ at $t = 10000$.

Table 2. Success rates

scheme	success rate $\alpha = 10$	success rate $\alpha = 30$
CBO	91,6 %	98,1 %
PB	100 %	100 %
wPB	100 %	100 %

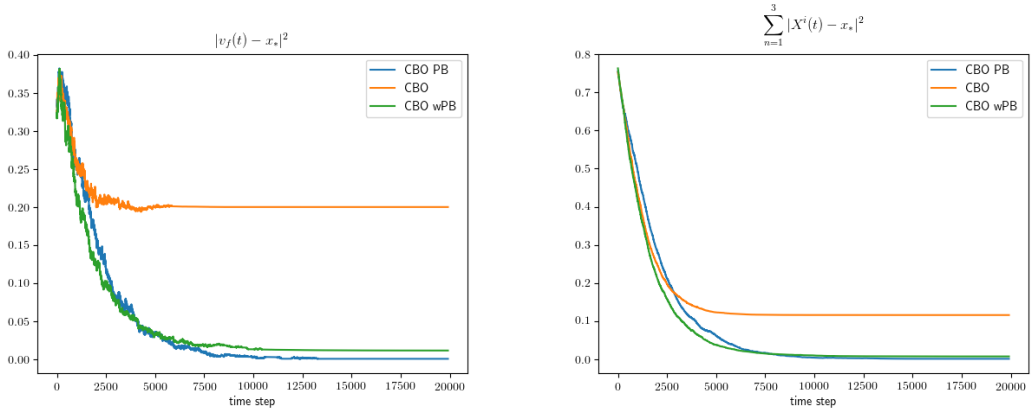

Figure 4. (IC2) All schemes find reasonable approximations of the minimizer. The results of the methods with personal best information have a better accuracy. The plot on the left shows the mean of the distances of $v_f(t)$ to the global minimizer. The plot on the right shows the mean energy $\sum_{i=1}^{3} |X^i(t) - x_*|^2$. The mean involves 1000 Monte Carlo runs. As expected the particles following the CBO scheme are concentrating very fast. The methods with personal best information need more time for stabilization. The one with weighted personal best is slightly faster than with one with true personal best values.

3. Well-posedness results

In the following we discuss well-posedness of the wPB model. We begin by considering CBO with component-wise diffusion, which was proposed in [11].

3.1. Well-posedness of CBO with component-wise diffusion

Theorem 1. *Let f be locally Lipschitz and $N \in \mathbb{N}$. Then system* (2.5) *with $\lambda(t, X) \equiv \lambda$, $\mu(t, X) \equiv 0$ admits a unique strong solution for any initial condition $\xi = (\xi_1, \dots, \xi_N)$ satisfying $\mathbb{E}\,|\xi|^2 < \infty$.*

A detailed proof can be found in the Appendix. Let us just emphasize that the estimates in the

proof of Theorem 1 are independent of the dimension, d, as was already highlighted in [11] for the mean-field setting. This is in contrast to [6, 7], where the estimates depend on d.

3.2. Well-posedness in case of weighted personal best

Next, we present an existence and uniqueness result for the proposed SDE model with weighted personal best and smoothed Heaviside functions. Note that the structure of the weighted personal best suggests the idea of introducing new variables for the numerator and the denominator. This reformulation converts the non-Markovian process into a Markovian system for times $t > 0$, but violates the initial condition. We therefore use different proofs for properties of SDEs with local Lipschitz conditions as well as path-dependent SDEs that can be found in the literature [23]. To the authors' knowledge none of them covers the case of path-dependent SDEs with local Lipschitz conditions. In the following we present a proof which combines the two techniques to obtain a well-posedness result.

We assume that the regularized Heaviside function H^ϵ satisfies the following conditions:

(A1) Let $0 \le H^\epsilon(x) \le 1$ for all $x \in \mathbb{R}$.

(A2) There exists a constant $C > 0$ such that

$$|H^\epsilon(x) - H^\epsilon(y)| \le \frac{C}{\epsilon}|x - y| \qquad \text{for all } x, y \in \mathbb{R}. \tag{3.1}$$

This corresponds to the following regularized problem

$$dX(t) = \left[-\lambda^\epsilon(t, X)(X(t) - v_f) - \mu^\epsilon(t, X)(X(t) - p_f^i)\right] dt + \sqrt{2}\sigma \mathrm{diag}(X(t) - v_f)\, dB_t, \tag{3.2a}$$

with

$$\lambda^\epsilon(t, X) = \mathrm{diag}\Big(\mathrm{H}^\epsilon(f(X^i(t)) - f(v_f))\, \mathrm{H}^\epsilon(f(p_f^i) - f(v_f))\Big)_{i=1,\dots,N} \in \mathbb{R}^{dN\times dN}, \tag{3.2b}$$

$$\mu^\epsilon(t, X) = \mathrm{diag}\Big(\mathrm{H}^\epsilon(f(X^i(t)) - f(p_f^i))\, \mathrm{H}^\epsilon(f(v_f) - f(p_f^i))\Big)_{i=1,\dots,N} \in \mathbb{R}^{dN\times dN}. \tag{3.2c}$$

Moreover, we assume that the objective function f satisfies the following properties:

(A3) Positivity: it holds $0 \le f(x)$ for all $x \in \mathbb{R}^d$,

(A4) Quasi-local Lipschitz condition: for any $n < \infty$ and $|x|, |y| \le n$ it holds

$$|f(x) - f(y)| \le L_f|x - y|,$$

with a constant $L_f > 0$ depending on n only.

Remark 2. *The well-known regularization of the Heaviside function*

$$H^\epsilon(x) = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{x}{\epsilon}\right)$$

satisfies the assumptions (A1) and (A2). Note that in the context of optimization problems, the positivity assumption on f is not too restrictive. Since f corresponds to a minimization functional it is naturally bounded from below and can be shifted to satisfy the positivity constraint.

The following proof is based on a combination of arguments of Theorem 3.17 and Theorem 3.27 in [23] - this yields well-posedness of (3.2). We begin with two lemmata providing necessary estimates. The first lemma is concerned with properties of the weighed averages. Note that the global best v_f depends on the current state of the process and the personal best p_f on the whole trajectory. We therefore write $v_f[\varphi(t)] = v_f(t)$ and $p_f[\varphi] = p_f$.

Lemma 1. *Let f satisfy (A3) and (A4), $N \in \mathbb{N}$ and $\varphi = (\varphi^1, \dots \varphi^N) \in C(\mathbb{R}_+, \mathbb{R}^{dN})$. Then*

$$\begin{aligned} &v_f[\varphi(t)] \in \mathbb{R}^d, && |v_f[\varphi(t)]| \le |\varphi(t)| \quad \text{for every } t, \\ &p_f[\varphi] \in C(\mathbb{R}_+, \mathbb{R}^{dN}), && |p_f[\varphi](t)| \le |\varphi|_t. \end{aligned} \tag{3.3}$$

Moreover, the averages satisfy the local Lipschitz conditions:

$$|p_f[\varphi](t) - p_f[\hat{\varphi}](t)|^2 = \sum_{i=1}^{N} \left|p_f^i[\varphi](t) - p_f^i[\hat{\varphi}](t)\right|^2 \le C_1 \|\varphi - \hat{\varphi}\|_t^2, \tag{3.4}$$

$$\left|v_f[\varphi(t)] - v_f[\hat{\varphi}(t)]\right|^2 \le C_2 |\varphi(t) - \hat{\varphi}(t)|^2 \tag{3.5}$$

for all $t \in [0, \infty)$ with $|\varphi|_t, |\hat{\varphi}|_t \le n$ with constants

$$C_1 = \left(1 + (1 + 2L_f)\beta n e^{\beta(\overline{f} - \underline{f})}\right), \text{ and } C_2 = \left(1 + \frac{\alpha n L_f e^{-\alpha \underline{f}}}{N} + n e^{\alpha(\overline{f} - \underline{f})}\left(\frac{1}{N} + \alpha n L_f\right)\right)^2 2^{N-1}. \tag{3.6}$$

Here L_f is the Lipschitz constant of f in $B_n = \{x \colon |x| \le n\}$ and $\underline{f}, \overline{f}$ correspond, respectively, to the minimal and maximal values of f on B_n.

The proof of Lemma 1 can be found in the Appendix. Using Lemma 1 we show that the drift and diffusion terms satisfy local Lipschitz and linear growth conditions. These properties allow us to apply the existence and uniqueness result later on.

Lemma 2. *Let (A1)–(A4) hold. Then*

$$b \colon [0, +\infty) \times C(\mathbb{R}_+, \mathbb{R}^{dN}) \to \mathbb{R}^{dN} \text{ and } \Sigma \colon [0, +\infty) \times C(\mathbb{R}_+, \mathbb{R}^{dN}) \to \mathbb{R}^{dN \times dN}$$

given by

$$b(t, \varphi) = -\lambda^\epsilon(t, \varphi)(\varphi(t) - v_f) - \mu^\epsilon(t, \varphi)(\varphi(t) - p_f)$$

and

$$\Sigma(t, \varphi) = diag\left((\varphi^i(t) - v_f)_{i=1,\dots,N}\right) \in \mathbb{R}^{dN \times dN}$$

with $\varphi^i = (\varphi_{(i-1)d+1}, \dots, \varphi_{(i-1)d+d})$ satisfy the following conditions for all $\varphi, \psi \in C(\mathbb{R}_+, \mathbb{R}^{dN})$ and all $R > 0$:

(i) $|b(t, \varphi) - b(t, \psi)| \le L_R \|\varphi - \psi\|_t$,
(ii) $|b(t, \varphi)| \le a_R \|\varphi\|_t$,
(iii) $|\Sigma(t, \varphi(t)) - \Sigma(t, \psi(t))| \le \ell_R |\varphi(t) - \psi(t)|$,
(iv) $|\Sigma(t, \varphi(t))| \le b_R |\varphi(t)|$,

where $L_R, \ell_R, a_R, b_R \in \mathbb{R}$.

Proof. To show (i) we calculate

$$|b^i(t,\varphi) - b^i(t,\psi)|^2 \le 2(I_1 + I_2 + I_3 + I_4), \tag{3.7}$$

where

$$\begin{aligned}
I_1 &:= \left|(\lambda^{\epsilon,i}(t,\varphi) - \lambda^{\epsilon,i}((t,\psi))(\varphi^i(t) - v_f[\varphi(t)])\right|^2 \\
&\le \frac{1}{2\epsilon} L_f(|\varphi^i(t)| + |v_f[\varphi(t)]|)\left(2|\varphi^i(t) - \psi^i(t)|^2 + 8|v_f[\psi(t)] - v_f[\varphi(t)]|^2 + 4|p_f^i[\varphi](t) - p_f^i[\psi](t)|^2\right), \\
I_2 &:= \left|\lambda^{\epsilon,i}((t,\psi)(|\varphi^i(t) - \psi^i(t)| + |v_f[\psi(t)] - v_f[\varphi(t)]|)\right|^2 \le 2|\varphi^i(t) - \psi^i(t)|^2 + 2|v_f[\psi(t)] - v_f[\varphi(t)|^2, \\
I_3 &:= \left|(\mu^{\epsilon,i}((t,\varphi) - \mu^{\epsilon,i}((t,\psi))(\varphi^i(t) - p_f^i[\varphi](t))\right|^2 \\
&\le \frac{1}{2\epsilon} L_f(|\varphi(t) - p_f^i[\varphi](t)|)\left(2|\varphi^i(t) - \psi^i(t)|^2 + 4|v_f[\psi(t)] - v_f[\varphi(t)]|^2 + 8|p_f^i[\varphi](t) - p_f^i[\psi](t)|^2\right), \\
I_4 &:= \left|\mu^{\epsilon,i}((t,\psi)(\varphi^i(t) - \psi^i(t) + p_f^i[\psi](t) - p_f^i[\varphi](t))\right|^2 \le 2|\varphi^i(t) - \psi^i(t)|^2 + 2|p_f^i[\varphi](t) - p_f^i[\psi](t)|^2.
\end{aligned}$$

From Lemma 1 we know that $|v_f[\psi(t)] - v_f[\varphi(t)]|^2 \le C_1|\varphi(t) - \psi(t)|^2$, and $|p_f^i[\varphi](t) - p_f^i[\psi](t)|^2 \le C_2\|\varphi^i - \psi^i\|_t^2$ with constants C_1 and C_2 given by (3.6) with $n = R$. This yields (i) since

$$|b(t,\varphi) - b(t,\psi)| = \left(\sum_{i=1}^N |b^i(t,\varphi) - b^i(t,\psi)|\right)^{1/2} \le L_R\|\varphi - \psi\|_t.$$

The Lipschitz bound (iii) follows from similar arguments using the diagonal structure of Σ:

$$\begin{aligned}
|\Sigma(t,\varphi(t)) - \Sigma(t,\psi(t))| &= \left(\sum_{i=1}^N |\Sigma_{ii}(t,\varphi(t)) - \Sigma_{ii}(t,\psi(t))|^2\right)^{1/2} \\
&\le \left(\sum_{i=1}^N 2|\varphi^i(t) - \psi^i(t)|^2 + 2|v_f[\varphi(t)] - v_f[\psi(t)]|^2\right)^{1/2} \le \ell_R|\varphi(t) - \psi(t)|.
\end{aligned}$$

The last two inequalities hold due to

$$\begin{aligned}
|b(t,\varphi)| &= \left(\sum_{i=1}^N b^i(t,\varphi)^2\right)^{1/2} \le \left(\sum_{i=1}^N 8|\varphi^i(t)|^2 + 4|v_f[\varphi(t)]|^2 + 2|p_f^i[\varphi](t)|^2\right)^{1/2} \le a_R\|\varphi\|_t, \\
|\Sigma(t,\varphi)| &= \left(\sum_{i=1}^N \Sigma_{ii}(t,\varphi)^2\right)^{1/2} \le \left(\sum_{i=1}^N 2\varphi^i(t)^2 + 2|v_f[\varphi(t)]|^2\right)^{1/2} \le b_R|\varphi(t)|.
\end{aligned}$$

□

Equipped with this lemma, we have everything at hand to prove the main theorem.

Theorem 2. *Let (A1)–(A4) be satisfied and $\xi \in L^0(\Omega, \mathcal{F}_0, \mathbb{P}, \mathbb{R}^{dN})$ with $\mathbb{E}|\xi|^p < \infty$ for each $p > 0$. Then, there exists a unique strong global solution to* (3.2)*. Moreover, there exists a constant C_{p,T,L_r,ℓ_R} such that*

$$\mathbb{E}\sup_{t\in[0,T]}|X(t)|^p \le C_{p,T,L_r,\ell_R}\mathbb{E}|\xi|^p.$$

The proof combines arguments of Theorem 3.17 (path-dependent SDE) and Theorem 3.27 (SDE with local Lipschitz coefficients) in [23].

Proof. We start by proving uniqueness. Let $X, \hat{X} \in S^0_{dN}$ be two solutions to (3.2) corresponding to initial data $\xi, \hat{\xi} \in L^0(\Omega, \mathcal{F}_0, \mathbb{P}, \mathbb{R}^{dN})$, respectively. Then it holds $\mathbb{E}\,\xi = \mathbb{E}\,\hat{\xi}$. Define the stopping time $\tau_n(\omega) = \inf\{t \geq 0 \colon |X_t(\omega)| + |\hat{X}_t(\omega)| \geq n\}$. Then the two solutions satisfy

$$X_{t\wedge\tau_n} = \xi + \int_0^t 1_{[0,\tau_n]}(s) b(s \wedge \tau_n, X) ds + \int_0^t 1_{[0,\tau_n]}(s) \sigma(s \wedge \tau_n, X_{s\wedge\tau_n}) dB_s,$$
$$\hat{X}_{t\wedge\tau_n} = \hat{\xi} + \int_0^t 1_{[0,\tau_n]}(s) b(s \wedge \tau_n, \hat{X}) ds + \int_0^t 1_{[0,\tau_n]}(s) \sigma(s \wedge \tau_n, \hat{X}_{s\wedge\tau_n}) dB_s,$$

and $\tau_n \to \infty$ for $n \to \infty$. Note that we have global Lipschitz constants for b and σ for all times $t \in [0, \tau_n]$ with n arbitrary but fixed. This allows us to use Theorem 3.8 in [23], see proof of Theorem 3.27 in [23] for more details, to obtain

$$\mathbb{E}\frac{\|e^{-V^R}(X_{\cdot\wedge\tau_n} - \hat{X}_{\cdot\wedge\tau_n})\|_T^p}{\left(1 + \|e^{-V^R}(X_{\cdot\wedge\tau_n} - \hat{X}_{\cdot\wedge\tau_n})\|_T^2\right)^{p/2}} \leq C_p \mathbb{E}\frac{|X_0 - \hat{X}_0|^p}{\left(1 + |X_0 - \hat{X}_0|^2\right)^{p/2}},$$

for some V^R depending on the local Lipschitz constants L_R, ℓ_R, a_R, b_R and $p \geq 2$ arbitrary. Hence, the uniqueness of the solution on $[0, \tau_n]$. As $\tau_n \to \infty$ for $n \to \infty$, this allows us to conclude the global uniqueness of $X \in S^0_{dN}$.

Next, we show the existence of solutions. Let $M \in \mathbb{N}^*$ and $0 = T_0 < T_1 < \cdots < T_M = \tau_n$ with $T_i = \frac{i\tau_n}{M}$. It holds

$$\alpha(\frac{\tau_n}{M}) := \sup_{0<s-t<\frac{\tau_n}{M}} \left(\int_t^s L_R\, dr\right)^p + \left(\int_t^s \ell_R^2\, dr\right)^{p/2} \longrightarrow 0 \quad \text{as } M \to \infty.$$

We employ a fixed point argument for the mapping $\Gamma \colon S^p_{dN}[0, T_1] \to S^p_{dN}[0, T_1]$ given by

$$\Gamma(U)_t = \xi + \int_0^t b(s, U) ds + \int_0^t \sigma(s, U_s) dB_s,$$

where we need no stopping times due to $t < \tau_n$ on $[0, T_1]$. Indeed, the mapping Γ is well-defined since for all $\varphi \in C(\mathbb{R}_+, \mathbb{R}^{dN})$

$$|b(t, \varphi)| \leq L_R \|\varphi\|_t, \text{ and } |\sigma(t, \varphi)| \leq \ell_R \|\varphi\|_t,$$

is satisfied. Because of the Lipschitz continuity, both stochastic processes $b(\cdot, U)$ and $\sigma(\cdot, U_s)$ are progressively measurable for all $U \in S^p_{dN}[0, \tau_n]$ and $b(\cdot, U) \in L^p(\Omega, L^1(0, \tau_n))$ and $\sigma(\cdot, U) \in \Lambda^p_{dN\times dN}(0, \tau_n)$. Therefore,

$$\int_0^\bullet b(r, U) dr\,, \int_0^\bullet \sigma(r, U_r) dB_r \;\in S^p_{dN}[0, \tau_n].$$

We will show that the operator Γ is a strict contraction on the complete metric space $S^p_{dN}[0, T_1]$ for sufficiently large M (where $S^p_{dN}[0, T_1]$ is equipped with the usual distance $d_{p,M}(U, V) = (\mathbb{E}\|U - V\|^p_{T_1})^{1/p\vee 1}$). Let $U, V \in S^p_{dN}[0, T_1]$. By the Burkholder-Davis-Gundy inequality we have

$$\mathbb{E}\|\Gamma(U) - \Gamma(V)\|^p_{T_1} \leq (1 \vee 2^{p-1})\, \mathbb{E} \sup_{s\in[T_0,T_1]} |\int_{T_0}^s b(r, U) - b(r, V) dr|^p$$

$$
\begin{aligned}
&+ (1 \vee 2^{p-1})\,\mathbb{E} \sup_{s\in[T_0,T_1]} \Big| \int_{T_0}^{s} \sigma(r, U_r)\sigma(r, V_r) dB_r \Big|^p \\
&\leq (1 \vee 2^{p-1}) \left[\mathbb{E}\left(\int_{T_0}^{T_1} L_R \|U - V\|_r dr \right)^p + \mathbb{E}\left(\int_{T_0}^{T_1} \ell_R^2 |U_r - V_r| dr \right)^{p/2} \right] \\
&\leq (1 \vee 2^{p-1})\, \alpha(\frac{\tau_n}{M})\, \mathbb{E}(\|U - V\|_{T_1}^p).
\end{aligned}
$$

Let $M_0 \in \mathbb{N}^*$ such that $(1 \vee 2^{p-1})\alpha(\frac{\tau_n}{M_0}) \leq \left(\frac{1}{2}\right)^{1\vee p}$. Then Γ is a strict contraction in $S_{dN}^p[0, T_1]$ and thus (3.2) has a unique solution $X \in S_{dN}^p[0, T_1]$. We extend the solution to the interval $[0, T_2]$ by defining a mapping, again, called $\Gamma\colon S_{dN}^p[0, T_2] \to S_{dN}^p[0, T_2]$:

$$
\Gamma(U)_t = \begin{cases} X_t, & \text{if } t \in [0, T_1], \\ X_{T_1} + \int_{T_1}^t b(s, U)ds + \int_{T_1}^t \sigma(s, U_s)ds, & \text{if } t \in (T_1, T_2]. \end{cases}
$$

We repeat the argument M_0 times to be valid the whole interval $[0, \tau_n]$. Since $\tau_n \to \infty$ almost surely, the uniqueness of the solution implies that

$$
[X_t^{n+1}(\omega) - X_t^n(\omega)] 1_{[0,\tau_n(\omega)]}(t) 1_{[0,\infty)}(\tau_n(\omega)) = 0.
$$

Hence, the process $X \in S_d^0$ is defined by $X_t(\omega) = X_t^n(\omega)$ if $0 \leq t \leq \tau_n(\omega)$ and $\tau_n(\omega) > 0$ is the unique solution to the regularized problem (3.2). □

4. Numerical results

The numerical simulations are based on the direct simulation of system (2.5) using the Euler-Maruyama scheme, in which we do not approximate the Heaviside function H. Note that the smoothing of H was only needed for analytic considerations. In practise, we want only one of the drift terms to effect the particles dynamics, which is why we have not pursued this option any further. The final time is set to $T = 15$, discretized into 3×10^4 time steps. All presented results are averaged over $M = 5000$ realizations. The standard deviation is set to $\sigma = 0.5$, while the number of agents depends on the dimension of the function space. In particular, we set the number of agents to 3, 5 or 10 times the space dimension. The initial positions of particles are drawn from a uniform distribution within a specific domain for each function. The parameters α and β to compute the global and personal best are set to

$$
\alpha = 10 \text{ and } \beta = 10.
$$

A realization is successful if the average mean is close to the function minimum $f_{\min}$, in particular

$$
|f(v_f(T)) - f(x_{\min})| < 0.1.
$$

We compare the performance of the CBO scheme with $\mu = 0$, PB and wPB for the following benchmark problems:

1. *Alpine [26]:* This non-convex differentiable function has a global minimum at $x_{\min} = (0, \ldots, 0)$

$$
f(x) = \sum_{i=1}^{d} |x_i \sin(x_i) + 0.1 x_i|. \tag{4.1}
$$

2. *Ackley [27]:* This function is continuous, non-differentiable and non-convex and has its global minimum at $x_{\min} = (0, \dots, 0)$.

$$f(x) = -20\exp(0.2\sqrt{\frac{1}{d}|x|^2}) - \exp(\frac{1}{d}\sum_{i=1}^{d}\cos(2\pi x_i)) + 20 + \exp(1). \tag{4.2}$$

3. *Rastrigin [28]:* The Rastrigin function is continuous, differentiable and convex, has lots of local minima and a global minimum at $x_{\min} = (0, \dots 0)$.

$$f(x) = 10d + \sum_{i=1}^{d}(x_i^2 - 10\cos(2\pi x_i)). \tag{4.3}$$

4. *Xinsheyang2: [26]* This function is continuous but not differentiable and non-convex with a global minimum at $x_{\min} = (0, \dots 0)$.

$$f(x) = \sum_{i=1}^{d}|x_i|\exp(-\sum_{i=1}^{d}\sin(x_i^2)). \tag{4.4}$$

The choice of these functions is based on the different characteristics they have, see Figure 5 for plots in 2D. Table 3 shows the results for the Alpine function (4.1) and the Ackley function (4.2). We observe that the success rate increases with the number of particles, and decreases for higher space dimension. Weighted personal best and personal best give comparable results, which is not surprising since wPB approximates PB for large values of β. A similar behavior can be seen in the case of the Rastrigin function (4.3) and the Xinsheyang function (4.4) in Table 4.

Table 3. Success rates of consensus based optimization (CBO), personal best (PB) and weighted personal best (wPB) scheme for Alpine (4.1) and Ackley (4.2) in space dimension d for different # of particles.

d	# par.	CBO	PB	wPB
1	3	0.8372	0.7184	0.8456
1	5	0.9626	0.8424	0.968
1	10	0.9986	0.9292	0.999
3	9	0.4868	0.4904	0.485
3	15	0.6294	0.6754	0.6458
3	30	0.847	0.8858	0.8404
5	15	0.3246	0.3266	0.3194
5	25	0.4206	0.4352	0.4226
5	50	0.5748	0.6092	0.5778

d	# par.	CBO	PB	wPB
1	3	0.8138	0.8164	0.8144
1	5	0.9598	0.9602	0.9598
1	10	0.9986	0.999	0.9986
3	9	0.902	0.917	0.9008
3	15	0.9908	0.9934	0.99
3	30	1	1	1
5	15	0.9886	0.9928	0.9908
5	25	0.9996	1	0.9998
5	50	1	1	1

We conclude by investigating a 2D version of the toy problem considered in Section 2.3:

$$f(x_1, x_2) = (x_1^2 - 1)^2 + 0.01x_1 + 0.5 + x_2^2.$$

This function has a global minimum at $x_{\min} = (-1.00125, 0)$ and a local minimum at $(0.998748, 0)$. We wish to explore the dynamics of this 2D version for different number of particles. In doing so we

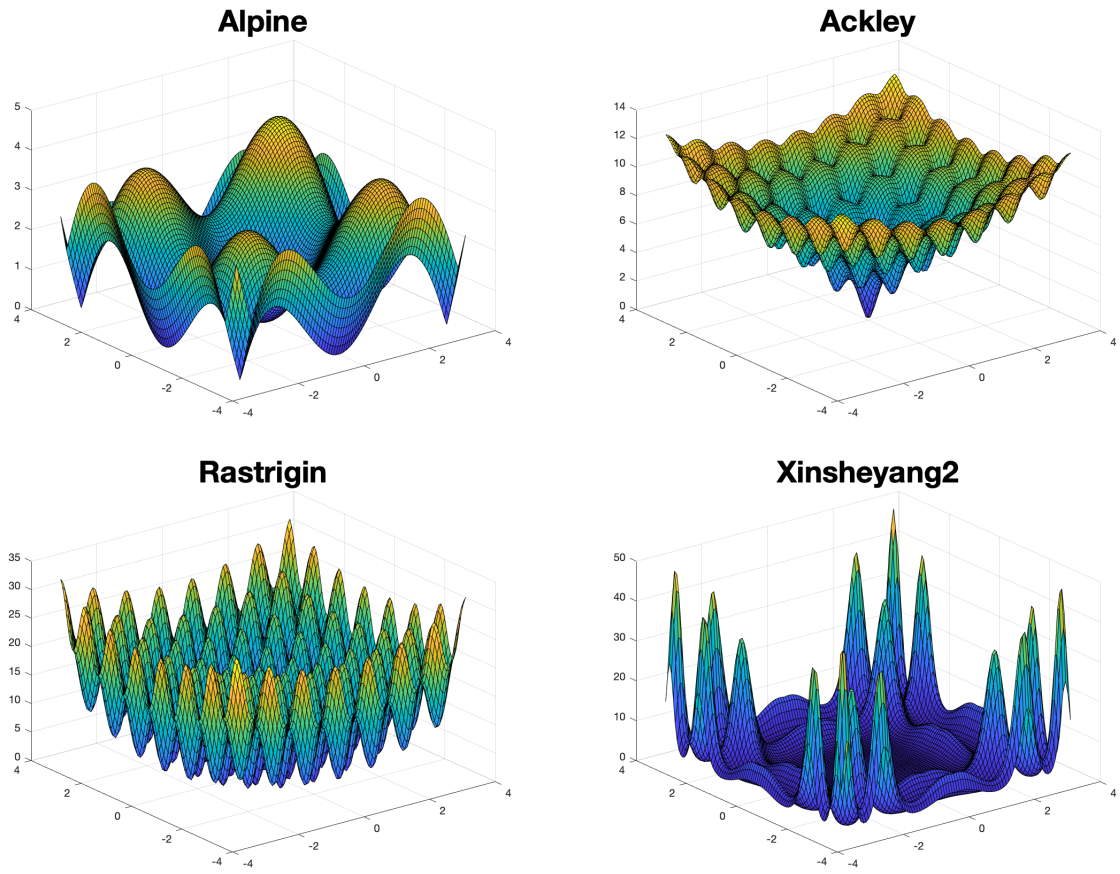

Figure 5. 2D plots of the benchmark functions.

Table 4. Success rates of consensus based optimization (CBO), personal best (PB) and weighted personal best (wPB) scheme for Rastrigin (4.3) and Xinsheyang2 (4.4) in space dimension d for different # of particles.

d	# par.	CBO	PB	wPB
1	3	0.7924	0.7992	0.8006
1	5	0.9528	0.953	0.9522
1	10	0.9984	0.999	0.9988
3	9	0.4634	0.4726	0.4624
3	15	0.6844	0.7112	0.6914
3	30	0.9138	0.9278	0.9204
5	15	0.5012	0.4944	0.5
5	25	0.7074	0.7296	0.7144
5	50	0.908	0.9172	0.9106

d	# par.	CBO	PB	wPB
1	3	0.8732	0.8914	0.8762
1	5	0.9786	0.9892	0.9802
1	10	0.9996	1	0.9998
3	9	0.8538	0.8634	0.8548
3	15	0.9268	0.9412	0.9298
3	30	0.9766	0.984	0.9792
5	15	0.481	0.45	0.4742
5	25	0.4996	0.4724	0.498
5	50	0.5318	0.4918	0.5216

consider 4, 8 or 16 particles and start the particle schemes with 2, 4 and 8 placed in each of the two wells with a random perturbation. Furthermore we set $\alpha = 10$ and $\beta = 20$. Table 5 shows the results as the number of particles increases. We observe that CBO and (w)PB perform equally well for large numbers of particles, and that (w)PB outperform CBO for few particles. This could be explained by the fact that the probability of all particles deviating from the global minimum decreases as their number increases.

Table 5. Success rates of consensus based optimization (CBO), personal best (PB) and weighted personal best (wPB) scheme for the 2D toy problem in space dimension 2 for different # of particles.

d	# par.	CBO	PB	wPB
2	4	0.6572	0.8212	0.7596
2	8	0.7168	0.78	0.7684
2	16	0.7768	0.7684	0.7736

5. Conclusion

In this paper we introduced a consensus based global optimization scheme, which includes the personal best information of each particle. The proposed generalization is motivated by the original works on particle swarm algorithms, in which particles adjust their position as a linear combination of moving towards the current global best and their personal best value.

We discussed how information about the personal best can be included in consensus based optimization schemes, leading to a system of functional stochastic differential equations. A well-posedness result for the respective regularized non-Markovian SDEs was presented. New features of the algorithm with personal best were illustrated and compared in computational experiments. The numerical results indicate that information about the personal best leads to higher success rates in the case of few particles and that the corresponding weighted means are better approximations of the global function minima.

Acknowledgments

The authors would like to thank Christoph Belak (TU Berlin), Stefan Grosskinsky (University of Warwick), Greg Pavliotis (Imperial College London) and Oliver Tse (TU Eindhoven) for the helpful discussions and constructive input. CT was partly supported by the European Social Fund and by the Ministry Of Science, Research and the Arts Baden-Württemberg. MTW was partly supported by the New Frontier's grant NST-0001 of the Austrian Academy of Sciences ÖAW.

Conflict of interest

All authors declare no conflicts of interest in this paper.

Appendix

Proof. [Theorem 1] In the following we denote in abuse of notation by $v_f[X]$ the vector $(v_f, \dots, v_f) \in \mathbb{R}^{dN}$. We rewrite (2.5) with $\lambda(X^i(t), v_f, p_f^i) \equiv \lambda, \mu(X^i(t), v_f, p_f^i) \equiv 0$ as

$$dX(t) = -\lambda(X(t) - v_f[X(t)])dt + \sqrt{2}\sigma \mathrm{diag}(X(t) - v_f[X(t)])dB_t$$

with $\mathrm{diag}(X(t) - v_f[X(t)]) \in \mathbb{R}^{dN\times dN}$ being the diagonal matrix with $d_k = diag(X^k(t) - v_f) \in \mathbb{R}^{d\times d}$ for $k = 1, \dots, N$ and dB_t a dN-dimensional Brownian motion. The argument follows the lines of the well-posedness in [7]. In fact, let $M[X(t)] = \mathrm{diag}(X(t) - v_f[X(t)])$ and $n \in \mathbb{N}$ arbitrary. We have to check that there exists a constant C_n such that

$$-2\lambda X(t) \cdot (X(t) - v_f[X(t)]) + 2\sigma^2 \mathrm{trace}(M[X(t)]M[X(t)]^T) \le C_n |X(t)|^2, \tag{5.1}$$

for every $|X(t)| \le n$. Note that $f(X(t))$ is bounded for $|X(t)| \le n$ due to its local Lipschitz continuity. Hence, the estimate for the first term on the left-hand side is identical to the one in [7]. Indeed, we have

$$-2\lambda X(t) \cdot (X(t) - v_f[X(t)]) \le 2\lambda \sqrt{N}|X(t)|^2.$$

For the component-wise drift we obtain

$$2\sigma^2 \mathrm{trace}\left(M[X(t)]M[X(t)]^T\right) = 2\sigma^2 \sum_{j=1}^{N}\sum_{k=1}^{d}[(X^j(t) - v_f)_k]^2 \le 4\sigma^2(1+N)|X(t)|^2.$$

Combining the two preceding estimates we obtain Eq (5.1). Now, employing [29, Ch 5.3,Thm 3.2] yields the desired result. □

[Lemma 1] We start by showing continuity. For $p_f[\varphi]$ we need to check continuity as $t \to 0$. In fact, l'Hospital's rule yields

$$\lim_{t\to 0} \frac{\int_0^t \varphi^i(s)e^{-\beta f(\varphi^i(s))}ds}{\int_0^t e^{-\beta f(\varphi^i(s))}ds} = \lim_{t\to 0} \frac{\varphi^i(t)e^{-\beta f(\varphi^i(t))}}{e^{-\beta f(\varphi^i(t))}} = \lim_{t\to 0} \varphi^i(t) = \varphi_0^i.$$

This directly implies the continuity of the vector $p_f[\varphi]$. Moreover, it is easy to see that

$$|v_f[\varphi(t)]|^2 = \left|\frac{\sum_{i=1}^N \varphi^i(t)e^{-\alpha f(\varphi^i(t))}}{\sum_{i=1}^N e^{-\alpha f(\varphi^i(t))}}\right|^2 \le |\varphi(t)|^2,$$

$$|p_f[\varphi](t)|^2 = \sum_{i=1}^N |p_f^i[\varphi](t)|^2 = \sum_{i=1}^N \left|\frac{\int_0^t \varphi^i(s)e^{-\beta f(\varphi^i(s))}ds}{\int_0^t e^{-\beta f(\varphi^i(s))}ds}\right|^2 \le \|\varphi\|_t^2.$$

We are left to show the local Lipschitz continuity. Therefore, we estimate

$$\left|p_f^i[\varphi](t) - p_f^i[\hat{\varphi}](t)\right| \le \left|\frac{\int_0^t (\varphi^i(s) - \hat{\varphi}^i(s))e^{-\beta f(\varphi^i(s))}ds}{\int_0^t e^{-\beta f(\varphi^i(s))}ds}\right| + \left|\frac{\int_0^t \hat{\varphi}^i(s)\left(e^{-\beta f(\varphi^i(s))} - e^{-\beta f(\hat{\varphi}^i(s))}\right)ds}{\int_0^t e^{-\beta f(\varphi^i(s))}ds}\right|$$

$$+\left|\frac{\int_0^t \hat{\varphi}^i(s)e^{-\beta f(\hat{\varphi}^i(s))}ds\left(\int_0^t \hat{\varphi}^i(s)e^{-\beta f(\hat{\varphi}^i(s))}ds - \int_0^t \varphi^i(s)e^{-\beta f(\varphi^i(s))}ds\right)}{\int_0^t e^{-\beta f(\varphi^i(s))}ds \int_0^t e^{-\beta f(\hat{\varphi}^i(s))}ds}\right|$$
$$=: I_1 + I_2 + I_3,$$

to obtain

$$I_1 \le \|\varphi^i - \hat{\varphi}^i\|_t, \quad I_2 \le \beta e^{\beta(\overline{f}-\underline{f})} L_f n \|\varphi^i - \hat{\varphi}^i\|_t, \text{ and } I_3 \le \beta e^{\beta(\overline{f}-\underline{f})}(1 + L_f n)\|\hat{\varphi}^i\|_t \|\hat{\varphi}^i - \varphi^i\|_t,$$

with L_f being the global Lipschitz constant of f on $B_n = \{x \in R^d \colon |x| \le n\}$ and $\underline{f}, \overline{f}$ are the minimal and maximal value of f on B_n, respectively. Altogether, this yields the estimate

$$\left|p_f^i[\varphi](t) - p_f^i[\hat{\varphi}](t)\right| \le \left(1 + (1 + 2L_f)\beta n e^{\beta(\overline{f}-\underline{f})}\right)\|\hat{\varphi}^i - \varphi^i\|_t.$$

For the vectors $p_f(t) = (p_f^i[\varphi](t))_{i=1,\dots,N}$ and $\hat{p}_f(t) = (\hat{p}_f^i[\varphi](t))_{i=1,\dots,N}$ this implies

$$|p_f(t) - \hat{p}_f(t)|^2 = \sum_{i=1}^N \left|p_f^i[\varphi](t) - p_f^i[\hat{\varphi}](t)\right|^2 \le \left(1 + (1 + 2L_f)\beta n e^{\beta(\overline{f}-\underline{f})}\right)\|\varphi - \hat{\varphi}\|_t^2.$$

Similarly, we have

$$\left|v_f[\varphi(t)] - v_f[\hat{\varphi}(t)]\right| \le \left|\frac{\sum_{i=1}^N (\varphi^i(t) - \hat{\varphi}^i(t))e^{-\alpha f(\varphi^i(t))}}{\sum_{i=1}^N e^{-\alpha f(\varphi^i(t))}}\right| + \left|\frac{\sum_{i=1}^N \hat{\varphi}^i(t)\left(e^{-\alpha f(\varphi^i(t))} - e^{-\alpha f(\hat{\varphi}^i(t))}\right)}{\sum_{i=1}^N e^{-\alpha f(\varphi^i(t))}}\right|$$
$$+\left|\frac{\sum_{i=1}^N \hat{\varphi}^i(t)e^{-\alpha f(\hat{\varphi}^i(t))}\left(\sum_{i=1}^N \hat{\varphi}^i(t)e^{-\alpha f(\hat{\varphi}^i(t))} - \sum_{i=1}^N \varphi^i(t)e^{-\alpha f(\varphi^i(t))}\right)}{\left(\sum_{i=1}^N e^{-\alpha f(\varphi^i(t))}\right)\left(\sum_{i=1}^N e^{-\alpha f(\hat{\varphi}^i(t))}\right)}\right|$$
$$=: J_1 + J_2 + J_3,$$

which satisfy

$$J_1 \le |\varphi(t) - \hat{\varphi}(t)|_1, \quad J_2 \le \frac{\alpha n L_f e^{-\alpha \underline{f}}}{N}|\varphi(t) - \hat{\varphi}(t)|_1, \text{ and } J_3 \le n e^{\alpha(\overline{f}-\underline{f})}\left(\frac{1}{N} + \alpha n L_f\right)|\hat{\varphi}(t) - \varphi(t)|_1.$$

Thus, we get

$$\left|v_f[\varphi(t)] - v_f[\hat{\varphi}(t)]\right| \le \left(1 + \frac{\alpha n L_f e^{-\alpha \underline{f}}}{N} + n e^{\alpha(\overline{f}-\underline{f})}\left(\frac{1}{N} + \alpha n L_f\right)\right)|\hat{\varphi}(t) - \varphi(t)|_1.$$

Taking squares leads to the estimate

$$\left|v_f[\varphi(t)] - v_f[\hat{\varphi}(t)]\right|^2 \le \left(1 + \frac{\alpha n L_f e^{-\alpha \underline{f}}}{N} + n e^{\alpha(\overline{f}-\underline{f})}\left(\frac{1}{N} + \alpha n L_f\right)\right)^2 2^{N-1}|\hat{\varphi}(t) - \varphi(t)|^2.$$

References

1. R. C. Eberhart, J. Kennedy, Particle swarm optimization, in *Proceedings of ICNN'95-International Conference on Neural Networks* IEEE, (1995), 1942–1948.
2. M. Dorigo, G. Di Caro, Ant colony optimization: a new meta-heuristic, in *Proceedings of the 1999 Congress on Evolutionary Computation* (Cat. No. 99TH8406 Vol. 2), IEEE, (1999), 1470–1477.
3. L. J. Fogel, A. J. Owens, M. J. Wash, *Artificial Intelligence through a Simulation Evolution*, John Wiley & Sons Inc, New York, 1966.
4. J. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Harbor, 1975.
5. R. Poli, J. Kennedy, T. Blackwell, Particle swarm optimization, *Swarm. Intell.*, **1** (2007), 33–57.
6. R. Pinnau, C. Totzeck, O. Tse, S. Martin, A consensus-based model for global optimization and its mean-field limit, *Math. Models Methods Appl. Sci.*, **27** (2017), 183–204.
7. J. A. Carrillo, Y.-P. Choi, C. Totzeck, O. Tse, An analytical framework for consensus-based global optimization method, *Math. Models Methods Appl. Sci.*, **28** (2018), 1037–1066.
8. S. Chatterhee, E. Seneta, Towards consensus: Some convergence theorems on repeated averaging, *J Appl. Probab.*, **14** (1977), 89–91.
9. R. Hegselmann, U. Krause, Opinion dynamics and bounded confidence models, analysis, and simulation, *JASSS*, **5** (2002), 1–33.
10. S. Motsch, E. Tadmor, Heterophilious dynamics enhances consensus, *SIAM Rev.*, **56** (2014), 577–621.
11. J. A. Carrillo, S. Jin, L. Li, Y. Zhu, A consensus-based global optimization method for high dimensional machine learning problems, *arXiv preprint* `arXiv:1909.09249`.
12. M. Fornasier, H. Huang, L. Pareschi, P. Sünnen, Consensus-Based Optimization on the Sphere I: Well-Posedness and Mean-Field Limit, *arXiv preprint* `arXiv:2001.11994`.
13. M. Fornasier, H. Huang, L. Pareschi, P. Sünnen, Consensus-based Optimization on the Sphere II: Convergence to Global Minimizers and Machine Learning, *arXiv preprint* `arXiv:2001.11988`.
14. S.-Y. Ha, S. Jin, D. Kim, Convergence and error estimates for time-discrete consensus-based optimization algorithms, *arXiv preprint* `arXiv:2003.05086`.
15. P. Butta, F. Flandoli, M. Ottobre, B. Zegarlinski, A non-linear model of self-propelled particles with multiple equilibria, *Kinet. Relat. Mod.*, **12** (2019), 791.
16. D. Crisan, C. Jangjigian, T. G. Kurtz, Particle representations for stochastic partial differential equations with boundary conditions, *Electron. J. Probab.*, **23** (2018), 65–94.
17. M. Wiedermann, J. F. Donges, J. Heitzig, J. Kurths, Node-weighted interacting network measures improve the representation of real-world complex systems, *Europhys. Lett.*, **102** (2013), 28007.
18. G. Wergen, Records in stochastic processes—theory and applications, *J. Phys. A: Math. Theor.*, **46** (2013), 223001.
19. S. Gadat, F. Panloup, Long time behaviour and stationary regime of memory gradient diffusions, *Ann. Inst. H. Poincaré Probab. Statist.*, **50** (2014), 564–601.

20. R. Zwanzig, *Nonequilibrium Statistical Mechanics*, Oxford University Press, New York, 2001.
21. H. Duong, G. Pavliotis, Mean field limits for non-Markovian interacting particles: convergence to equilibrium, GENERIC formalism, asymptotic limits and phase transitions, *Commun. Math. Sci.*, **16** (2018), 2199–2230.
22. A. Kuntzmann, Convergence in distribution of some self-interacting diffusions, *J. Probab. Stat.*, **2014** (2014), 1–13.
23. E. Pardoux, A. Răşcanu, *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*, Springer, Cham Heidelberg New York Dordrecht London, 2014.
24. A. Dembo, O. Zeitouni, *Large deviations techniques and applications*, Springer Science & Business Media, 2009.
25. S. Kirkpatrick, C. D. Gelatt Jr, M. P. Vecchi, Optimization by Simulated Annealing, *Science*, **220** (1983), 671–680.
26. M. Jamil, X.-S. Yang, A literature survey of benchmark functions for global optimisation problems, *Int. J. Math. Mod. Num. Opt.*, **4** (2013), 150–194.
27. D. H. Ackley, *A Connectionist Machine for Genetic Hillclimbing*, Kluwer Academic Publishers, Boston, 1987.
28. L. A. Rastrigin, *Systems of extremal control*, Nauka, Moscow, 1974.
29. R. Durrett, *Stochastic calculus: a practical introduction*, CRC press, Boca Raton, Florida, 1996.

AIMS Press